THE UNIVERSITY *of* EDINBURGH

Bayesian inference for ocean transport and diffusivity fields from Lagrangian trajectory data

Ying, Yik Keung

Doctor of Philosophy
University of Edinburgh
May 2020

Acknowledgement

Throughout my 4-year PhD journey, I have received support from many kind and helpful people. Noting that it's impossible to name them all, I would like to dedicate my special gratitude to a subset of them.

Firstly, I would like to express my deepest appreciation to my PhD supervisors James Maddison and Jacques Vanneste. Without their patience and guidance, I would have fallen in many rabbit holes and would not have been able to produce a coherent thesis over such a short time. I will always remember the big whiteboard in James' office portraying the interplay of numerics, fluid dynamics and inverse problems.

I would also like to thank the examiners – Colin Cotter and Aretha Teckentrup – for their careful review and insightful comments on my thesis. It has been a great pleasure to discuss the research at such great details with experts from different areas.

My fellows – Saleh Baqer, Josaine Alejandro Zarco Roldan, Bryan Tope and Miles Savva – should also be addressed. Without them, the PhD journey would not have been such fun and many conversations would not have embarked. Thank you for always listening to and sharing your thoughts with me, especially when I was frustrated. Good luck with your endeavours!

I would also love to thank the many good friends I met in West Mains Roads, particularly – Cammie, Cassie, Bomb, James, Haruka, Hasibi, MacLean, Diego, Eddie, Peter, Ben, Miguel and many more. Special thanks to Tahirah Yasmin, whom I am particularly thankful for her companionship over the full 4 years! I am sure we will meet again, sooner or later.

Good old friends in Hong Kong should not be forgotten – Tom Kwok, Tina Kwok and Catherine Lam. Thank you very much for keeping me connected with our home-town. From the bottom of my heart, I would also like to acknowledge the many courageous HKers who sacrificed their time, freedom and even their lives to defend for our home in the past year. May justice prevail. Glory be to thee Hong Kong.

I would be remiss not to acknowledge my beloved family. Thanks to my Dad for offering me the freedom and supporting me to explore the world and thanks to my mum for reminding me to take care of myself all the time. Thanks to my younger brother, Chiu, for helping me take care of the never-ending home duties when I am abroad. May we all stay healthy and well.

Last but not least, I would like to thank my wife, Tracy Tang, for supporting me all along. Your advice and cheering have always been my greatest source of support and hope. It is not an exaggeration to say 'without you I am nothing'. Happiness with you is all I want for the rest of my life.

Lay summary

Calm as it seems, the ocean is never quiet. Scrutiny will reveal the abundance of eddies at all scales in the oceanic flows. The swirling and self-sustaining circular eddies transport heat and material around the globe, and in turn, regulate the climate.

Modern computational capacity does not allow climate prediction models to fully resolve the turbulent motion of mesoscale eddies, living at the typical length scale of 10 km to 200 km, despite their significance redistributing heat. To overcome this technological constraint, a common approach is to decompose the turbulent ocean motion into a well-resolved mean and unresolved eddy component, with the latter modelled by a diffusion process. This leads to the notion of eddy diffusivity, which relates the rate of unresolved transport to resolved quantities.

Estimating the eddy diffusivity is however an outstanding challenge. In this thesis, we develop a Bayesian framework to infer the eddy diffusivity from the data of flow-following fluid parcels, with the vision to apply the method to the drifter data in the ocean. The Bayesian approach not only allows an estimation of the eddy diffusivity, but also the quantification of uncertainty.

Under the umbrella of the Bayesian framework, different implementations are developed to overcome the various limitation in the inference. Each of the implementations has its strengths and weaknesses and is applicable in different geographical ocean basins. In this thesis, we apply the inference to synthetic trajectory data, and manage to obtain reasonable diffusivity estimates.

Abstract

Eddy diffusion is commonly used to characterise subgrid-scale mixing of tracer quantities in geophysical fluid simulations. Limited by data availability, estimating eddy diffusivity from observational data remains challenging. This thesis describes a novel Bayesian framework to infer ocean transports and diffusivity fields from Lagrangian trajectories data. Modelling the Lagrangian trajectories by a stochastic differential equation whose transition density is given by the advection–diffusion equation, this framework produces a posterior probability distribution for the parameters defining the transport and diffusivity fields, enabling uncertainty quantification.

In this thesis, Lagrangian trajectories from a three-layer quasigeostrophic double-gyre configuration are used to test the inference schemes. The double-gyre is a classic idealisation of large-scale ocean circulation. Mesoscale eddies are continuously produced and dissipated, leading to a complicated time-dependent flow which can plausibly be coarse-grained as diffusion and suitably used to validate the new inference schemes.

Different approaches are implemented based on the Bayesian framework. In a local approach, the ocean domain is divided into an array of cells. Cell-wise defined linear velocity and constant diffusivity fields are inferred using the displacement data of Lagrangian particles originating from the cell. This approach proves capable of estimating the diffusivity in areas with a slow flow such that particles remain in the neighbourhood of their originating cell, but it fails to account for the particle trajectories straddling multiple cells in the considered time interval.

An approach to correct the inference for particles straddling two cells is devised using large deviation theory, assuming the dominance of advective transport over diffusive transport. This approach successfully infers piecewise constant velocity and diffusivity using synthesised trajectories, especially in the limit of increasing dominance of advection.

A global approach is then developed to infer velocity and diffusivity fields defined on the entire domain. This resolves the locality restriction on the trajectory data in the local approach. A naïve implementation of the global inference, however, involves an exceedingly large number of solutions to the advection–diffusion equation. A data coarse-graining approach is applied to overcome the computational challenge. The impact of the data coarse-graining procedures is quantified in the limit of large data. The global approach is applied to the double-gyre simulation data, using a finite volume method for the solution of advection–diffusion equation. The results demonstrate that the global approach enables a robust inference of the mean flow and diffusivity fields at varying sampling interval.

Contents

Chapter 1

Introduction

1.1 Large-scale ocean circulation and mesoscale eddies

The large-scale ocean circulation plays an important role for mankind through its impact on climate and maritime transport. Over the past century, theories and mathematical models have actively been developed to explain the patterns observed from multiple sources, such as sailing logbooks, research vessels, surface buoys, automated drifters and ocean monitoring satellites. The development benefited from the rapid advancement of computational technology, which allows ocean flows to be scrutinised at ever-increasing resolution.

A cornerstone in geophysical fluid dynamics is the theory of geostrophic flow, which is found to capture well the global scale ocean circulation, that is, at length scales of $\mathcal{O}(1000$ km). Geostrophic flows are characterised by the simultaneous maintenance of geostrophic and hydrostatic balance. Geostrophic balance is an equilibrium state where the pressure gradient is laterally balanced by Coriolis force, and hydrostatic balance is the equilibrium between vertical pressure gradient and gravity. A key characteristic of geostrophic flows is that the motion is incompressible and two-dimensional, with the horizontal spatial scales and velocities orders of magnitude greater than the vertical ones.

At the mesoscale of $\mathcal{O}(100$ km), the ocean flows are dynamic and turbulent. At the heart of interest in oceanography are mesoscale eddies, which are ubiquitous in the global ocean. Example hotspots of mesoscale eddies are at the mid-latitudes Gulf stream and Kuroshio or high-latitude Antarctica Circumpolar Current. Typically identified from the surface as a coherent and rotating ring of radius 10 km to 200 km, mesoscale eddies also extend in depth up to 1 km. Fluid properties, such as temperature, salinity or surface elevation, within an individual eddy can also contrast sharply with the background flow (see e.g. Vallis (2006) for a theoretical discussion and e.g. Lilly et al. (2003) for observation of eddies in the real ocean).

The generation of mesoscale eddies can be attributed to different processes. At the larger scales ($\gtrsim 100$ km), barotropic instability is found to be responsible for the emergence of eddies (see e.g. Waterman et al., 2011). Under this regime, a large horizontal velocity gradient in the fluid enables small perturbations from the equilibrium flow to grow into persistent eddies and propagate laterally. Another process, dominating the smaller scales (10 km to 100 km), is baroclinic instability. In this regime, growth in perturbations are triggered by the misalignment in the constant pressure and constant density surfaces, so that denser water tends to displace lighter water to restore the lateral geostrophic balance. We refer readers to Charney (1947); Eady (1949); Pedlosky

(1987); Vallis (2006) for the details on the mechanism.

Mesoscale eddies actively exchange momentum and interact with flows at other length scales. On one hand mesoscale eddies extract energy from the instability of large-scale flows and cascade the energy into smaller sub-mesoscale dynamics (Ferrari and Wunsch, 2009). On the other hand, the mesoscale eddies feed energy back to the large-scale flow and rectify the circulation patterns (Marshall and Radko, 2003; Hallberg and Gnanadesikan, 2006). The intrinsically multiscale nature of mesoscale eddies complicates any attempt to separately break down their net effects to the large-scale flows. However, this is pressingly needed in climate prediction.

Constrained by computational resources, climate models can rarely resolve the mesoscale eddies. An outstanding problem is how the turbulent mixing induced by mesoscale eddies can be represented in a low resolution climate model. It has been long recognised that the mesoscales eddies enhance mixing and transport of material (Lee et al., 1997; Jayne and Marotzke, 2002). On sufficiently large time and length scales, it is hypothesised that the sub-grid scale mixing can be modelled by a diffusion process. This diffusion process, known as *eddy diffusion*, is expected to capture the statistical properties of unresolved advective transport by a diffusive transport so that the net large scale flow remains correctly budgeted. While in this thesis we only address the eddy diffusivity associated with *passive tracers*, which are massless materials are advected by the flows, the concept of eddy diffusion applies also to *active tracers* which interact with the flow fields.

The strength of the diffusion is measured by an *eddy diffusivity*. Its determination from observational data is the central theme of this thesis. While the eddy diffusivity is generally a three-dimensional tensor (see e.g. Redi, 1982; Gargett, 1984), we focus on the lateral components not only for the reason of simplification, but also that the mesoscale transports are approximately in geostrophic balance, hence quasi-horizontal.

In the remaining sections of this chapter, we introduce the physical picture of eddy diffusion (section 1.2), with some rigorous derivations (section 1.3) and outline a list of existing eddy diffusivity diagnostics (section 1.4). Each of these sections is presented from both *Lagrangian* and *Eulerian* perspectives. In section 1.5 we introduce a quasi-geostrophic double gyre flow which serves as a simulator for mesoscale oceanic flows and further application in this thesis. The final section 1.6 outlines the plan of this thesis.

1.2 Eddy diffusivity

Eddy diffusivity can be interpreted from both the Eulerian and Lagrangian viewpoints. This section outlines a number of heuristics of eddy diffusivity in highly idealistic scenarios. Formal treatments in specific cases are discussed in upcoming sections.

1.2.1 Eulerian viewpoint

The Eulerian interpretation, which specifies the flow over a fixed reference frame, of eddy diffusivity has a root in Reynolds averaging for turbulent flows. For a passive tracer c, the Eulerian representation expresses the field in terms of a fixed spatial and temporal coordinates $c = c(\mathbf{x}, t)$.

Consider a passive tracer c advected by a time-varying incompressible velocity field $\mathbf{u}(\mathbf{x}, t) = \overline{\mathbf{u}}(\mathbf{x}) + \mathbf{u}'(\mathbf{x}, t)$ with a steady mean $\overline{\mathbf{u}}$ and a rapidly-revolving fluctuation $\mathbf{u}'$.

The evolution of the passive tracer profile is governed by the advection equation

$$\frac{\partial c}{\partial t} + \mathbf{u} \cdot \nabla c = 0.$$

Suppose the passive tracer $c(\mathbf{x}, t) = \overline{c}(\mathbf{x}, t) + c'(\mathbf{x}, t)$ can also be decomposed into a slowly-varying mean $\overline{c}$ with primarily large-scale structures and a rapidly-revolving fluctuation component c' with fine-grained features. Reynolds averaging the advection equation gives

$$\frac{\partial \overline{c}}{\partial t} + \overline{\mathbf{u}} \cdot \nabla \overline{c} = -\nabla \cdot (\overline{\mathbf{u}'c'}). \tag{1.1}$$

One option for closing the equation is to set up a functional relation between the eddy flux $\overline{\mathbf{u}'c'}$ and the mean scalar $\overline{c}$. A general model is a general linear relation between flux and gradient (e.g Andrews et al., 1987):

$$\overline{\mathbf{u}'c'} = -\mathsf{K}\nabla \overline{c} + \psi_{\mathrm{E}} \nabla^{\perp} \overline{c} + \nabla \times \boldsymbol{\Theta}, \tag{1.2}$$

where K is the eddy diffusivity, ψ_{E} is an eddy-induced streamfunction, $\nabla^{\perp} = (-\partial_y, \partial_x)^{\mathrm{T}}$ is the skew gradient and $\boldsymbol{\Theta}$ is a rotational flux. Note here that 1) the eddy diffusivity is restricted to be symmetric positive-definite, to represent a downgradient Fickian flux in the irreversible turbulent mixing, 2) the eddy-induced streamfunction ψ_{E} is responsible to the net drift induced by eddies, and 3) the rotational flux $\boldsymbol{\Theta}$ is gauge-free under the divergence operator on the eddy flux. Introducing this into (1.1) yields a coarse-grained advection–diffusion equation for the mean tracer $\overline{c}$

$$\frac{\partial \overline{c}}{\partial t} + \left(\overline{\mathbf{u}} + \nabla^{\perp} \psi_{\mathrm{E}}\right) \cdot \nabla \overline{c} = \nabla \cdot (\mathsf{K} \nabla \overline{c}). \tag{1.3}$$

The Eulerian approach introduces eddy diffusivity in an extra diffusion term $\nabla \cdot (\mathsf{K}\nabla\overline{c})$ after averaging the advection equation. The eddy diffusion acts to mix and homogenise the mean tracer over time, with the statistical rate of mixing characterised the eddy diffusivity K.

1.2.2 Lagrangian viewpoint

The Lagrangian interpretation, which observes the flow in a flow-following reference frame, of eddy diffusivity has a root in the studies of turbulent dispersion. The Lagrangian interpretation tracks the motion of individual passive tracer particles.

Consider a single passive particle with position $\mathbf{X}(t)$ at time t. Suppose the particles are advected by a time-varying incompressible velocity field $\mathbf{u}(\mathbf{x}, t)$. The trajectory of a particle is given by the ordinary differential equation

$$\mathrm{d}\mathbf{X}(t) = \mathbf{u}(\mathbf{X}(t), t)\mathrm{d}t,$$

subject to the time of deployment t_0 and initial position $\mathbf{X}(t_0)$.

Consider a statistically stationary flow field $\mathbf{u}(\cdot, \cdot)$ with an ensemble of particles differing by initial positions and/or deployment time. Eddy diffusivity in the Lagrangian viewpoint characterises the statistical dispersion rate of particles. The seminal work of Taylor (1922) shows that in a homogeneous, mean zero and statistically stationary turbulent flow, the variance of particle displacements $\mathbf{S}(t) = \mathbf{X}(t_0 + t) - \mathbf{X}(t_0)$ grows linearly in time beyond the velocity decorrelation time. The half growth rate is known as the absolute diffusivity $\mathsf{K}_{\mathrm{abs}}$. The absolute diffusivity $\mathsf{K}_{\mathrm{abs}}$ is given by

$$\mathsf{K}_{\mathrm{abs}}(\tau) = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\tau} \left\langle \mathbf{S}(\tau) \otimes \mathbf{S}(\tau) \right\rangle, \tag{1.4}$$

where $\otimes$ is the outer product, $\langle\cdot\rangle$ is an ensemble average over all sample displacement $\mathbf{S}(\cdot)$, and τ is the length of the considered time window. As $\tau \to \infty$, the absolute diffusivity converges to a constant and characterises the asymptotic growth rate of dispersion.

A key insight in Taylor (1922) is that the absolute diffusivity is determined by the autocovariance of Lagrangian velocity $\dot{\mathbf{X}}$. Noting that the particle displacement can be represented by $\mathbf{S}(\tau) = \int_0^\tau \dot{\mathbf{X}}(s)\mathrm{d}s$, the absolute diffusivity becomes

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\tau}\left\langle \mathbf{S}(\tau)\otimes\mathbf{S}(\tau)\right\rangle = \left\langle \dot{\mathbf{X}}(\tau)\otimes\mathbf{S}(\tau)\right\rangle = \int_0^\tau \left\langle \dot{\mathbf{X}}(\tau)\otimes\dot{\mathbf{X}}(s)\right\rangle \mathrm{d}s.$$

Note that the outer product $C(\tau,s) = \left\langle \dot{\mathbf{X}}(\tau)\otimes\dot{\mathbf{X}}(s)\right\rangle$ is the velocity autocovariance at times s and τ. The absolute diffusivity $\mathsf{K}_{\mathrm{abs}}$, is therefore given by integrating the autocovariance

$$\mathsf{K}_{\mathrm{abs}}(\tau) = \int_0^\tau C(\tau,s)\mathrm{d}s. \tag{1.5}$$

Note that the absolute diffusivity is a *single-particle* diagnostic, meaning it only involves statistics along an individual trajectory. In principle, the absolute diffusivity is sufficiently determined by the long-time trajectory data of one particle. Should more particles be available, an average over particles can be taken in the average operator $\langle\cdot\rangle$ as well.

For a τ exceeding the Lagrangian velocity decorrelation time scale τ_{D}, the absolute diffusivity converges to a constant K. The linear growth rate in covariance suggests that a stochastic differential equation with diffusivity K may be appropriate to describe the temporal evolution of a particle position $\mathbf{X}(t)$

$$\mathrm{d}\mathbf{X}(t) = \sqrt{2\mathsf{K}}\,\mathrm{d}\mathbf{W}(t), \tag{1.6}$$

where $\mathbf{W}$ is a Brownian process. This is will be formally treated in the next section.

The Lagrangian approach introduces eddy diffusivity through a stochastic noise in the governing equation for particle position. As diffusion acts to model the random walk of particles, the eddy diffusivity characterises the statistical spread rate of particles due to the randomness.

1.2.3 Reconciling the Eulerian and Lagrangian viewpoints

In the previous sections we note that the coarse-grained equations in the Eulerian and Lagrangian viewpoints are respectively an advection–diffusion equation and a stochastic differential equation. The two are in fact closely related. Suppose we start from the Lagrangian viewpoint to consider an 'anti-Itô' stochastic differential equation for the trajectory of a particle $\mathbf{X}(t)$

$$\mathrm{d}\mathbf{X}(t) = \mathbf{U}(\mathbf{X}(t))\,\mathrm{d}t + \sqrt{2\mathsf{K}(\mathbf{X}(t))}\star\mathrm{d}\mathbf{W}(t), \tag{1.7}$$

where $\mathbf{U}(\mathbf{x})$ is a velocity field and $\mathsf{K}(\mathbf{x})$ is a diffusivity field, and $\star\mathrm{d}\mathbf{W}(t)$ is the anti-Itô stochastic integral defined by the right-hand limit

$$\int_0^T g(\mathbf{X}(t))\star\mathrm{d}\mathbf{W}(t) = \lim_{\Delta t\to 0}\sum_0^N g(\mathbf{X}(t_{j+1}))\left(\mathbf{W}(t_{j+1})-\mathbf{W}(t_j)\right), \tag{1.8}$$

on a refining temporal mesh $0 = t_0 < t_1 < \ldots t_N = T$, $\Delta t = t_{j+1} - t_j$. Note that anti-Itô integral evaluates the function g at the final point of the interval $[t_j, t_{j+1}]$,

in contrast with the convention of Itô integration which uses the starting point. The anti-Itô equation can be transformed into the Itô form by introducing an extra drift

$$\mathrm{d}\mathbf{X}(t) = (\mathbf{U}(\mathbf{X}(t)) + \nabla \cdot \mathsf{K}(\mathbf{X}(t)))\ \mathrm{d}t + \sqrt{2\mathsf{K}(\mathbf{X}(t))}\,\mathrm{d}\mathbf{W}(t). \tag{1.9}$$

Subject to sufficient regularity condition on the velocity $\mathbf{U}$ and diffusivity K, the transition density $\pi(\mathbf{x}, t|\mathbf{y})$ for a particle starting at $\mathbf{X}(0) = \mathbf{y}$ to be found at position $\mathbf{x}$ at time t satisfies the Fokker–Planck equation (e.g. Evans, 2013; Pavliotis, 2014)

$$\frac{\partial \pi}{\partial t} + \nabla \cdot (\mathbf{U}\pi) = \nabla \cdot (\mathsf{K}\nabla\pi), \tag{1.10a}$$

$$\pi\left(\mathbf{x}, 0|\mathbf{y}\right) = \delta\left(\mathbf{x} - \mathbf{y}\right), \tag{1.10b}$$

which contains a standard Fickian diffusion. We note that the Fokker–Planck equation coincides with the advection–diffusion equation, if the velocity is incompressible $\nabla \cdot \mathbf{U} = 0$.

Now returning to the coarse-grained advection–diffusion equation (1.3) defined in the Eulerian viewpoint

$$\frac{\partial \overline{c}}{\partial t} + \left(\overline{\mathbf{u}} + \nabla^{\perp}\psi_{\mathrm{E}}\right) \cdot \nabla \overline{c} = \nabla \cdot (\mathsf{K}\nabla\overline{c}),$$

we note that by normalising the profile $\overline{c}$, the coarse-grained advection–diffusion equation (1.3) can be interpreted as the probability density function of the position of an individual passive particle, whose trajectory $\mathbf{X}(t)$ is governed by the stochastic differential equation

$$\mathrm{d}\mathbf{X}(t) = \left(\overline{\mathbf{u}}(\mathbf{X}(t)) + \nabla^{\perp}\psi_{\mathrm{E}}(\mathbf{X}(t)) + \nabla \cdot \mathsf{K}(\mathbf{X}(t))\right)\ \mathrm{d}t + \sqrt{2\mathsf{K}(\mathbf{X}(t))}\,\mathrm{d}\mathbf{W}(t), \tag{1.11}$$

noting $\overline{\mathbf{u}}$ is assumed incompressible. The stochastic differential equation can hence be interpreted as the coarse-grained equation in the Lagrangian viewpoint.

1.3 Homogenisation theory

An eddy diffusivity can be formally derived for systems with time or spatial scale separation. This is achieved using homogenisation theory – a general class of asymptotic methods. Homogenisation theory is applied to the advection by a rapidly varying velocity field (e.g. Majda and McLaughlin, 1993; Majda and Kramer, 1999). The method applies to a variety of problems ranging from Markov chains, ordinary, stochastic to partial differential equations. To illustrate the key principles, we consider two examples - one for the advection–diffusion equation in the Eulerian frame and one for a chaotic ordinary differential equation in the Lagrangian frame.

1.3.1 Eulerian example

Consider the two-dimensional advection–diffusion equation for a passive scalar $c = c(\mathbf{x}, t)$. The incompressible velocity field $\mathbf{u}$ is composed of a uniform mean component $\tilde{\mathbf{u}}$ and an eddy component $\mathbf{u}_{\mathrm{E}}(\mathbf{x})$

$$\mathbf{u}(\mathbf{x}) = \tilde{\mathbf{u}} + \mathbf{u}_{\mathrm{E}}(\mathbf{x}),$$

and the small-scale diffusivity κ is constant. We further assume that the eddy component is 1-periodic along both spatial dimensions. The advection–diffusion equation is therefore

$$\frac{\partial c}{\partial t} + (\tilde{\mathbf{u}} + \mathbf{u}_{\mathrm{E}}(\mathbf{x})) \cdot \nabla c = \kappa \nabla^2 c. \tag{1.12}$$

Since the mean velocity is uniform, by introducing a moving coordinate

$$\mathbf{x} = \tilde{\mathbf{x}} + t\tilde{\mathbf{u}}. \tag{1.13}$$

and defining $\tilde{c}(\tilde{\mathbf{x}}, t) = c(\tilde{\mathbf{x}} + t\tilde{\mathbf{u}}, t)$, an advection–diffusion equation with mean flow eliminated follows

$$\frac{\partial \tilde{c}}{\partial t} + \mathbf{u}_{\mathrm{E}}(\tilde{\mathbf{x}} + t\tilde{\mathbf{u}}) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c} = \kappa \nabla^2_{\tilde{\mathbf{x}}} \tilde{c}, \tag{1.14}$$

where the subscripts denote derivatives with respect to the indicated variable.

To seek a solution to the transformed advection–diffusion equation (1.14) at large length and time scale is seek, the *diffusive rescaling* $\tilde{\mathbf{x}} \to \tilde{\mathbf{x}}/\epsilon$ and $t \to t/\epsilon^2$ with a small bookkeeping parameter ϵ is considered. Define the large scale solution

$$\tilde{c}^\epsilon(\tilde{\mathbf{x}}, t) = \tilde{c}\left(\frac{\tilde{\mathbf{x}}}{\epsilon}, \frac{t}{\epsilon^2}\right).$$

The rescaled advection–diffusion equation becomes

$$\frac{\partial \tilde{c}^\epsilon}{\partial t} + \frac{1}{\epsilon}\mathbf{u}_{\mathrm{E}}\left(\frac{\tilde{\mathbf{x}}}{\epsilon} + \frac{t\tilde{\mathbf{u}}}{\epsilon^2}\right) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}^\epsilon = \kappa \nabla^2_{\tilde{\mathbf{x}}} \tilde{c}^\epsilon. \tag{1.15}$$

Denote an auxiliary variable $\mathbf{y} = \tilde{\mathbf{x}}/\epsilon + t\tilde{\mathbf{u}}/\epsilon^2$ so that the differential operators are transformed as

$$\frac{\partial}{\partial t} \mapsto \frac{\partial}{\partial t} + \frac{1}{\epsilon^2}\tilde{\mathbf{u}} \cdot \nabla_{\mathbf{y}}, \quad \nabla_{\tilde{\mathbf{x}}} \mapsto \nabla_{\tilde{\mathbf{x}}} + \frac{1}{\epsilon}\nabla_{\mathbf{y}}, \quad \nabla^2_{\tilde{\mathbf{x}}} \mapsto \nabla^2_{\tilde{\mathbf{x}}} + \frac{2}{\epsilon}\nabla_{\tilde{\mathbf{x}}} \cdot \nabla_{\mathbf{y}} + \frac{1}{\epsilon^2}\nabla^2_{\mathbf{y}}.$$

We introduce a multiscale ansatz for equation (1.15)

$$\tilde{c}^\epsilon(\tilde{\mathbf{x}}, t) = \tilde{c}_0(\tilde{\mathbf{x}}, t, \mathbf{y}) + \epsilon \tilde{c}_1(\tilde{\mathbf{x}}, t, \mathbf{y}) + \epsilon^2 \tilde{c}_2(\tilde{\mathbf{x}}, t, \mathbf{y}) + \dots, \tag{1.17}$$

and assume that all $\tilde{c}_i(\tilde{\mathbf{x}}, \mathbf{y}, t)$ are 1-periodic in $\mathbf{y}$, as is the eddy velocity $\mathbf{u}_{\mathrm{E}}$. The $\mathbf{y}$-variable is here seen as an independent variable which can be rigorously justified in the limit of $\epsilon \to 0$ (chapter 21 of Pavliotis and Stuart, 2008). By matching the terms in the order respectively in $\mathcal{O}\left(\epsilon^{-2}\right)$, $\mathcal{O}\left(\epsilon^{-1}\right)$ and $\mathcal{O}(1)$, the multiscale ansatz leads to a hierarchy of equations

$$-\tilde{\mathbf{u}} \cdot \nabla_{\mathbf{y}} \tilde{c}_0 + \kappa \nabla^2_{\mathbf{y}} \tilde{c}_0 = 0, \tag{1.18a}$$

$$-(\tilde{\mathbf{u}} + \mathbf{u}_{\mathrm{E}}(\mathbf{y})) \cdot \nabla_{\mathbf{y}} \tilde{c}_1 + \kappa \nabla^2_{\mathbf{y}} \tilde{c}_1 = -\mathbf{u}_{\mathrm{E}}(\mathbf{y}) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}_0 + 2\kappa \nabla_{\mathbf{y}} \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}_0, \tag{1.18b}$$

$$-(\tilde{\mathbf{u}} + \mathbf{u}_{\mathrm{E}}(\mathbf{y})) \cdot \nabla_{\mathbf{y}} \tilde{c}_2 + \kappa \nabla^2_{\mathbf{y}} \tilde{c}_2 = \frac{\partial \tilde{c}_0}{\partial t} + \mathbf{u}_{\mathrm{E}}(\mathbf{y}) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}_1 - \kappa\left(\nabla^2_{\tilde{\mathbf{x}}} \tilde{c}_0 + 2\nabla_{\tilde{\mathbf{x}}} \cdot \nabla_{\mathbf{y}} \tilde{c}_1\right). \tag{1.18c}$$

The $\mathcal{O}\left(\epsilon^{-2}\right)$ equation (1.18a), equipped with a periodic boundary condition in y, only has solution independent of $\mathbf{y}$, thus

$$\tilde{c}_0 = \tilde{c}_0(\mathbf{x}, t).$$

Introducing $\tilde{c}_0$ into the right-hand side of the $\mathcal{O}\left(\epsilon^{-1}\right)$ equation (1.18b) and assuming a separable solution to $\tilde{c}_1$ gives

$$\tilde{c}_1(\mathbf{x}, t, \mathbf{y}) = \boldsymbol{\chi}(\mathbf{y}) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}_0(\mathbf{x}, t), \tag{1.19}$$

where the function $\boldsymbol{\chi}(\mathbf{y})$ satisfies the *cell problem*

$$-\left(\tilde{\mathbf{u}} + \mathbf{u}_{\mathrm{E}}\left(\mathbf{y}\right)\right) \cdot \nabla_{\mathbf{y}} \boldsymbol{\chi}(\mathbf{y}) + \kappa \nabla_{\mathbf{y}}^2 \boldsymbol{\chi}(\mathbf{y}) = \mathbf{u}_{\mathrm{E}}(\mathbf{y}). \tag{1.20}$$

Now we note that the same operator acting on $\tilde{c}_1$ and $\tilde{c}_2$ in the left-hand sides of the $\mathcal{O}\left(\epsilon^{-1}\right)$ and $\mathcal{O}\left(1\right)$ equations has a one-dimensional null space. To uniquely determine χ and $\tilde{c}_2$, we demand a *centering* condition in the $\mathcal{O}\left(\epsilon^{-1}\right)$ equation

$$\int_{\mathbb{T}^2} \boldsymbol{\chi}(\mathbf{y}) \mathrm{d}\mathbf{y} = \mathbf{0}, \tag{1.21}$$

a *solvability condition* (a result known as Fredholm alternative) in the $\mathcal{O}\left(1\right)$ equation,

$$\int_{\mathbb{T}^2} \frac{\partial \tilde{c}_0}{\partial t} + \mathbf{u}_{\mathrm{E}}(\mathbf{y}) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}_1 - \kappa \left(\nabla_{\tilde{\mathbf{x}}}^2 \tilde{c}_0 + 2 \nabla_{\tilde{\mathbf{x}}} \cdot \nabla_{\mathbf{y}} \tilde{c}_1\right) \mathrm{d}\mathbf{y} = 0. \tag{1.22}$$

The solvability condition leads to a coarse-grained equation for $\tilde{c}_0$,

$$\begin{aligned} \frac{\partial \tilde{c}_0}{\partial t} = \kappa \nabla_{\tilde{\mathbf{x}}}^2 \tilde{c}_0 + \int_{\mathbb{T}^2} & -\mathbf{u}_{\mathrm{E}}(\mathbf{y}) \cdot \nabla_{\tilde{\mathbf{x}}} \left(\boldsymbol{\chi}(\mathbf{y}) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}_0(\mathbf{x}, t)\right) \\ & + 2\kappa \nabla_{\tilde{\mathbf{x}}} \cdot \nabla_{\mathbf{y}} \left(\boldsymbol{\chi}(\mathbf{y}) \cdot \nabla_{\tilde{\mathbf{x}}} \tilde{c}_0(\mathbf{x}, t)\right) \mathrm{d}\mathbf{y}. \end{aligned}$$

Upon algebraic manipulations (similar to section 13.6.2 in Pavliotis and Stuart, 2008), the coarse-grained equation simplifies into a diffusion equation with eddy diffusivity K

$$\frac{\partial \tilde{c}_0}{\partial t} = \mathsf{K} : \nabla_{\mathbf{x}}^2 \tilde{c}_0, \quad \text{where} \quad \mathsf{K} = \kappa \mathsf{I} + \int_{\mathbb{T}^2} \mathbf{u}_{\mathrm{E}}(\mathbf{y}) \otimes \boldsymbol{\chi}(\mathbf{y}) \mathrm{d}\mathbf{y}, \tag{1.23}$$

with the Frobenius inner product :, the Hessian matrix $\nabla_{\mathbf{x}}^2$ and an identity matrix I.

Inverse transforming the moving coordinates recovers an advection–diffusion equation for the leading order solution $c_0 = c_0(\mathbf{x}, t)$,

$$\frac{\partial c_0}{\partial t} + \tilde{\mathbf{u}} \cdot \nabla_{\mathbf{x}} c_0 = \mathsf{K} : \nabla_{\mathbf{x}}^2 c_0, \tag{1.24}$$

which effectively coarse-grains the small-scale eddy velocity $\mathbf{u}_{\mathrm{E}}$ into the eddy diffusivity K. Note that the eddy diffusivity is enhanced from the small-scale diffusivity $\kappa\mathsf{I}$. The impact of the small-scale velocity field is exactly diffusive in the asymptotic limit $\epsilon \to 0$. This is an example of rigorous justification of what is otherwise heuristic. More examples of homogenisation theory applied to flows with non-zero mean can be found in Pavliotis (2002) and references therein.

1.3.2 Lagrangian example

The previous section shows that a velocity field that varies rapidly in space can be parameterised by an eddy diffusion. We now show that this is also true for a velocity field that varies rapidly in space. Consider an ordinary differential equation with a slow ($t \sim \mathcal{O}(1)$) mean component $\mathbf{u}_0(\mathbf{x})$ and a fast ($t \sim \mathcal{O}(1/\epsilon)$) and rapidly evolving eddy component $\mathbf{u}_1(\mathbf{x}, t/\epsilon^2)$

$$\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} = \frac{1}{\epsilon} \mathbf{u}_1\left(\mathbf{X}, t/\epsilon^2\right) + \mathbf{u}_0(\mathbf{X}), \tag{1.25}$$

where the capitalised $\mathbf{X}(t)$ refers to the position of a particle.

Following a multiscale asymptotic expansion, it is shown (e.g. chapter 11 of Pavliotis and Stuart, 2008) that the temporal evolution of the $\mathbf{X}(t)$ at $t \sim \mathcal{O}(1)$ can be approximated by the stochastic differential equation

$$\mathrm{d}\mathbf{X} = \mathbf{U}(\mathbf{X})\,\mathrm{d}t + \sqrt{2\mathsf{K}(\mathbf{X})}\,\mathrm{d}\mathbf{W}, \tag{1.26}$$

where the mean drift $\mathbf{U}(\mathbf{X})$ and eddy diffusivity $\mathsf{K}(\mathbf{X})$ are given by

$$\mathbf{U}(\mathbf{x}) = \mathbf{u}_0(\mathbf{x}) + \int_0^\infty \overline{\nabla_{\mathbf{x}}\mathbf{u}_1(\mathbf{x}, s/\epsilon^2)\mathbf{u}_1(\mathbf{x}, (s+t)/\epsilon^2)}\mathrm{d}s \tag{1.27a}$$

$$\mathsf{K}(\mathbf{x}) = \int_0^\infty \overline{\mathbf{u}_1(\mathbf{x}, s/\epsilon^2) \otimes \mathbf{u}_1(\mathbf{x}, (s+t)/\epsilon^2)}\,\mathrm{d}s. \tag{1.27b}$$

where the overline $\overline{\cdot}$ is the time-average operator $\overline{S} = \lim_{T\to\infty}\int_0^T S(t)\,\mathrm{d}t/T$ for a function $S(t)$.

While the detailed derivations are omitted, we highlight the physical interpretation of the derived mean drift and eddy diffusivity. The mean drift $\mathbf{U}(\mathbf{x})$ is the sum of a time-independent slow motion $\mathbf{u}_0(\mathbf{x})$ and a net eddy-driven drift. The net eddy-driven drift is controlled by the correlation between the spatial gradient $\nabla_{\mathbf{x}}\mathbf{u}_1(\mathbf{x}, s/\epsilon^2)$ and temporal fluctuation $\mathbf{u}_1(\mathbf{x}, (s+t)/\epsilon^2)$ in the eddy velocity component. On the other hand, the eddy diffusivity K is effectively determined by integrating the autocovariance of the eddy velocity $\mathbf{u}_1(\mathbf{x}, s/\epsilon^2)$ – a result analogous to the absolute diffusivity (1.28) firstly defined by Taylor (1922).

Through this example, we note how an originally deterministic Lagrangian dynamical system can formally be coarse-grained to a stochastic one. Here we also formally justify the link between the eddy diffusivity and the autocorrelation of eddy velocity.

1.4 Oceanographic diffusivity diagnostics

In the previous section, we demonstrated, with examples from both Eulerian and Lagrangian perspectives, that homogenisation theory can be applied to justify and derive eddy diffusivities for advection by multiscale velocity fields. We however stress that such formal treatments for eddy diffusivity are rarely justified in practice. In the two examples, the fast components in velocity are respectively periodic and rapidly evolving at fixed scales well separated from those of the mean flow. In the ocean this is however not the case, as oceanic flows simultaneously interact at a continuous spectrum of scales.

Nonetheless, the use of eddy diffusivities are commonly employed, and there is still a range of approaches to estimate an eddy diffusivity for oceanographic data. We again broadly classify them into two major groups - Eulerian and Lagrangian.

1.4.1 Eulerian diagnostics

Eulerian diagnostics can be built upon the Reynolds averaging technique introduced in section 1.2. We recall that the eddy flux is modelled by

$$\overline{\mathbf{u}'c'} = -\mathsf{K}\nabla\overline{c} + \psi_{\mathrm{E}}\nabla^{\perp}\overline{c} + \nabla\times\boldsymbol{\Theta},$$

with an eddy diffusivity K, eddy-induced streamfunction ψ_{E} and a rotational flux $\boldsymbol{\Theta}$. The central idea in Eulerian flux diagnostics is to determine K (and ψ_{E}) from observational data for $\mathbf{u}'$, c' and $\overline{c}$. This task is nonetheless complicated by the ambiguity in the rotational flux (Marshall and Shutts, 1981).

There are many approaches to link K to different physical quantities. One approach is through mixing length arguments to directly relate the eddy diffusivity K to observed eddy length and velocity scales (e.g Green, 1970). This approach is generalised to account for the relative motion of eddies to the mean flow and it is related to Lagrangian diagnostics (discussed in section 1.4.2) in recent studies (Ferrari and Nikurashin, 2010; Klocker et al., 2012a). These studies however only consider the diffusivity in the cross-stream direction orthogonal to a priori known mean flows.

Another approach to estimate K is to focus on the advection of specific active tracers, such as potential vorticity or buoyancy, and rely on additional physical and geometric constraints to uniquely determine an eddy diffusivity beyond the gauge-freedom (e.g Eden and Greatbatch, 2008; Maddison et al., 2015; Mak et al., 2016). This line of approach may not necessarily distinguish between the up-gradient stirring and down-gradient mixing, leading to a negative eddy diffusivity in practice. This can complicate the physical interpretation of the diffusivity estimates.

A third approach to infer K is by tracking the temporal evolution of specific tracer properties. The effective diffusivity defined in Nakamura (1996); Marshall et al. (2006) tracks the temporal evolution of contour lengths and areas in tracer concentration profile to quantify the diffusive transport in the across-stream direction. Bachman et al. (2015, 2020); Haigh et al. (2020) formulate optimisation problems to fit the statistics of artificially deployed passive tracers to eddy diffusivity tensor.

The Eulerian diagnostics are applied when full knowledge of the spatio-temporal fluctuations of flow and tracer fields are available. This is typically the case in a numerical simulation but not real observations. In practice, sea surface satellite altimetry may be used to reconstruct the instantaneous geostrophic flows and simulate tracer dispersion (e.g Zhai and Greatbatch, 2006; Abernathey and Marshall, 2013). This approach relies on an assumption of geostrophic balance that is not always well satisfied, especially at the surface.

1.4.2 Lagrangian diagnostics

Lagrangian diagnostics can be further categorised into two major sub-groups: *single-particle* and *multi-particle*. The former is derived from statistics of the temporal evolution of individual trajectories and the latter is derived from statistics of the relative displacement among trajectories.

Single-particle diffusivity

The Lagrangian diagnostics are built upon the idea of quantifying the dispersion rates of particles based on the characterisation of Taylor (1922), presented in section 1.2. In contrast to the scenario considered by Taylor (1922), turbulent flow in the ocean is highly heterogeneous. Adaptations are therefore needed to generalise the absolute diffusivity in section 1.2.

One approach to define an eddy diffusivity in a non-uniform flow field is to measure the particle spread rates relative to of the mean flow. This can be performed by correcting the observed displacement with respect to the mean displacement, and defining the absolute diffusivity (e.g. Sallée et al., 2008; Rypina et al., 2012),

$$\mathsf{K}_{\mathrm{abs}}(\tau) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\tau}\left\langle (\mathbf{S}_i(\tau) - \langle \mathbf{S}_i(\tau)\rangle) \otimes (\mathbf{S}_i(\tau) - \langle \mathbf{S}_i(\tau)\rangle)\right\rangle , \tag{1.28}$$

with an ensemble average $\langle\cdot\rangle$ over a set of particles in the same neighbourhood to equip (1.28) with a spatial correspondence. However, the resulting diffusivity makes no

distinctions between the dispersion caused by velocity shear, velocity strain or turbulent transport. Shear and strain dispersion grow quadratically and exponentially in time respectively, differing from the linear rate in diffusion. It is therefore not necessary that such $\mathsf{K}_{\text{abs}}(\tau)$ would converge to any constant as $\tau \to \infty$, and be interpreted as the rate of the irreversible diffusive process.

Another approach to define an eddy diffusivity is to generalise the formula associated with eddy velocity autocovariance (1.5). This will necessarily involve decomposing the observed Lagrangian velocity $\dot{\mathbf{X}}_i$ into a mean and eddy component. One may expect that the Lagrangian mean velocity could be applied here for the mean-eddy decomposition, but the determination of Lagrangian mean velocity in itself can be challenging. Davis (1987) instead uses the time-averaged Eulerian mean velocity $\overline{\mathbf{u}}(\mathbf{x})$:

$$\dot{\mathbf{X}}_i'(s) = \dot{\mathbf{X}}(s) - \overline{\mathbf{u}}(\mathbf{X}_i(s)), \tag{1.29}$$

for the mean-eddy decomposition. We however emphasise that there are multiple proposals on the optimal way to perform this decomposition, for example, by fitting the trajectory data with time-series models (Bauer et al., 2002; Sykulski et al., 2016) or stochastic processes (Lilly et al., 2017). Once the Lagrangian velocity is decomposed, a diffusivity can be defined by the integral of eddy velocity autocovariance

$$\mathsf{K}_{\text{abs}}(\tau) = \int_0^\tau \left\langle \dot{\mathbf{X}}_i'(s) \otimes \dot{\mathbf{X}}_i'(t+s) \right\rangle \mathrm{d}s, \tag{1.30}$$

with an ensemble average $\langle\cdot\rangle$ over a targeted set of particles.

While the eddy diffusivities defined as (1.28) and (1.30) offer a quantification of the particle dispersion rate, it is unclear how they are related to Lagrangian coarse-grained models. A systematic approach to derive an eddy diffusivity for a coarse-grained equation is presented by Davis (1987). In his approach, he assumes 1) a scale of separation between the eddy and mean velocity, 2) the eddy velocity is approximately Gaussian in probability distribution, 3) a finite decorrelation time scale on successive particle displacements exists, 4) the spatial variability of the mean flow is small relative to the particle displacement at the considered time scale. Note these assumptions may not necessarily hold in the ocean (Davis, 1991; Swenson and Niiler, 1996; Bracco et al., 2000).

Davis derives a generalised advection–diffusion equation

$$\begin{aligned}\frac{\partial}{\partial t}\overline{c}(\mathbf{x},t) + \mathbf{u}(\mathbf{x})\cdot\nabla\overline{c}(\mathbf{x},t) = \nabla\cdot\left(\mathsf{K}_{\text{Davis}}(\mathbf{x},\tau)\nabla\overline{c}(\mathbf{x},t)\right) + \\ + \nabla\cdot\left(\int_0^t \frac{\partial}{\partial\tau}\mathsf{K}_{\text{Davis}}(\mathbf{x},t-\tau)\nabla\left(\overline{c}(\mathbf{x},t) - \overline{c}(\mathbf{x},\tau)\right)\mathrm{d}\tau\right) \\ + \nabla\cdot\left(\int_0^t \frac{\partial}{\partial\tau}\mathsf{K}_{\text{Davis}}(\mathbf{x},\tau)\nabla\overline{c}(\mathbf{x},t-\tau)\mathrm{d}\tau\right),\end{aligned} \tag{1.31}$$

(see equation (2.10) of Davis, 1987)) for which a spatially dependent eddy diffusivity

$$\mathsf{K}_{\text{Davis}}(\mathbf{x};\tau) = \int_{-\tau}^0 \left\langle \dot{\mathbf{X}}_i'(s) \otimes \dot{\mathbf{X}}_i'(t+s) \right\rangle_{\{\mathbf{X}_i(t)=\mathbf{x}\}} \mathrm{d}s, \tag{1.32}$$

is defined. The conditional ensemble average $\langle\cdot\rangle_{\{\mathbf{X}_i(t)=\mathbf{x}\}}$ is over all trajectories passing through the position $\mathbf{x}$ at some time t. The Davis (1987) diffusivity is widely used in oceanography studies for both synthetic and actual drifters (e.g Oh et al., 2000;

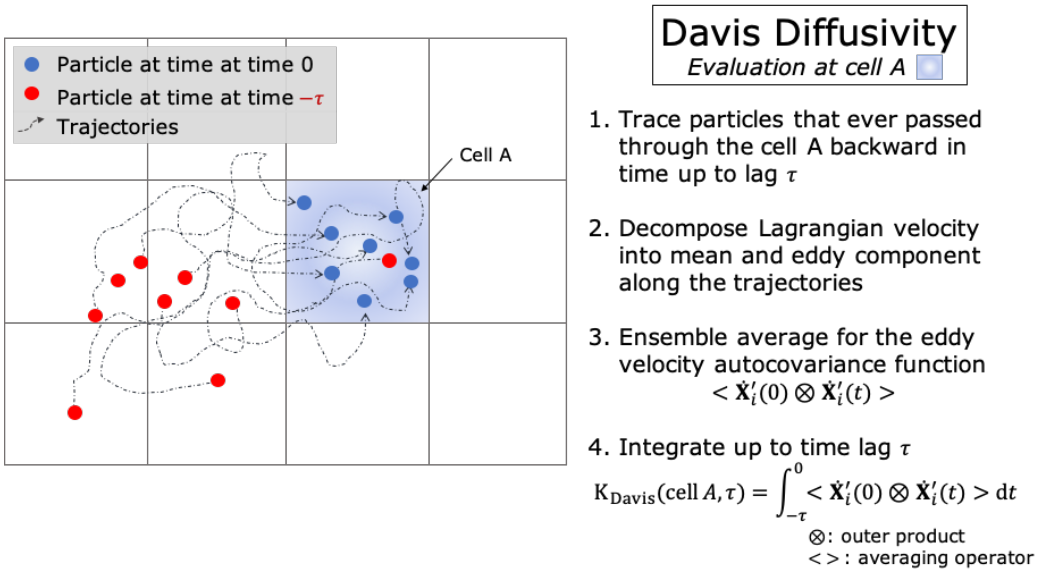

Figure 1.1: A schematic explanation of the evaluation of the Davis (1987) diffusivity in a cell from trajectory data. Note for this illustration we assume the time $t = 0$ is when particles are found in the cell.

Zhurbas and Oh, 2003, 2004; Griesel et al., 2010; Abernathey et al., 2013; Chiswell, 2013; Peng et al., 2015; Rühs et al., 2018) to quantify the spatial variability of mixing rates in different geographical areas of the ocean.

While the Davis (1987) diffusivity formally captures the spatial variations through the specific conditional averaging $\langle\cdot\rangle_{\{\mathbf{X}_i(t)=\mathbf{x}\}}$, it is not common for particle trajectories, either synthetic or actual ones, to arrive at any common point. In practice, local binning of particles in some predefined cells are necessary to approximate the conditional average by the ensemble average of all particles arriving the predefined cells. A schematic sketch of the evaluation of the Davis (1987) diffusivity is shown in figure 1.1.

Analogous to the eddy diffusivity (1.27b) derived in a rapidly varying multiscale dynamics, the Davis (1987) diffusivity at point $\mathbf{x}$ is also defined by integrating the eddy velocity autocovariance over time. A subtle difference between (1.27b) and the Davis (1987) diffusivity is that the latter uses the history of trajectories before their arrival at point $\mathbf{x}$, instead of the trajectory departed from the point.

In the coarse-grained equation (1.31) derived by Davis (1987), the diffusive flux is defined based the sum of a Fickian down-gradient term $\mathsf{K}_{\mathrm{Davis}}\nabla\bar{c}$, and two terms involving the history on tracer at point $\mathbf{x}$. If the $\mathsf{K}_{\mathrm{Davis}}$ reaches an asymptotic value quickly at small lag $\tau \ll t$, the history terms become negligible and the coarse-grained equation (1.31) reduces into the Fickian advection–diffusion equation, and can be interpreted as the eddy diffusivity defined by Taylor (1922). Real ocean drifter data does not always signal a sharp asymptotic value for $\mathsf{K}_{\mathrm{Davis}}$. Often the value of $\mathsf{K}_{\mathrm{Davis}}(\mathbf{x}, \tau)$ at specific τ, such as at the first zero-crossing of autocovariance (e.g Lumpkin et al., 2002) or at the decorrelation time (e.g. Griesel et al., 2010; Klocker et al., 2012b) is reported.

Multi-particle diffusivity

Multiple-particle diagnostics are derived from the cross-statistics of trajectory ensembles. One example is the *two-particle* relative diffusivity (see e.g. LaCasce, 2008)

$$\mathsf{K}_{\mathrm{rel}}(\tau) = \frac{1}{4}\frac{\mathrm{d}}{\mathrm{d}\tau}\left\langle (\mathbf{X}_{i_1}(\tau) - \mathbf{X}_{i_2}(\tau)) \otimes (\mathbf{X}_{i_1}(\tau) - \mathbf{X}_{i_2}(\tau)) \right\rangle, \tag{1.33}$$

where the average $\langle\cdot\rangle$ is taken over all distinct pairs ($i_1 \neq i_2$) of a set of particles. The relative diffusivity characterises the rate of spreading between pairs of particles and coincides with the absolute diffusivity in the long time limit, if the flow is homogeneous and statistical stationary. If, furthermore, the flow is incompressible, Bennett (1984) shows that the relative diffusivity scales with the separation scale D by $\mathsf{K}_{\mathrm{rel}} \sim D^{(\alpha+1)/2}$, where α is the slope in the power-law of the turbulent fluid energy spectrum, which makes the relative diffusivity a tool to analyse the energy transfer across scales (e.g. Koszalka et al., 2009; Lumpkin and Elipot, 2010; Sansón et al., 2017).

Before concluding this section, we highlight that novel diffusivity estimates are actively being developed. Some examples include Wolfram and Ringler (2017), which develops a Lagrangian tracking method to emulate the contour evolution of passive tracer and estimate the effective diffusivity (Marshall et al., 2006); Ivanov and Chu (2019) apply wavelet analysis techniques in signal processing to extract the length and time scales required for an eddy diffusivity in mixing length theory. In this thesis, we also formulate a novel and distinctive Lagrangian approach to infer the eddy diffusivity.

We highlight that the Eulerian and Lagrangian diagnostics are not necessarily equivalent, although one may expect some are equivalent in certain limits. The two types of diffusivity diagnostics generally agree in the order of magnitude (Klocker et al., 2012b; Qian et al., 2019). It is also important to note that each existing diffusivity diagnostic is defined for its purpose, and not all would ultimately point to a coarse-grained advection–diffusion equation, where the eddy diffusivity is defined to quantify the unresolved eddy transport.

1.5 Quasigeostrophic double gyre

This thesis concerns the development of a novel Bayesian inference to diagnose eddy diffusivity in the ocean. To test the inference in an oceanographically relevant context, Lagrangian trajectories in a quasigeostrophic double gyre paradigm are considered.

1.5.1 Quasigeostrophic dynamics

The quasigeostrophic equations are a set of simplified equations for ocean dynamics. They capture the near-geostrophic dynamics of large-scale circulations, that is, flows with pressure gradient dominantly balanced by Coriolis force. A small Rossby number $\mathrm{Ro} = U/fL$ justifies the use of the quasigeostrophic approximation, where U is the velocity scale, f is the latitude-dependent Coriolis parameter and L is the length scale. Another assumption for the quasigeostrophic equations is that the depth variation of the fluid should be small relative to the mean depth. This is justified by having the scale of motion L to be comparable with the deformation radius $L_d = \sqrt{gH}/f$, where g is the gravitational acceleration and H is the mean depth of the fluid. We refer readers to Pedlosky (1987); Vallis (2006) for the details of quasigeostrophic dynamics.

The mesoscale dynamics in mid-latitudes are well-captured by quasigeostrophic equations, where the Rossby number is $\mathrm{Ro} \sim 0.01$ and deformation radius $L_d \sim 100$ km.

The quasigeostrophic equations are defined on a tangential Cartesian plane projected from Earth surface. At the mid-latitudes, the variations of Coriolis parameter f is small relative to the length scale of the mesoscale eddies. This enables a β-plane approximation to the Coriolis parameter, in which the Coriolis parameter $f = f_0 + \beta y$ is linearised with respect to the meridional coordinate y, on top of a reference f_0 and a meridional gradient β. A detailed justification of the β-plane approximation can be found in Chapter 2 of Pedlosky (1987).

In this thesis a three-layer quasigeostrophic model is considered. The thickness H_i of each layer increases from top to bottom, where the subscript $i = 1$ to 3 is the top to bottom layer. Wind forcing and bottom friction are respectively imposed in the top layer and bottom layer to drive and dampen the fluid flow, enabling the fluid to reach a forced-dissipative statistically steady state.

The quasigeostrophic equations are a set of prognostic equation describing the time evolution of the potential vorticity q_i, a dynamically conserved quantity

$$\partial_t q_i + \nabla \cdot (\mathbf{u}_i q_i) = \nu \nabla^2 \omega_i - r \delta_{i,3} \omega_i + \delta_{i,1} Q_w \quad \text{for } i \in \{1, 2, 3\}\,, \tag{1.34}$$

where the potential vorticity q_i is related to the streamfunction ψ_i by the diagnostic elliptic equations

$$q_1 = \nabla^2 \psi_1 + \beta y + s_1^+ (\psi_2 - \psi_1), \tag{1.35a}$$
$$q_2 = \nabla^2 \psi_2 + \beta y + s_2^+ (\psi_3 - \psi_2) + s_2^- (\psi_1 - \psi_2), \tag{1.35b}$$
$$q_3 = \nabla^2 \psi_3 + \beta y + s_3^- (\psi_2 - \psi_3). \tag{1.35c}$$

The velocity $\mathbf{u}_i$ and relative vorticity ω of the flow are solely determined from ψ_i by

$$\mathbf{u}_i = \nabla^\perp \psi_i = (-\partial_y \psi_i, \partial_x \psi_i)^{\mathrm{T}} \qquad \text{and} \qquad \omega_i = \nabla^2 \psi_i,$$

interlinking the prognostic equation (1.34) and the diagnostic equation (1.35).

The left-hand side of the prognostic equation (1.34) is the material derivative of the potential vorticity $\mathrm{D}_t q_i$. The right-hand side is a net of viscous dissipation term $\nu \nabla^2 \omega_i$ controlled by the Laplacian viscosity coefficient ν, a bottom friction $r\omega_i$ controlled by an inverse time scale r and a prescribed surface wind forcing Q_w.

The diagnostic elliptic equations (1.35) define the underlying relation between the potential vorticity q_i and the streamfunction ψ_i. The potential vorticity is a sum of relative vorticity $\omega_i = \nabla^2 \psi_i$, referenced planetary vorticity βy and stretched vorticity $s_i^\pm (\psi_{i\pm1} - \psi_i)$. The stratification parameter $s_i^\pm = f_0^2 / (H_i g_{i\pm1/2})$ is a proportionality constant that links the cross-layer difference in streamfunction (hence the fluid layer thickness) to the stretched vorticity. In practice often what is given is knowledge of the instantaneous potential vorticity q_i, and the streamfunctions ψ_i are inverted from q_i, a process known as potential vorticity inversion.

On the domain boundary, partial-slip conditions (Haidvogel et al., 1992) $\alpha \nabla^2 \psi_i = -\nabla \psi_i \cdot \hat{\mathbf{n}}$ are applied, with $\hat{\mathbf{n}}$ being an outward unit normal on the boundary, and $1/\alpha = 120$ km being a partial-slip length scale. This choice of boundary consideration is identical to similar simulations in Karabasov et al. (2009); Maddison et al. (2015); Mak et al. (2016).

1.5.2 Double gyre circulation

A double gyre circulation is considered for the quasigeostrophic model. The well-studied configuration (e.g. Berloff et al., 2007; Karabasov et al., 2009; Marshall et al.,

Parameter	Symbol	Value(s)
Domain size	L	$3,840$ km
Top layer mean depth	H_1	0.25 km
Middle layer mean depth	H_2	0.75 km
Bottom layer mean depth	H_3	4 km
Laplacian viscosity coefficient	ν	100 m^2 s^{-1}
Bottom friction inverse time scale	r	4×10^{-8} s^{-1}
Meridional planetary vorticity gradient	β	2×10^{-11} m^{-1}s^{-1}
Stratification parameters	$s_1^+ H_1 = s_2^- H_2$	2.965×10^{-7} m^{-1}
	$s_2^+ H_2 = s_3^- H_3$	5.603×10^{-7} m^{-1}
Wind stress	τ_0	0.08 N m^{-2}
Reference density	ρ_0	1000 kg m^{-3}
Partial-slip length scale	$1/\alpha$	120 km
Time step	t_{QG}	$1,800$ s

Table 1.1: Parameters used in the idealised quasigeostrophic double gyre simulations.

2012; Maddison et al., 2015; Haigh et al., 2020) leads to a flow consisting of a western boundary current that detaches to form a strong jet and eddies.

Sitting in a square domain of length L with a flat bottom, an asymmetrical wind forcing profile is applied to the top layer of the quasigeotrophic model to drive the circulation. The profile is given by (also equation (26) Maddison et al., 2015)

$$Q_w = \begin{cases} -\frac{2\pi\tau_0}{\rho_0 H_1 L} A \sin\left(\pi \frac{L/2+y_v}{L/2+y_m}\right) & \text{if } y_v < y_m \\ \frac{2\pi\tau_0}{\rho_0 H_1 L} \frac{1}{A} \sin\left(\pi \frac{y_v-y_m}{L/2-y_m}\right) & \text{otherwise} \end{cases}, \tag{1.36}$$

where $A = 0.9$, $B = 0.2$, $y_v = y - L/2$ and $y_m = B(x - L/2)$. A partial-slip boundary condition $\alpha\nabla^2\psi_i = -\nabla\psi_i \cdot \hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is an outward unit normal on the boundary, is applied. The parameter $1/\alpha$ is a partial-slip length scale that controls not only the width of the boundary current, but also the strength of the jet, the geometry of the gyres and the energetics of the system (Haidvogel et al., 1992). A widely used and tested value of α is applied here. A full list of the physical parameters is given in Table 1.1.

The circulation is characterised by an eastward meandering jet at the interface of a sub-polar cyclonic gyre and a sub-tropical anticyclonic gyre (left panel of figure 1.2). Around the intensified western boundary current and eastward jet, well-developed eddies are continuously produced and dissipated (middle panel of figure 1.2). Despite a reported decadal variability mode (Berloff et al., 2007), at a lower time scale the flows can be seen as statistical stationary. The eddy kinetic energy concentrates around the eastward jet and the western boundary (right panel of figure 1.2).

The quasigeostrophic equations are solved using a finite difference method with details given in Appendix C.2.

1.6 Plan of the thesis

The central theme of this thesis is a novel Lagrangian diffusivity diagnostic based on Bayesian inference. We take a data-driven approach to seek the possible mean velocity $\mathbf{U}(\mathbf{x})$ and eddy diffusivity $\mathsf{K}(\mathbf{x})$ implied by the trajectory data, assuming the particle

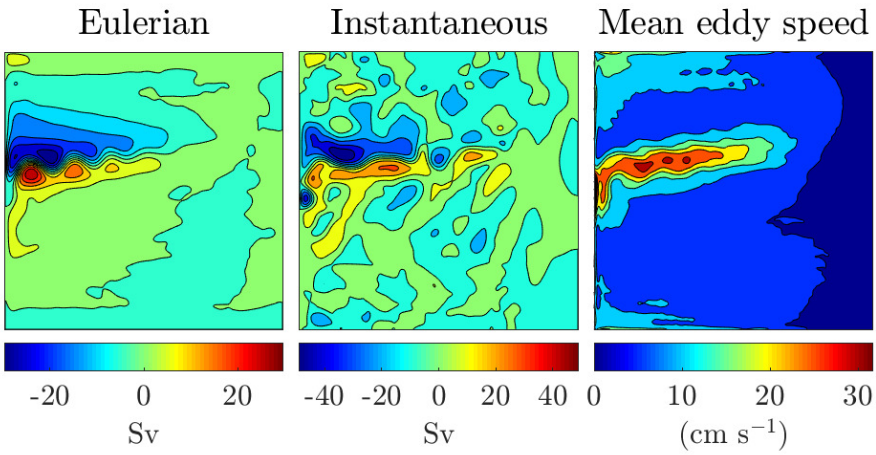

Figure 1.2: Left panel: middle layer 10-year Eulerian time-averaged streamfunction, multiplied by the layer depth. Middle panel: an instantaneous middle layer streamfunction, multiplied by layer depth. Right panel: the magnitude of 10-year time-averaged eddy velocity.

position $\mathbf{X}$ is governed by the stochastic differential equation

$$\mathrm{d}\mathbf{X} = \left[\mathbf{U}\left(\mathbf{X}\left(t\right)\right) + \nabla \cdot \mathsf{K}\left(\mathbf{X}\left(t\right)\right)\right]\mathrm{d}t + \sqrt{2\,\mathsf{K}\left(\mathbf{X}\left(t\right)\right)}\mathrm{d}\mathbf{W},$$

whose Fokker–Planck equation coincides with the advection–diffusion equation (see 1.2.3)

$$\frac{\partial \pi}{\partial t} + \nabla \cdot (\mathbf{U}\pi) = \nabla \cdot (\mathsf{K}\nabla\pi).$$

Note that Fickian diffusion $\nabla \cdot (\mathsf{K}\nabla\pi)$ leads to the up diffusivity gradient $\nabla \cdot \mathsf{K}$ drift in the stochastic differential equation.

Bayesian inference establishes a *posterior distribution*, enabling a probabilistic estimate of $\mathbf{U}$ and K simultaneously. The joint probability posterior distribution assign a value for each combination of $\mathbf{U}$ and K to quantify its plausibility. This also provides a quantification of the uncertainty in the estimation. The method and the theoretical basis of the Bayesian inference are explained in Chapter 2, *Bayesian inference for stochastic differential equations*, with a dedicated discussion on inferring from time-series data.

Chapter 3, *Inference applied to cellular flows*, applies the Bayesian approach to infer the eddy diffusivity in a multiscale cellular flow. The results are compared against the rigorous estimates derived from homogenisation theory, enabling strengths and weaknesses of the inference scheme to be identified. This chapter also verifies, as pointed out by literature, the necessity to downsample the trajectory data at a sufficiently long time interval for the inference to correctly identify the eddy diffusivity.

Chapter 4, *Localised inference applied to an idealised oceanic flow*, presents a first attempt to infer ocean eddy diffusivity in the idealised quasigeostrophic double gyre. Due to computational constraints, the Bayesian method developed in Chapter 2 cannot be applied directly to an oceanic flow. This chapter develops a localised inference, by binning the trajectory data and the ocean, to infer a cell-wise mean velocity and

diffusivity field. This approach gives an explicit formula for the posterior distribution and results in a computationally feasible inference. However, the localised treatment is also its bottleneck. At the time scale where eddy diffusivity becomes relevant, particle trajectories may not remain confined within a cell, hindering the general applicability of the method.

Chapter 5, *Global inference applied to idealised oceanic flows*, explores a new approach to overcome the locality limitation in the localised inference presented in Chapter 4. A data coarse-grained inference is proposed that discards information of exact particle positions in the trajectory data. While the data coarse-grained approach also requires the trajectory data to be binned to local cells, the fields inferred from them are globally defined, so that particles transported across cells can be accounted for. This approach drastically reduces the number of partial differential equations that need to be solved in evaluating the posterior distribution and results in a computationally feasible inference. We show the validity of the inference is not significantly affected by information discarded when sufficient particle trajectory data are supplied. The data coarse-grained approach is applied to a synthetic test case and the double gyre flow considered in Chapter 4. The inference results are validated against the ground truth in the synthetic test case and tracer experiments in the double gyre flow.

Chapter 6, *Overcoming non-locality – a large-deviation approach*, takes an analytical route to reduce the computational burden in an inference. This chapter applies large deviation theory to derive an explicit formula to approximate the posterior distribution, while allowing particles to be transport across areas with distinct flow characteristics. The explicit formula allows the Bayesian inference to be efficiently performed. The method is tested on synthetic piecewise constant velocity and diffusivity fields on two sub-planes and shows a promising potential to be adapted for inference in the ocean.

Results in chapter 3 and 4 have been published in Ying et al. (2019).

Chapter 2

Bayesian inference for stochastic differential equations

2.1 The Bayesian framework for parameter estimation

Bayesian inference is a paradigm to perform statistical inference, from which properties of the probability distribution for a set of model parameters are deduced. The Bayesian paradigm combines a *prior* distribution of a priori knowledge and a *likelihood* function resulted from a statistical model to assign a probability measure, conditioned on the observed data, on unknown parameters of the model. The resulting distribution, known as the *posterior*, follows from Bayes' theorem and is proportional to the product of the prior and likelihood function.

Parameter estimation is a typical area where Bayesian inference is applied. Knowing that a set of observed data R is governed by a statistical model parameterised by some unknown finite-dimensional $\boldsymbol{\theta} \in \mathbb{R}^n$, Bayesian inference attempts to estimate the parameter $\boldsymbol{\theta}$ through the posterior distribution

$$p\left(\boldsymbol{\theta}|R\right) \propto p\left(R|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right), \tag{2.1}$$

with a prior $p\left(\boldsymbol{\theta}\right)$ and likelihood $p\left(R|\boldsymbol{\theta}\right)$. The prior distribution reflects existing and subjective knowledge about the plausibility of a given set parameter $\boldsymbol{\theta}$ in the absence of data. Its relative importance is expected to diminish as the number of data points in R increases. The likelihood, set by the statistical model, measures the probability of reproducing the data R at a given $\boldsymbol{\theta}$. The posterior gathers information from the prior and likelihood to construct a probability distribution of the parameter $\boldsymbol{\theta}$. Instead of producing single-valued *point estimates* for $\boldsymbol{\theta}$ in the parameter space, the Bayesian approach sees all $\boldsymbol{\theta}$ as potential candidates and accounts for the plausibility of each through a probability measure.

Point estimates to $\boldsymbol{\theta}$ can be computed from the posterior distribution. Two widely used summary statistics are the posterior mean $\overline{\boldsymbol{\theta}}$ and the maximum a posteriori (MAP) $\boldsymbol{\theta}_{\mathrm{MAP}}$ estimates, which are respectively

$$\overline{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p\left(\boldsymbol{\theta}|R\right) \mathrm{d}\boldsymbol{\theta}, \tag{2.2}$$

and

$$\boldsymbol{\theta}_{\mathrm{MAP}} = \arg \sup_{\boldsymbol{\theta}} \; p\left(\boldsymbol{\theta}|R\right). \tag{2.3}$$

The posterior mean is the expectation of $\boldsymbol{\theta}$ under the posterior and the MAP estimate is the candidate that is most plausible under the posterior distribution.

The probabilistic inference also enables quantification of uncertainty. There are a large class of methods to quantify the plausibility of some estimates. A few examples include a credible region construction, posterior covariance evaluation and posterior modality examination. The credible region $D(\alpha)$ is the domain where $\boldsymbol{\theta}$, if distributed as the posterior, falls with a given probability α, that is,

$$D(\alpha): \int_{D(\alpha)} p\left(\boldsymbol{\theta}|R\right) \mathrm{d}\boldsymbol{\theta} = \alpha. \tag{2.4}$$

The credible region is uniquely determined only when additional constraints are imposed and is usually applied to low-dimensional parameter space. A typical example is to construct a 95% credible interval from the marginal posterior of a parameter component by aligning its MAP estimates at the interval centre.

The posterior covariance $\mathrm{Cov}(\boldsymbol{\theta}, \boldsymbol{\theta})$ is the covariance matrix of $\boldsymbol{\theta}$ under the posterior, defined as

$$\mathrm{Cov}(\boldsymbol{\theta}, \boldsymbol{\theta}) = \int \left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\right) \otimes \left(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\right) p\left(\boldsymbol{\theta}|R\right) \mathrm{d}\boldsymbol{\theta}, \tag{2.5}$$

here $\otimes$ is the vector outer product. The posterior covariance gives an estimate of the spread of the posterior around its mean (diagonal terms) and the correlations among components of $\boldsymbol{\theta}$ (off-diagonal terms).

The landscape of the posterior also controls the uncertainty. As the posterior may contain multiple local maxima in its parameter space, an examination on the modality offers insights on both the suitability and stability of a point estimate to $\boldsymbol{\theta}$.

Standard reference for the Bayesian statistics and inverse problems can be found in Robert and Casella (2005); Kaipio and Somersalo (2005).

2.2 Computational sampling

2.2.1 Integral evaluation

With Bayes' formula (2.1), the posterior distribution is abstractly defined on a parameter space and is determined up to a normalisation constant. In practice, often of interest is the integrals of various quantities against the posterior – that is, the evaluation of

$$\mathbb{E}\left(f\left(\boldsymbol{\theta}\right)\right) = \int f\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|R\right) \mathrm{d}\boldsymbol{\theta}, \tag{2.6}$$

for some quantity of interest $f\left(\boldsymbol{\theta}\right)$. Examples include the posterior mean (2.2) and the posterior covariance (2.5).

Only a small class of posterior allows analytical and closed-form computation of the integral (2.6). While quadrature rules may apply in low-dimensional parameter space, the number of grid points required increases geometrically with dimension size, making the brute force approach computationally infeasible for a high-dimensional parameter space. Moreover, the normalisation constant of the posterior needs to be determined, which is as computationally intensive as evaluating the quantity of interest.

To overcome the computational challenge, sampling methods are more often used to evaluate the quantity of interest. With a numerical algorithm to generate samples $\boldsymbol{\theta}^{(k)} : k = 1, 2 \ldots$, collectively distributed under the posterior $p\left(\boldsymbol{\theta}|R\right)$, the sample mean of $f(\boldsymbol{\theta}^{(k)})$ approximates the integral (2.6) as sample size increases. This transforms the

integral evaluation into a sampling problem, where t suffices to simulate a large number of samples under a target distribution $p(\boldsymbol{\theta}|R)$.

There are a variety of sampling algorithms. Commonly used algorithms include the rejection sampler and importance sampler, where the samples under a known proposal distribution are discarded or reweighted to mimic the target distribution. These methods are simple to implement and are backed by formal convergence analysis. However, they require knowledge of the posterior normalisation constant and their efficiencies heavily depend on the chosen proposal distribution. Details and analysis of these algorithms can be found in (e.g. Ross, 2006; Asmussen and Glynn, 2007).

Another class of sampling algorithm is the Markov chain Monte Carlo methods. The underlying principle is to simulate a Markov chain, governed by a transition kernel $T(\boldsymbol{\theta}^{(k+1)}|\boldsymbol{\theta}^{(k)})$, whose stationary distribution coincides with the target posterior distribution. The sample mean of a long record of sample states $\{\boldsymbol{\theta}^{(k)} : k = 1, 2 \dots\}$ can be used to estimate the integrals in (2.6).

2.2.2 The Metropolis–Hastings algorithm

A commonly used Markov chain Monte Carlo method is the *Metropolis–Hastings* algorithm (see also a thorough introduction by Liu, 2008). The algorithm constructs a transition kernel $T(\boldsymbol{\theta}'|\boldsymbol{\theta})$ that satisfies the detailed balance condition for any $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$

$$T(\boldsymbol{\theta}'|\boldsymbol{\theta})p(\boldsymbol{\theta}|R) = T(\boldsymbol{\theta}|\boldsymbol{\theta}')p(\boldsymbol{\theta}'|R), \tag{2.7}$$

so that the Markov chain is guaranteed to converge to a stationary distribution given by $p(\cdot|R)$. To achieve this, an independent proposal transition kernel $P(\boldsymbol{\theta}'|\boldsymbol{\theta})$ is composited with an acceptance/rejection decision to form a transition kernel $T(\boldsymbol{\theta}'|\boldsymbol{\theta})$. After generating a sample $\boldsymbol{\vartheta}$ from the proposal transition kernel $P(\boldsymbol{\vartheta}|\boldsymbol{\theta})$, the proposed $\boldsymbol{\vartheta}$ is accepted to be a new sample $\boldsymbol{\theta}'$ with a probability

$$\alpha = \begin{cases} \frac{p(\boldsymbol{\vartheta}|R)}{p(\boldsymbol{\theta}|R)} \frac{P(\boldsymbol{\theta}|\boldsymbol{\vartheta})}{P(\boldsymbol{\vartheta}|\boldsymbol{\theta})}, & \text{if } p(\boldsymbol{\vartheta}|R)P(\boldsymbol{\theta}|\boldsymbol{\vartheta}) \leq p(\boldsymbol{\theta}|R)P(\boldsymbol{\vartheta}|\boldsymbol{\theta}) \\ 1, & \text{otherwise,} \end{cases}$$

so that $T(\boldsymbol{\vartheta}|\boldsymbol{\theta}) = P(\boldsymbol{\vartheta}|\boldsymbol{\theta})\alpha$ and satisfies the detailed balance condition (2.7). Note that the definition of the acceptance probability α involves only the ratio of posterior value at $\boldsymbol{\vartheta}$ and $\boldsymbol{\theta}$, and is thus independent of the proportional constant in Bayes' formula (2.1). This makes the algorithm particularly suitable for Bayesian inference computation.

The algorithm is outlined as follows

1. Set $k = 0$. Choose a proposal density $P(\cdot|\cdot)$ and take an initial value for the parameter $\boldsymbol{\theta}^{(0)}$.

2. Iterate:

   (a) randomly draw a candidate parameter $\boldsymbol{\vartheta}$ with probability $P(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{(k)})$,

   (b) compute the proposal transition kernel $P(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{(k)})$ and $P(\boldsymbol{\theta}^{(k)}|\boldsymbol{\vartheta})$,

   (c) compute the posterior value $p(\boldsymbol{\theta}^{(k)}|R)$ and $p(\boldsymbol{\vartheta}|R)$ (up to an irrelevant proportionality constant) from Bayes' formula (2.1)

   (d) let

$$\boldsymbol{\theta}^{(k+1)} = \begin{cases} \boldsymbol{\vartheta} & \text{with probability } \alpha, \\ \boldsymbol{\theta}^{(k)} & \text{with probability } 1 - \alpha, \end{cases} \tag{2.8}$$

where

$$\alpha = \min\left(1, \frac{p(\boldsymbol{\vartheta}|R)}{p(\boldsymbol{\theta}^{(k)}|R)} \frac{P(\boldsymbol{\theta}^{(k)}|\boldsymbol{\vartheta})}{P(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{(k)})}\right), \tag{2.9}$$

(e) increment $k \mapsto k+1$.

Typically the proposal transition kernel $P(\boldsymbol{\vartheta}|\boldsymbol{\theta})$ is chosen so that it is cost-effective to simulate. A classical choice for the proposal kernel is a Gaussian distribution with mean $\boldsymbol{\theta}$ and fixed covariance matrix, which leads to the *random–walk Metropolis–Hastings*. Sampling from a Gaussian distribution is straightforward using standard methods such as the Box-Muller transform or inverse transform. A Gaussian proposal also has the advantage that it is symmetric, that is $P(\boldsymbol{\vartheta}|\boldsymbol{\theta}) = P(\boldsymbol{\theta}|\boldsymbol{\vartheta})$, which simplifies the evaluation of the acceptance probability α in (2.9).

A limitation to the random-walk Metropolis–Hastings is that the proposal Gaussian distribution needs to be tuned for optimal efficiency. Small covariance will lead to highly correlated samples at successive steps while big covariance will lead to frequent rejection of the proposal. Formally neither of these affects the validity of the algorithm, in practice, they lead to a slow exploration of the landscape of the posterior and inefficient use of computational resources. A rule-of-thumb, suggested by Roberts et al. (1997) in idealised studies for sampling a multivariate Gaussian distribution, is that the acceptance ratio of a proposal should be approximately 0.234 for optimal efficiency.

Another undesirable property of the random-walk Metropolis–Hastings is the dimensional dependence of the optimal proposal covariance. To achieve the desired acceptance ratio, the proposal covariance would converge to zero as the dimension of the parameter space increases (Roberts et al., 1997). This means the landscape of posterior will necessarily be explored slower as dimension grows, even at the optimal acceptance ratio. As the parameter spaces considered in this thesis come with a dimension ranging from 5 to more than $1,000$, it is unlikely that the random-walk Metropolis–Hastings would suffice in all considered cases.

One sampling algorithm to eliminate the dimensional dependence is the preconditioned Crank–Nicolson (pCN) algorithm (Cotter et al., 2013; Vollmer, 2015; Hosseini, 2019). The algorithm assumes a Gaussian prior with a predetermined covariance C_0. Samples are proposed under the Gaussian distribution centred at $\sqrt{1-\beta^2}\boldsymbol{\theta}^{(k)}$ and covariance $\beta\mathsf{C}_0$, characterised by a free step-size parameter β. This choice of proposal, which refers to the prior distribution, turns out to scale independently of the dimension (Cotter et al., 2013; Hairer et al., 2014). For a posterior possessing a non-Gaussian prior, one can re-weight the posterior density over a Gaussian one (assuming absolute continuity) to obtain an adapted posterior that is defined with respect to a Gaussian measure. This will involve choosing a suitable mean and covariance C_0, which ideally approximate the adapted posterior for computational efficiency. For the inference considered in this thesis, we always consider uniform prior on the parameter θ to confine θ in a range to satisfy physical constraints. Knowledge of the range and spread of the posterior is a priori deficient so it is difficult and speculative to choose the reweighting Gaussian density. As this thesis chiefly serves as a viability assessment of the Bayesian approach to infer diffusivity, we decided not to pursue the pCN algorithm to step away from this highly technical and potentially misleading tuning procedures. Instead, we apply simpler and provably robust random-walk sampling algorithm, despite the undesirably excessive computational cost.

To address the high-dimensionality, we apply a special form of random–walk Metropolis–Hastings, where updates to individual components in the parameter $\boldsymbol{\theta}$ are

proposed sequentially, using a one-dimensional Gaussian proposal distribution with mean at the current value of the component. Such a sampling strategy is known as *Metropolis–within–Gibbs*, *single-site update* or *Gibbs sampler* (e.g. Metropolis et al., 1953; Roberts and Rosenthal, 2006). Specifically, the proposal density $P(\boldsymbol{\vartheta}|\boldsymbol{\theta})$ is taken as

$$P(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{(k)}) = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{\sqrt{2\pi V_j}}\exp\left(-\frac{\left(\vartheta_j-\theta_j^{(k)}\right)^2}{2V_j}\right)\prod_{i\neq j}\delta\left(\theta_i^{(k)}-\vartheta_i\right), \tag{2.10}$$

where $j = 1, \cdots, J$ labels the components of $\boldsymbol{\theta}$ and the variances V_j are tuned individually for each component. This strategy eases the tuning procedure at the expense of producing highly correlated consecutive samples. Nonetheless, numerical experiments show this strategy works well for the inference in this thesis.

2.2.3 Burn-in and tuning

A *burn-in* process needs to be taken before working with the generated samples. Since only the stationary distribution of the Markov chain converges to the target posterior, initial states of samples generated from the algorithm are typically discarded as they often over-represent the region of low posterior density. Details on the thresholds for 'burn-in' are given in the following chapters alongside with the application.

Upon the completion of the burn-in, we proceed to tune the variances V_j in the Gaussian proposal distribution. The sample $\boldsymbol{\theta}_{\mathrm{MAP}}$ that maximises the posterior density $p(\boldsymbol{\theta}|R)$ during the 'burn-in' is used as the initial condition for tuning V_j and post-'burn-in' sampling. To tune V_j, the algorithm is re-run with a certain number of additional steps, during which the fraction of $\boldsymbol{\theta}^{(k)}$ accepted is recorded. If the acceptance fraction exceeds 0.5, V_j is multiplied by 4/3; if it is lower than 0.15, V_j is multiplied by 2/3. The tuning process is repeated for certain number of times and stops once the acceptance fraction falls in the range of $[0.15, 0.50]$.

With the tuned variance V_j the Metropolis–Hasting algorithm is re-run with an initial condition $\boldsymbol{\theta}_{\mathrm{MAP}}$ and the samples $\boldsymbol{\theta}^{(k)}$ thereafter are used for inference.

2.2.4 Gelman–Rubin convergence test

To determine the number of MCMC steps needed for the sample distribution of $\boldsymbol{\theta}^{(k)}$ to converge to the target posterior $p(\boldsymbol{\theta}|R)$, the Gelman–Rubin convergence test (Gelman and Rubin (1992); Brooks and Gelman (1998), also section 11.4 of Gelman et al. (2013)), which compares multiple chains of $\boldsymbol{\theta}^{(k)}$ under different initial conditions $\boldsymbol{\theta}^{(0)}$, is performed.

The diagnostic attempts to determine the number of samples N_{mh} needed to approximate the posterior $p(\boldsymbol{\theta}|R)$ robustly. This is done by shooting the Metropolis–Hastings algorithm with multiple initial conditions and comparing the summary statistics among different sets of drawn samples.

Suppose M different Markov chains with identical length N_{mh} are generated in the MCMC algorithm using distinct initial states $\boldsymbol{\theta}^{(0)}$. For simplicity, the convergence of the Markov chain of the multivariate $\boldsymbol{\theta}$ is assessed in a component-wise manner. In other words, each component of $\boldsymbol{\theta}$ is independently assessed and an overall convergence is said to be achieved when all components pass the convergence test.

Denote $\overline{\theta}^{(m)}$ and $s^{(m)}$ to be the sample mean and variance of a component in the m-th Markov chain. Define the overall sample mean $\overline{\theta}$, between-chain variance B and

within-chain variance W:

$$B = \frac{1}{M-1}\sum_{m=1}^{M}\left(\overline{\theta}^{(m)} - \overline{\theta}\right)^2, \text{ where } \overline{\theta} = \frac{1}{M}\sum_{m=1}^{M}\overline{\theta}^{(m)},$$
$$W = \frac{1}{M}\sum_{m=1}^{M} s^{(m)}.$$

The potential scale reduction factor (PSRF) R_c is defined by the root ratio between the pooled variance $(1 - 1/N_{\text{mh}})B + W$ and W :

$$R_c = \sqrt{\left(1 - \frac{1}{N_{\text{mh}}}\right) + \frac{B}{W}}. \tag{2.12}$$

The Markov chains are said to have converged if R_c is close enough to 1. In this thesis, we choose the cut off at 1.2 recommended by Brooks and Gelman (1998).

2.2.5 Alternative sampling algorithms

In section 2.2.2 we introduce the random–walk Metropolis–Hastings algorithm. The rule of thumb acceptance ratio ≈ 0.234 indicates that more than three-quarters of proposed samples are discarded. Should each proposal be computationally expensive, such low acceptance ratio implies a huge effort to generate independent samples.

Novel sampling algorithms are actively developed to improve efficiency. In this section, we outline the principles of two alternative algorithms. While these algorithms are not applied in this thesis (for reasons to be explained later in this section), it is worthwhile to keep in mind of them as they open up new avenues for further research.

Metropolis-adjusted Langevin algorithm

One commonly applied sampling algorithm is based on the overdamped Langevin dynamics, governed by the stochastic differential equation

$$\mathrm{d}\boldsymbol{\theta} = \frac{1}{2}\nabla \log p(\boldsymbol{\theta}|R)\,\mathrm{d}t + \mathrm{d}\mathbf{W}. \tag{2.13}$$

The invariant distribution of $\boldsymbol{\theta}(t)$ is precisely the target posterior $p(\cdot|R)$, so that the long-time trajectory of $\boldsymbol{\theta}(t)$ can be seen as samples from the posterior.

In theory, all position records in the trajectory of the Langevin dynamics (2.13) can be accepted as samples. In practice, numerical simulations of the Langevin dynamics typically involve errors when marching the system with a finite time step.

The *Metropolis-adjusted Langevin algorithm* (MALA) numerically solves the Langevin dynamics using a Euler–Maruyama method at a fixed time step. At each time step, the proposed position update is subject to a probabilistic decision, with an acceptance probability determined by (2.9). This ultimately retains the detailed balance of the simulated trajectory and ensures convergence into the correct stationary distribution. For optimal sampling efficiency of the MALA, the step size in the Euler–Maruyama stepping should be tuned such that the acceptance ratio is ≈ 0.574 (Roberts and Rosenthal, 1998), considerably higher than the 0.234 in random–walk Metropolis–Hastings.

Hamiltonian Monte Carlo

Another set of algorithms is based on Hamiltonian dynamics. Define a Hamiltonian by

$$\mathcal{H}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \log(p(\boldsymbol{\theta}|R)) + \frac{1}{2}\boldsymbol{\phi}^{\mathrm{T}}\mathsf{M}^{-1}\boldsymbol{\phi},$$

with an auxiliary 'momentum' $\boldsymbol{\phi}$ sharing the same dimension as the parameter $\boldsymbol{\theta}$ and an arbitrary symmetric positive definite matrix M. Consider the Hamiltonian dynamics

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = -\nabla_{\boldsymbol{\phi}}\mathcal{H}, \qquad \frac{\partial \boldsymbol{\phi}}{\partial t} = +\nabla_{\boldsymbol{\theta}}\mathcal{H}. \tag{2.14}$$

The marginal invariant distribution over the trajectory $\boldsymbol{\theta}(t)$ coincides with the target posterior $p(\boldsymbol{\theta}|R))$, hence it suffices to simulate a long-time trajectory $\boldsymbol{\theta}(t)$ to sample the target posterior.

The *Hamiltonian Monte Carlo* (HMC) algorithm numerically simulates the Hamiltonian system (2.14). This can be performed by, for example, a leapfrog time-stepping scheme, usually also supplemented with an operator splitting scheme (see e.g. Leimkuhler and Reich, 2004). Convergence to the consistent stationary distribution is again guaranteed by an acceptance probability (2.9). For optimal sampling efficiency of the HMC, the step size in the leapfrog stepping should be tuned such that the acceptance ratio is ≈ 0.651 (Beskos et al., 2013), considerably higher than that of random–walk Metropolis–Hastings and MALA.

Both MALA and HMC are widely applied. However, their usage requires the gradient of posterior to be computed. For the Bayesian inference problems considered in this thesis, the gradient of the posteriors typically lacks analytical expression. While the gradients could be numerically evaluated using the algorithmic differentiation toolkit, this will involve substantial testing and validation of the implementation. As in this thesis, the main subject is the development and quality assessment of the inference, we choose to apply the less efficient yet simpler random–walk sampler throughout. Should the inference be proven to work, more sophisticated samplers can be applied to boost the computational efficiency.

2.3 Bayesian inference for stochastic differential equations

From Chapter 3 to 6, we apply the paradigm of Bayesian parameter estimation discussed in the preceding sections to infer the flow properties from time-series data of Lagrangian trajectories. In this section, we specify the data, the stochastic model and hence the full inference problem.

2.3.1 Data

Lagrangian time-series data are considered for all inference problem in this thesis. In particular, the time-series data is a set of N distinguishable sequences of particle positions $\mathbf{X}_i(t)$, passively evolving in time following a deterministic or stochastic dynamical regime, and are observed at a P different discrete-time t_j. The data set R can be summarised as

$$R = \{(i, \mathbf{X}_i(t_j), t_j) : i = 1, \cdots, N,\ j = 1, \cdots, P\}, \tag{2.15}$$

here $\mathbf{X}_i(t_j)$ is the position of the ith particle at the jth observation time. The position data are assumed to be exact without any measurement error.

Oceanic Lagrangian time-series are obtained with sophisticated devices deployed in the ocean, which follow the ocean flow and report their positions at times. Examples include the drifters deployed in ARGO (Argo, 2000) or the global drifter program (Laurindo et al., 2017). In this thesis we consider synthetic Lagrangian trajectories generated from numerical simulations only.

2.3.2 Stochastic models

There are multiple stochastic models for Lagrangian trajectories in the ocean with different strengths and levels of complexity. Berloff and McWilliams (2002) assess a hierarchy of Langevin-typed stochastic differential equations to model the long-term dispersion of particles. Borgas et al. (1997); Veneziani et al. (2004) investigate the use of a two-dimensional linear stochastic differential equation to model the spinning of particles inside eddies motion. Lilly et al. (2017) propose a stochastic model using a Matern process which matches the observed Lagrangian velocity spectrum.

In this thesis we focus the stochastic differential equation (SDE) in the form of

$$\mathrm{d}\mathbf{X} = \left[\mathbf{U}\left(\mathbf{X}\left(t\right)\right) + \nabla \cdot \mathsf{K}\left(\mathbf{X}\left(t\right)\right)\right]\mathrm{d}t + \sqrt{2\,\mathsf{K}\left(\mathbf{X}\left(t\right)\right)}\mathrm{d}\mathbf{W}, \tag{2.16}$$

to model the Lagrangian trajectories. Here the eddy diffusivity tensor K and its square root are symmetric positive-definiteness, and $\mathbf{W}$ is a two-dimensional Brownian process.

The SDE (2.16) is a coarse-grained model for position $\mathbf{X}(t)$. It is expected to capture only the statistics of $\mathbf{X}(t)$ over sufficiently long time scales, instead of mimicking the full trajectories of $\mathbf{X}(t)$.

The evolution of $\mathbf{X}(t)$ according to (2.16) is entirely characterised by the transition probability density $\pi\left(\mathbf{x}, t|\mathbf{x}_0\right)$ which defines the probability of finding the particle in the neighbourhood of $\mathbf{x}$ at time t given it is initially at $\mathbf{x}_0$. The transition probability evolves under the Fokker–Planck equation (e.g. Evans, 2013; Pavliotis, 2014)

$$\frac{\partial \pi}{\partial t} + \nabla \cdot \left(\mathbf{U}\pi\right) = \nabla \cdot \left(\mathsf{K}\nabla\pi\right), \tag{2.17}$$

with initial condition $\pi\left(\mathbf{x}, 0|\mathbf{x}_0\right) = \delta\left(\mathbf{x} - \mathbf{x}_0\right)$. This is the advection–diffusion equation, and hence (2.16) is a natural stochastic model for advective and diffusive processes.

2.3.3 Field parameterisations and likelihood

The velocity and diffusivity $\mathbf{U}$ and K are fields defined over the entire spatial domain. For practical computations it is necessary to discretise and parameterise these fields with a finite-length vector of parameters $\boldsymbol{\theta}$

$$\mathbf{U}(\mathbf{x}) = \mathbf{U}(\mathbf{x}; \boldsymbol{\theta}) \quad \text{and} \quad \mathsf{K}(\mathbf{x}) = \mathsf{K}(\mathbf{x}; \boldsymbol{\theta}), \tag{2.18}$$

where $\boldsymbol{\theta}$ denotes the degrees of freedom for both $\mathbf{U}$ and K.

There is no universal recipe for the parameterisation. One framework is to assume the fields of $\mathbf{U}$ and K live in a finite-dimensional subspace of a function space, with $\boldsymbol{\theta} \in \mathbb{R}^N$ being the degrees of freedom. One example is a finite element discretisation. This approach allows the theoretical analysis over for the infinite-dimensional Bayesian inverse problem (e.g. Dashti and Stuart, 2017) defined to be directly linked to the finite-dimensional inference problem. In this thesis, we are not concerned with the detailed analysis for the posterior measure under infinite-dimensional inference, such as stability or the convergence properties. Instead, we adhere to a fixed discretisation, motivated

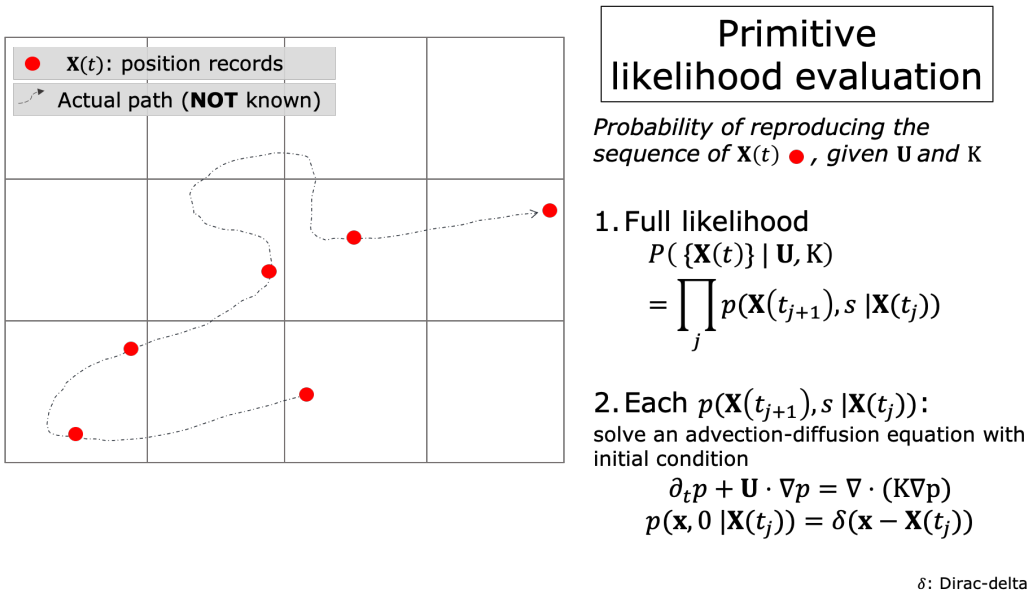

Figure 2.1: A schematic explanation of the likelihood evaluation from a single-particle trajectory.

by the physical length scales of concern and subject to availability of computational resources, and examine the behaviour of the posterior under varying configurations, data sizes and physical parameters.

Hereafter the dependence of quantities on $\boldsymbol{\theta}$ is omitted, but it should be borne in mind that most objects of interest, the transition probability π for instance, have such a dependence. The problem of estimating the discretised velocity and diffusivity fields now reduces to the estimation of $\boldsymbol{\theta}$. In the Bayesian-inference approach we adopt, the entire probability distribution of $\boldsymbol{\theta}$, and hence a probability distribution for $(\mathbf{U}, \mathsf{K})$, is estimated from trajectory data.

As the SDE is Markovian and the initial condition $\mathbf{X}_i(t_0)$ is deterministic and known a priori, the data set R can be replaced with

$$R = \{(\mathbf{X}_i(t_j), \mathbf{X}_i(t_{j+1}), t_{j+1} - t_j) : i = 1, \cdots, N,\ j = 1, \cdots, P-1\}. \tag{2.19}$$

That is, the data consist of the start and end positions of each particle between consecutive pairs of observations, and the time separation between the observations. We denote each data point as a *transition*. If the transitions are independent, the likelihood associated with the data is given by the product of individual transitions,

$$p(R|\boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{j=1}^{P-1} \pi(\mathbf{X}_i(t_{j+1}), t_{j+1} - t_j | \mathbf{X}_i(t_j)), \tag{2.20}$$

where π satisfies the Fokker–Planck equation (2.17). The likelihood $p(R|\boldsymbol{\theta})$ is, therefore, the probability that particles evolving according to (2.16), and with $(\mathbf{U}, \mathsf{K})$ fixed by $\boldsymbol{\theta}$, have positions matching R. Note that each likelihood evaluation for a parameter $\boldsymbol{\theta}$ involves solving $N(P-1)$ independent Fokker–Planck equations. We show in figure 2.1 a schematic illustration of the evaluation of the likelihood.

2.3.4 Choice of prior

The choice of prior inevitably affects the posterior and the inference. In this thesis, we apply uniform priors on all parameters $\boldsymbol{\theta}$, typically over a wide range on each component. This serves to firstly ensure a data-driven inference as the posterior is proportional to the likelihood in the constrained range of interest, and secondly enforces specific properties on the inferred field, for example, the diffusivity tensor should be positive definite. The choice of prior depends specifically on the parameterisation and details are supplied in the coming sections alongside with each application.

2.4 Discussion and summary

In this chapter, we presented a Bayesian approach to estimate the velocity and diffusivity fields in a stochastic differential equation from time-series data. This approach is applied in the appearing chapters of this thesis. The fields are parameterised by a finite number of degrees of freedom, differing from the approach taken by Crommelin and Vanden-Eijnden (e.g. 2006); Froyland et al. (e.g. 2014) where fields are reconstructed by a spectral consideration, and a posterior distribution is formulated to account for the plausibility of parameters under the given data and prior. We emphasise that the interpretation of the posterior is model-specific, with trajectory data assumed to be generated from a perfectly parameterised model.

While a posterior distribution is abstractly defined, information can be extracted using a sampling algorithm, which generates samples that are distributed by the posterior and enables evaluation of quantities of interest. In this thesis we specifically apply the Random–walk Metropolis–Hastings algorithm. We note that, to generate a new sample, each transition data point leads to one Fokker–Planck equation to be solved. This may lead to a potentially very high computational intensity – a concern to be addressed in the following chapters.

Chapter 3

Inference applied to cellular flows

This chapter concerns the application of the Bayesian inference in Chapter 2 to an idealised cellular flow. The goal of the inference is not on recovering the exact cellular flow but its coarse-grained approximation, whose existence is guaranteed by homogenisation theory presented in section 1.3. This chapter also demonstrates a key principle this thesis relies on – for the Bayesian approach to correctly infer the coarse-grained dynamics, particle trajectories need to be downsampled at a sufficiently long time interval (see Pavliotis and Stuart, 2007; Cotter et al., 2013).

Section 3.1 outlines the considered cellular flow and its coarse-grained dynamics. The detailed field parameterisation and computational configurations are elaborated in section 3.2, before moving on to the inference results in section 3.4. Section 3.5 summarises the key messages in this chapter.

3.1 Taylor–Green vortices with a background flow

A highly idealised model of oceanic eddies in a background current is constructed by superimposing a constant mean flow $\mathbf{u}_\mathrm{M}$ on top of Taylor–Green vortices $\mathbf{u}_\mathrm{TG}$, leading to the two-dimensional and doubly-periodic steady velocity field

$$\mathbf{u}_\mathrm{TG}(\mathbf{x}) = u_\mathrm{TG} \begin{pmatrix} -\sin(2\pi x/l)\cos(2\pi y/l) \\ \cos(2\pi x/l)\sin(2\pi y/l) \end{pmatrix}, \tag{3.1a}$$

$$\mathbf{u}_\mathrm{M} = u_\mathrm{M} \begin{pmatrix} \cos\phi_\mathrm{M} \\ \sin\phi_\mathrm{M} \end{pmatrix}, \tag{3.1b}$$

where u_TG is the maximum vortex speed, u_M is a background flow speed, and ϕ_M is the angle of the background flow to the x-axis. The small-scale advection–diffusion of particles according to

$$\mathrm{d}\mathbf{X} = (\mathbf{u}_\mathrm{TG}(\mathbf{X}(t)) + \mathbf{u}_\mathrm{M})\,\mathrm{d}t + \sqrt{2\kappa}\,\mathrm{d}\mathbf{W}, \tag{3.2}$$

is considered, where $\kappa \neq 0$ is here a small-scale scalar diffusivity. Note that κ, which governs the small-scale motion of the particles, is not the object to be inferred in this problem. Rather we seek to infer information about a large-scale effective diffusivity, which governs the long-time behaviour.

Homogenisation theory provides rigorous coarse-graining results for this problem. Specifically, over scales much larger than the vortex period l, the motion of particles is approximated by the SDE (2.16) with a uniform mean velocity $\mathbf{U} = \mathbf{u}_\mathrm{M}$ and an eddy diffusivity tensor K.

Parameter	Symbol	Value(s)
Spatial period	l	100 km
Maximum vortex speed	u_{TG}	40 cm s^{-1}
Background flow speed	u_{M}	20 cm s^{-1}
Background flow angle	ϕ_{M}	30°
Small-scale diffusivity	κ	50 m^2 s^{-1}
Peclet number	$\mathrm{Pe} = u_M l/\kappa$	400
Particle integration time step size	–	84.3750 s
Total particle integration time	–	256 days
Number of particles	–	256
Data sampling interval	s	3 hours, 6 hours, … 120 days
Markov chain Monte Carlo iterations	N_{mh}	10^5
Number of independent Markov chains	M	3

Table 3.1: Parameters used in the idealised Taylor–Green vortex configuration.

The effective diffusivity tensor K can be computed by solving the two-dimensional elliptic problem known as the 'cell problem' for a doubly periodic vector-valued function $\boldsymbol{\chi}(\mathbf{y})$ (1.20) on the rescaled coordinates $\mathbf{y} = \mathbf{x}/l$ in $\mathbb{T}^2$

$$-\left(\mathbf{u}_{\mathrm{M}} + \mathbf{u}_{\mathrm{TG}}\left(\mathbf{y}\right)\right) \cdot \nabla_{\mathbf{y}} \boldsymbol{\chi}(\mathbf{y}) + \kappa \nabla_{\mathbf{y}}^2 \boldsymbol{\chi}(\mathbf{y}) = \mathbf{u}_{\mathrm{TG}}(\mathbf{y}), \quad \text{with} \int_{\mathbb{T}^2} \boldsymbol{\chi}(\mathbf{y}) \mathrm{d}\mathbf{y} = \mathbf{0}, \tag{3.3}$$

and the eddy diffusivity K is given by

$$\mathsf{K} = \kappa \mathsf{I} + \int_{\mathbb{T}^2} \mathbf{u}_{\mathrm{TG}}(\mathbf{y}) \otimes \boldsymbol{\chi}(\mathbf{y}) \mathrm{d}\mathbf{y}, \tag{3.4}$$

here the gradient $\nabla_{\mathbf{y}}$, Laplacian $\nabla_{\mathbf{y}}^2$ and integral $\int_{\mathbb{T}^2} \cdot \mathrm{d}\mathbf{y}$ are defined with respect to the rescaled variable $\mathbf{y}$, I is the identity matrix and $\otimes$ is the outer product. A derivation can be found in section 1.3 by realising that the Fokker–Planck equation of the SDE (3.2) takes the form of the advection–diffusion equation (1.12). The coarse-grained equation for the cellular flow (3.2) is given by

$$\mathrm{d}\mathbf{X} = \mathbf{u}_{\mathrm{M}} \, \mathrm{d}t + \sqrt{2\mathsf{K}} \, \mathrm{d}\mathbf{W}. \tag{3.5}$$

3.2 Parameterisation

We apply Bayesian inference to this problem for the uniform velocity and diffusivity

$$\mathbf{U} = \mathbf{U}\left(\boldsymbol{\theta}\right) = U_0 \begin{pmatrix} \cos \Phi_0 \\ \sin \Phi_0 \end{pmatrix}, \tag{3.6a}$$

$$\mathsf{K} = \mathsf{K}\left(\boldsymbol{\theta}\right) = \mathsf{R}\left(\Phi_{\mathsf{K}}\right) \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix} \mathsf{R}\left(\Phi_{\mathsf{K}}\right)^{\mathrm{T}}, \tag{3.6b}$$

where

$$\mathsf{R}\left(\Phi_{\mathsf{K}}\right) = \begin{pmatrix} \cos \Phi_{\mathsf{K}} & -\sin \Phi_{\mathsf{K}} \\ \sin \Phi_{\mathsf{K}} & \cos \Phi_{\mathsf{K}} \end{pmatrix} \tag{3.7}$$

is a rotation matrix. Thus the parameters to infer are

$$\boldsymbol{\theta} = \left(U_0, \Phi_0, \Gamma_1, \Gamma_2, \Phi_{\mathsf{K}}\right)^{\mathrm{T}}. \tag{3.8}$$

Parameter $\boldsymbol{\theta}$	Initial value $\boldsymbol{\theta}^{(0)}$	Proposal s.d. $\sqrt{V_j}$	Prior range
U_0	0 m s^{-1}	0.001 m s^{-1}	[0 m s^{-1}, 10 m s^{-1}]
Φ_0	0 rad	0.05 rad	unbounded
Γ_1	1,000 m^2 s^{-1}	100 m^2 s^{-1}	[1 m^2 s^{-1}, 10^5 m^2 s^{-1}]
Γ_2	500 m^2 s^{-1}	50 m^2 s^{-1}	[1 m^2 s^{-1}, 10^5 m^2 s^{-1}]
Φ_{K}	0 rad	0.05 rad	unbounded

Table 3.2: Parameters used to initialise the Metropolis–Hastings Algorithm in the 'burn-in' phase.

The representation (3.6b) of the diffusivity K is motivated by its eigendecomposition, and guarantees that it is symmetric positive-definite when the eigenvalues Γ_1 and Γ_2 are positive, regardless of the orientation Φ_{K} of the eigenvector associated with Γ_1.

Parameters used in this example are provided in Table 3.1. The domain size, flow speeds, and small-scale diffusivity are chosen so as to yield an ocean-like regime. Particle trajectory data are generated by solving the SDE (3.2) for 256 particles initially located on a uniform square grid in the doubly-periodic domain $(x, y) \in [-l, l]^2$. The SDE is solved numerically using the Euler–Maruyama method with a small timestep size of 84.3750 s. For the purposes of the Bayesian inference, their positions are sampled with a sampling interval $s = t_j - t_{j-1}$ over a total time of 256 days.

3.3 Posterior evaluation

For the uniform velocity and diffusivity (3.6), the Fokker–Planck equation can be solved analytically (see B.3), yielding the Gaussian transition probability density

$$\pi\left(\mathbf{X}_i\left(t_{j+1}\right), s | \mathbf{X}_i\left(t_j\right)\right) = \frac{1}{2\pi\sqrt{\det \Sigma_s}} \exp\left(-\frac{1}{2}\left\|\mathbf{X}_i\left(t_{j+1}\right) - \mathbf{m}_s\left(\mathbf{X}_i\left(t_j\right)\right)\right\|^2_{\Sigma_s^{-1}}\right), \quad (3.9)$$

where

$$\mathbf{m}_s(\mathbf{x}) = \mathbf{x} + \mathbf{U}s, \quad \Sigma_s = 2s\mathsf{K} \quad (3.10)$$

and, for a suitably sized vector $\mathbf{v}$, the square norm $\|\mathbf{v}\|^2_{\Sigma_s^{-1}}$ is defined to be

$$\|\mathbf{v}\|^2_{\Sigma_s^{-1}} = \mathbf{v}^{\mathrm{T}} \Sigma_s^{-1} \mathbf{v}. \quad (3.11)$$

This gives an explicit expression for the likelihood (2.20).

In order to perform the Bayesian inference a prior must be chosen. In the absence of strong constraints, the priors for the angles Φ and Φ_{K} are set equal to the uniform distribution and priors for the remaining parameters are uniform in the ranges $U_0 \in \left[0, 10 \text{ m s}^{-1}\right]$ and $\Gamma_1, \Gamma_2 \in \left[1 \text{ m}^2 \text{ s}^{-1}, 10^5 \text{ m}^2 \text{ s}^{-1}\right]$, and zero elsewhere (see Table 3.2 for a summary). The wide and conservative prior ranges are meant to capture the order of magnitude of different quantities, and cover the reported values in previous studies of actual and simulated ocean diffusivity such as Marshall et al. (2006); Mak et al. (2016). The wide range would also avoid overconfidence in the uncertainty estimates. If more information was available about the problem, more restrictive priors could be chosen.

The posterior is evaluated, up to some unknown proportionality constant, as the product of the likelihood and the prior, noting that the proportionality constant is not required by the Metropolis–Hastings algorithm. In total 3 independent set of 100,000 samples $\boldsymbol{\theta}^{(k)}$ are drawn, and it is verified that the Gelman–Rubin diagnostic criterion (see section 2.2.4) is satisfied.

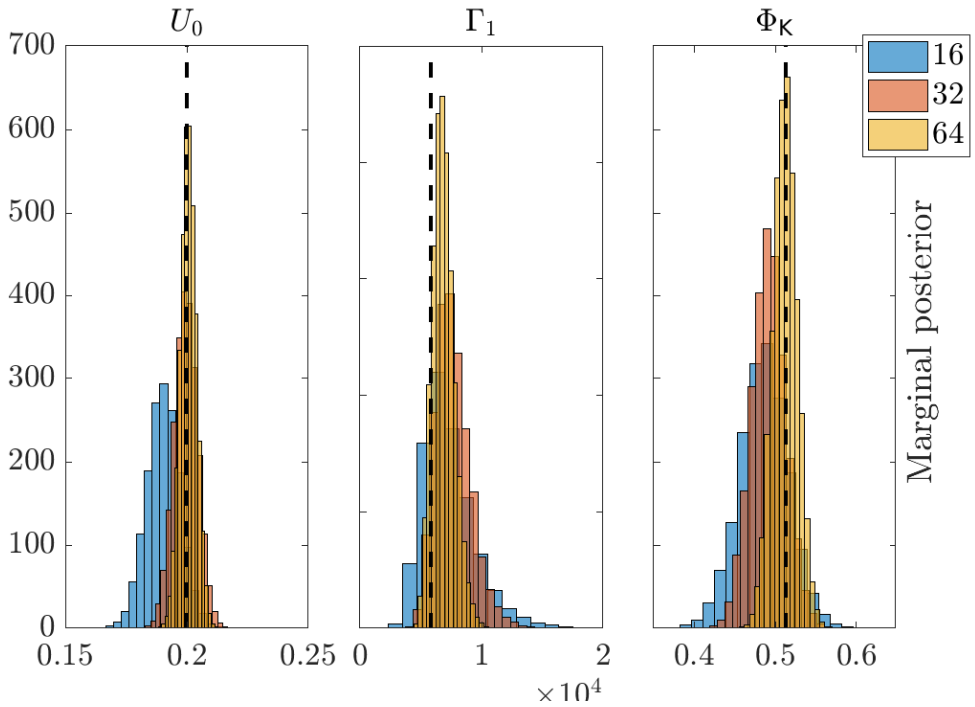

Figure 3.1: Marginal posterior distributions of the mean flow speed U_0, the leading eigenvalues Γ_1 and the orientation Φ_K of the leading eigenvector of the diffusivity tensor for the periodic flow (3.2) using a varying number of particles 16, 32 and 64 of the same temporal length of 256 days. The trajectory data is downsampled at an interval of 120 days. The exact magnitude of the background flow speed and the computed effective diffusivity components of homogenisation theory are shown as black dashed lines.

3.4 Results

For reference the effective diffusivity K of homogenisation theory is computed by solving the elliptic 'cell problem' (3.3) and evaluating (3.4). The equations are solved using degree-one continuous Lagrange finite elements using the FEniCS system (Logg et al., 2012; Alnæs et al., 2015) version 2018.1.0. A finite element mesh is formed via a 512×512 structured and uniform square mesh, with each square divided along the lower-left to upper-right diagonal to form a triangle mesh.

To examine the dependence of the posterior on data size, the marginal posterior on the inferred velocity magnitude U_0, leading eigenvalue of the diffusivity tensor Γ_1 and orientation of the leading eigenvector Φ_K are shown in figure 3.1. The number of particles considered is reduced to 16 and 32 and 64, instead of 256 as listed in table 3.1, for a clearer visualisation of the spread in the posterior. The trajectories are downsampled at a sampling interval of 120 days. For reference the corresponding values derived from homogenisation theory are shown in black dashed lines. The spreads of the marginal posteriors shrink to concentrate around the values predicted by homogenisation theory as the particle number increases. This hints posterior consistency – the capability for the Bayesian inference to recover the coarse-grained dynamics as data size increases.

Another important parameter for the Bayesian inference is the time interval at which trajectories are sampled. Figure 3.2 shows the inference results against the sampling intervals on a fixed trajectory. The posterior mean velocity components (left panel

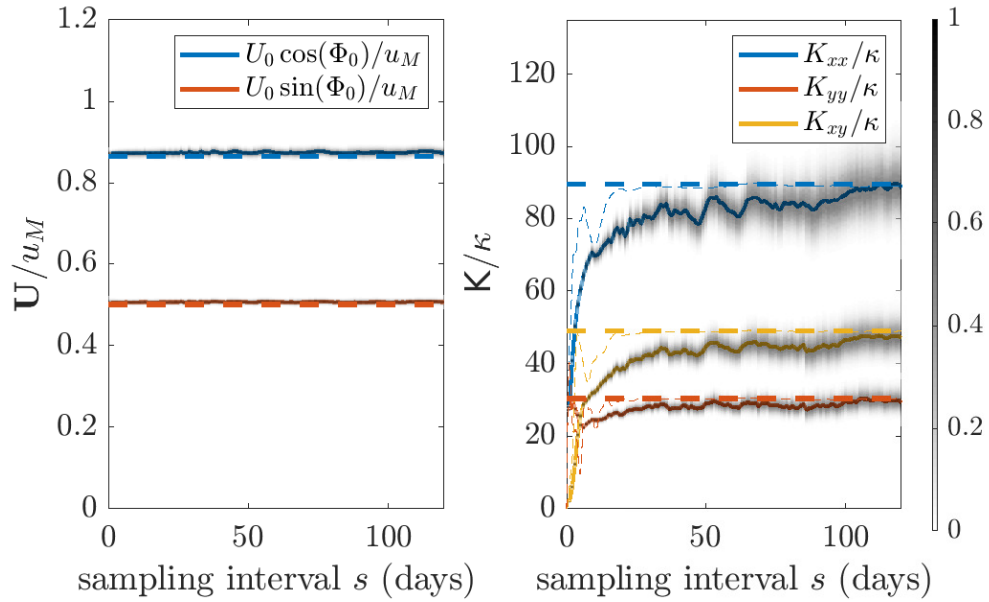

Figure 3.2: Results of the Bayesian inference for the mean flow and diffusivity tensor components for the periodic flow (3.2) under varying sampling intervals. Left panel: the Cartesian component of the inferred mean flow. Right panel: the Cartesian component of the inferred diffusivity tensor. In both panels the marginal posterior distributions are shown with shading, normalised so that the maximum value at each sampling interval is 1. The posterior means are shown as solid lines. The computed effective diffusivity components of homogenisation theory are shown as thick dashed lines, and the estimated absolute diffusivity at different lag τ are shown as thin dash lines.

of figure 3.2) show little variability with sampling interval and agree excellently with the background flow. The posterior mean diffusivity components (right panel of figure 3.2) show much greater variability. For example, over short time scales the particles experience only local small-scale dynamics, and hence short sampling intervals are associated with low values of inferred diffusivity. The diffusivity components increase with increasing sampling interval and approach a stable value in good agreement with the computed values from homogenisation theory or absolute diffusivity (1.28) at large lag τ. As the sampling interval increases, the number of particle positions used in the inference decreases (since the same length of particle trajectory is considered in all cases). As a result, the uncertainty of the inference increases, leading to a widening of the posterior distribution.

A clear message shown in figure 3.2 is that in a bid to accurately estimate the effective diffusivity K, it is insufficient to collect a large amount of particle trajectory data. Instead, the trajectory should be properly downsampled at a sufficiently long sampling interval before proceeding to the Bayesian machinery. This echos the conclusions in Pavliotis and Stuart (2007) and Cotter and Pavliotis (2009) for inference of coarse-grained dynamics in a multiscale system.

3.5 Discussion and summary

In this chapter, we applied the Bayesian inference approach to infer the coarse-grained mean velocity and eddy diffusivity in a multiscale cellular flow. We demonstrated the quality of the inference is affected by two important parameters – data size and sampling interval.

In this example, an increase in data size leads to a less dispersed posterior distribution and provides sharper estimates of the inferred quantity. This hints that the parameters considered are *identifiable* in the sense that the posterior contracts to a Dirac-delta distribution as data size increases. In fact, by noting the spatial homogeneity of the coarse-grained stochastic differential equation, one can interpret each transition $(\mathbf{X}_i(t_j), \mathbf{X}_i(t_{j+1}))$ as a displacement of $\mathbf{X}_i(t_{j+1}) - \mathbf{X}_i(t_j)$ from the origin. The displacement follows a Gaussian distribution with mean $s\mathbf{U}$ and covariance $2s\mathsf{K}$, so that $\mathbf{U}$ and K can be uniquely identified by, for example, the maximum likelihood estimates.

We verify, as literature has suggested, that the choice of sampling interval is equally, if not more, important in estimating the eddy diffusivity. The trajectory should be downsampled at a suitable temporal frequency to properly infer the coarse-grained quantities, particularly the eddy diffusivity. This inevitably discards a large amount of data. In practice, the optimal sampling interval, which balances the loss of data and accuracy of the estimation, is often unknown. We therefore take a trial-and-error approach, by examining the convergence of statistics in the posterior against varying sampling interval. Should the summary statistics become moderately insensitive to the sampling interval, we deem it a suitable sampling interval. This is a key principle applied in the upcoming chapters.

Chapter 4

Localised inference applied to an idealised oceanic flow

This chapter concerns the application of Bayesian inference approach to an idealised eddy-resolving quasigeostrophic double gyre flow. The configuration mimics the large-scale circulation in northern mid-latitude Atlantic, with a cyclonic and an anticyclonic gyre respectively in the north and southern half of the domain. A highly turbulent and eddy emitting eastward jet is maintained at the gyre interface. Over time the flow is statistically stationary, giving room for one to expect an advection–diffusion model to capture the average transport properties over a large temporal and spatial scale.

The goal of the inference is not to recover the time-varying turbulent flow but to produce a stationary and coarse-grained advection–diffusion model which picks up the large-scale properties in the flow.

In chapter 3 the Bayesian inference is applied to infer a spatially homogeneous mean velocity and diffusivity field, where a closed-form expression of the transition probability is available for likelihood evaluations. This is however not the case for the double gyre flow in this chapter. The mean velocity and diffusivity fields are expected to be spatially-varying in a way that explicit formulas for the transition density are not likely to present. While numerical methods can be applied to solve the Fokker–Planck (advection–diffusion) equation (2.17) for the transition density, such a brute-force approach involves a computational cost scaling with data size, and is infeasible given the trajectory data considered.

In this chapter, we develop a localised inference to avoid the need for a numerical differential equation solver. This approach divides the domain into cells, in which locally-defined mean velocity and diffusivity fields are inferred. By choosing linear mean velocity and constant diffusivity fields in each cell, analytic solutions of the Fokker–Planck (2.17) are available for likelihood evaluation. The cell-wise fields are decoupled so that inference in each can be independently performed.

A valid local inference demands conditions on the flow fields and data to be satisfied. An important condition is a locality requirement – the particle displacement at the considered time interval should be small relative to the cell size and variability of the mean velocity and diffusivity. This implies limitations on the resolution and characteristics of the fields and challenges the general applicability in an oceanographic context.

Section 4.1 outlines the configurations for the simulation of particle trajectories in the quasigeostrophic double gyre. Section 4.2 introduces the local approach to perform Bayesian inference. Section 4.5 shows the results of the local inference. The results of

this chapter is also found from the published article Ying et al. (2019).

4.1 Particle trajectory data

The numerical model of the quasigeostrophic double gyre simulation presented in section 1.5 is applied to drive the motion of particle. We consider only particle advection, with no explicit small-scale diffusivity, within the middle layer of the quasigeostrophic model. This layer experiences no direct wind forcing or bottom linear drag. A time-averaged streamfunction in the middle layer is shown in the right panel of figure 4.1.

676 particles are distributed at the centre of a 26×26 uniform-sized grid partitioning the square domain. This number is chosen so as to resemble the typical number of ARGO drifters available in the North Atlantic (Argo, 2000). After a 100 year [1] spinup, the particles are advected for a further 10 years, using a geometric integration approach summarised in the next paragraph with details in Appendix D. Their positions are recorded daily. The resulting trajectories for 50 arbitrarily selected particles are shown in figure 4.1.

The particle advection algorithm is motivated by the geometric integrator described in Ham et al. (2006) and Ham (2006). A piecewise linear streamfunction is constructed from the finite-difference grid point values by dividing each square cell corner-to-corner to yield four isosceles triangles, bi-linearly interpolating to yield a value at the centre vertex, and then linearly interpolating within the triangles. The time-dependent streamfunction is further linearly interpolated in time. The approximate streamfunction allows particle trajectories to be computed exactly and maintain the exact area-preserving property. Initial starting cells are determined using a quad-tree based search (Samet, 1984) using code derived from libsupermesh (Panourgias and Maddison, 2016), after which they are advected along contours of the discrete streamfunction. Note that care needs to be taken to ensure that the particle advection – which is a two-dimensional computational geometry problem – is solved in a precision-robust manner. A useful property of the particle advection scheme is that, given a streamfunction which is constant on the boundary, particles are guaranteed to never leave the bounding domain (see Ham et al., 2006). Hence the particle advection scheme requires no further consideration of boundary conditions.

4.2 Local Bayesian inference

The domain of interest is partitioned into a coarse mesh, and we seek to obtain information on the flow velocity and diffusivity for each mesh cell. The result of the inference is expected to be dependent on the choice of mesh, and in particular on the mesh cell size. This is consistent with the coarse-graining paradigm – the eddy diffusivity obtained is dependent upon the spatial scales.

Note that a meaningful eddy diffusivity is only realised after a decorrelation time scale. Over short time scales, correlated advection associated with the so-called "ballistic" regime (e.g. Pasquero et al., 2007; Rypina et al., 2012) dominates and is incompatible with the diffusive model (2.16). It is therefore necessary to ensure that the pairs of observed particle positions employed are separated by a sufficient time interval – a principle noted in a multi-scale system in Pavliotis and Stuart (2007) (see also Cotter

[1] Julian years are used throughout.

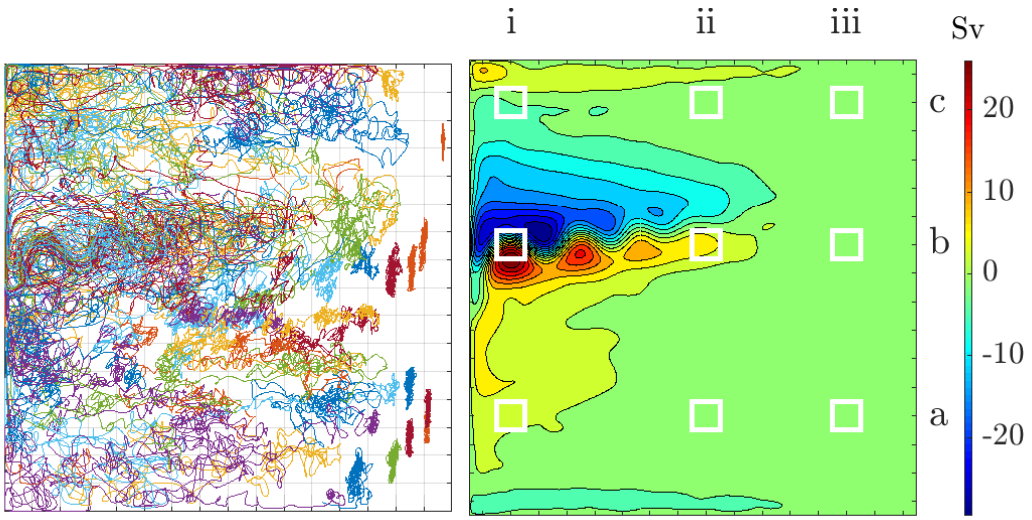

Figure 4.1: Left panel: 10-year trajectories for 50 arbitrarily selected particles in the middle layer of the quasigeostrophic double-gyre system. The division of the domain into a 16×16 array of square elements is shown in grey. Right panel: 10-year time-averaged streamfunction, multiplied by the layer thickness, in the middle layer. Selected cells of the 16×16 array referred to in the main text are highlighted in white and labelled by a letter/numeral coordinate.

and Pavliotis, 2009, for an application to eddy diffusivity). An optimal sampling interval, which discards the minimum number of position records while preserving the validity of the model (2.16), is rarely known a priori. In practice the inference is performed with varying sampling intervals and the convergence of the various estimates is examined. In the local inference approach we take here it is also necessary for the particles to remain in (or at least close to) the cell considered over the sampling interval. There is therefore a trade-off between two competing requirements: the sampling interval must be long enough that the particles do decorrelate, and short enough that they are not transported far from the considered cell. One must therefore take care to choose an appropriate sampling interval between observations, and be aware that this may not always exist. The possibility for a more advanced "non-local" inference, which alleviates this difficulty, is discussed in the conclusions.

4.3 Parameterisation

The domain is partitioned into a 16×16 array of square cells with 240 km side lengths. Within each cell the velocity is represented as a linearly varying non-divergent field, and the diffusivity as a constant symmetric positive definite tensor,

$$\mathbf{U}(\mathbf{x}) = \mathbf{U}(\mathbf{x};\boldsymbol{\theta}) = \mathsf{A}(\mathbf{x} - \mathbf{x}_0) + \mathbf{U}_0, \tag{4.1a}$$

$$\mathsf{K} = \mathsf{K}(\boldsymbol{\theta}) = \mathsf{R}(\Phi_{\mathsf{K}}) \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix} \mathsf{R}(\Phi_{\mathsf{K}})^{\mathrm{T}}, \tag{4.1b}$$

where

$$\mathbf{U}_0 = \mathbf{U}_0(\boldsymbol{\theta}) = U_0 \begin{pmatrix} \cos\Phi_0 \\ \sin\Phi_0 \end{pmatrix}, \tag{4.2a}$$

$$\mathsf{A} = \mathsf{A}(\boldsymbol{\theta}) = R(\Phi_\mathsf{A}) \begin{pmatrix} 0 & \Upsilon_2 + \Upsilon_1 \\ \Upsilon_2 - \Upsilon_1 & 0 \end{pmatrix} R(\Phi_\mathsf{A})^\mathrm{T}. \tag{4.2b}$$

Here R is a rotation matrix as in (3.7) and $\mathbf{x}_0$ is the centre of the cell. Thus the parameters to infer in each cell are

$$\boldsymbol{\theta} = (U_0, \Phi_0, \Upsilon_1, \Upsilon_2, \Phi_\mathsf{A}, \Gamma_1, \Gamma_2, \Phi_\mathsf{K})^\mathrm{T}. \tag{4.3}$$

The parameters Υ_1 and Υ_2 are related to the vorticity and strain associated with the velocity gradient tensor A, as explained in Appendix B.

The linear velocity field introduces additional degrees of freedom compared with the uniform velocity field used in section 3.2. It is motivated by the large shears that are present in the jet region of the simulation and that can severely bias the inferred diffusivity if unresolved (Oh et al., 2000; Griesel et al., 2010). It should be noted that the cell-wise linear velocity field is only a local approximation to the coarse-grained mean flow. The validity of the approximation can be justified on a sufficiently short time-scale when the dispersion of particles remains dominated by the linear components in the shear.

4.4 Posterior evaluation

The Fokker–Planck equation can be solved analytically for the velocity and diffusivity (4.1), yielding the Gaussian transition probability density

$$\pi\left(\mathbf{X}_i(t_{j+1}), s | \mathbf{X}_i(t_j)\right) = \frac{1}{2\pi\sqrt{\det(\Sigma_s)}} \exp\left(-\frac{1}{2}\left\|\mathbf{X}_i(t_{j+1}) - \mathbf{m}_s(\mathbf{X}_i(t_j))\right\|^2_{\Sigma_s^{-1}}\right), \tag{4.4}$$

where

$$\mathbf{m}_s(\mathbf{x}) = \mathbf{m}_s(\mathbf{x};\boldsymbol{\theta}) = e^{\mathsf{A}s}\mathbf{x} + \int_0^s e^{\mathsf{A}t}\mathrm{d}t\,(\mathbf{U}_0 - \mathsf{A}\mathbf{x}_0), \tag{4.5a}$$

$$\Sigma_s = \Sigma_s(\boldsymbol{\theta}) = 2\int_0^s e^{\mathsf{A}t}\mathsf{K}e^{\mathsf{A}^\mathrm{T}t}\mathrm{d}t, \tag{4.5b}$$

and $\|\mathbf{v}\|_{\Sigma_s^{-1}}$ is defined in (3.11), see also appendix B.3

$$\|\mathbf{v}\|^2_{\Sigma_s^{-1}} := \mathbf{v}^\mathrm{T}\Sigma_s^{-1}\mathbf{v}.$$

This gives an explicit expression for the likelihood (2.20).

We take again simple uniform priors for $p(\boldsymbol{\theta})$: the angles Φ_0, Φ_A, and Φ_K are uniform, and remaining parameters are uniformly distributed in the ranges $U_0 \in \left[0, 10\text{ m s}^{-1}\right]$, $\Upsilon_1, \Upsilon_2 \in \left[-10^{-5}\text{ s}^{-1}, 10^{-5}\text{ s}^{-1}\right]$, $\Gamma_1, \Gamma_2 \in \left[1\text{ m}^2\text{ s}^{-1}, 10^5\text{ m}^2\text{ s}^{-1}\right]$ and are zero elsewhere. It has been verified that the results would be unaffected if these ranges were extended. The results would of course be affected were much more restrictive priors imposed.

The posterior is evaluated, up to some unknown proportionality constant, as the product of the likelihood and the prior. In total 10 independent chains of 4×10^5 samples

Parameter $\boldsymbol{\theta}$	Initial value $\boldsymbol{\theta}^{(0)}$	Proposal s.d. $\sqrt{V_j}$	Prior range
U_0	0 m s^{-1}	0.001 m s^{-1}	[0 m s^{-1}, 10 m s^{-1}]
Φ_0	0 rad	0.05 rad	unbounded
Υ_1	0 s^{-1}	2.5×10^{-8} s^{-1}	[-10^{-5} s^{-1}, 10^{-5} s^{-1}]
Υ_2	0 s^{-1}	2.5×10^{-8} s^{-1}	[-10^{-5} s^{-1}, 10^{-5} s^{-1}]
Φ_{A}	0 rad	0.05 rad	unbounded
Γ_1	1,000 m^2 s^{-1}	100 m^2 s^{-1}	[1 m^2 s^{-1}, 10^5 m^2 s^{-1}]
Γ_2	500 m^2 s^{-1}	50 m^2 s^{-1}	[1 m^2 s^{-1}, 10^5 m^2 s^{-1}]
Φ_{K}	0 rad	0.05 rad	unbounded

Table 4.1: Parameters used to initialise the Metropolis–Hastings Algorithm in the 'burn-in' phase and the choice of prior for the local inference.

$\boldsymbol{\theta}^{(k)}$ are then drawn using the Metropolis–Hastings algorithm, with a summary of the configuration given in Table 4.1. The Gelman and Rubin (1992) diagnostic (see section 2.2.4) to test convergence is applied to verify convergence of the sample Markov chain. This process is performed separately for each cell of the model domain. We consider the sampling intervals $s = 1, 2, 4, 8, 16, 24, 32, \cdots, 128$ days. The samples of each of the independent chains are combined to approximate the posterior distribution.

4.5 Results

The upper middle and right panels of figure 4.2 show the maximum a posteriori estimate (MAP) for the middle-layer velocity field, together with the Eulerian mean flow computed over the 10-year data collection window on the left. Note only the MAP estimates are presented to help visualise the inferred posterior. The MAP here is estimated by the single MCMC sample maximising the (unnormalised) posterior, picked from the millions sample obtained. We acknowledge that, compared with a direct numerical optimisation search, the search of MAP using MCMC is an highly inefficient and the estimate may converge slowly to the theoretical value. Yet since we only demand a crude representation of the posterior, this readily available estimate is sufficient here. The MAP estimate of $\boldsymbol{\theta}$ is the maximiser for the posterior $p(\boldsymbol{\theta}|R)$ and indicates the most likely combination of mean flow and diffusivity fields [2] to recover the trajectory data. In all cases described here the MAP estimate is approximated by the sample $\boldsymbol{\theta}^{(k)}$ that maximises the posterior $p(\boldsymbol{\theta}^{(k)}|R)$ over all MCMC steps k and over all chains. For a short sampling interval $s = 1$ day, the MAP flow velocity is comparable to the Eulerian mean velocity. Over longer sampling intervals, where Lagrangian-mean effects are expected to play an increasing role, the inferred flow deviates from the Eulerian mean. The presence of local noise in the MAP estimate is attributed to the multi-modal marginal distributions for mean flow magnitude in some cells. Note that the posterior mean velocity at cell centres, computed as a mean over all samples and shown in the bottom panels of figure 4.2, has a smoother profile.

The bottom left panel of figure 4.2 shows a Lagrangian eddy decorrelation time, specifically the e-folding scale T_e introduced by Garraffo et al. (2001) (see Appendix A.2). This is computed using 10,000 particle trajectories over the same 10-year period. The figure indicates that typical decorrelation time scales are of the order of about 10 days.

[2] instead of the independent MAP estimates in the marginal posterior of the mean flow and diffusivity

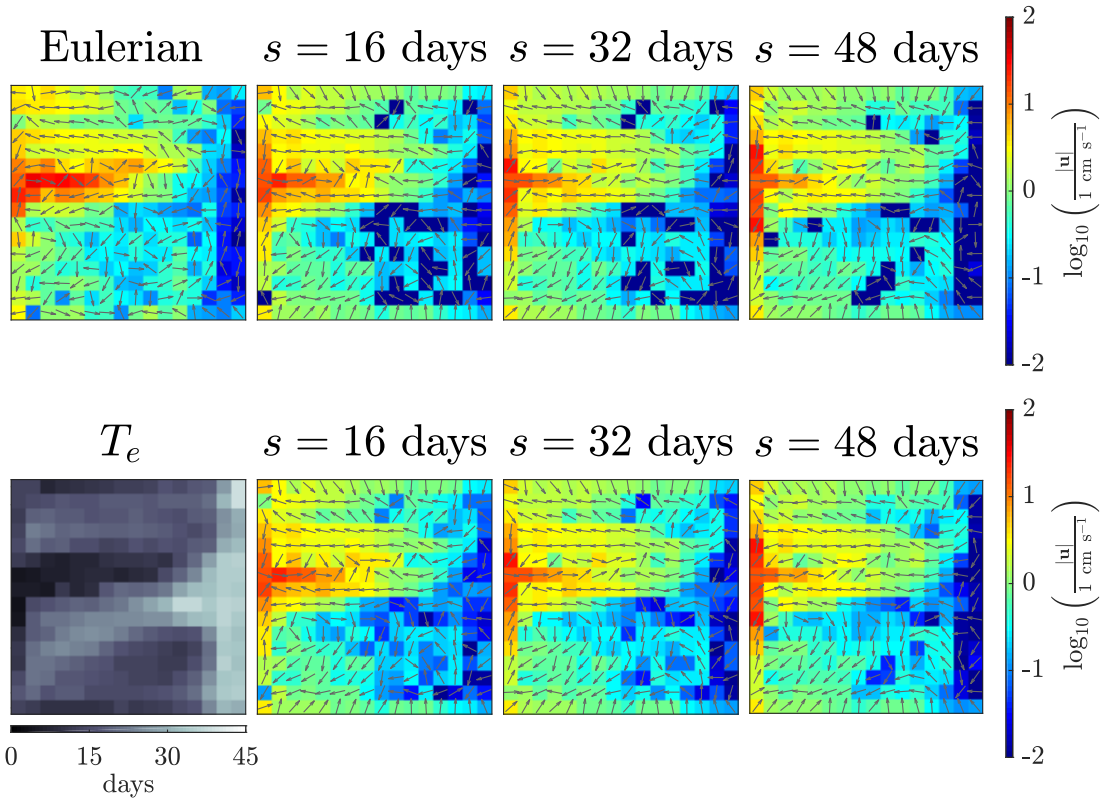

Figure 4.2: Top left: 10-year Eulerian mean velocity in the middle layer at the cell centres. Top Middle and right: Posterior mean for the cell-centre middle layer velocity using particle positions observed at sampling intervals of $s = 1$ day and $s = 32$ days. Bottom left: e-folding time T_e estimated from Lagrangian data. Bottom middle and right: Posterior mean for the cell-centre middle layer velocity using particle positions observed at sampling intervals of $s = 1$ day and $s = 32$ days. The magnitude of the velocity is shown using a logarithmic colour scale, and the velocity direction is indicated by equal-length arrows.

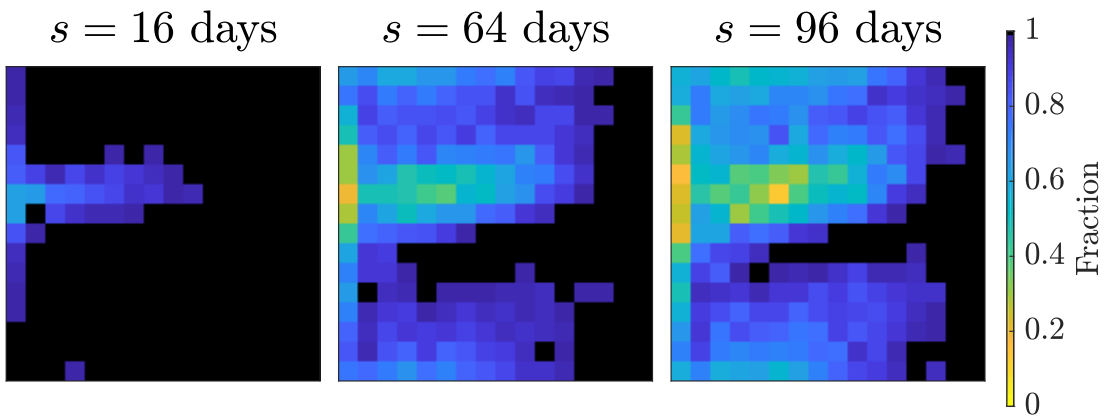

Figure 4.3: Fractions of particles found in their origin cell, or in the eight cells surrounding the origin cell, at the end of the sampling interval. For sampling intervals shorter than 4 days all particles remain in this neighbourhood.

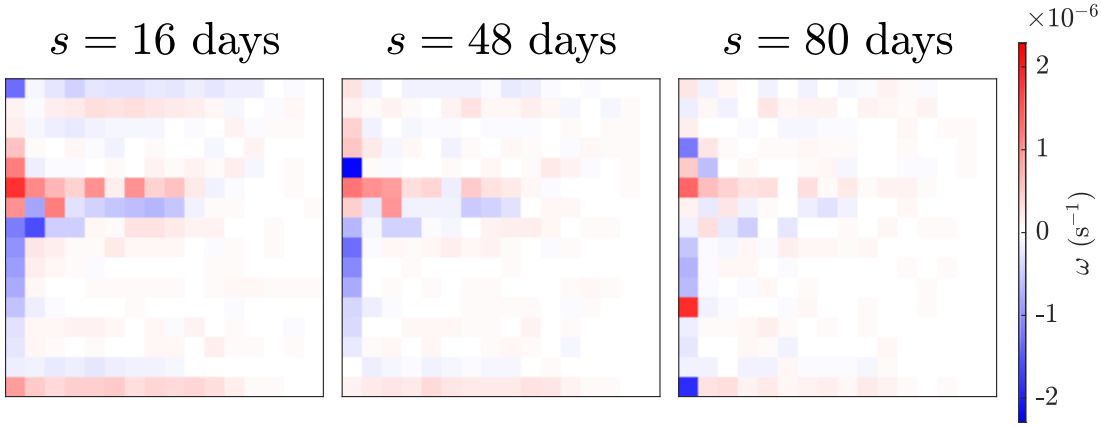

Figure 4.4: MAP estimate of the vorticity $\omega = \partial_x v - \partial_y u$ field at different sampling intervals s.

Figure 4.3 shows the fraction of particles which are found in their cell of origin or in one of the eight surrounding cells at the end of the sampling interval (regardless of the intermediate trajectory). This indicates the validity of the locality assumption inherent in the local inference approach. For short sampling intervals this fraction is high, but, as expected, drops as the sampling interval increases; in particular, it drops to very low values in the jet and on the western boundary. There is therefore a potential misattribution of the spatial location of flow properties in these regions. This is a significant issue on the western boundary, where particles flow rapidly from the boundary into the jet, and rapidly change direction from a northward or southward flow, to an eastward flow.

At short sampling intervals ($\lesssim 16$ days), strong shears are inferred along the jet and on the northern, western, and southern boundaries. This is indicated by the large local vorticity $\omega = \partial_x v - \partial_y u$, corresponding to the off-diagonal elements of $\mathsf{A} - \mathsf{A}^{\mathrm{T}}$, shown in figure 4.4. The inferred diffusivity in these areas is significantly reduced (not shown) when the spatial gradients of the mean flow are resolved, by permitting a non-zero linear shear. For the large sampling intervals the inferred shear tensor is smaller, as may be expected for a Lagrangian average of the flow over these time scales. Hence for the large sampling intervals the inferred diffusivity is largely unaffected by the inclusion of shear in $\mathbf{U}(\mathbf{x}; \boldsymbol{\theta})$, and inference with a locally constant velocity would yield similar results.

Figure 4.5 visualises the MAP estimate for the middle layer diffusivity tensor for differing sampling intervals. The "diffusivity ellipses" in figure 4.5 outline the orientations of contours of a passive tracer if it undergoes pure diffusion with a Dirac-delta initial profile, characterising the directions of the anisotropy of the eddy diffusion tensor. The diffusivity magnitude, defined as the half trace of the diffusivity tensor, is visualised using the colour scale. The inferred diffusivity is largest in the jet region, and strengthens with increasing sampling interval. There is a region of very weak inferred diffusivity in the eastern part of the southern half of the domain. At large sampling interval the anisotropic diffusion has a preferential east-west orientation in the gyres and the core of the jet. Near the western boundary the anisotropic diffusivity is tilted towards the direction of the jet – this is attributed to 1) non-local effects, as particles are rapidly transported into the jet from this region and 2) the absence of any boundary conditions in the inferred field, i.e. the inferred field is theoretically defined

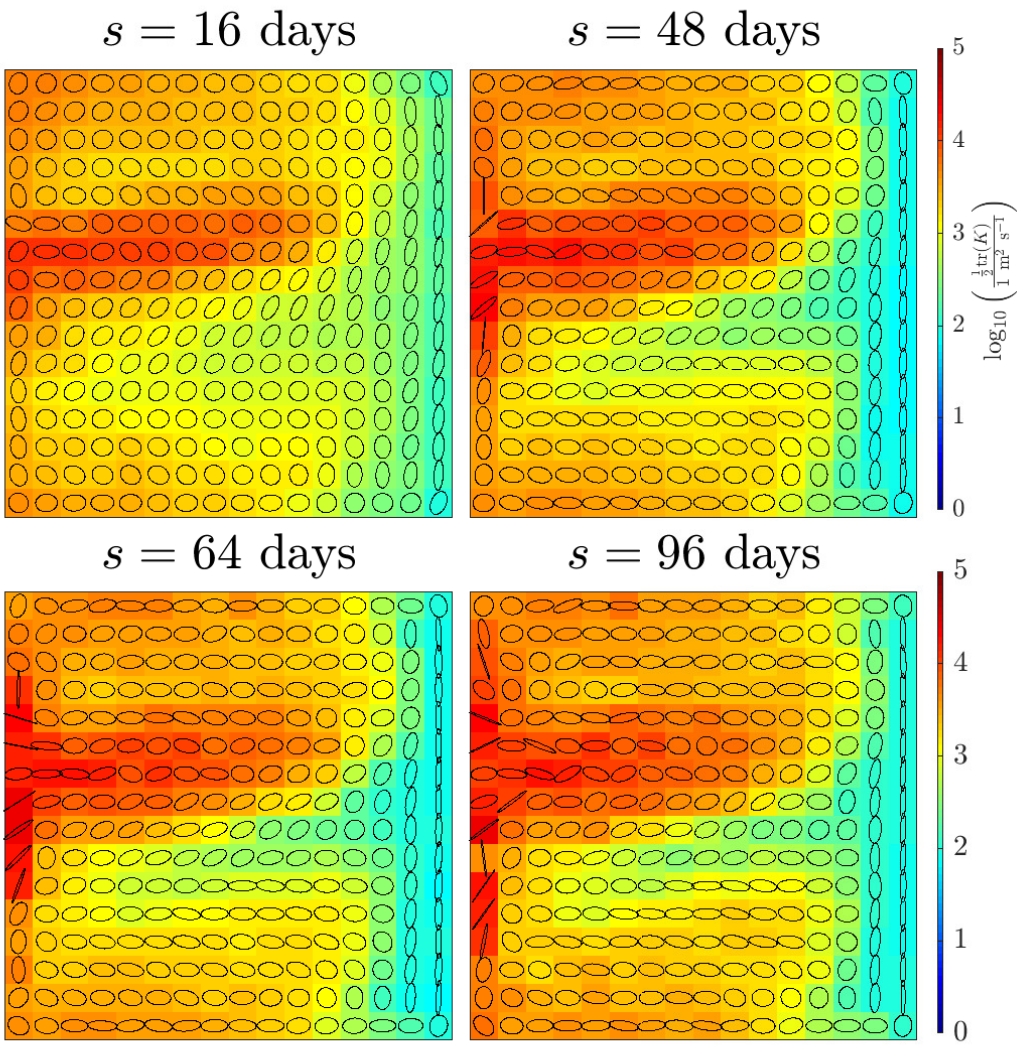

Figure 4.5: MAP estimate of the middle layer diffusivity field at different sampling intervals s. The (logarithmic) colour scale gives the half trace of the diffusivity tensor K, which is also the arithmetic mean of the eigenvalues, to characterise the magnitude of diffusivity. The ellipses visualise the directions and the relative magnitude of the two eigenvectors of the diffusivity tensor in each cell.

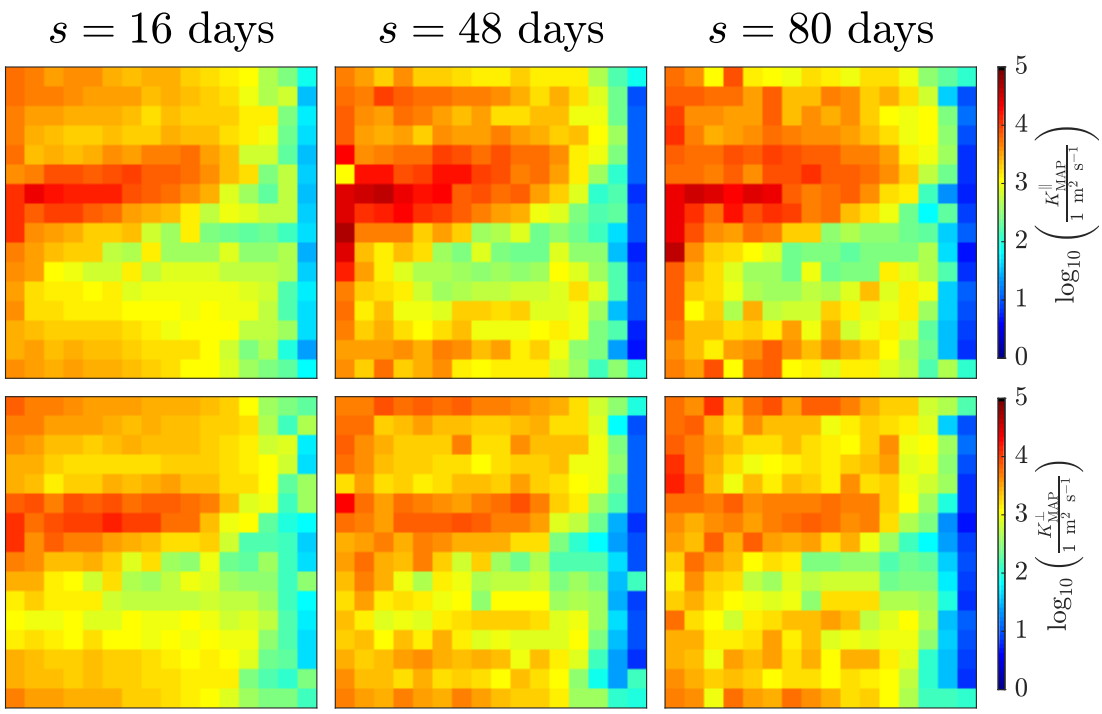

Figure 4.6: MAP estimate for the diffusivity components in the middle layer with different sampling intervals. Upper panels: along-stream diffusivity; lower panels: cross-stream diffusivity. A logarithmic colour scale is used.

on an unbounded $\mathbb{R}^2$-plane. The second has particularly resulted in a systematic bias in the mean flow and thus an overshoot in the estimation of diffusivity.

The Metropolis–Hastings algorithm samples the joint posterior distribution of the velocity and diffusivity and so makes it possible to infer quantities that depend on both fields. In particular, we can construct distributions for the cross-stream and along-stream diffusivity components $K^{\perp}$ and $K^{\parallel}$ by projecting for each sample k, the sample diffusivity $\mathsf{K}^{(k)}$, in directions perpendicular to and parallel to the sample velocity $\mathbf{U}^{(k)}$. The resulting MAP estimates are shown in figure 4.6. For comparison, the cross-stream and along-stream Davis (1987) diffusivity, defined with respect to the 10 year Eulerian mean flow at the cell centre, are shown in figure 4.7. A total number of $40,000$ particles, initially deployed uniformly across the domain, are used for the Davis (1987) diffusivity evaluation. Details of the evaluation can be found in Appendix A.1.

The two diagnostic approaches generally agree well in order of magnitude and spatial structure, with increased diffusivity in the region of the jet and reduced diffusivity on the eastern boundary and in the region south of the jet. There is some disagreement in detail, for example near the northern and southern boundaries. Note that the Davis (1987) diffusivity as computed here is not a symmetric positive definite (or even symmetric) quantity in general, leading to some regions of missing data indicated in white in figures 4.7. Note again MAP estimates here are derived from all MCMC samples, this is effectively millions of times computationally costlier than the Davis diffusivity, which is derived from numerically integrating autocovariance of the trajectory data (see equation 1.31). Should only the MAP estimates be sought after, an algorithmic optimiser can be applied to the posterior.

To analyse our results in more detail, we now focus on the 9 cells highlighted in figure 4.1 and labelled (i)–(iii) with increasing x coordinate, and (a)–(c) with increasing

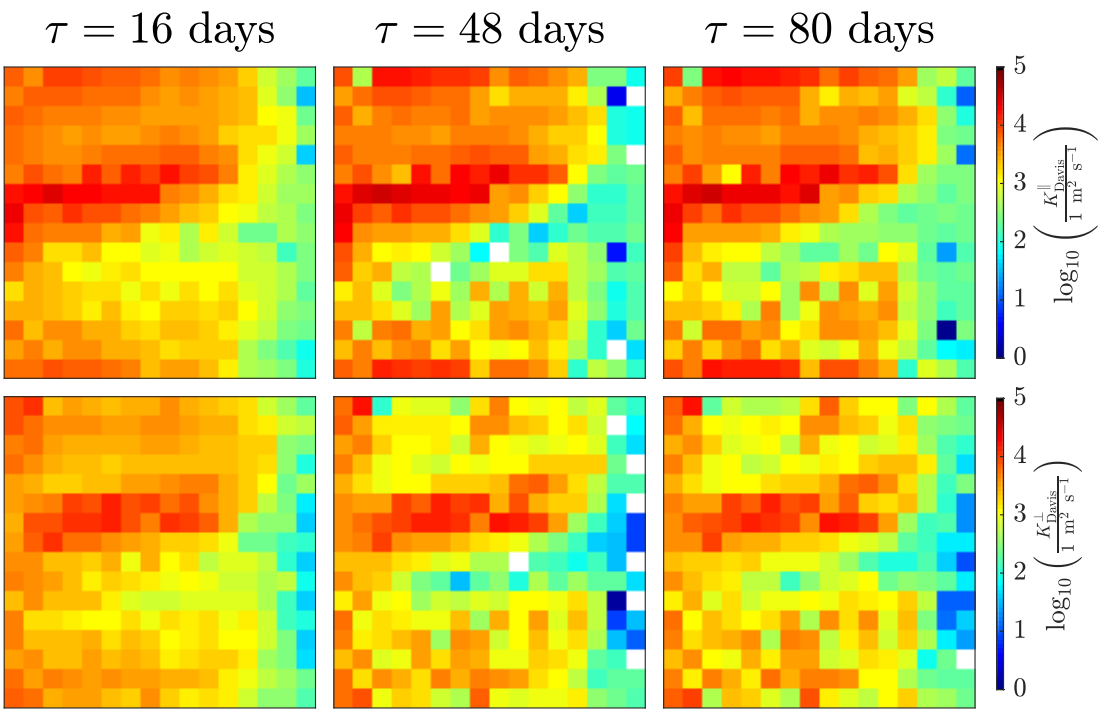

Figure 4.7: Davis (1987) diffusivity components in the middle layer with different time lags. Upper panel: along-stream diffusivity; lower panel: cross-stream diffusivity. A logarithmic colour scale is used. Missing data, shown in white, correspond to negative values of the Davis (1987) diffusivity components.

y coordinate. Figures 4.8 and 4.9 show the MAP of the middle layer cross-stream and along-stream diffusivity in these cells as functions of the sampling interval s. The Davis (1987) diffusivity is shown for comparison; the time lag τ and sampling interval s are shown on a common scale even though the two parameters are not strictly comparable. The MAP diffusivities do demonstrate a degree of convergence at larger sampling intervals, and agree in order of magnitude, at larger sampling intervals, with the large time-lag Davis (1987) diffusivity. The approximate convergence takes place for values τ and s that are roughly similar and comparable to an estimate of the Lagrangian decorrelation time of the velocity. The MAP diffusivities are never negative, as a consequence of the choice of prior, and while some variation is observed with sampling interval, the Bayesian diffusivity estimates are generally more stable in magnitude than the Davis (1987) diffusivity values.

One of the advantages of the Bayesian approach is that it provides a multidimensional probability distribution, rather than single estimates for $\boldsymbol{U}$ and K, and hence allows for a quantification of the uncertainty. This is illustrated in Figures 4.8 and 4.9 which also show the marginal posterior probability density for the two diffusivity components at each sampling interval s. The probability densities are shown as shading and normalised by their maximum value at each value of s. Broadly speaking, the figures suggest that the range of plausible values is reasonably well constrained, with low probabilities for values more than a factor of, say 2, away from the MAP. Nevertheless, relatively long tails of the posterior distribution indicate that there is a significant probability of diffusivities of much larger magnitude than the MAP values. There are cases of multi-modality, for example in the middle and top right panel of figure 4.8 and figure 4.9, with in this cases a MAP value which switches between the

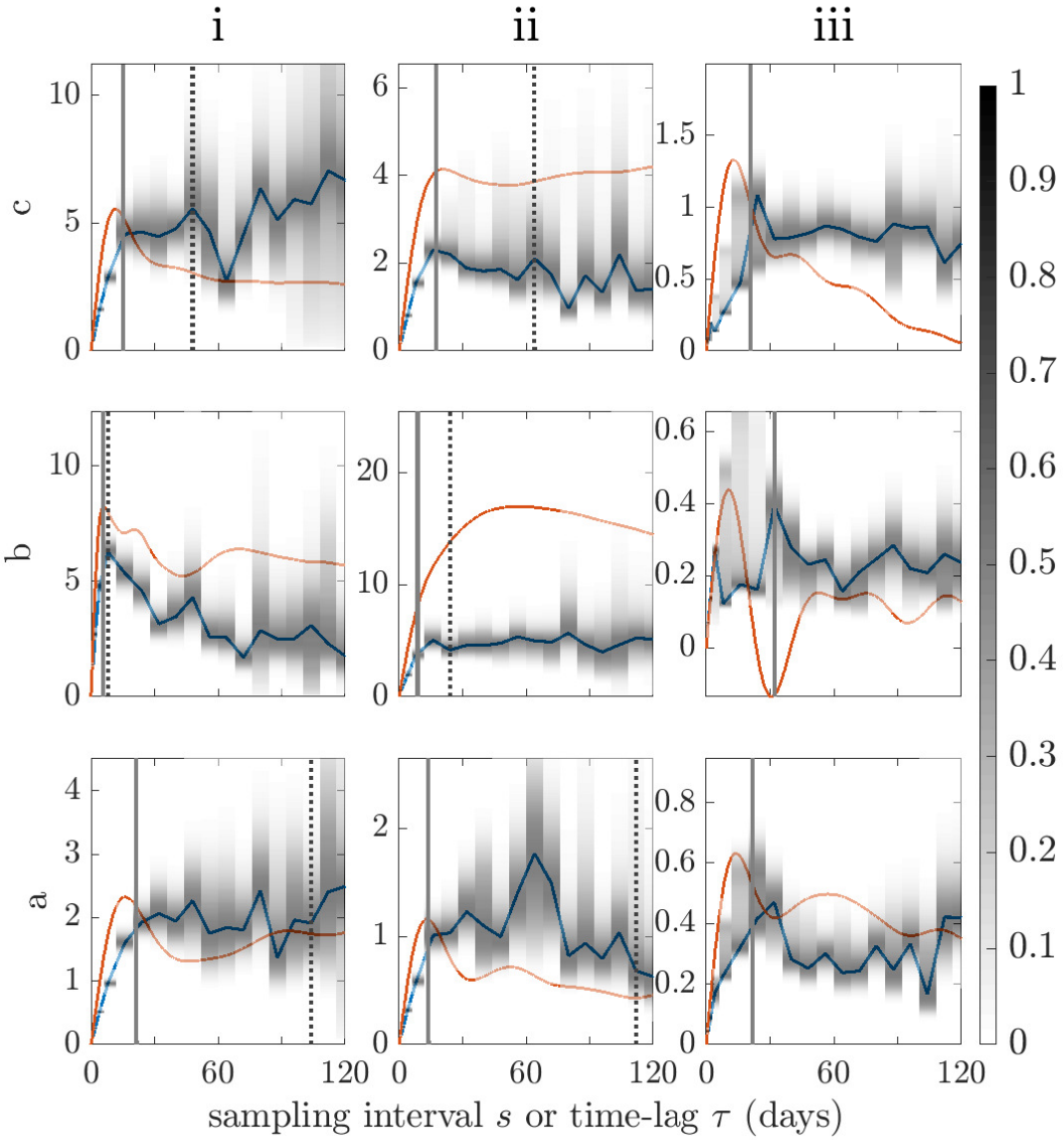

Figure 4.8: Cross-stream diffusivity $K^{\perp}$ (in 1000 m^2 s^{-1}) in the middle layer against sampling interval s or time-lag τ in selected cells, labeled on the top of each column and left of each row (see figure 4.1). The blue lines are the MAP estimates of the Bayesian inference; the red lines correspond to the Davis (1987) diffusivity. The grey shading shows the marginal posterior density for $K^{\perp}$, normalised by its maximum values for each s. The dash-dot vertical lines indicate the time taken for 10 percent of particles to exit the origin and its neighbouring 8 cells. The solid grey vertical lines show the e-folding scale estimated from Lagrangian trajectories as described in Appendix A.1. Note that the vertical lines are not shown if they correspond to times beyond 128 days.

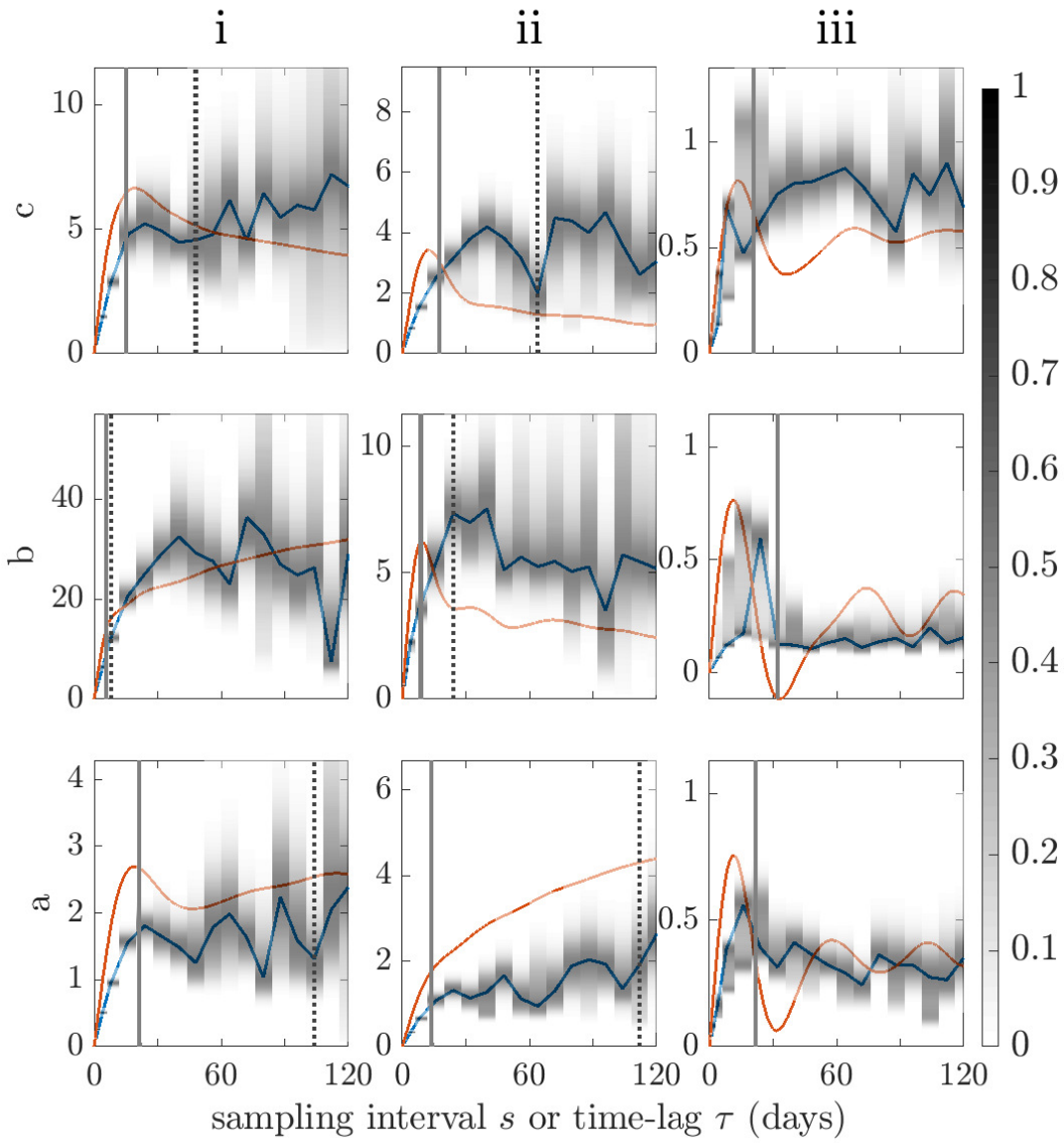

Figure 4.9: Same as figure 4.8 but for the along-stream diffusivity $K^{\parallel}$ in the middle layer.

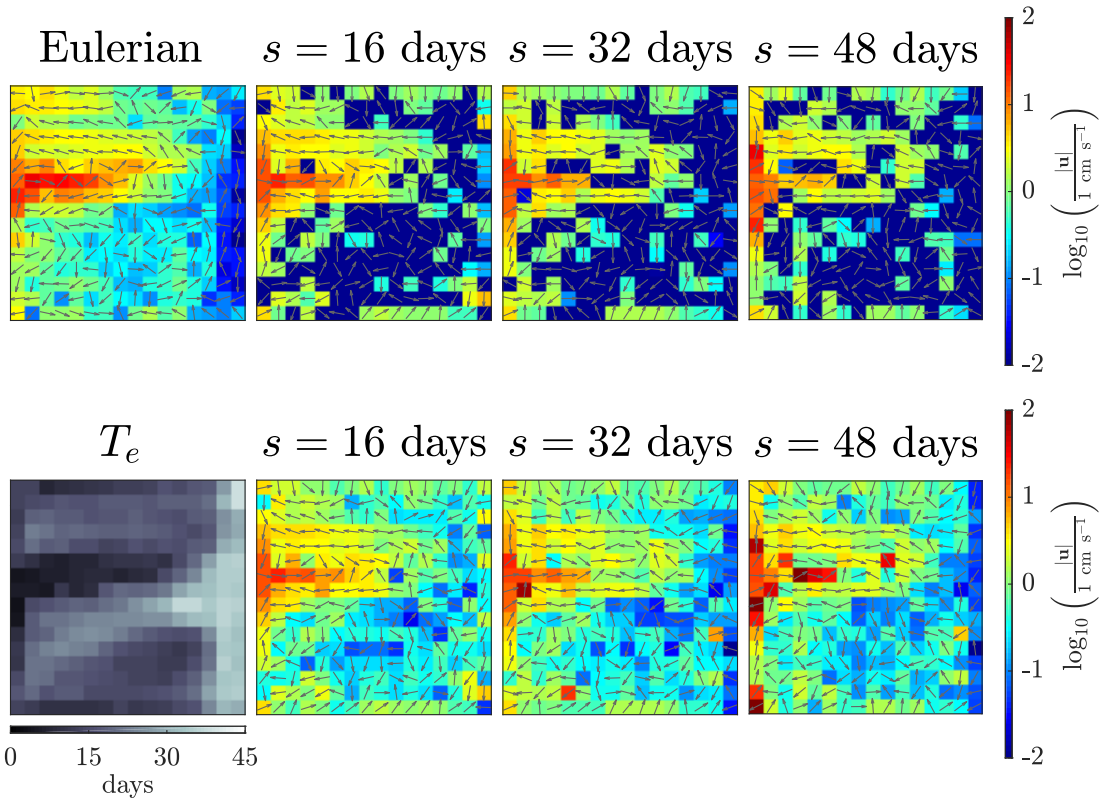

Figure 4.10: Mean velocities and e-folding time. Same as figure 4.2 but with 169 particles.

two local maxima. We attribute this to the weakness of the flow in these regions which leads to an ambiguity in the flow direction and hence in the decomposition between along-stream and cross-stream diffusivity.

We assessed the robustness of the inference by carrying out computations under different settings. First, we tested the sensitivity to data size by changing the number of particles from 676 to 169 and 2704 while retaining 10 years of data. The results (see figure 4.10 to 4.13) indicate that the MAP estimates of both the mean flow and diffusivity are sensitive in cells with weak mean flows and small numbers of start and end positions records when the number of particles decreases from 676 to 169, but they do not change markedly when the number of particles increases from 676 to 2704. As expected, the posterior distributions narrow as the data size increases. With only 169 particles, the distribution is very broad, so that caution should be exercised when interpreting the inferred diffusivity. This reflects a fundamental limitation in the amount of information provided by the data, relative to the prior.

Second, we tested the effect of flow strength by analysing the top layer and the bottom layer of the quasigeostrophic simulation, using again the trajectories of 676 particles. In the bottom layer (see figure 4.15 and 4.17) the diffusivity estimates converge in most of the domain but typically for sampling intervals s considerably larger than in the middle layer. This is attributed to long decorrelation times and the weakness of the flow; the latter improves the validity of the locality assumption. In the top layer (see figure 4.14 and 4.16) the particles rapidly escape from any given cell. This makes the validity of the locality assumption questionable there.

Third, we changed the number of cells in which the domain is partitioned from

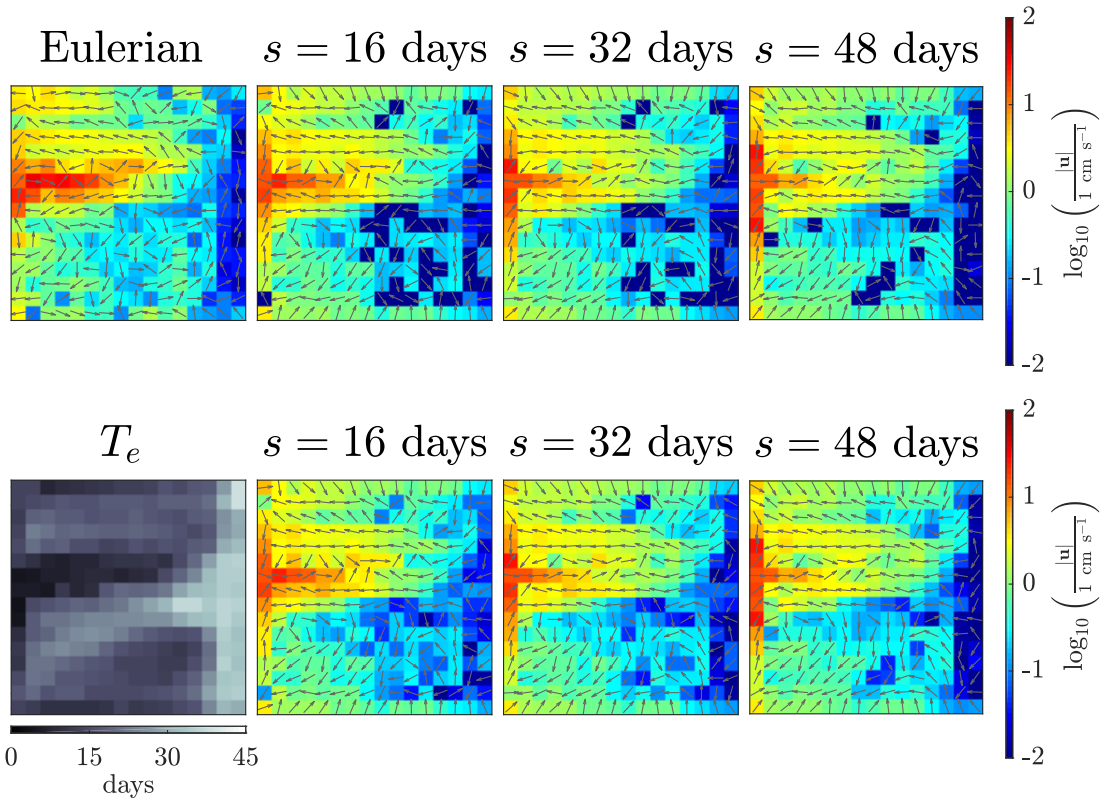

Figure 4.11: Mean velocities and e-folding time. Same as figure 4.2 but with 2704 particles.

16×16 to 8×8 and 32×32 cells. For 8×8 cells (see figure 4.18 and 4.18), the gyre is poorly resolved in the inferred mean flow and the diffusivity values (converging over longer sampling times) are larger. For 32×32 cells (see figure 4.19 and 4.19), particles escape cells over timescales that are short compared with that required for the diffusivity estimate to converge, violating the locality assumption. Moreover, very few particle start positions are found in some cells due to their small size, in which the posterior is almost entirely determined by the prior. We emphasise that the dependence of the inferred diffusivity on the number of cells is expected since different cell sizes correspond to different coarse-graining scales.

Fourth, we also examined prior sensitivity. By doubling and halving the ranges of uniform distributions, we confirmed that the MAP estimates in most of the cells are largely unaffected. Furthermore, except in the cells with a small number of trajectory records (that is, near the eastern boundary), the moments of the posterior sample are also barely affected. Similar results are obtained by using a Gaussian distribution with means and variances identical to the adopted uniform priors.

4.6 Discussion and summary

In this chapter, we develop a localised inference to infer the velocity and eddy diffusivity fields from Lagrangian trajectories. This approach divides the domain of interest into cells and infers a locally defined linear velocity and constant diffusivity field. The localisation enables the likelihood to be evaluated using analytic solutions to the Fokker–Planck equations and is computationally efficient for sampling algorithms.

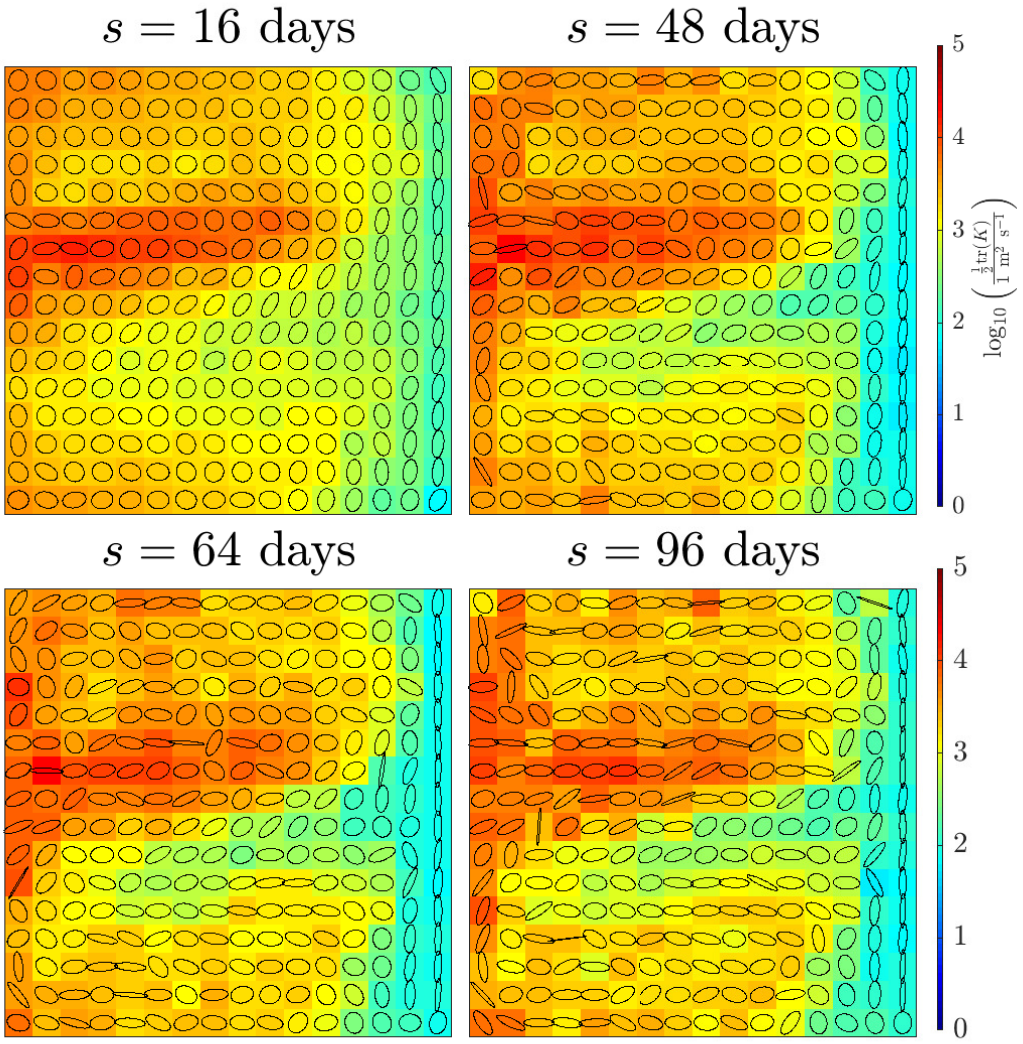

Figure 4.12: MAP estimate of the middle layer diffusivity field at different sampling intervals s using 169 particles. The (logarithmic) colour scale gives the half trace of the diffusivity tensor K, which is also the arithmetic mean of the eigenvalues, to characterise the magnitude of diffusivity.

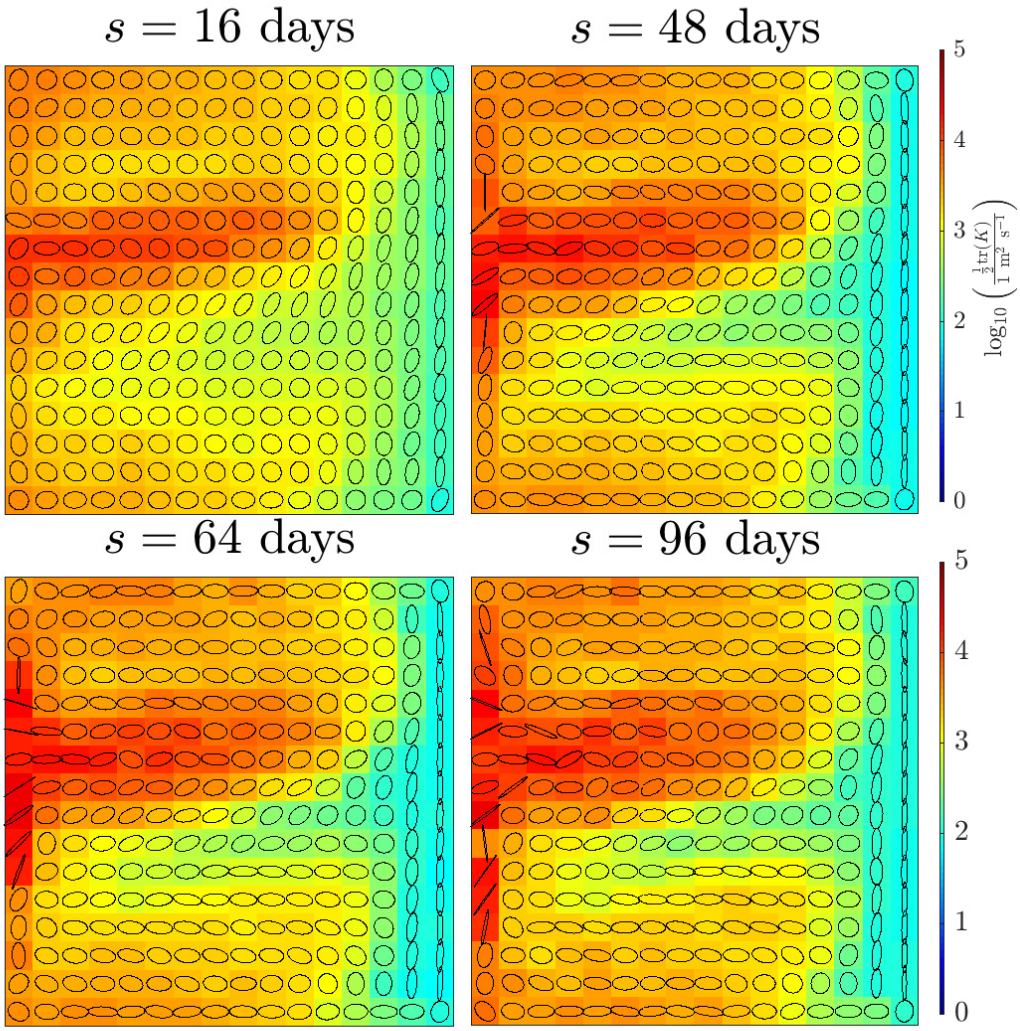

Figure 4.13: MAP estimates of the middle layer diffusivity at various sampling intervals. Same as figure 4.5 but with 2704 particles.

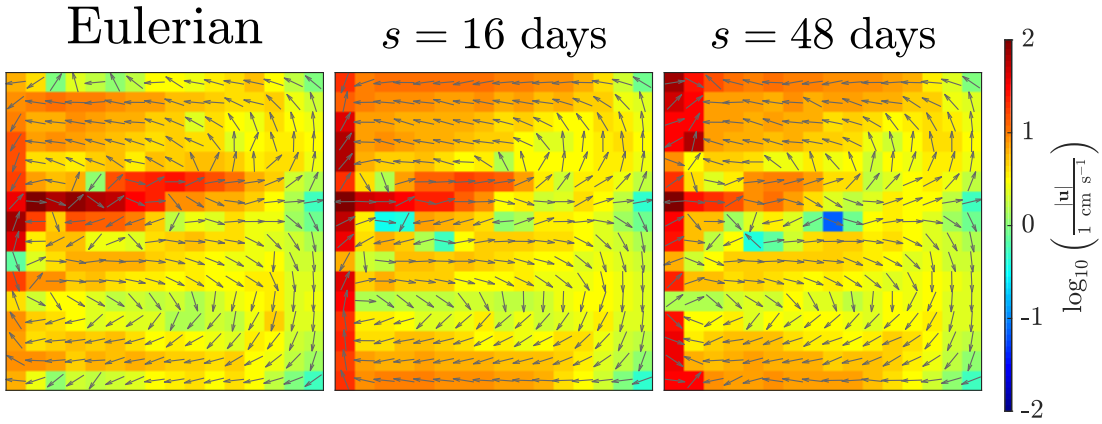

Figure 4.14: Same as the top panels in figure 4.2 but in the top layer. Left panel: Eulerian Mean velocity. Middle and right panel: MAP estimates of the inferred mean velocity at sampling intervals of 16 days and 48 days. The magnitude of the velocity is shown using a logarithmic colour scale, and the velocity direction is indicated by equal-length arrows.

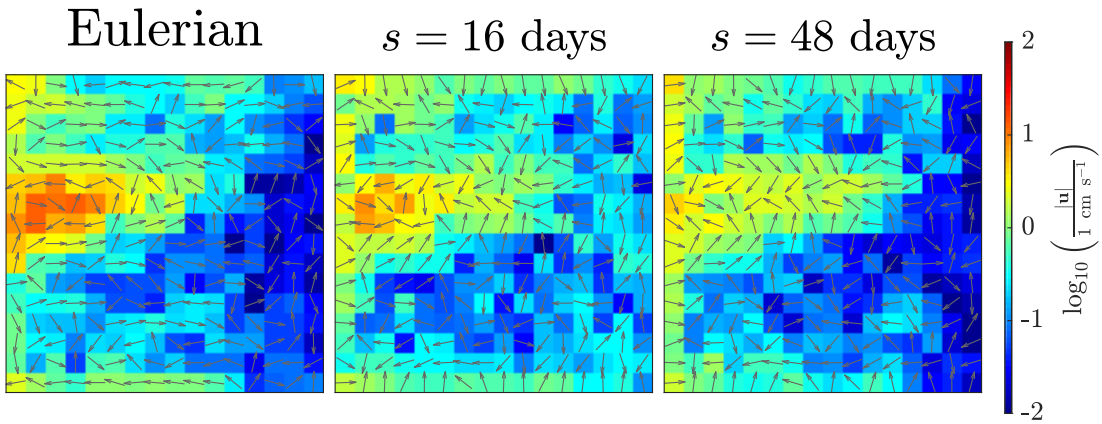

Figure 4.15: Same as figure 4.14 but in the bottom layer.

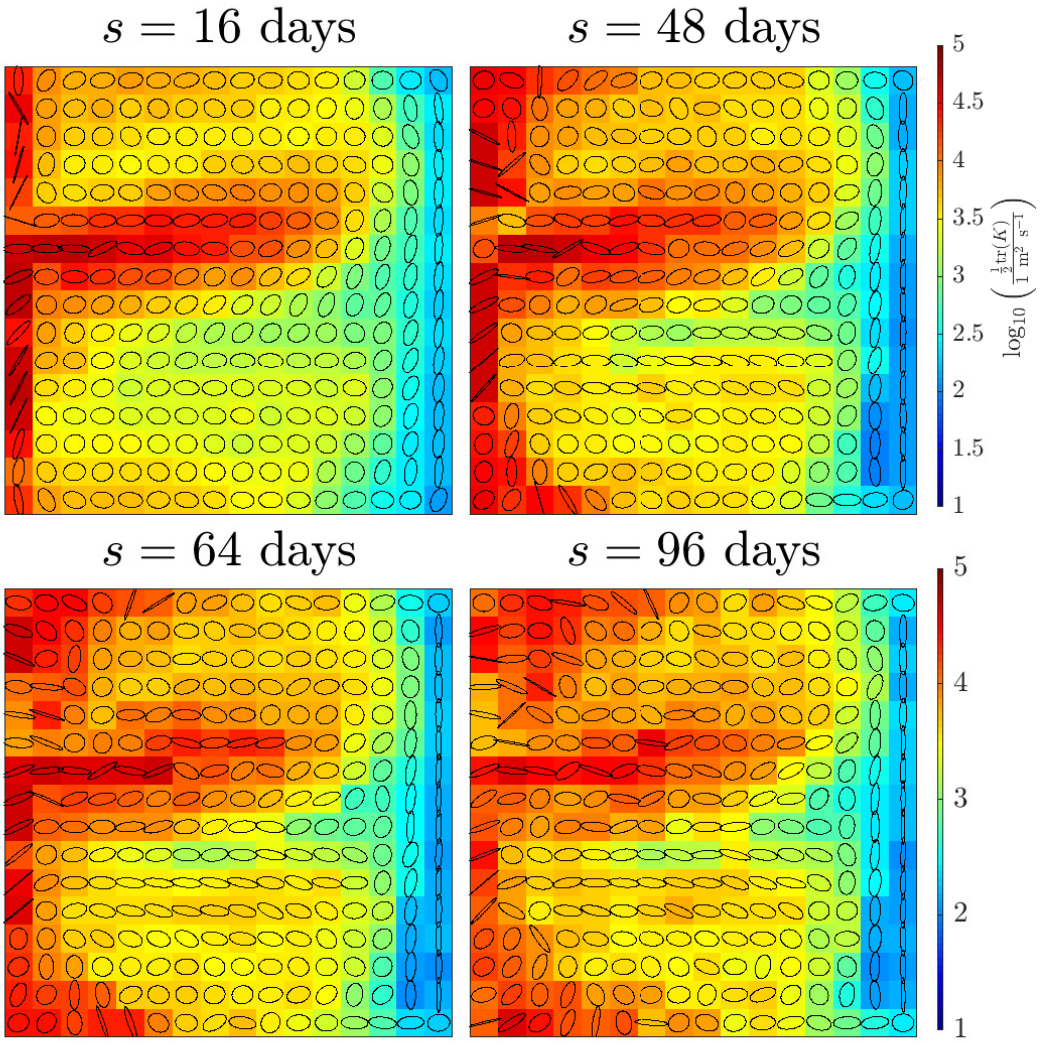

Figure 4.16: MAP estimate of the top layer diffusivity field at different sampling intervals s. The (logarithmic) colour scale gives the half trace of the diffusivity tensor K, which is also the arithmetic mean of the eigenvalues, to characterise the magnitude of diffusivity.

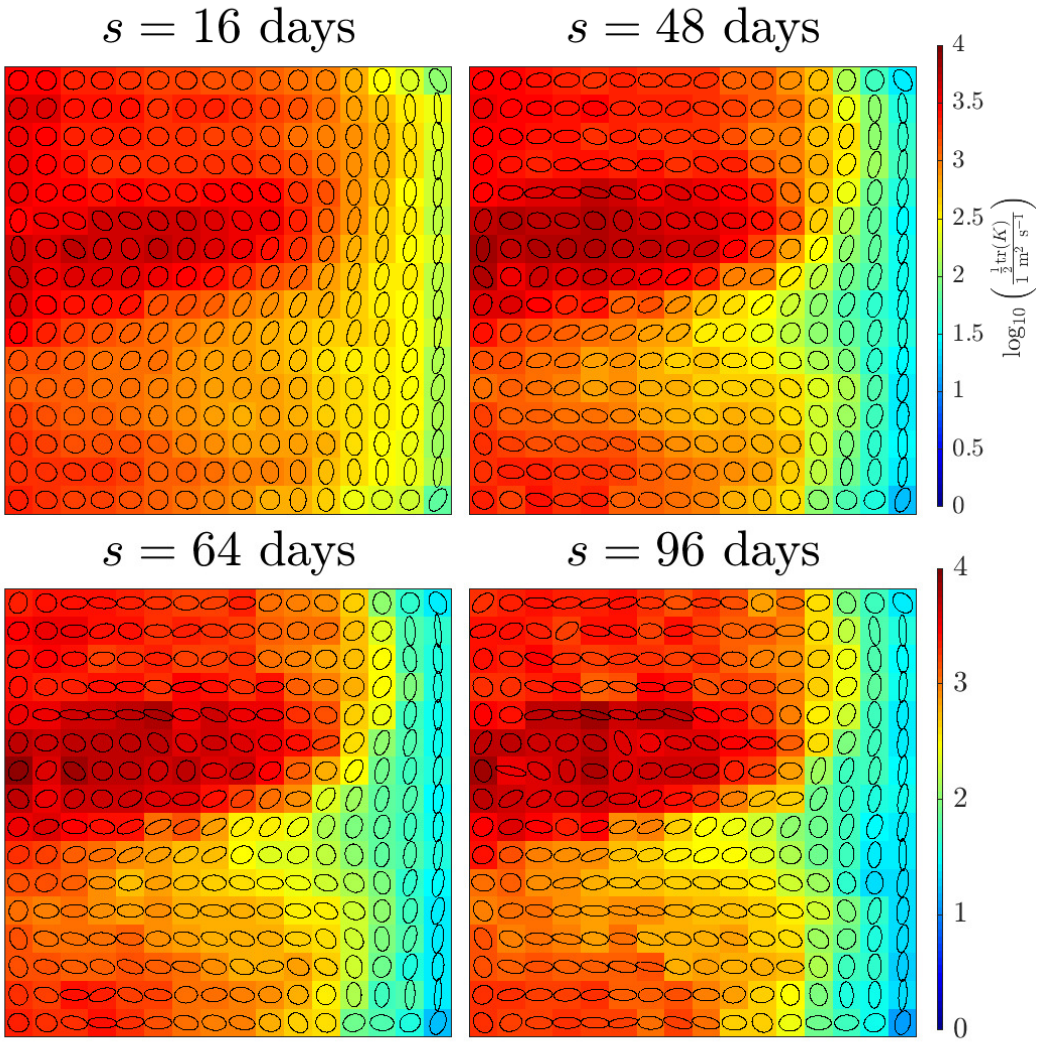

Figure 4.17: MAP estimates of diffusivity at various sampling intervals. Same as figure 4.16 but in the bottom layer. Note here a different colour scale is applied.

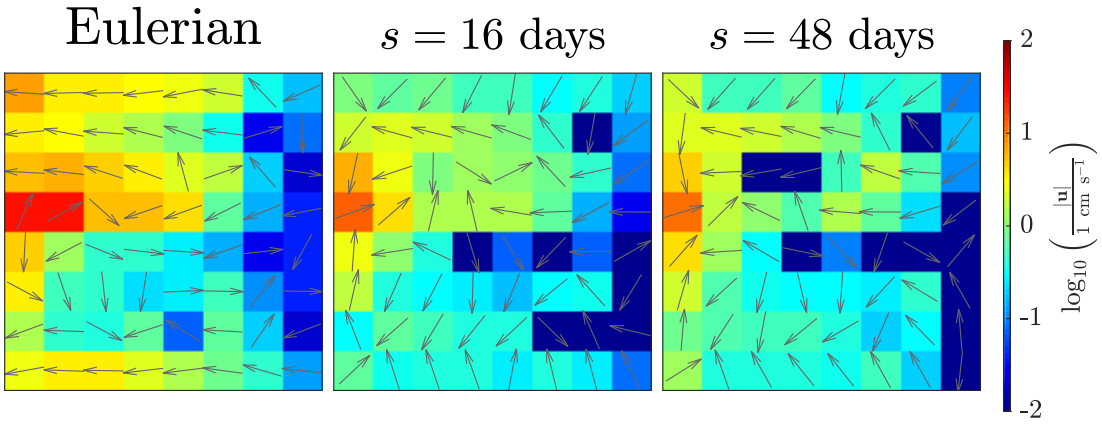

Figure 4.18: Same as the top panels in figure 4.2 but on a 8×8 cell array. Left panel: Eulerian Mean velocity. Middle and right panel: MAP estimates of the inferred mean velocity at sampling intervals of 16 days and 48 days. The magnitude of the velocity is shown using a logarithmic colour scale, and the velocity direction is indicated by equal-length arrows.

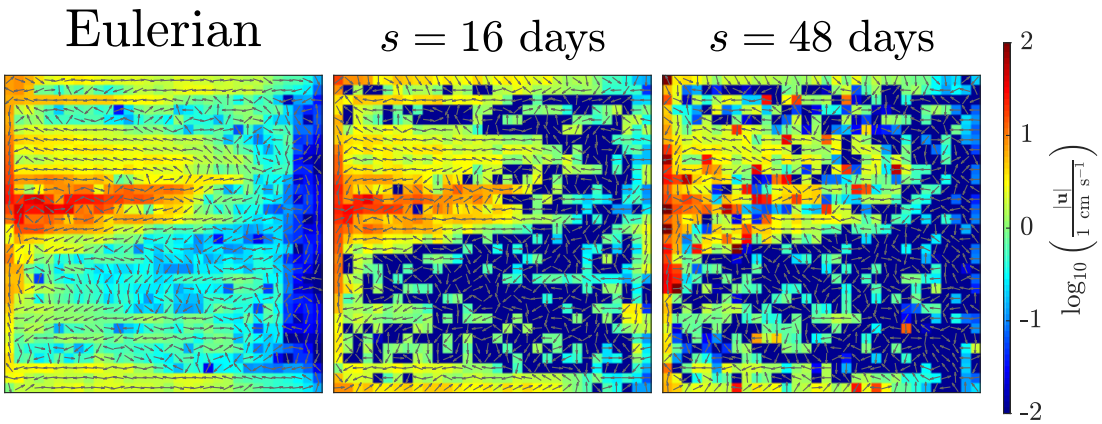

Figure 4.19: Same as figure 4.18 but on a 32×32 cell array.

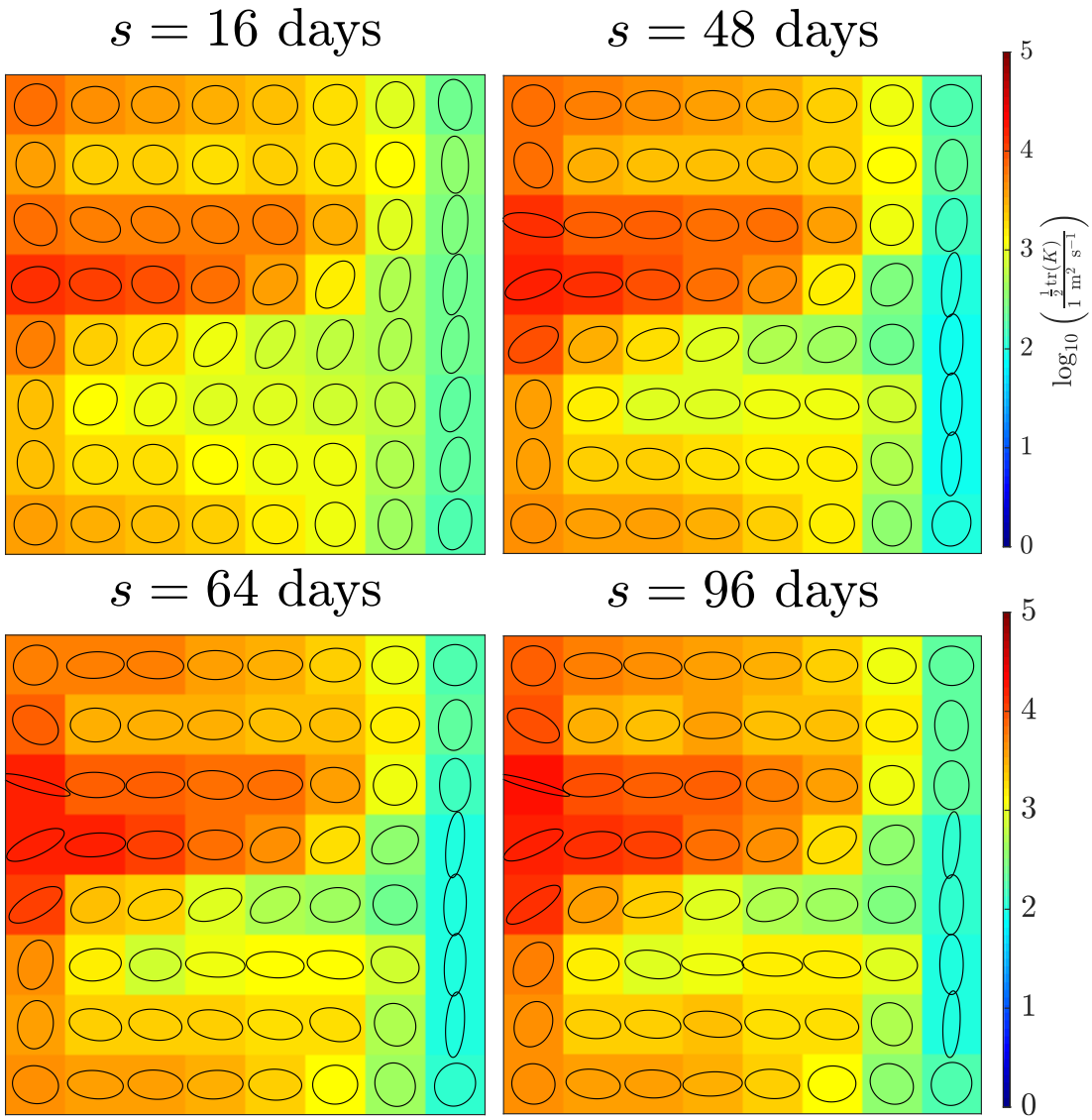

Figure 4.20: MAP estimate of the middle layer diffusivity field on a 8×8 cell arrays at different sampling intervals s. The (logarithmic) colour scale gives the half trace of the diffusivity tensor K, which is also the arithmetic mean of the eigenvalues, to characterise the magnitude of diffusivity.

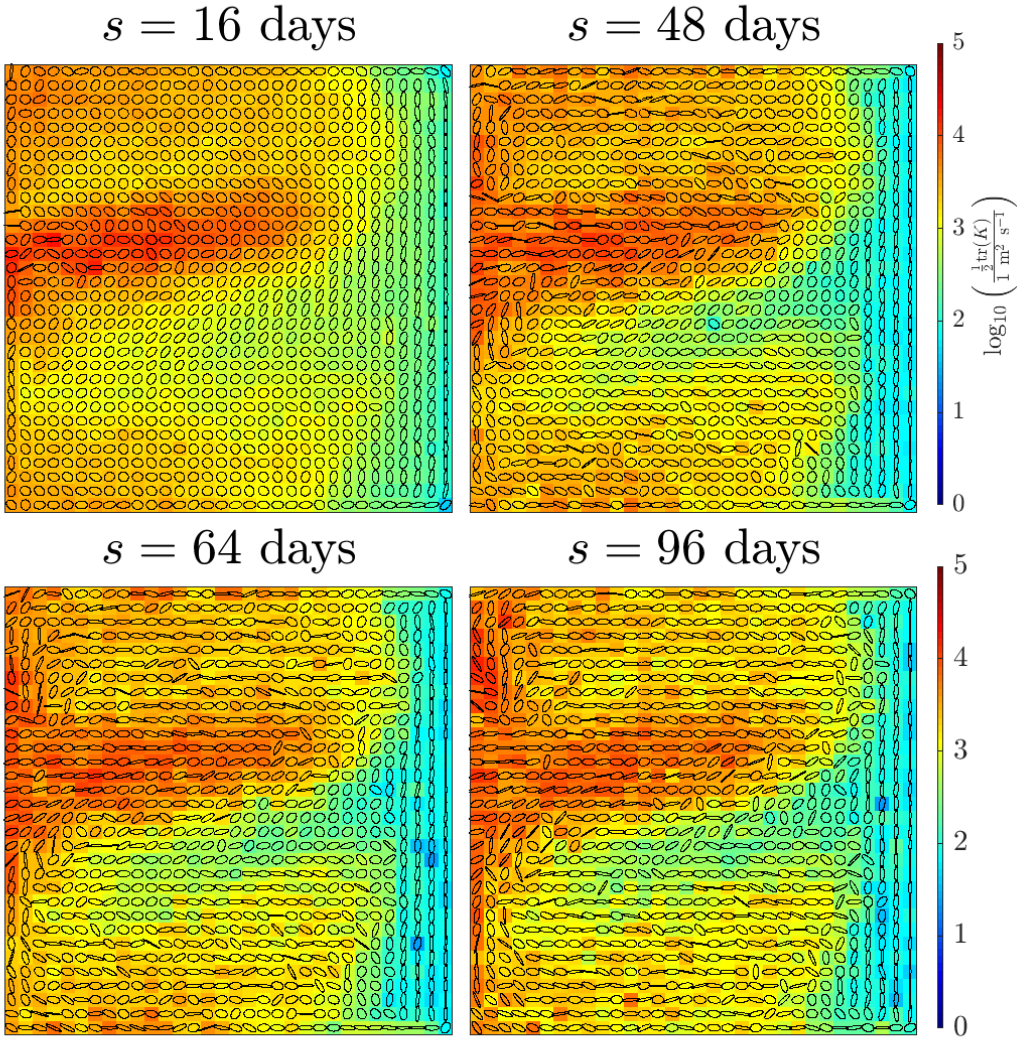

Figure 4.21: Same as figure 4.20 but on a 32×32 cell arrays.

This approach is applied to a modest amount of 676 decade-long Lagrangian particle trajectories in a three-layer quasigeostrophic double gyre flow, with a focus on the middle layer dynamics. On a 16×16 cell array, the localised inference is capable of resolving the spatial variability in the velocity and diffusivity fields, and constraining both estimates within an order of magnitude in most of the domain. The values found become relatively insensitive to the sampling time when this exceeds 30 days and are roughly comparable to the Davis (1987) diffusivity.

The localisation is however a double-edged sword. Particles observed in a given cell, of size 240 km comparable to the length scale of eddies, do not necessarily experience the same flow velocity and diffusivity over the considered sampling interval. The cells near the western boundary and in the region of the separated jet are especially problematic as particles typically straddle several cells at the decorrelation time scale, despite the relatively large size of cells of 240 km. This limitation may jeopardise the hope to apply the localised approach to infer fields at a higher resolution, especially when applied to ocean basin such as in the Gulf stream or Antarctic Circumpolar Current where the flows are fast and energetic.

It is worth examining if the approach is applicable to infer ocean diffusivity in the area with weak advection. However, the mixing rate in such areas is typically not high, and so diffusivity estimates in such an area may attract much interest for further use in, for example, climate model parameterisation or oceanographic studies. For a generally applicable diffusivity diagnostic, it is essential to resolve the locality limitation. This motivates the studies in Chapter 5 where globally defined fields are inferred.

Chapter 5

Global inference applied to idealised oceanic flows

In Chapter 4, the Bayesian inference is implemented using a localised approach for a quasigeostrophic double gyre flow. The domain is divided into multiple cells to independently infer locally defined velocity and diffusivity fields. This allows likelihood evaluation by closed-form expressions of the transition density. The localised approach is, however, limited by the locality assumption that the particle displacements should be small relative to the cell-size and variability of the flow properties.

Over a decorrelation time scale where Fokker–Planck models become legitimate, ocean particles could have travelled large distances. Cells large enough to accommodate the long displacement may, on the other hand, be too big to resolve the spatial variability of the mean velocity and diffusivity fields.

A brute-force resolution to the problem is to solve the Fokker–Planck equation (2.17) for the transition probabilities under globally-defined fields, which we denote as *primitive* inference. However, this involves solving a Fokker–Planck equation for each transition data point, which is computationally prohibitive for the considered trajectory data.

In this chapter we develop a *data coarse-grained* method to strategically discard trajectory information and infer globally-defined fields at a computational cost independent of the data size. We show also that, in the large data limit, the maximum likelihood estimate in the data coarse-grained inference is consistent with that in the primitive inference.

Section 5.1 introduces the data coarse-grained framework. Section 5.2 gives the technical derivations for the coarse-grained inference and shows its relation with the primitive inference. Section 5.3 and 5.4 apply the data coarse-grained method to infer the velocity and diffusivity fields using synthetic particle trajectories respectively from an SDE and the quasigeostrophic double gyre considered in Chapter 4.

5.1 Methods

5.1.1 A data coarse-grained framework

The inference presented in chapter 4 can be formally expressed through the *primitive* likelihood of the full trajectory data R observed at discrete time

$$R = \{(i, \mathbf{X}_i(t_j), t_j) : 1 \leq i \leq N, 1 \leq j \leq P\},$$

is the product of the $N(P-1)$ transition densities $\pi(\cdot,\cdot|\cdot)$

$$p(R|\boldsymbol{\theta}) = \prod_{i=1}^{N}\prod_{j=1}^{P-1} \pi(\mathbf{X}_i(t_{j+1}), t_{j+1} - t_j | \mathbf{X}_i(t_j)). \tag{5.1}$$

Each transition density $\pi(\cdot,\cdot|\cdot)$ satisfies a Fokker–Planck equation with an initial Dirac–delta profile

$$\frac{\partial \pi}{\partial t} + \nabla \cdot (\mathbf{U}\pi) = \nabla \cdot (\mathsf{K}\nabla\pi)\,, \tag{5.2a}$$

$$\pi\left(\mathbf{x}, 0|\mathbf{x}_0\right) = \delta\left(\mathbf{x} - \mathbf{x}_0\right). \tag{5.2b}$$

Inference based on (5.1) is called a *primitive* inference in this thesis.

A closed-form solution to $\pi(\cdot,\cdot|\cdot)$ is unavailable in general mean flow velocity and diffusivity fields. The highly irregular Dirac–delta profile further complicates the numerical evaluation of $\pi(\cdot,\cdot|\cdot)$. In a typical ocean basin, the number of particles N and temporal observations P are both in the order of $\mathcal{O}(100)$ (e.g Argo, 2000). These imply a total of $N(P-1) \sim \mathcal{O}(10^4)$ independent Fokker–Planck equations should be solved per likelihood evaluation. If a brute-force approach is taken to numerically solve all Fokker–Planck equations, a single likelihood evaluation is computationally intensive, let alone for the tens of thousands of them needed for Metropolis–Hastings sampling.

We propose a 'binning' approach to coarse-grain the trajectory data. The domain is divided into a set of non-overlapping cells. Each *transition*, a pair of successive positions data over a uniform interval $s = t_{j+1} - t_j$, is attributed to the cell where the start position is found. The precise location of the start position is then replaced by the cell index c and collectively form the data set of forward transitions R_f^c for each cell c

$$R_f^c = \{(c, \mathbf{X}_i(t_{j+1}), s) | \mathbf{X}_i(t_j) \in c, 1 \le i \le N, 1 \le j \le P\}. \tag{5.3}$$

Note here that the full distinguishability of individual particles is lost. Each 3-tuple data $(c, \mathbf{X}_i(t_{j+1}), s)$ is seen as a sample of $\mathbf{X}(t)$ under the SDE (2.16) at time $t = s$, whose exact initial position $\mathbf{X}(0)$ lies in cell c.

An SDE (2.16) with an incompressible mean velocity field has a uniform invariant distribution (see appendix E). This motivates modelling the initial position $\mathbf{X}(0)$ by a cell-wise uniform probability distribution. The probability density reads $\rho_c(\mathbf{y}) = \mathbf{1}_c(\mathbf{y})/|c|$, where $\mathbf{1}_c$ and $|c|$ are the indicator function for and area of cell c. The likelihood to observe an end position $\mathbf{X}_i(t_{j+1})$ given $\mathbf{X}_i(t_j) \in c$ is then given by integrating the starting position in the transition density $\pi(\cdot,\cdot|\mathbf{y})$ with respect to $\rho_c(\mathbf{y})$

$$\phi_c(\mathbf{x}, s) = \int \pi(\mathbf{x}, s|\mathbf{y})\rho_c(\mathbf{y})\ \mathrm{d}\mathbf{y} = \frac{1}{|c|}\int \pi(\mathbf{x}, s|\mathbf{y})\mathbf{1}_c(\mathbf{y})\ \mathrm{d}\mathbf{y}, \tag{5.4}$$

and evaluating at the end position $\mathbf{x} = \mathbf{X}_i(t_{j+1})$. Note that the transition probability $\phi_c(\mathbf{x}, t)$ solves the Fokker–Planck equation

$$\frac{\partial \phi_c}{\partial t} + \nabla \cdot (\mathbf{U}\phi_c) = \nabla \cdot (\mathsf{K}\nabla\phi_c)\,, \tag{5.5a}$$

$$\phi_c(\mathbf{x}, 0) = \frac{1}{|c|}\mathbf{1}_c(\mathbf{x}), \tag{5.5b}$$

differing from the primitive (5.2) by the initial condition. We define the likelihood of the coarse-grained data set for cell c to be

$$q(R_f^c|\boldsymbol{\theta}) = \prod_{\mathbf{X}_i(t_j)\in c} \phi_c(\mathbf{X}_i(t_{j+1}), s). \tag{5.6}$$

For uniform sampling interval s, Equation (5.6) indicates only one Fokker–Planck equation is needed to be solved per likelihood evaluation, regardless of the amount of coarse-grained data points in R_f^c. We note that formally the likelihood (5.6) should no longer be seen as a likelihood for the trajectory data, but on a set of uncorrelated transition data. For the stochastic differential equation (2.16), the transition data derived from a trajectory are necessarily independently thanks to the Markov property so that the trajectory and transition data are equivalent. For non-Markovian processes, extra caution must be taken to account for both the spatial and temporal correlation among transitions, when devising the likelihood for the transition data converted from trajectory data.

The data sets R_f^c in different cells are independent, so that the likelihood $q(R|\boldsymbol{\theta})$ for the full coarse-grained trajectory data is given by the joint-product of likelihoods from each cell

$$q(R|\boldsymbol{\theta}) = \prod_c q(R_f^c|\boldsymbol{\theta}) = \prod_c \prod_{\mathbf{X}_i(t_j)\in c} \phi_c(\mathbf{X}_i(t_{j+1}), s). \tag{5.7}$$

By reordering the products, the likelihood can be equivalently interpreted as a modified likelihood on the original data set R

$$q(R|\boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{j=1}^{P-1} \sum_c \mathbf{1}_c(\mathbf{X}_i(t_j))\phi_c(\mathbf{X}_i(t_{j+1}), s), \tag{5.8}$$

through the indicator function $\mathbf{1}_c(\cdot)$ to partition the start positions in each transition. We call the likelihood (5.8) the *data coarse-grained* likelihood, from which *data coarse-grained* inference is performed. We emphasise that once we fix the cells, the computational cost per likelihood evaluation become independent of data size.

The physical interpretation of the data coarse-grained inference is intuitive. In each cell, particles are collectively seen as sample particles of a *test tracer* $\phi_c(\mathbf{x}, 0) = \mathbf{1}_c(\mathbf{x})/|c|$. The profile of the test tracer $\phi_c(\mathbf{x}, s)$ tells us the probability distribution of the tracer particles after a time interval s, if initially they are uniformly distributed in the cell and consequently undergo transports driven by a velocity $\mathbf{U}$ and diffusivity K field. In this sense, the data coarse-grained approach essentially infers the fields by comparing the particles distribution and the test tracer profile at time s after the release. We show in figure 5.1 a schematic sketch of the data coarse-grained procedures.

For the data coarse-grained inference, the number of Fokker–Planck equations needed to be solved per likelihood evaluation depends only on the number of test tracers used. This drastically alleviates the computational burden for Bayesian inference in contrast with a primitive inference. To partition an ocean basin such as the Northern Atlantic using cells of side length $\sim \mathcal{O}(100$ km) will merely introduce a total amount of $\mathcal{O}(100)$ cells, whereas the number of particle transitions is of size typically $\mathcal{O}(10^4)$. The computation burden for the data coarse-grained inference would therefore be reduced by at least one to two orders of magnitude.

In the next section 5.2, we show that the primitive likelihood $p(R|\boldsymbol{\theta})$ in (5.1) and the coarse-grained $q(R|\boldsymbol{\theta})$ in (5.8) share a common maximiser $\boldsymbol{\theta}_*$ in the limit of large data. We also prove that, if equipped with the same prior, the posterior induced by the coarse-grained likelihood $q(R|\boldsymbol{\theta})$ is no sharper than the exact one $p(R|\boldsymbol{\theta})$, suggesting the data coarse-grained treatment would not introduce false certainty.

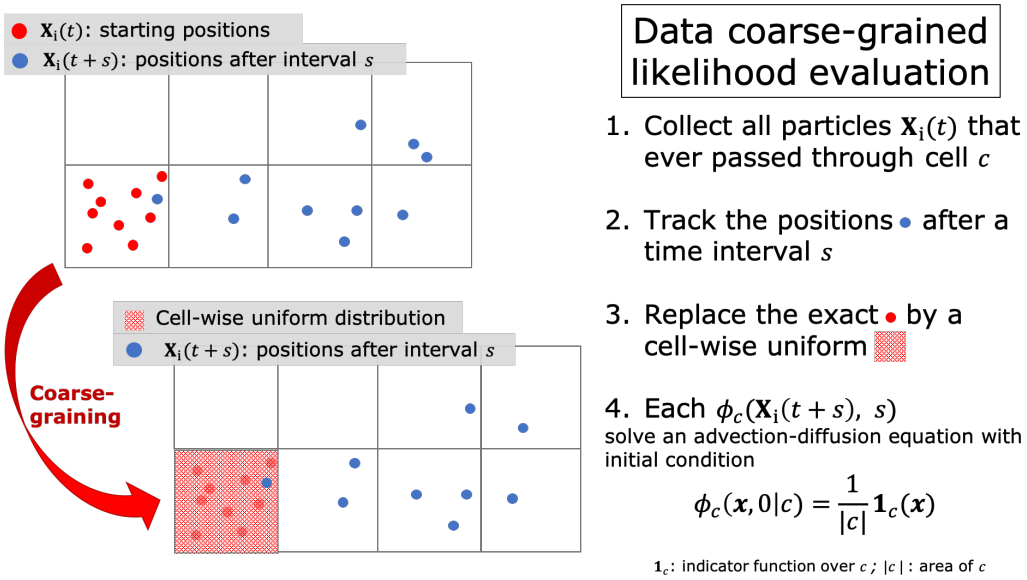

Figure 5.1: A schematic explanation of the data coarse-grained likelihood in a cell c.

5.2 Validity of the coarse-grained inference

5.2.1 Notations

Consider a pair of successive observation $(\mathbf{X}(t), \mathbf{X}(t+s))$ sampled at a uniform interval s where the start position $\mathbf{X}(t)$ follows a distribution with a known invariant density $\rho(\cdot)$. Suppose the time evolution of the process $\mathbf{X}(t)$ is governed by an SDE parameterised by $\boldsymbol{\theta}$ with a transition density for $\mathbf{X}(t+s)|(\mathbf{X}(t)=\mathbf{y})$ given by $\pi(\cdot, s|\mathbf{y}, \boldsymbol{\theta})$. The joint density of $(\mathbf{X}(t)=\mathbf{y}, \mathbf{X}(t+s)=\mathbf{x})$ is the product of the transition density and initial distribution

$$\phi(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) = \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta})\rho(\mathbf{y}), \tag{5.9}$$

where here the dependence on sampling interval s is omitted for brevity.

The closed domain is partitioned into a set of finite-sized cells. For each cell c, we seek to derive an expression for the probability density of $\mathbf{X}(t+s)$ in the event of $\mathbf{X}(t) \in c$, which is the likelihood function associated with a data point in the coarse-grained inference.

Firstly consider the conditional probability for $\mathbf{X}(t)$, given that $\mathbf{X}(t)$ lies within the cell c. The density for $\mathbf{X}(t)|(\mathbf{X}(t) \in c)$ is determined by normalising the density $\rho(\mathbf{y})$ over the probability of $\mathbf{X}(t) \in c$, denoted $\overline{\rho}_c$, such that

$$\begin{aligned}\rho_c(\mathbf{y}) = \rho(\mathbf{y}|\mathbf{X}(t) \in c) &= \frac{\rho(\mathbf{y})\mathbf{1}_c(\mathbf{y})}{\int \rho(\mathbf{y})\mathbf{1}_c(\mathbf{y})\mathrm{d}\mathbf{y}} \\ &= \frac{1}{\overline{\rho}_c}\rho(\mathbf{y})\mathbf{1}_c(\mathbf{y}), \qquad \text{where } \overline{\rho}_c = \int \rho(\mathbf{y})\mathbf{1}_c(\mathbf{y})\mathrm{d}\mathbf{y},\end{aligned}$$

Secondly consider the conditional joint probability of $(\mathbf{X}(t), \mathbf{X}(t+s))$, given that $\mathbf{X}(t)$ lies within the cell c. The density for $(\mathbf{X}(t), \mathbf{X}(t+s))|(\mathbf{X}(t) \in c)$ is determined by normalising the full joint density (5.9) over the probability of $\mathbf{X}(t) \in c$, such that

$$\begin{aligned}\phi_c(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) &= \phi(\mathbf{y}, \mathbf{x}|\mathbf{X}(t) \in c, \boldsymbol{\theta}) \\ &= \frac{1}{\overline{\rho}_c}\phi(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})\mathbf{1}_c(\mathbf{y}) = \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta})\rho_c(\mathbf{y}).\end{aligned} \tag{5.10}$$

Marginalising $\phi_c(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})$ over $\mathbf{y}$ gives the probability density of $(\mathbf{X}(t+s) = \mathbf{x}, \mathbf{X}(t) \in c)$

$$\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}) = \int \phi_c(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})\mathrm{d}\mathbf{y} = \int \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta})\rho_c(\mathbf{y})\mathrm{d}\mathbf{y}, \tag{5.11}$$

which satisfies the Fokker–Planck equation

$$\frac{\partial \overline{\phi}_c}{\partial t} + \nabla \cdot \left(\mathbf{U}\overline{\phi}_c\right) = \nabla \cdot \left(\mathsf{K}\nabla\overline{\phi}_c\right), \tag{5.12a}$$

$$\overline{\phi}_c(\mathbf{x}, 0) = \rho_c(\mathbf{x}). \tag{5.12b}$$

Thirdly consider the conditional probability on $\mathbf{X}(t)$ given that $\mathbf{X}(t+s) = \mathbf{x}$ and $\mathbf{X}(t) \in c$. The density for $\mathbf{X}(t)|(\mathbf{X}(t+s) = \mathbf{x}, \mathbf{X}(t) \in c)$ is determined by normalising the conditional joint density for $(\mathbf{X}(t), \mathbf{X}(t+s))$ in (5.10) over the marginal probability density of $(\mathbf{X}(t+s), \mathbf{X}(t) \in c)$ in (5.11), such that

$$\mu_c(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \frac{\phi_c(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})}{\int \phi_c(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})\mathrm{d}\mathbf{y}} = \frac{\phi_c(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})}{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta})}. \tag{5.13}$$

Note now $\int \mu_c(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\mathrm{d}\mathbf{y} = 1$ for any $\mathbf{x}$. μ_c allows one to trace back the precise probability of an initial position $\mathbf{X}(t) = \mathbf{y}$ within cell c, from knowledge of the final position $\mathbf{X}(t+s) = \mathbf{x}$. This quantity plays a significant role in analysing the discrepancy between the primitive and data coarse-grained inference in section 5.2.4.

5.2.2 Primitive log-likelihood in the large data limit

In this section we consider a single trajectory evolved under an SDE with the mean velocity and diffusivity fields parameterised by an exact $\boldsymbol{\theta}^*$. We furthermore assume the SDE is ergodic and supports an invariant distribution with density $\rho(\cdot)$.

Consider a single trajectory $R = \{(\mathbf{X}(t_j), t_j)\}$ sampled at a uniform time interval $s = t_{j+1} - t_j$. The log-likelihood in the primitive inference is a sum of log-transition density $\log(\pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta}))$ for velocity and diffusivity fields parametrised by $\boldsymbol{\theta}$

$$\log p\left(R|\boldsymbol{\theta}\right) = \sum_{j=1}^{N} \log \pi(\mathbf{X}(t_{j+1}), s|\mathbf{X}(t_j), \boldsymbol{\theta}). \tag{5.14}$$

The ergodic assumption of the SDE implies that (see e.g. Pavliotis, 2014), as the length of trajectory N increases, the distribution of the $\mathbf{X}(t_j)$ converges (almost surely) to the invariant distribution, so that the joint density pair of successive record $(\mathbf{X}(t_j), \mathbf{X}(t_{j+1})) = (\mathbf{y}, \mathbf{x})$ converges (almost surely) to $\phi(\mathbf{y}, \mathbf{x}) = \rho(\mathbf{y})\pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta}^*)$. It follows that the log-likelihood converges (almost surely) to a double integral

$$\begin{aligned}\log p\left(R|\boldsymbol{\theta}\right) &\sim N \iint \phi(\mathbf{y}, \mathbf{x}) \log \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}, \\ &= N \int \left(\int \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta}^*) \log \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta})\mathrm{d}\mathbf{x} \right) \rho(\mathbf{y})\mathrm{d}\mathbf{y},\end{aligned} \tag{5.15}$$

The $\mathbf{x}$-integral can be expressed in terms of a Kullback–Leibler divergence $D_{\mathrm{KL}}\left(\cdot||\cdot\right)$ and Shannon entropy $H(\cdot)$

$$\begin{aligned}&\int \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta}^*) \log \pi(\mathbf{x}, s|\mathbf{y}, \boldsymbol{\theta})\mathrm{d}\mathbf{x} \\ =& - D_{\mathrm{KL}}\left(\pi(\cdot, s|\mathbf{y}, \boldsymbol{\theta}^*)||\pi(\cdot, s|\mathbf{y}, \boldsymbol{\theta})\right) - H(\pi(\cdot, s|\mathbf{y}, \boldsymbol{\theta}^*)),\end{aligned}$$

where

$$D_{\mathrm{KL}}\left(\pi(\cdot,s|\mathbf{y},\boldsymbol{\theta}^*)||\pi(\cdot,s|\mathbf{y},\boldsymbol{\theta})\right) = -\int \pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta}^*)\log\left(\frac{\pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta})}{\pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta}^*)}\right)\mathrm{d}\mathbf{x},$$
$$H(\pi(\cdot,s|\mathbf{y},\boldsymbol{\theta}^*)) = -\int \pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta}^*)\log\pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta}^*)\mathrm{d}\mathbf{x}.$$

Keeping only the terms that depend on $\boldsymbol{\theta}$, this leads to

$$p\left(R|\boldsymbol{\theta}\right) \asymp \exp\left(-N\int D_{\mathrm{KL}}\left(\pi(\cdot,s|\mathbf{y},\boldsymbol{\theta}^*)||\pi(\cdot,s|\mathbf{y},\boldsymbol{\theta})\right)\rho(\mathbf{y})\mathrm{d}\mathbf{y}\right),$$

where the log-equality sign '$\asymp$' is defined

$$a(N)\asymp b(N) \iff \lim_{N\to\infty}\log a(N) = \lim_{N\to\infty}\log b(N). \tag{5.17}$$

Since the Kullback–Leibler divergence is non-negative and vanishes only when the two arguments are equal, if the transition density $\pi(\cdot,s|\mathbf{y},\boldsymbol{\theta})$ is injective with respect to $\boldsymbol{\theta}$, the unique likelihood maximiser will converge to the exact parameter in the limit of infinite number of position records $N\to\infty$.

In the preceding analysis, the partition of the domain is not considered. To examine the likelihood associated with each cell, we split the $\mathbf{y}$-integral by cells and apply the formula (5.10)

$$\begin{aligned}\int \phi(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\log\pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta})\mathrm{d}\mathbf{y} &= \sum_c \int \phi(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\log\pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta})\mathbf{1}_c(\mathbf{y})\mathrm{d}\mathbf{y},\\ &= \sum_c \overline{\rho}_c\int \phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\log\pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta})\mathrm{d}\mathbf{y}.\end{aligned}$$

Noting $\log\pi(\mathbf{x},s|\mathbf{y},\boldsymbol{\theta}) = \log\left(\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})/\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\right) + \log\left(\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)/\rho_c(\mathbf{y})\right)$ from equation (5.10) and introducing the Kullback–Leibler divergence $D_{\mathrm{KL}}\left(\cdot||\cdot\right)$ and conditional entropy $H(\cdot|\cdot)$

$$D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right) = -\iint \phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\log\left(\frac{\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)}\right)\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}, \tag{5.18a}$$
$$H(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)|\rho_c(\cdot)) = -\iint \phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\log\left(\frac{\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)}{\rho_c(\mathbf{y})}\right)\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}, \tag{5.18b}$$

the exact log-likelihood (5.15) can be expressed as a sum of Kullback–Leibler divergence and conditional entropy weighted by $\overline{\rho}_c$

$$\log p\left(R|\boldsymbol{\theta}\right) \sim -N\sum_c \overline{\rho}_c\left[D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right) + H(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)|\rho_c(\cdot))\right].$$

Keeping only the terms that depends on $\boldsymbol{\theta}$, the likelihood $p\left(R|\boldsymbol{\theta}\right)$ is proportional to

$$p\left(R|\boldsymbol{\theta}\right) \asymp \exp\left(-N\sum_c \overline{\rho}_c D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right)\right). \tag{5.19}$$

We emphasise that the form (5.19) is a limiting case of the primitive likelihood (5.1) on the assumptions of large data and data-model consistency. The global likelihood maximiser $\boldsymbol{\theta}^*$ to (5.19) is also the global likelihood maxmimiser to (5.1).

5.2.3 Coarse-grained log-likelihood in the large data limit

Now consider the coarse-grained inference. The invariant distribution $\rho(y)$ is assumed to be known a priori. Using the renormalised $\rho_c(\cdot)$ defined for each of the cell c as the initial condition of Fokker–Planck solver (see equation (5.11)), the data coarse-grained likelihood of the transition data derived from the full trajectories R is given by

$$\log q\left(R|\boldsymbol{\theta}\right) = \sum_{j=1}^{N} \sum_{c} \mathbf{1}_c\left(\mathbf{X}(t_j)\right) \log \overline{\phi}_c(\mathbf{X}(t_{j+1})|\boldsymbol{\theta}). \tag{5.20}$$

Noting that the transition $(\mathbf{X}(t_j), \mathbf{X}(t_{j+1})$ possesses a joint probability density (5.9), the sum in (5.20) converges (almost surely) to a double integral according to the law of large numbers

$$\begin{aligned}\log q\left(R|\boldsymbol{\theta}\right) &\sim N \iint \sum_{c} \mathbf{1}_c(\mathbf{y}) \phi(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}^*) \log \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}) \mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x} \\ &= N \sum_{c} \overline{\rho}_c \int \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*) \log \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}) \mathrm{d}\mathbf{x}.\end{aligned}$$

With the expression of the joint density (5.9), introducing $\log \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}) = \log\left(\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta})/\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)\right) + \log \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)$, the Kullback–Leibler divergence $D_{\mathrm{KL}}\left(\cdot||\cdot\right)$ and Shannon entropy $H(\cdot)$

$$D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right) = -\int \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*) \log\left(\frac{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta})}{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)}\right) \mathrm{d}\mathbf{x}, \tag{5.21a}$$

$$H(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)) = -\int \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*) \log \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*) \mathrm{d}\mathbf{x}, \tag{5.21b}$$

the expression of the primitive log-likelihood (5.15) can be represented as a weighted sum of Kullback–Leibler divergence and Shannon entropy with weights $\overline{\rho}_c$

$$\log q\left(R|\boldsymbol{\theta}\right) \sim -N \sum_{c} \overline{\rho}_c \left[D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right) + H(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*))\right].$$

Keeping only the terms that depend on $\boldsymbol{\theta}$, the likelihood $q\left(R|\boldsymbol{\theta}\right)$ is given by

$$q\left(R|\boldsymbol{\theta}\right) \asymp \exp\left(-N \sum_{c} \overline{\rho}_c D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right)\right). \tag{5.22}$$

Hence again the likelihood is maximised when $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ such that the Kullback–Leibler divergence $D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right)$ vanishes. Note however that the likelihood maximiser may not be unique.

5.2.4 Difference between the primitive and data coarse-grained likelihood

Now we examine the difference between the primitive likelihood $p\left(R|\boldsymbol{\theta}\right)$ and the data coarse-grained likelihood $q\left(R|\boldsymbol{\theta}\right)$. From (5.19) and (5.22) it suffices to compare the Kullback–Leibler divergence terms in each cell c.

Firstly recall from equation (5.11) that $\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)$ is a marginal of $\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)$ which allows the term $D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right)$ in the coarse-grained likelihood to be expressed in terms of a double integral

$$\begin{aligned} D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right) &= -\int\left(\int \phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\mathrm{d}\mathbf{y}\right)\log\left(\frac{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta})}{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)}\right)\mathrm{d}\mathbf{x}, \\ &= -\iint \phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\log\left(\frac{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta})}{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)}\right)\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}, \end{aligned}$$

and the difference between $D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right)$ and $D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right)$ follows

$$\begin{aligned} &D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right) - D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right) \\ &\qquad = \iint \phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)\log\left(\frac{\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}^*)}\,\frac{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)}{\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta})}\right)\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x} \\ &\qquad = \iint \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)\mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}^*)\log\left(\frac{\mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{\mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}^*)}\right)\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}, \end{aligned} \tag{5.23}$$

where the last line follows from the definition (5.13) of the conditional density $\mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) = \phi_c(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})/\overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta})$.

Now note that, for a fixed $\mathbf{x}$, $\mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ is a probability density with respect to $\mathbf{y}$ so that the $\mathbf{y}$-integral in (5.23) is the negative Kullback–Leibler divergence of $\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta}^*)$ from $\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta})$

$$D_{\mathrm{KL}}\left(\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta}^*)||\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta})\right) = -\int \mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}^*)\log\left(\frac{\mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{\mu_c(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}^*)}\right)\mathrm{d}\mathbf{y}.$$

This leads to

$$\begin{aligned} &D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right) - D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right) \\ &\qquad = -\int \overline{\phi}_c(\mathbf{x}|\boldsymbol{\theta}^*)D_{\mathrm{KL}}\left(\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta}^*)||\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta})\right)\mathrm{d}\mathbf{x}. \end{aligned} \tag{5.24}$$

As a Kullback–Leibler divergence has a property of being negative, noting $D_{\mathrm{KL}}\left(\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta}^*)||\mu_c(\cdot|\mathbf{x},\boldsymbol{\theta})\right) \geq 0$ leads to the inequality

$$-D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right) \geq -D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right). \tag{5.25}$$

The Kullback-Leiber divergences on both sides of the inequality (5.25) share a common maximiser $\boldsymbol{\theta}^*$ with maximum zero. The left panel of figure 5.2 shows an example schematic of the two Kullback-Leiber divergences. If imposing a uniform prior, and assuming both the primitive likelihood $p\left(R|\boldsymbol{\theta}\right)$ (5.19) and coarse-grained likelihood $q\left(R|\boldsymbol{\theta}\right)$ (5.22) are normalisable with respect to $\boldsymbol{\theta}$, the induced posteriors are proportional to the likelihood functions. The peak values of the posteriors are controlled by normalisation constants, denoted C_{prim} and C_{cg} for the primitive and coarse-grained posterior, are respectively given by

$$\begin{aligned} C_{\mathrm{prim}} &= \frac{1}{\int \exp\left(-N\sum_c \overline{p}_c D_{\mathrm{KL}}\left(\phi_c(\cdot,\cdot|\boldsymbol{\theta}^*)||\phi_c(\cdot,\cdot|\boldsymbol{\theta})\right)\right)\mathrm{d}\boldsymbol{\theta}}, \\ C_{\mathrm{cg}} &= \frac{1}{\int \exp\left(-N\sum_c \overline{p}_c D_{\mathrm{KL}}\left(\overline{\phi}_c(\cdot|\boldsymbol{\theta}^*)||\overline{\phi}_c(\cdot|\boldsymbol{\theta})\right)\right)\mathrm{d}\boldsymbol{\theta}}. \end{aligned}$$

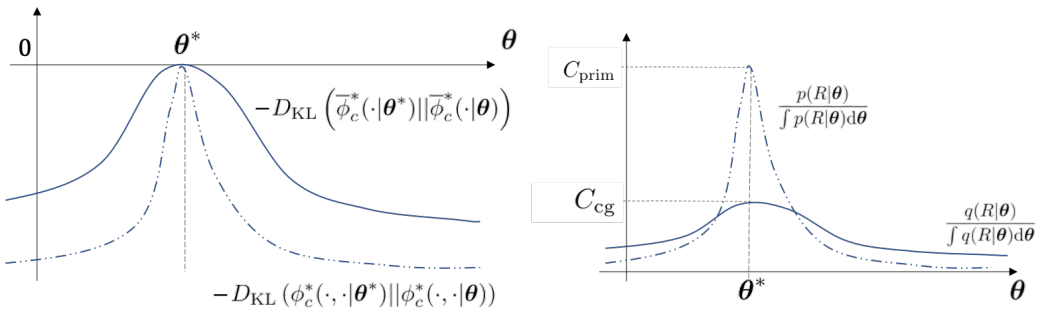

Figure 5.2: Schematic sketches of the Kullback–Leibler divergences and likelihood in the primitive and coarse-grained inference. Solid lines: primitive inference. Dotted lines: coarse-grained inference. Left panel: The Kullback–Leibler divergence. Right panel: The posterior distributions induced by the primitive and coarse-grained likelihoods under a uniform prior.

A direct corollary of the inequality (5.25) is $C_{\mathrm{cg}} \leq C_{\mathrm{prim}}$ is that the primitive posterior has a sharper peak than the coarse-grained posterior, assuming the global maximiser is unique. The right panel of figure 5.2 shows an example schematic of the posterior distributions, where the spread of posterior increases as the peak lowers. This shows the coarse-grained posterior distribution, as the primitive one, collapses onto the exact θ_* and that the uncertainty for a finite data size N is larger for the coarse-grained approach, as one would expect in view of the information discarded.

5.3 Idealised example

In this section, we test the data coarse-grained inference to infer the velocity and diffusivity fields of an SDE, using synthetic trajectory data simulated from it. The test sheds lights on the key properties of the inference.

5.3.1 Configuration

Lagrangian particles trajectories in a square domain of side $l = 3840$ km are simulated. The trajectories are governed by the SDE (2.16) with artificially constructed mean flow $\mathbf{u}(\mathbf{x}) = \nabla^{\perp}\psi(\mathbf{x})$ and eigendecomposed diffusivity $\mathsf{K}(\mathbf{x}) = \mathsf{R}(\phi(\mathbf{x}))\mathrm{diag}(\gamma_1(\mathbf{x}), \gamma_2(\mathbf{x}))\mathsf{R}(\phi(\mathbf{x}))^{\mathrm{T}}$ given by

$$\psi(\mathbf{x}) = u_0 l \exp\left((3x - y)/2l\right) \sin\left(\pi x/l\right) \sin\left(2\pi y/l\right), \tag{5.26a}$$

$$\gamma_1(\mathbf{x}) = \kappa_0 \cos^2\left(\pi(x - 2y)/2l\right), \tag{5.26b}$$

$$\gamma_2(\mathbf{x}) = \kappa_0 \cos^2\left(\pi x/3l\right), \tag{5.26c}$$

$$\phi(\mathbf{x}) = (\pi/2) \sin\left(\pi x/l\right) \sin\left(\pi y/l\right), \tag{5.26d}$$

where u_0 and κ_0 are the characteristic scales of the mean flow and diffusivity, $\mathsf{R}(\cdot)$ is the rotation matrix parameterised by an angle argument and $\mathrm{diag}(a, b)$ is the diagonal matrix with entries a and b.

Parameters used in this example are provided in Table 5.1. The domain size, flow and diffusivity regimes are chosen to yield a double vortex configuration. Reflecting boundary conditions are imposed to keep all particle in the domain, justified as the

Parameter	Symbol	Value(s)
Domain size	l	$3,840$ km
Flow speed scale	u_0	$500/96 \approx 5.21$ cm s^{-1}
Diffusivity scale	κ_0	$10,000$ m^2 s^{-1}
Depth thickness	–	250 m
Particle integration time step size	–	2 hours
Total particle integration time	–	3653 days
Number of particles	N_p	169, 676, 2704
Data sampling interval	s	32 days, 48 days, ... 96 days
Markov Chain Monte Carlo iterations	N_{mh}	$292,000$
Number of independent Markov Chains	M	3
Number of cells per direction	N_c	8
Cell side length	h_c	240 km
Finite-volume solver grid resolution	Δx	60 km
Number of solver grid per direction	–	64
Time step	Δt	adaptive

Table 5.1: Parameters used in the idealised incompressible flow configuration.

mean flow has no cross-boundary flux $\mathbf{u} \cdot \hat{n} = 0$ and the diffusivity tensor is diagonal (in the $x - y$ coordinates) on the boundaries.

Particles trajectories are simulated by solving the SDE (2.16)

$$\mathrm{d}\mathbf{X} = [\mathbf{U}(\mathbf{X}(t)) + \nabla \cdot \mathsf{K}(\mathbf{X}(t))]\,\mathrm{d}t + \sqrt{2\,\mathsf{K}(\mathbf{X}(t))}\mathrm{d}\mathbf{W},$$

with fields given by (5.26) using the Euler–Maruyama method. N_p particles are initially located at the centres of a $\sqrt{N_p} \times \sqrt{N_p}$ uniform square grid and their trajectories are evaluated up to time 10 Julian years using a small time step of 2 hours. The full trajectories are sampled at various sampling interval s to form the data set R for the Bayesian inference.

5.3.2 Parametrisation

The square domain is divided into a $N_c \times N_c = 8 \times 8$ array of square cells, each denoted c and of side length $h_c = l/N_c$. Denote the nodes $(x_i, y_j) = (ih_c, jh_c)$ at cell corners with indices $i, j = 0, 1, \cdots N_c$.

A piecewise bilinear streamfunction Ψ and diffusivity tensor field K are inferred. The degrees of freedom in the streamfunction are the nodal values at cell corners, whereas in the diffusivity are the two eigenvalues $\Gamma_{1,ij}$ and $\Gamma_{2,ij}$ and the tilting angle Φ_{ij} through

$$\Psi(x_i, y_j) = \Psi_{ij}, \tag{5.27a}$$

$$\mathsf{K}(x_i, y_j) = \mathsf{R}(\Phi_{ij}) \begin{pmatrix} \Gamma_{1,ij} & 0 \\ 0 & \Gamma_{2,ij} \end{pmatrix} \mathsf{R}(\Phi_{ij})^{\mathrm{T}}. \tag{5.27b}$$

Each Cartesian components in the tensor K is piecewise bilinearly interpolated to construct the full tensor field. Thus the constituting parameters in the inference are

$$\boldsymbol{\theta} = (\Psi_{ij}, \Gamma_{1,ij}, \Gamma_{2,ij}, \Phi_{ij})^{\mathrm{T}}, \tag{5.28}$$

for all $i, j \in 0, 1, \cdots N_c$. Figure 5.3 shows a schematic of the representation of the fields using the finite-length parameter $\boldsymbol{\theta}$.

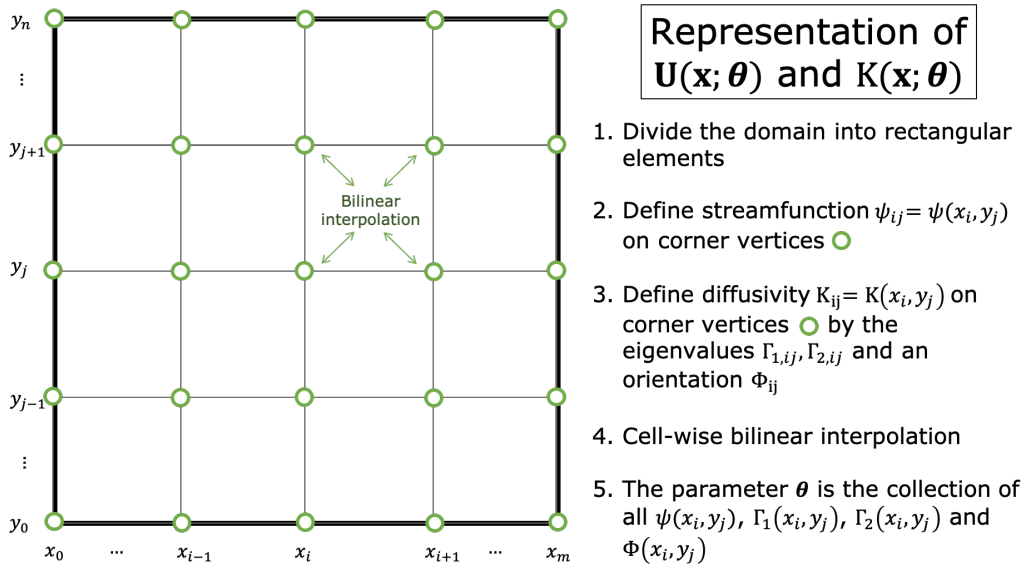

Figure 5.3: Schematic of the parameterisation of piecewise bilinear fields of streamfunction and diffusivity.

To impose the no-flux boundary condition, the streamfunction on boundary nodes are set to be zero. This results in a total number of $(N_c - 1)^2 + 3(N_c + 1)^2 = 292$ components to infer.

5.3.3 Posterior evaluation

The data coarse-grained approach described in section 5.1, which replaces the exact start position in each transition by the cell index, is applied. Each cell defines a test tracer $\mathbf{1}_c(\cdot)/|c|$ to be an initial condition for a Fokker–Planck equation. The solutions of Fokker–Planck equations at time s determine the likelihood (5.6).

A finite-volume method is implemented to solve the Fokker–Planck equations

$$\frac{\partial \phi_c}{\partial t} + \nabla \cdot \left(\phi_c \nabla^{\perp} \Psi \right) = \nabla \cdot (\mathrm{K} \nabla \phi_c) \tag{5.29a}$$

$$\phi_c(\mathbf{x}, 0) = \frac{1}{|c|} \mathbf{1}_c(\mathbf{x}), \tag{5.29b}$$

on a 64×64 uniform rectangular grid. A second-order Strang operator-splitting scheme, where at every time step we take a half-step with diffusion, a full-step with advection and finally another half-step with diffusion, is applied to march the solution in time. The monotonized central flux limiter (Van Leer, 1977) is applied in the advection step. Heun's method is used for the diffusion step. The time step Δt is adaptively determined by

$$\Delta t = \max \left\{ \delta t : \frac{s}{\delta t} \in \mathbb{N}, \frac{\|\mathbf{u}\|_{\max} \delta t}{\Delta x} \leq 0.2 \right\},$$

so that the Courant number $\|\mathbf{u}\|_{\max} \Delta t / \Delta x$ is bounded from above by 0.2. Here $\|\mathbf{u}\|_{\max} = \max_{i,j} \{ |\Psi_{i(j+1)} - \Psi_{ij}|/h_c, |\Psi_{(i+1)j} - \Psi_{ij}|/h_c \}$ is the maximum speed on any edge of rectangular cells. The solver grid spacing $\Delta x = l/64$ should not be confused with the cell spacing h_c between neighbouring nodes for discrete streamfunction Ψ_{ij}. Details of the finite volume solver can be found in Appendix C.1.

Parameter $\boldsymbol{\theta}$	Initial value $\boldsymbol{\theta}^{(0)}$	Proposal s.d. $\sqrt{V_j}$	Prior range
Ψ_{ij}	0 m^2 s^{-1}	0.001 m^2 s^{-1}	unbounded
$\Gamma_{1,ij}$	1,000 m^2 s^{-1}	100 m^2 s^{-1}	[1 m^2 s^{-1}, 10^5 m^2 s^{-1}]
$\Gamma_{2,ij}$	1,000 m^2 s^{-1}	100 m^2 s^{-1}	[1 m^2 s^{-1}, 10^5 m^2 s^{-1}]
Φ_{ij}	0 rad	0.25 rad	unbounded

Table 5.2: Parameters used to configure the Metropolis–Hastings Algorithm. 'Proposal s.d.' stands for the standard deviation of the Gaussian proposal distribution before tuning the algorithm.

To ensure the inferred diffusivity is not corrupted by the numerical ones, we conducted an experiment to determine the numerical diffusivity at various levels of Peclet numbers. We firstly consider a cellular flow given by streamfunction $\psi = (1/\pi)\sin(\pi x)\sin(\pi y)$ and a constant diffusivity κ on a square domain $[0,1]^2$. We then simulate the evolution of passive scalars, with an initial Gaussian profile centering at the domain centre at a unit diagonal covariance, using the numerical recipe discussed above. By varying κ one can control the grid Peclet number, defined as $\mathrm{Pe} = \kappa/(1\cdot\Delta x)$. The effective diffusivity κ_{eff}, which is the sum of a numerical diffusivity and κ, can be estimated using the diagnostic

$$\kappa_{\mathrm{eff}} = \frac{\frac{\mathrm{d}}{\mathrm{d}t}\int \phi_c^2 \mathrm{d}\mathbf{x}}{2\int \|\nabla\phi_c\|^2 \mathrm{d}\mathbf{x}},$$

where the integrals cover the complete domain. We confirm that at the level of grid Peclet number relevant to the inference, κ_{eff} differs from the exact κ by less than 0.1% and conclude that numerical diffusion is insignificant to the inference quality.

Uniform priors are imposed on each of the parameter elements and are listed in Table 5.2. A bounded range of [1 m^2 s^{-1}, 10^5 m^2 s^{-1}] is imposed on the diffusivity eigenvalues Γ_1 and Γ_2 reflected an expected range for diffusivity. The upper cap at 10^5 m^2 s^{-1} also guarantees stability of the explicit Heun's time integrator in the diffusion step.

The posterior distribution is sampled using the Metropolis–within–Gibbs scheme (see section 2.2.2). In total 3 independent set of 292,000 samples were drawn, after discarding 146,000 samples from inference 'burn-in'.

5.3.4 Results

The maximum a posteriori (MAP) estimate, formally the maximiser to the posterior $q(\boldsymbol{\theta}|R)$, indicates the most probable fields of mean flow and diffusivity under the given prior and trajectory data. Using samples from the MCMC algorithms, the MAP estimate is approximated by the sample $\boldsymbol{\theta}^{(k)}$ with maximal posterior value $p(\boldsymbol{\theta}^{(k)}|R)$ in all chains. Estimates of the mean flow velocities and diffusivities at this MAP are the primary focus of this section and are compared with the exact values.

The convergence of the MAP estimates from the data coarse-grained approach against data size is examined. Trajectories over a fixed 10-year temporal duration of variable particle number $N_p = 169,\ 676,\ 2704,\ 10816$ and 43262 are considered. The streamfunctions and diffusivity fields at the MAP at a sampling interval of 64 days are visualised in figure 5.4 and 5.5. A reference streamfunction and diffusivity field are created by L^2-optimally projecting the exact fields onto the 8×8 piecewise bilinear function space where the inferred fields are defined (see the top left panels of figure 5.4 and 5.5 for the exact fields, and figure 5.9 and 5.10 for the projected reference). We

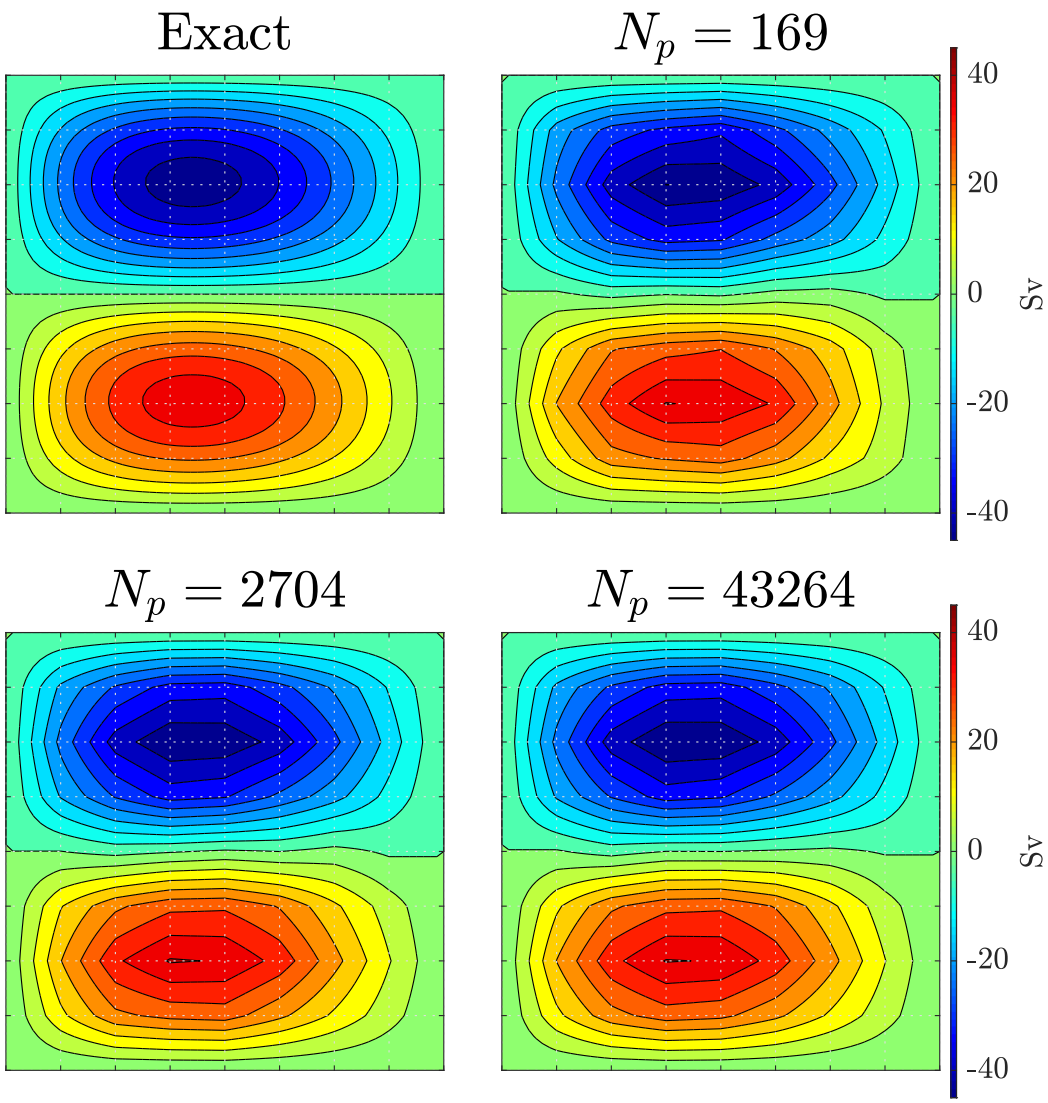

Figure 5.4: Streamfunctions multiplied by a depth thickness of 250 m. Top left panel: exact streamfunction used to simulate the particle trajectories. Remaining panels: MAP estimates using 10-year trajectories of varying amount of particles (see titles) sampled at an interval of 64 days.

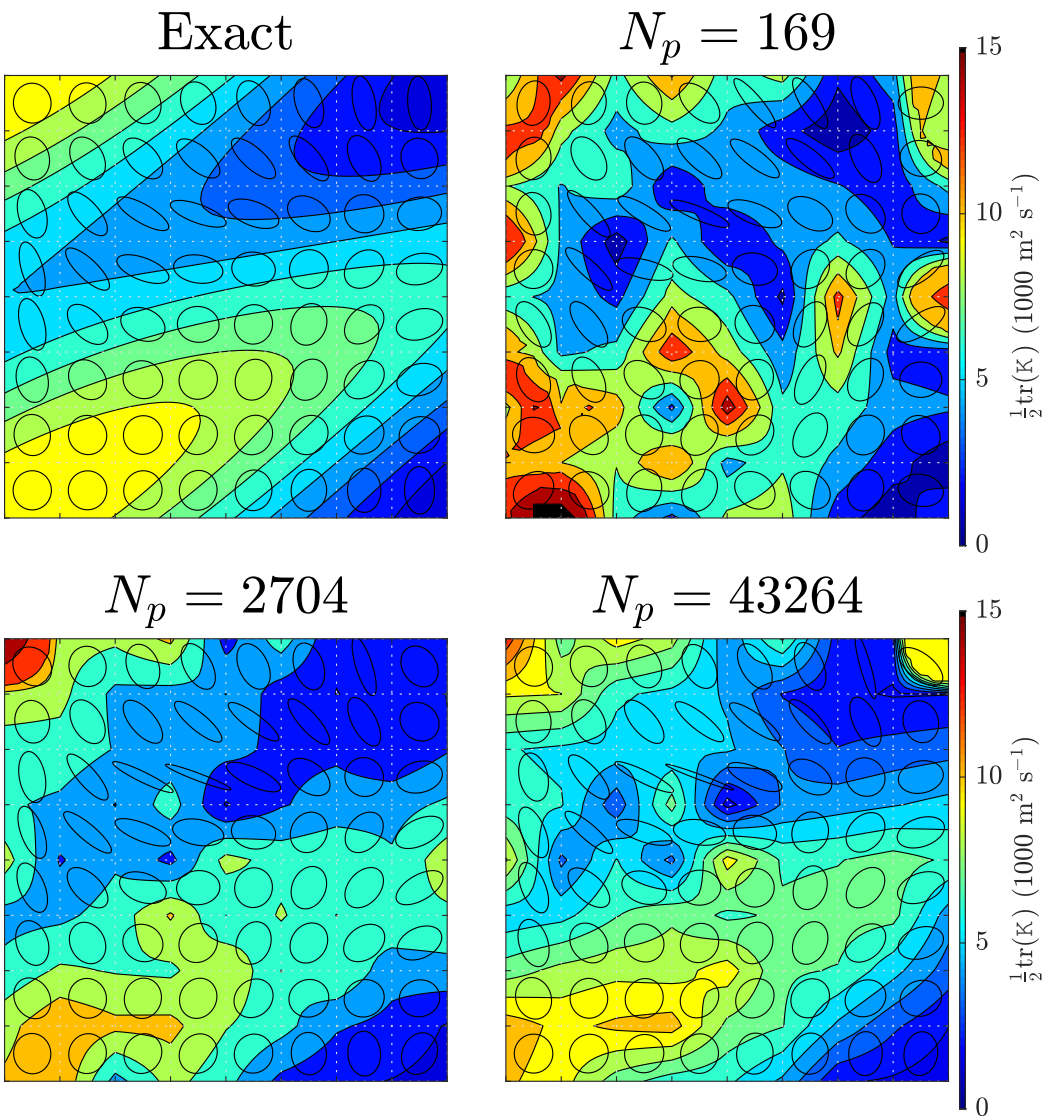

Figure 5.5: Comparison of the exact and inferred diffusivity tensor fields. The colour scale the half trace of the tensor, characterising the strength of the diffusivity. Values exceeding $15,000$ m^2 s^{-1} are marked in black. The diffusivity ellipses at the centres of an 8×8 uniform square grid are drawn to characterise the relative orientation (but not magnitude) of the anisotropies in the diffusivity tensor. Top left panel: exact diffusivity field used to simulate the particle trajectories. Remaining: MAP estimates using 10-year trajectories of varying amount of particles (see titles) sampled at an interval of 64 days.

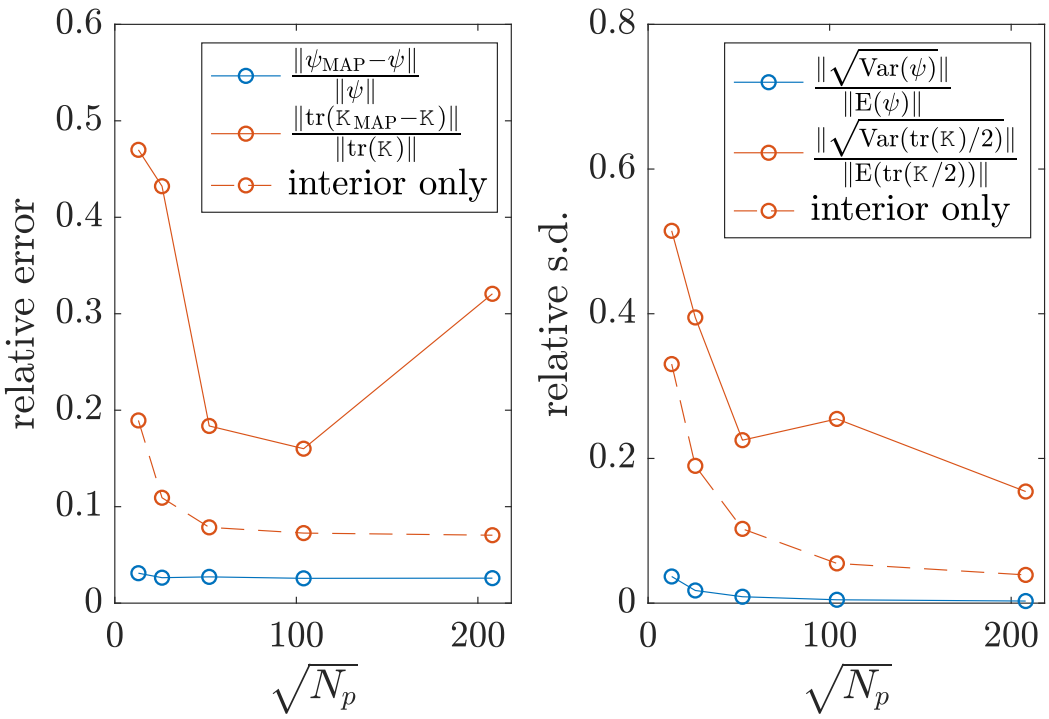

Figure 5.6: Left panel: Relative error of the MAP estimates from the reference fields, measured in L^2-norm, using the 10-year trajectories of varying particle data sizes sampled at 64-day intervals. Right panel: Relative standard deviation of the inferred fields derived from the posterior distribution, measured in L^2-norm, using the 10-year trajectories of varying particle data sizes sampled at 64-day intervals. In both panels the blue and red lines correspond to the streamfunction and diffusivity fields respectively.

emphasise that, even in the large data limit, the reference (optimal projection of the exact fields) may not coincide with the MAP estimates (optimal piecewise linear fields to reproduce the observed data). One may hope that in the limit of both large data and refined field resolution, both the reference and MAP estimates would converge to the exact fields, although we did not verify this.

The relative errors of the MAP estimates, defined as the relative L^2-error from the reference, are shown in the left panel of figure 5.6. The streamfunction converges as data size increases, which results in the insensitive error. The error of the diffusivity at MAP from the reference generally declines as data size increases, although it also spikes at the largest considered data size ($\sqrt{N_p} = 204$). Examination reveals the spike is mostly attributed to the error at the top right corner (see the bottom right panel of figure 5.5). If the boundary cells are excluded in the error calculations, the error of in diffusivity will exhibit a similar pattern as the streamfunction.

To quantify the uncertainty in the inference, we define the posterior expectation $\mathrm{E}(f)(\mathbf{x})$ and variance $\mathrm{Var}(f)(\mathbf{x})$ of a field of interest $f(\mathbf{x};\boldsymbol{\theta})$ by

$$\mathrm{E}(f)(\mathbf{x}) = \int f(\mathbf{x};\boldsymbol{\theta}) q(\boldsymbol{\theta}|R) \mathrm{d}\boldsymbol{\theta}, \tag{5.30a}$$

$$\mathrm{Var}(f)(\mathbf{x}) = \int (f(\mathbf{x};\boldsymbol{\theta}) - \mathrm{E}(f)(\mathbf{x}))^2 q(\boldsymbol{\theta}|R) \mathrm{d}\boldsymbol{\theta}. \tag{5.30b}$$

We then define a relative standard deviation to quantify the uncertainty by a ratio between the L^2-norms of standard deviation $\|\sqrt{\mathrm{Var}(f)}\|$ and expectation $\|\mathrm{E}(f)\|$. Examples are shown in the right panel of figure 5.6. The uncertainty generally declines as more data are supplied. When the boundary cells are excluded in the L^2-norms, the uncertainty in diffusivity sharply decreases.

This experiments shows that while the data coarse-grained inference is generally capable of estimating the spatial variability in the velocity and diffusivity fields, it struggles near the boundaries and especially at the corners.

We conjecture this is due to the loss of distinguishablity in individual particle displacements. In the primitive approach, the velocity and diffusivity fields near the boundary can be inferred by the local statistics of individual particle displacement. In the data coarse-grained approach, which matches the evolved profile of test tracers with particle distributions, the particles displaced in and out a cell carry information for the inference to adjust the velocity and diffusivity fields. As no particle in the data is transported across the boundary, the coarse-grained approach struggles to identify the fields on the boundary. Instead, it can only rely on the particles displaced in other directions to extrapolate the diffusivity on the boundary. We conjecture such information asymmetry between the interior and a boundary cell is the cause of the relatively larger uncertainty of boundary diffusivity.

The conjecture can be demonstrated on a pathological case where the diffusivity on a boundary becomes non-identifiable. Suppose that a piecewise bilinear scalar diffusivity is inferred on a corner cell, as shown in the schematic sketch in figure 5.7. The velocity is known to be zero everywhere and the diffusivity on the non-boundary edges (edge 1 and 4) is known to be a non-negative constant ϵ, so that diffusivity in the fields is fully controlled by the degree of freedom κ_c at the top right corner. In the limit of $\epsilon \to 0$, edges 1 and 4 acts like barriers and stops particles from travelling across. In the primitive inference the diffusivity κ_c can well be estimated by the statistics of the (small) individual particle displacements. In the data coarse-grained inference, however, the test tracer will become increasingly insensitive to κ_c, as it tends to remain

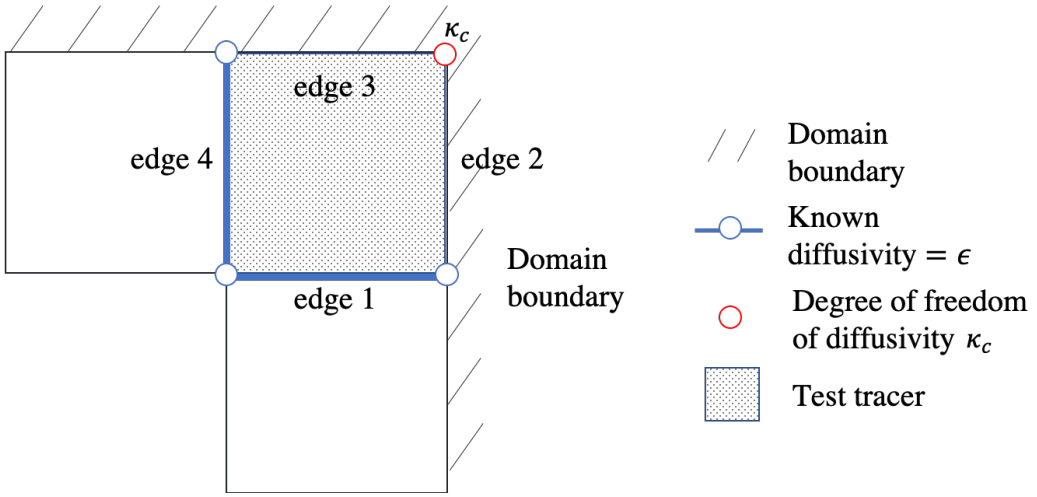

Figure 5.7: Schematic of the pathological example where the diffusivity in a corner cell becomes non-identifiable.

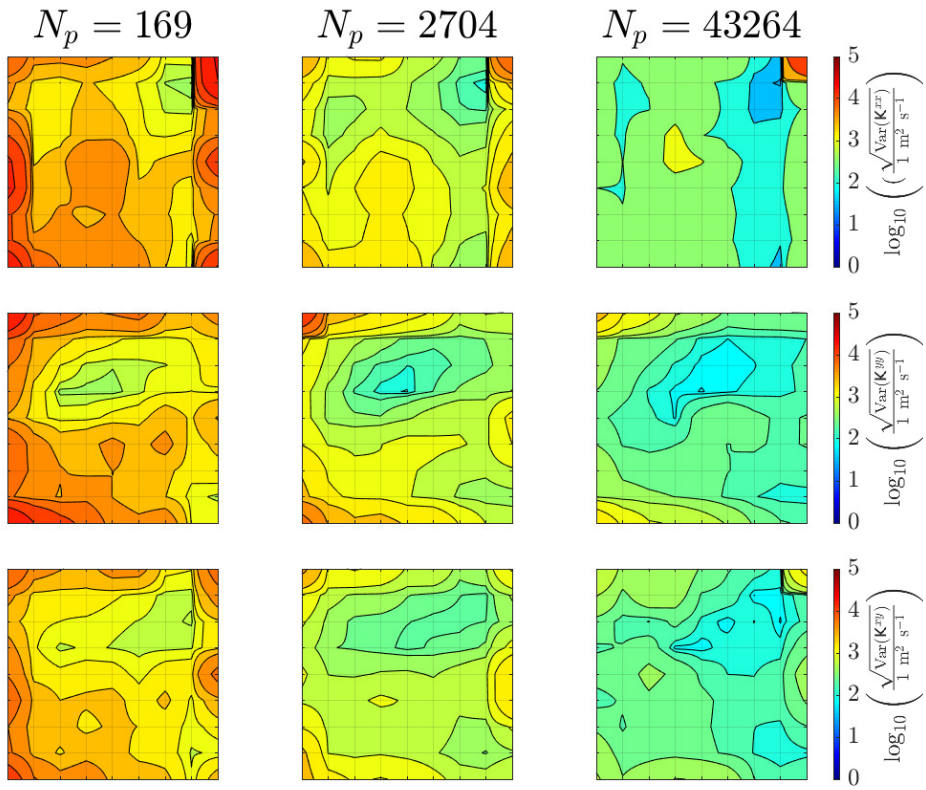

Figure 5.8: Standard deviation of diffusivity components (top to bottom: xx, yy and xy) inferred from 10-year trajectory of 676 particles sampled at a 96-day interval.

uniform in the cell. In other words, the diffusivity at the corner κ_c becomes increasingly non-identifiable as $\epsilon \to 0$, and the posterior for κ_c becomes increasingly dominated by the prior distribution. We speculate this causes the local spike of diffusivity at the top right corner in the lower right panel in figure 5.5, where both the mean drift and diffusivity are so weak that a large portion of particles are retained in the corner cell. Figure 5.8 shows that the posterior variance of the diffusivity at some corners may remain large relative to the rest of the domain, even when a large number (43264) of particles are used. The struggle of the data coarse-grained inference to constrain the values of diffusivity at these corners is reflected by the large uncertainty.

The inference at different sampling intervals is also examined by downsampling the 10-year trajectory of 676 particles in time. Note that an increase in sampling interval effectively reduces the number of transition data points for the inference. Figure 5.9 and 5.10 show the streamfunctions and diffusivity fields at MAP estimates under different sampling intervals. While the streamfunction at MAP is relatively insensitive to sampling interval, the diffusivity at MAP increasingly deviates from the reference. The error is particularly large near the domain boundaries at the larger sampling intervals.

The relative error of the MAP estimates and relative standard deviations at different sampling intervals are shown in the left and right panels of figure 5.11. Both errors and uncertainties increase with sampling interval, expectedly, as the effective data size reduces. After neglecting the diffusivity on the boundary cells, the recomputed relative error and standard deviations (shown in red dash lines in figure 5.11) drop significantly by roughly a factor of 2. This shows that a majority of error and uncertainty in the diffusivity field can be attributed to the domain boundary, which is conjectured to be caused by the coarse-graining of information on the individual trajectory.

The localised approach presented in chapter 4 is also applied for this problem on the same array of cells. A cellwise constant velocity and diffusivity tensor field are inferred using the Metropolis–Hastings algorithm configured as in Table 3.2. At the lowest considered sampling interval of 32 days, only a small fraction of ≈ 0.25 of transitions are intra-cell, indicating the breakdown of the locality assumptions. The relative errors associated with the localised approach, displayed also in the left panel of figure 5.11, are several times in size to that of the global approach at any considered sampling interval. The errors in the localised approach also increase with a steeper slope with sampling intervals. The qualitative difference of the inference at various sampling intervals demonstrate that the global approach successfully overcomes the locality limitation.

To summarise, the data coarse-grained method manages to infer a globally-defined mean flow and diffusivity tensor fields, given sufficient trajectory data size. However, the domain boundaries appear to be a hotbed of errors and uncertainties. We believe this is caused by the loss of information when coarse-graining trajectory data.

5.4 Quasigeostropic double gyre

The idealised example tested in the preceding section demonstrated the convergence properties of the data coarse-grained method in the global inference. In a practical oceanographic scenario, an inference for diffusivity is often limited by the fixed amount of trajectory data and incomplete knowledge of the mean velocity field. The global inference is now tested using the Lagrangian trajectories in the quasigeostropic double gyre calculations considered in Chapter 4.

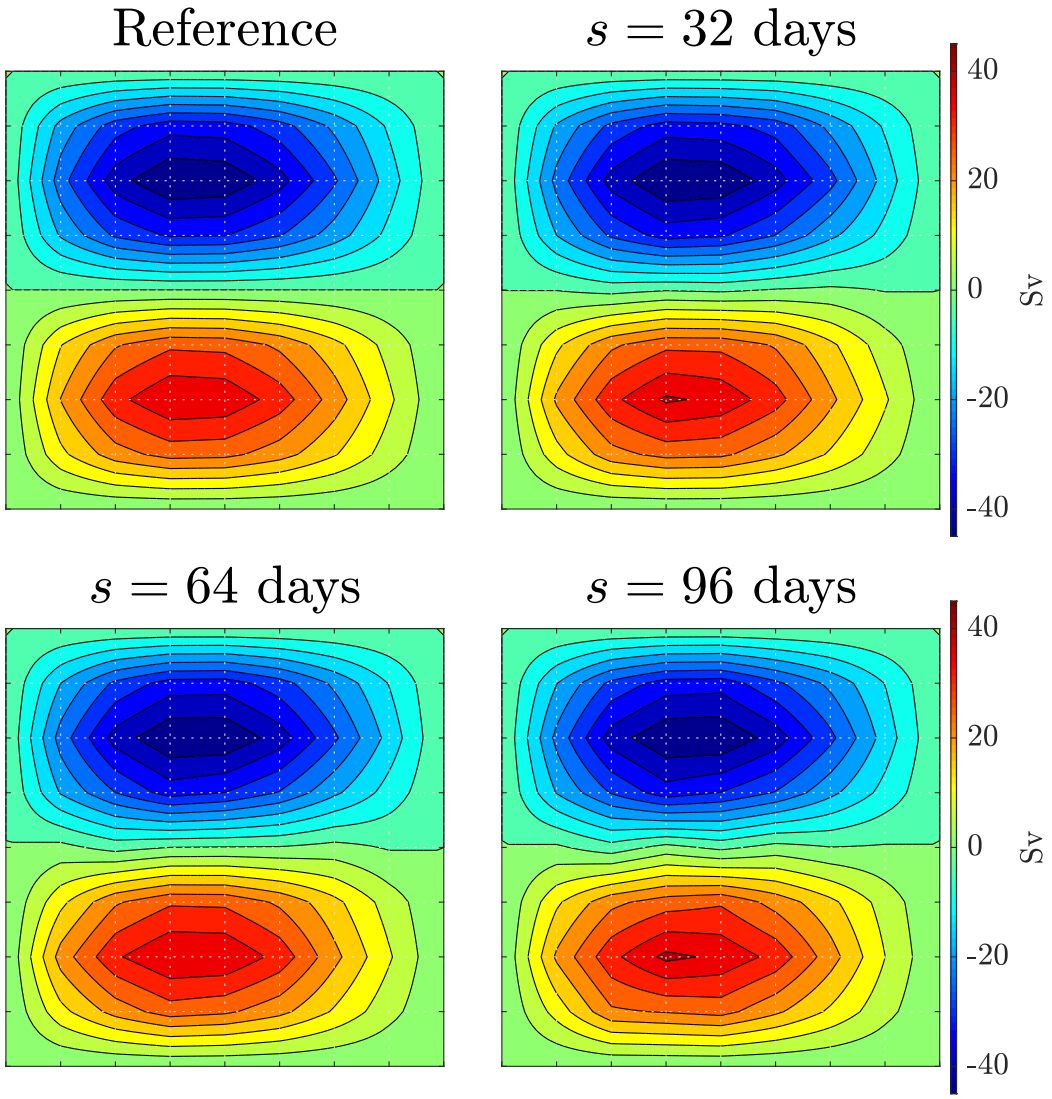

Figure 5.9: Streamfunctions multiplied by a depth thickness of 250 m. Top left panel: L^2-optimal projection of the exact streamfunction on an 8×8 cell array. Remaining: MAP estimates inferred from 10-year trajectory of 676 particles under different sampling intervals (see titles).

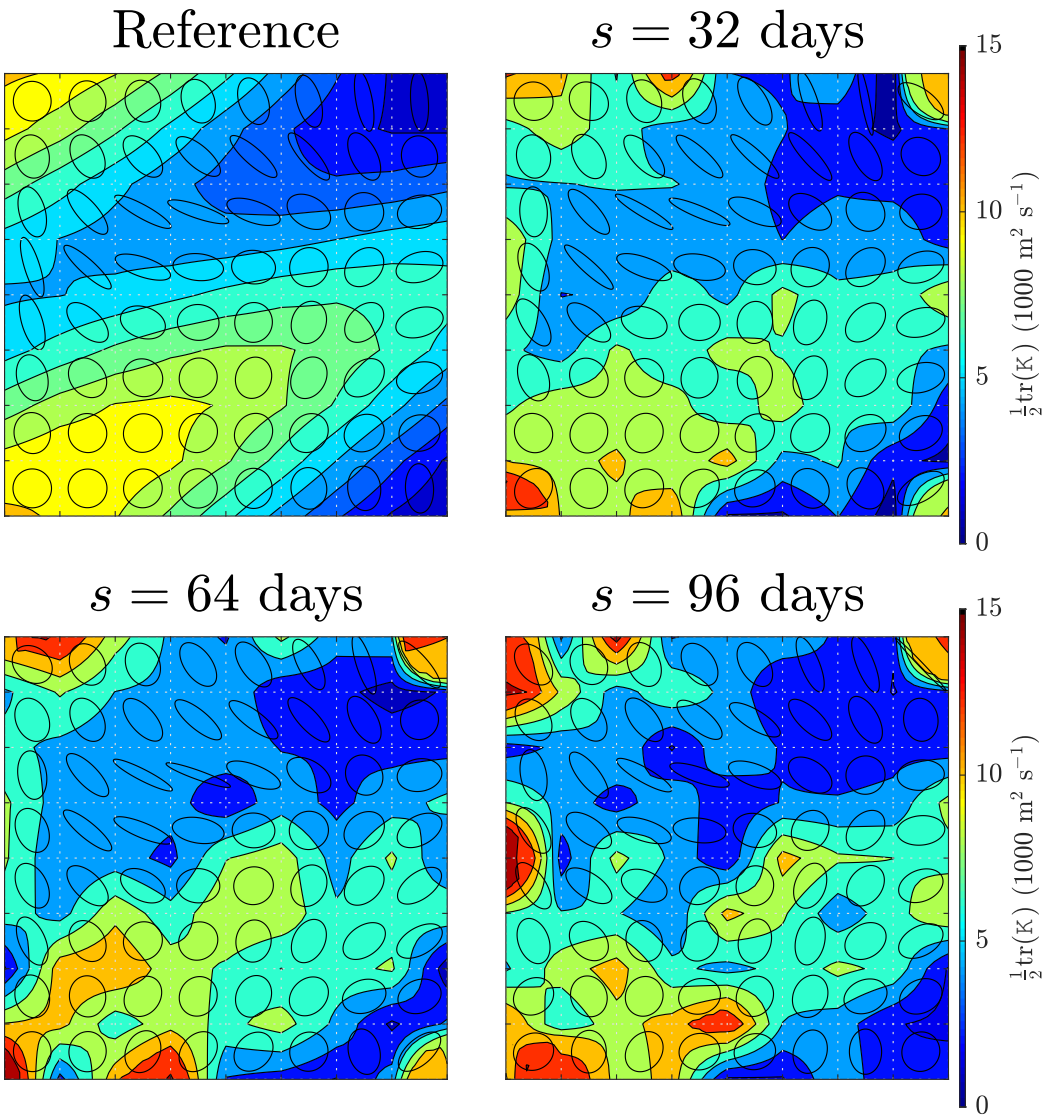

Figure 5.10: Comparison of the reference and inferred diffusivity tensor fields. The colour scale represents the half trace of the tensor, characterising the strength of the diffusivity. The diffusivity ellipses at the centres of an 8×8 uniform square grid are drawn to characterise the relative orientation (but not magnitude) of the anisotropies in the diffusivity tensor. Top left panel: L^2-optimal projection of the exact diffusivity on an 8×8 cell array. Remaining panels: MAP estimates using the 10-year trajectory of 676 particles at different sampling intervals (see titles).

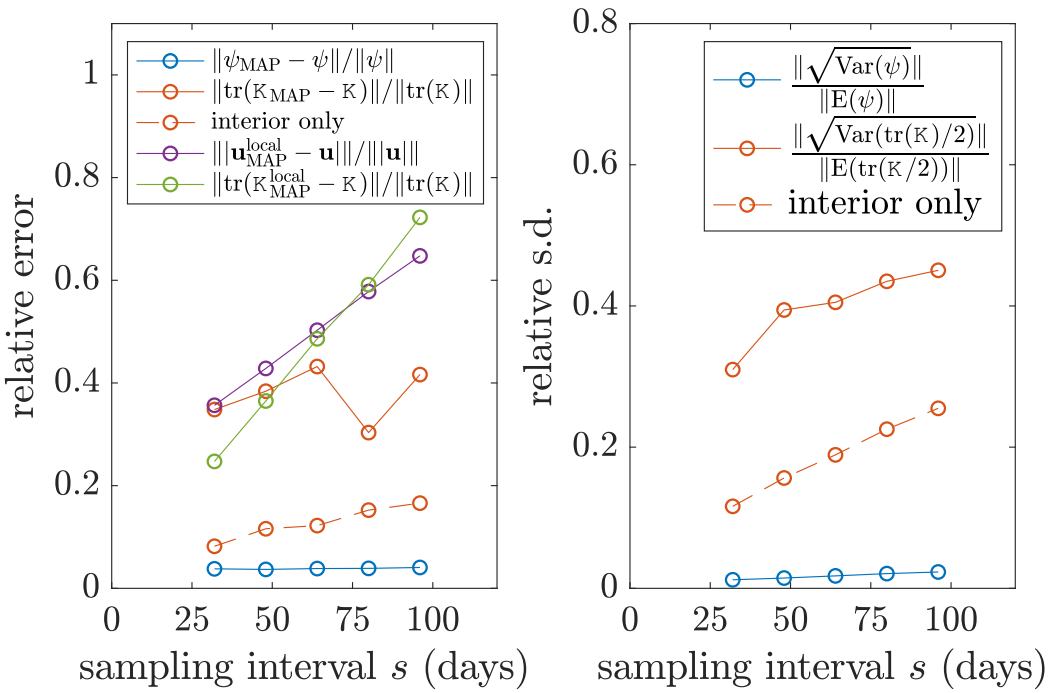

Figure 5.11: Left panel: Relative errors of the MAP estimates from the exact fields, measured in L^2-norm, using the 10-year trajectory of 676 particles sampled at varying sampling intervals. The blue and red solid lines correspond to the streamfunction and the trace of diffusivity respectively. The red dash line corresponds to the relative error of the diffusivity trace with the boundary cells excluded. The purple and green lines are the relative error of the MAP estimates of the cell-centred velocity and diffusivity tensor inferred using the localised approach presented in chapter 4. Right panel: Relative standard deviation of the inferred fields derived from the posterior distribution, measured in L^2-norm, using the 10-year trajectory of 676 particles sampled at varying sampling intervals. The blue and red solid lines correspond to the streamfunction and diffusivity fields respectively. The red dash line corresponds to the relative standard deviation of the half trace of diffusivity tensor with the boundary cells excluded.

5.4.1 Lagrangian trajectory data

The same set of Lagrangian trajectories data as in section 4.1 is used for the inference. Specifically, the 10-year trajectories of 676 particles in the middle layer are considered in this section.

5.4.2 Parameterisation

The square domain of side length $l = 3840$ km is divided into $N_c^2 = 16^2$ identical square cells. As in the idealised flow example in section 5.3.2, piecewise bilinear streamfunction and diffusivity tensor fields, both parameterised by nodal values at cell corners, are inferred. The parameter components therefore include

$$\boldsymbol{\theta} = (\Psi_{ij}, \Gamma_{1,ij}, \Gamma_{2,ij}, \Phi_{ij})^{\mathrm{T}}, \tag{5.31}$$

for all $i, j \in \{0, 1, \cdots 16\}$.

The boundary values of the streamfunction are set to be zero to enforce the no-flux condition. Hence a total number of $(N_c - 1)^2 + 3(N_c + 1)^2 = 1,092$ components of the parameter $\boldsymbol{\theta}$ are be inferred.

5.4.3 Posterior evaluation

The data coarse-grained approach outlined in section 5.1 is applied. Per each likelihood evaluation, a total number of ${N_c}^2 = 256$ Fokker–Planck equations (5.29) are solved using the same finite-volume solver on a 64×64 grid described in section 5.1.

The same choice of prior is imposed on the parameters as in section 5.1 (see Table 5.2). In particular, uniform priors are chosen for all components, with bounds only on the eigenvalues of diffusivity $\Gamma_{1,ij}$, $\Gamma_{2,ij} \in [1\ \mathrm{m}^2\ \mathrm{s}^{-1}, 10^5\ \mathrm{m}^2\ \mathrm{s}^{-1}]$.

The Metropolis–within–Gibbs algorithm is applied to sample the posterior distribution. In each of the 10 independent chains, every component is sampled $1,000$ times after discarding 500 in the burn-in phase and 500 in the tuning process. These add up to a total number of samples of $1,029 \times 1,000 = 1,029,000$ samples per chain.

5.4.4 Results

The maximum a posteriori (MAP) estimate, again approximated by the sample $\boldsymbol{\theta}^k$ with maximal posterior value $p(\boldsymbol{\theta}^k|R)$ in all chains, is the primary focus in this section.

Recall from Chapter 3 that estimating the eddy diffusivity in a multiscale dynamical system using a likelihood-based inference is only legitimate when the trajectory data is downsampled at sufficiently large temporal intervals (also Pavliotis and Stuart (2007); Cotter and Pavliotis (2009)). It is expected that the eddy diffusivity in the quasi-geostrophic double gyre system carries a similar property. In this study we therefore consider a sampling interval not less than 32 days. The performance of the global inference in 12 cells around the central jet is focused on and their locations are illustrated in the top left panel of figure 5.12. Along the zonal jet (row b in the top left panel of figure 5.12), the cells capture different dynamical regimes as particles rapidly leave the meridional boundary current, enter the eastward jet and get expelled out (from column i to iii). On row a and c, the westward circulation of the gyres are examined. In cells to the east of the jet tail (column iv) where the flow is relatively steady and slow, the global inference is expected to provide similar results to the localised inference.

The streamfunction at MAP under varying sampling intervals are shown in figure 5.12. The marginal posterior and the MAP estimates of the velocity components at

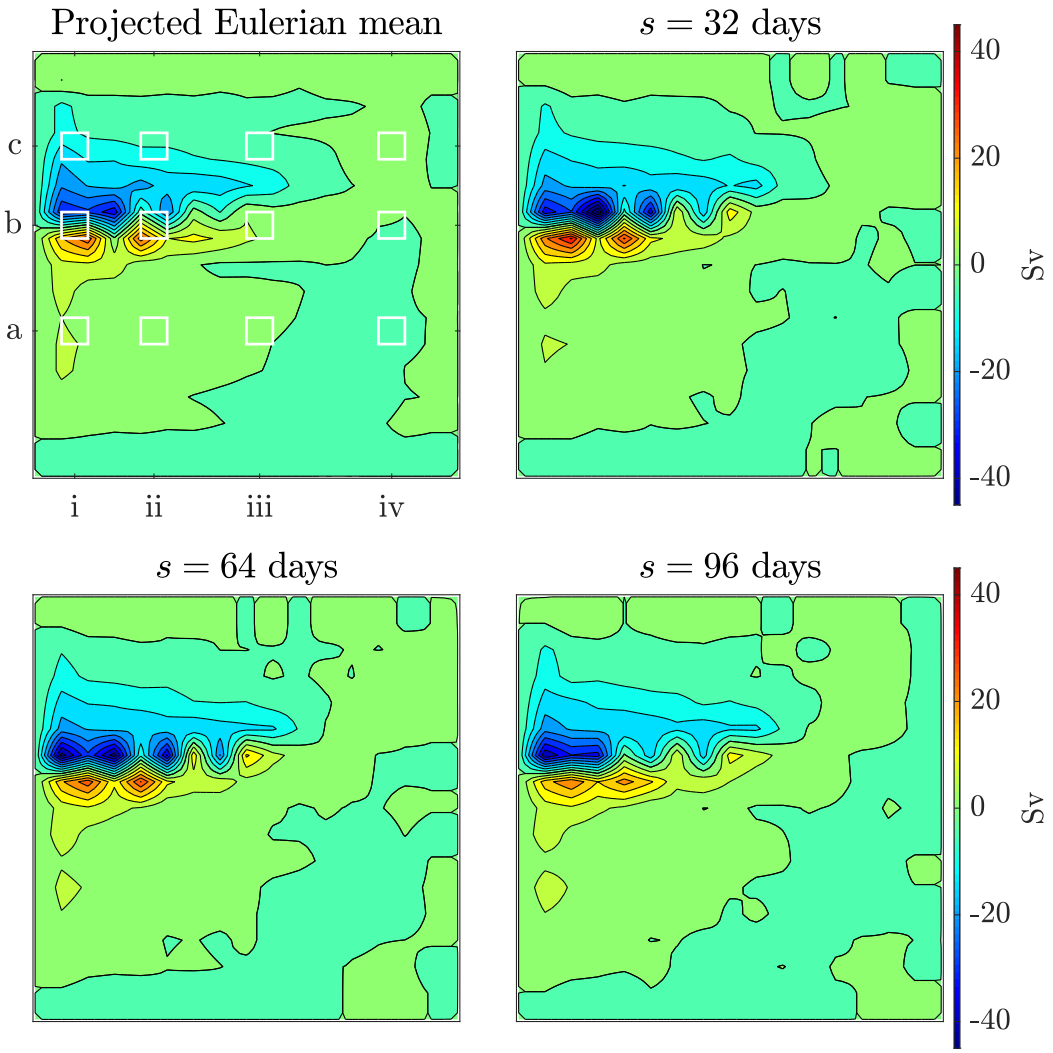

Figure 5.12: Top left panel: The 10-year Eulerian time average streamfunction L^2-optimally projected onto a 16×16 piecewise linear function space. Remaining panels: Maximum a posteriori (MAP) estimate for the middle layer streamfunction, weighted by the layer thickness, in the quasigeostrophic double gyre example under varying sampling intervals s. Top left panel highlights in white boxes the selected bins, which are labelled by a letter/numerical coordinates.

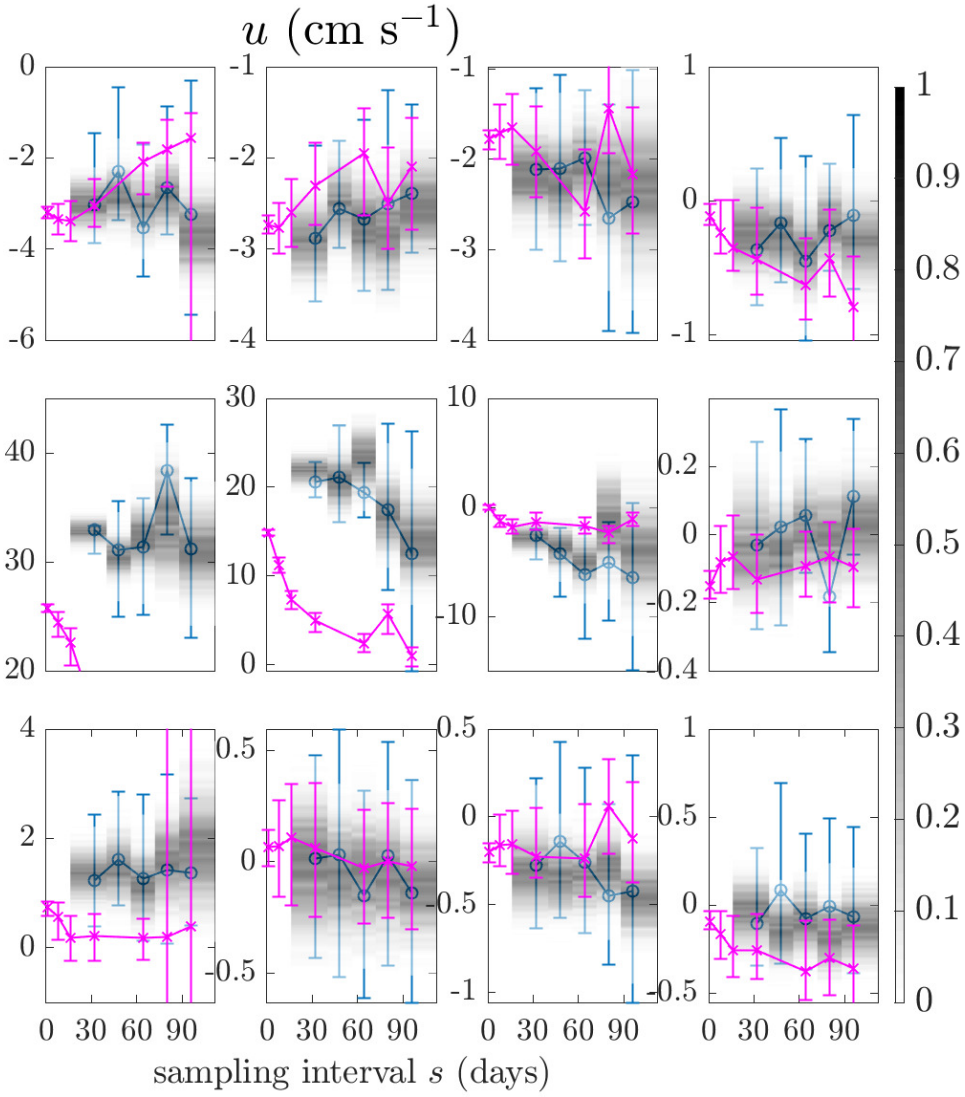

Figure 5.13: Middle layer zonal velocity (in cm s^{-1}) against sampling intervals s in the selected cells. The blue lines and the grey shadings are the estimates at the MAP and the marginal posterior (normalised by its peak for each s) resulting from the global approach. The pink lines are the estimates at the MAP resulting from the localised approach. blue and pink error respectively shows a 90%-credible interval (2.4), centred at the medians, for the global and local inference.

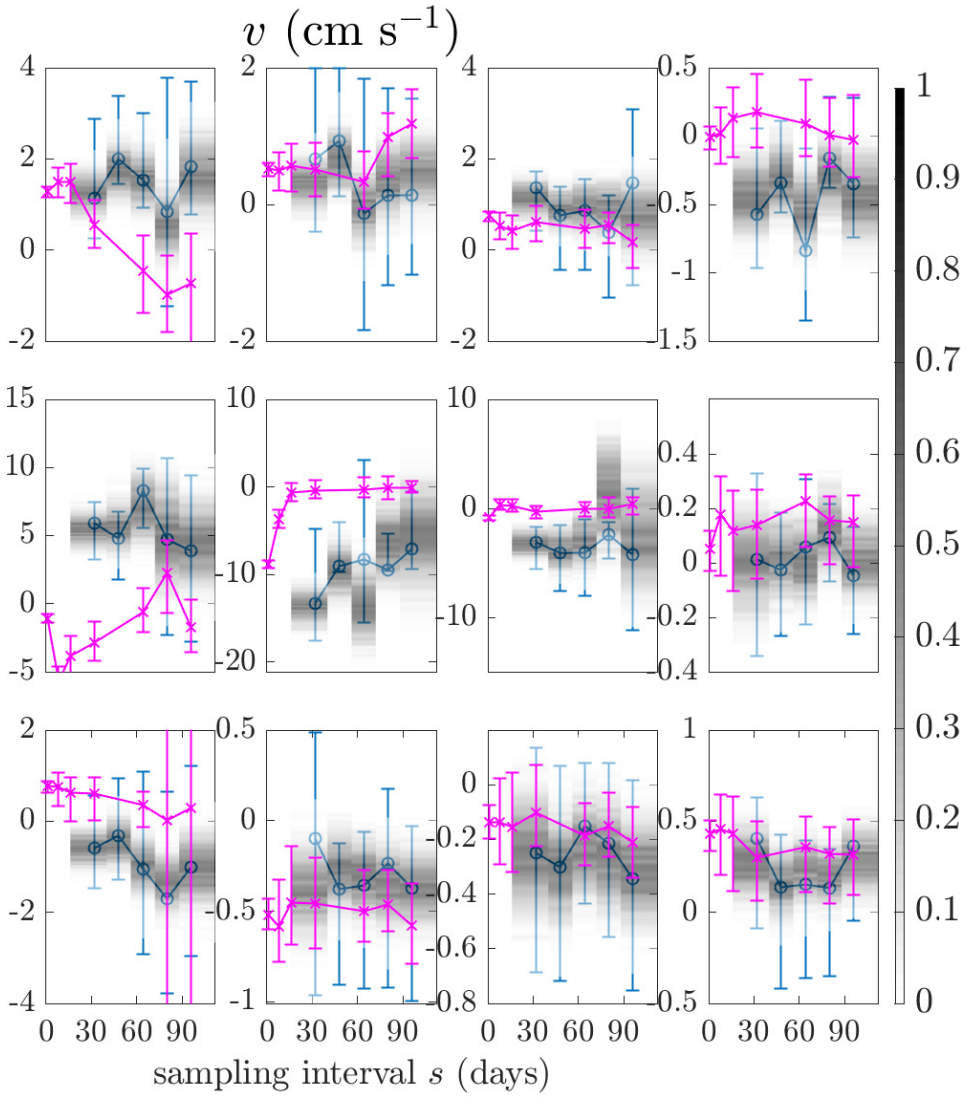

Figure 5.14: Middle layer meridional velocity (in cm s^{-1}) against sampling intervals s in the selected cells. The blue lines and the grey shadings are the estimates at the MAP and the marginal posterior (normalised by its peak for each s) resulting from the global approach. The pink lines are the estimates at the MAP resulting from the localised approach. blue and pink error bars respectively shows a 90%-credible interval (2.4), centred at the medians, for the global and local inference.

the cell centres are shown in figure 5.13 and 5.14 respectively. Note that the estimated MAP in general does not coincide with the maximum of the marginal posterior, even after taking into account the uncertainty. For cells in the fast flow region, such as near the western boundary (column i) and along the jet (row b) where particles rarely remain in the same cell in the considered time intervals, the MAP estimates under the global approach (blue lines) differ significantly from that under the localised approach (pink lines), and are relatively insensitive to sampling interval. The detachment of the meridional western boundary currents that forms the eastward jet is fully resolved in the global approach. At other cells with slow flow, the estimates from both approaches are comparable.

The diffusivity at MAP under varying sampling intervals are shown in figure 5.15. The cross-stream and along-stream components of the diffusivity are shown in figure 5.17. The inferred diffusivity fields peak along the western boundary and the central jet and reach the trough in the eastern part of the domain. Near the western boundary where the localised approach tends to misattribute the velocity and diffusivity estimates, the global approach gives relatively stable estimates of diffusivity at different sampling intervals compared with the MAP estimates in the localised approach (see figure 4.5). On the eastern boundary, however, bizarrely large values of diffusivity are inferred. Scrutiny reveals that in the eastern boundary cells, particles are driven by slowly evolving wavy motion and is purely transported meridionally. This is akin the pathological case discussed in section 5.3.4, and that the low values of diffusivity on the left edges of the boundary cells act as a barrier to prevent the test tracer from spreading in the zonal direction. As a result, the zonal-zonal diffusivity in the boundary cell is unconstrained. We verify that the posterior of the zonal-zonal component of diffusivity K^{xx} largely coincides with the prior for the diffusivity eigenvalues. A comparison between the marginal posterior standard deviation of the zonal-zonal(xx), meridional-meridional(yy) and zonal-meridional (xy) diffusivity, illustrated in figure 5.16, clearly shows the large uncertainty is only associated with the zonal-zonal component. A potential approach to regularise the highly unsmooth diffusivity field in this area is to impose prior on the smoothness of the inferred functions. For example, one may specify a Gaussian prior on the gradient of the inferred function. A technical constraint of this approach, however, is that one would require a very good guess on the penalty parameter to avoid under- or over-fitting, which is practically highly speculative or an ocean diffusivity inference.

The MAP estimates of global approach inference demonstrate a good agreement with the Davis (1987) diffusivity (shown in figure 5.17) estimated from the same 10-year particle trajectory data (see Appendix A.1). However it should be noted that estimating the Davis (1987) diffusivity requires knowledge of the Eulerian mean flow and Lagrangian eddy velocities along trajectories. Both the Davis (1987) diffusivity and MAP estimates are capable of resolving the anisotropy in the eddy diffusion, although positive-definiteness or symmetry are not guaranteed under the Davis (1987) diffusivity. Negative valued Davis (1987) diffusivity is recorded in a small number of cells (marked in white in figure 4.7).

The marginal distributions of along-stream $\mathsf{K}^{\parallel}$ and cross-stream $\mathsf{K}^{\perp}$ diffusivity are constructed by projecting the individual samples of diffusivity in the along and cross-stream velocity directions. The marginal distribution and MAP estimates in the selected cells are shown in figure 5.18 and 5.19. For comparison, the Davis (1987) diffusivity, estimated at different time lag τ using a 10-year Eulerian mean flow and trajectories of 676 particles, as well as the MAP estimates under the localised approach in chapter

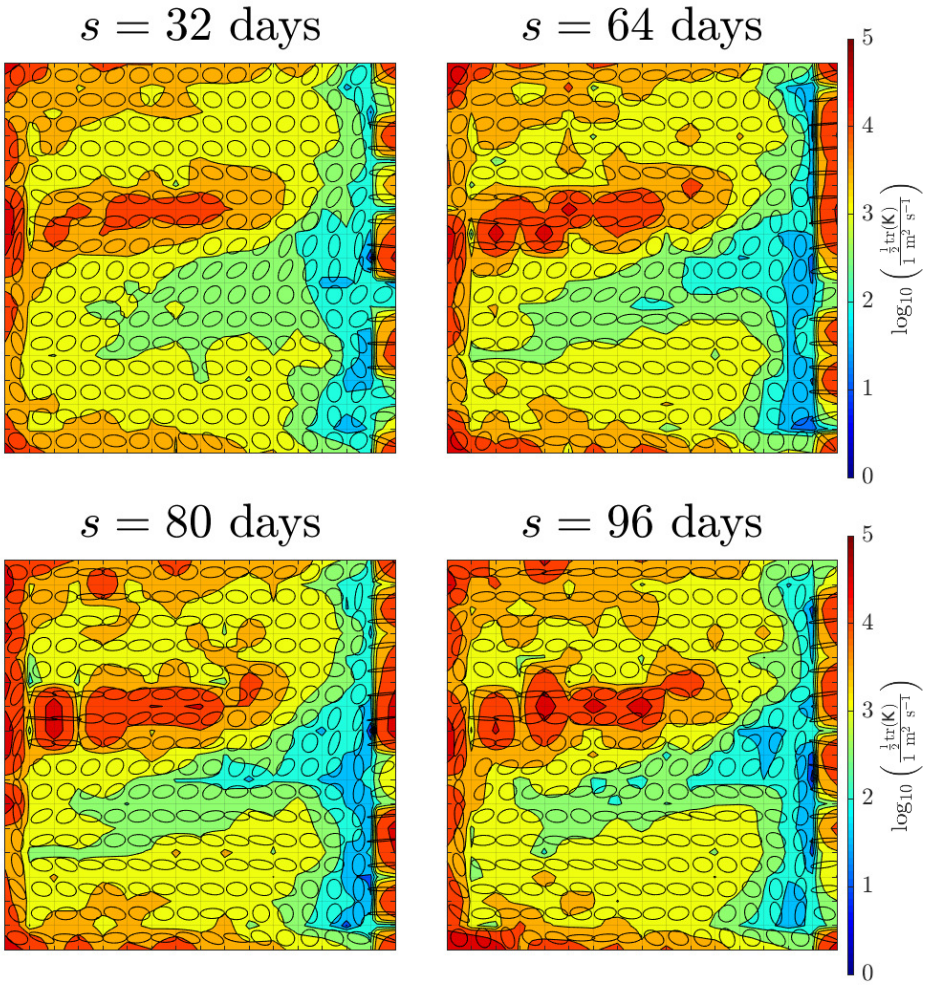

Figure 5.15: Maximum a posteriori (MAP) estimate for the middle layer diffusivity in the quasigeostrophic double gyre example under varying sampling intervals s. The logarithmic colour scales represent the half trace of the diffusivity tensor K and the diffusivity ellipses at cell centres represent the relative magnitude and orientations of the principal components of the diffusivity tensor.

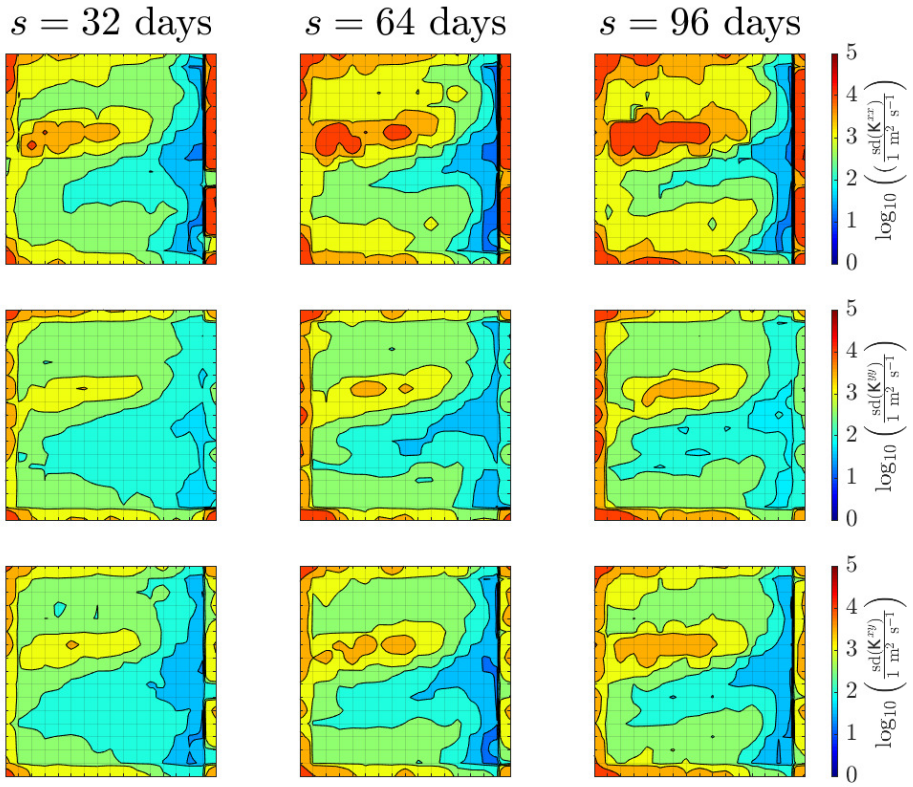

Figure 5.16: Marginal posterior standard deviation of the zonal-zonal(xx), meridional-meridional(yy) and zonal-meridional(xy) diffusivity (from top to bottom) in the middle layer of the quasigeostrophic double gyre example at varying sampling intervals s. A logarithmic colour scales is applied.

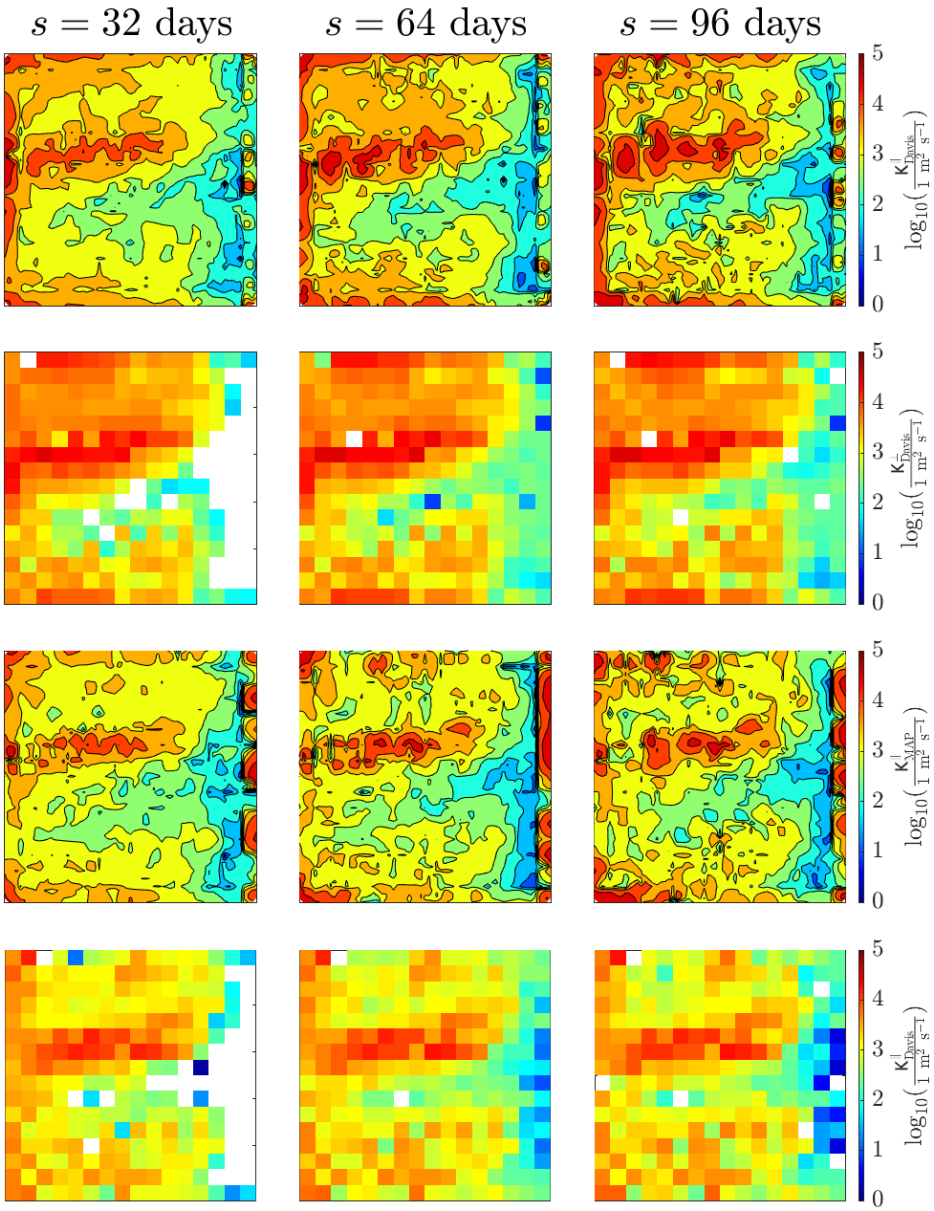

Figure 5.17: Row 1 and 3: Maximum a posteriori (MAP) estimate for the middle layer diffusivities in the along-stream (row 1) and cross-stream (row 3) directions in the quasigeostrophic double gyre example under varying sampling intervals s. Row 2 and 4: Davis diffusivities for the same layer in the along-stream (row 2) and cross-stream (row 4) directions under varying lag τ. The logarithmic colour scales represent the magnitude of the inferred diffusivity components. White colour indicates a negative value.

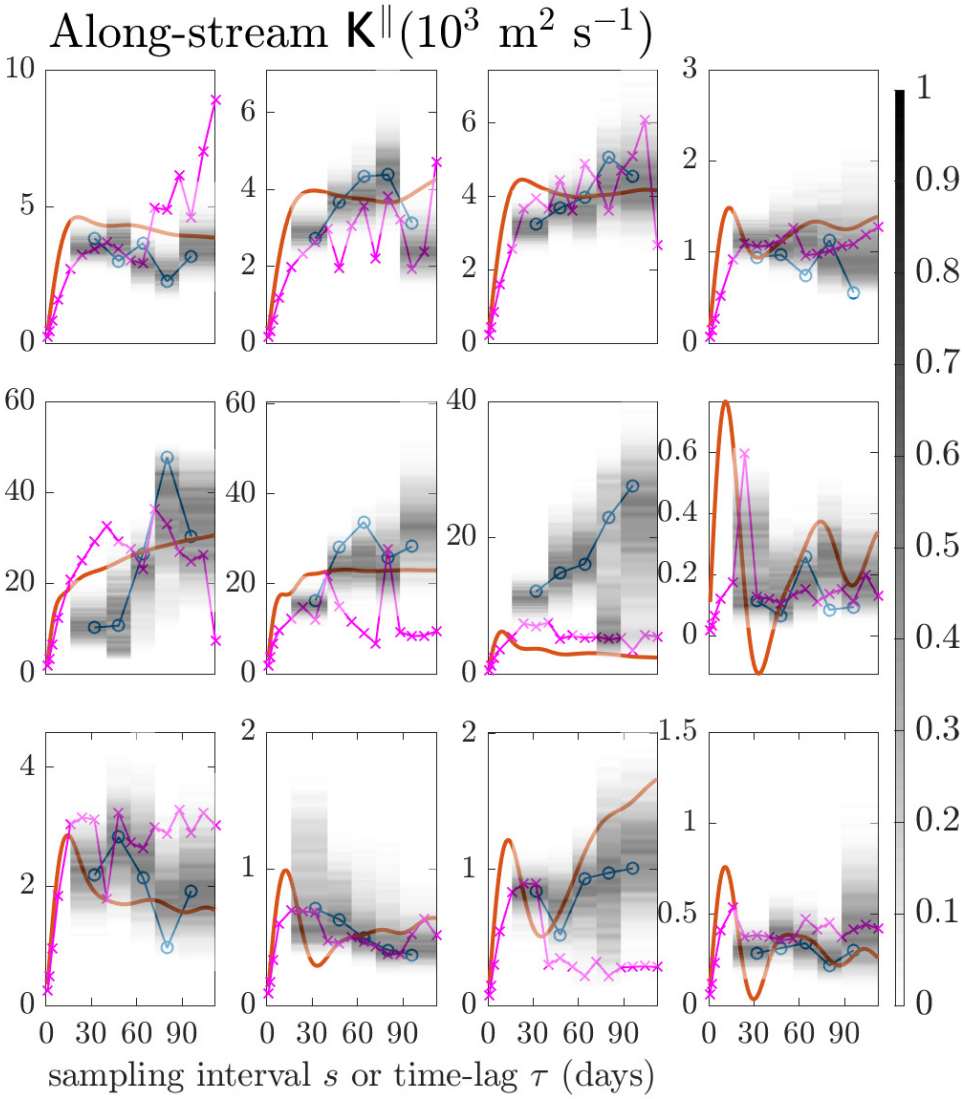

Figure 5.18: Middle layer along-stream diffusivity against sampling intervals s in the selected cells. The blue lines and the grey shadings are the MAP estimates and the marginal posterior (normalised by its peak for each s) resulting from the global approach. The pink lines are the MAP estimates resulting from the localised approach. The red lines are the same component of Davis diffusivity at different time lags τ.

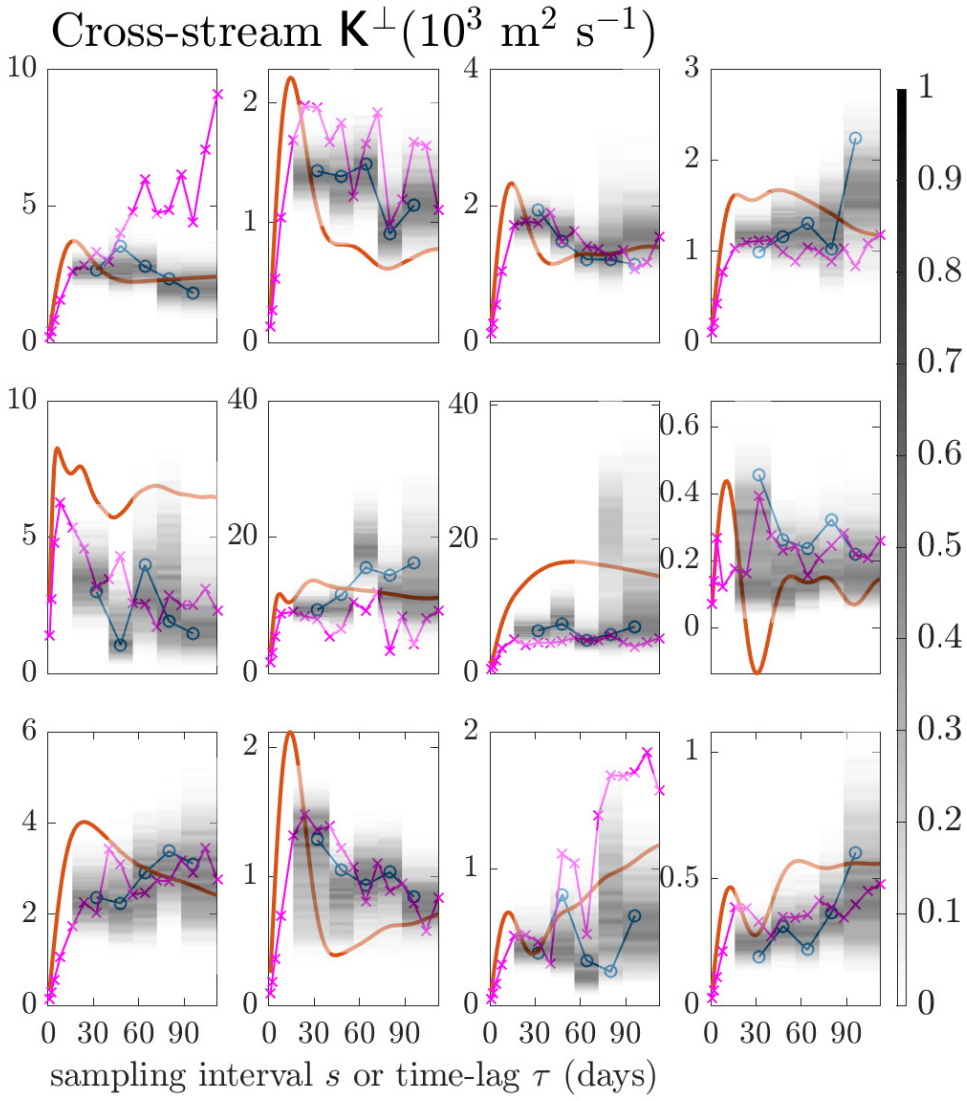

Figure 5.19: Middle layer cross-stream diffusivity against sampling intervals s in the selected cells. The blue lines and the grey shadings are the MAP estimates and the marginal posterior (normalised by its peak for each s) resulting from the global approach. The pink lines are the MAP estimates resulting from the localised approach. The red lines are the same component of Davis diffusivity at different time lags τ

.

4 are also shown in the same figure. Note that the time lag τ for the Davis (1987) diffusivity and sampling intervals s for the Bayesian inference are not strictly comparable despite being placed on the same scale. The global Bayesian approach shows a degree of convergence of diffusivity estimates at large sampling intervals and a general agreement with the Davis (1987) diffusivity at large lags. Unlike the velocity components, the diffusivity estimated by the global and localised approaches remains largely consistent in most of the cells, except in the cells near the western boundary (column i) and along the jet (row b) where the local flow speed is large to violate the locality assumption in the localised approach. Overall the global approach tends to produce diffusivity estimates better aligned with the Davis (1987) diffusivity than its localised approach counterpart. The global approach is capable of restricting the diffusivity estimates within a range of a factor of approximately 2 from the MAP estimates and Davis (1987) diffusivity.

While in this section we only presented MAP estimates and marginal distributions of the mean velocity and diffusivity components, the high-dimensional posterior, jointly depending on both $\mathbf{U}$ and K, can also be used to study the correlation between $\mathbf{U}$ and K. In figure 5.20 we visualise the correlation between the diffusivity components $\mathsf{K}^{\perp}$, $\mathsf{K}^{\parallel}$ and the magnitude of mean velocity $|\mathbf{U}|$ at different geographical locations. With the scatter plots of the posterior samples, we note generally that both diffusivity components generally increases with velocity across geographical locations. We highlight one outlying exception – the green points in the middle column, corresponding to the jet head, signal a clear diffusivity anisotropy in that the cross-stream diffusivity is suppressed, hinting a mixing barrier. This shows the potential for a detailed posterior analysis to systematically reveal and classify different flow regimes and be further used to address oceanographic questions.

5.5 Validation

To validate and assess the performance of the inference, we carry out a series of tracer dispersion experiments. A benchmark is created by averaging ensembles of tracer, identically initialised but released at different times, that are advected in the high-resolution quasigeostrophic simulation. In addition, a prediction of the tracer distribution can be made by solving the advection–diffusion equation with the inferred MAP estimates of velocity and diffusivity fields. By comparing the prediction against the benchmark, the quality of the MAP estimates is assessed.

5.5.1 Benchmark and performance metric

Tracers are recurrently released with the same initial profile in the high-resolution quasigeostrophic simulations at different times. Their profiles on day 96 after the release are recorded as a sample. For each initial profile, the samples are averaged to define the benchmark.

A total number of 36 initial conditions are considered. Each of the profiles is Gaussian with diagonal covariance with entries of 120^2 km^2, but differs by the location of its peak. We assign the peaks at the centres of a uniform 6-by-6 grid dividing the square domain (shown in the first column of figure 5.23).

The simultaneous release of tracers begins upon the completion of the simulation of 10-year Lagrangian trajectories used for the inference. After 96 days from the release, we record the sample tracer distribution and re-initialise the tracer in its initial Gaus-

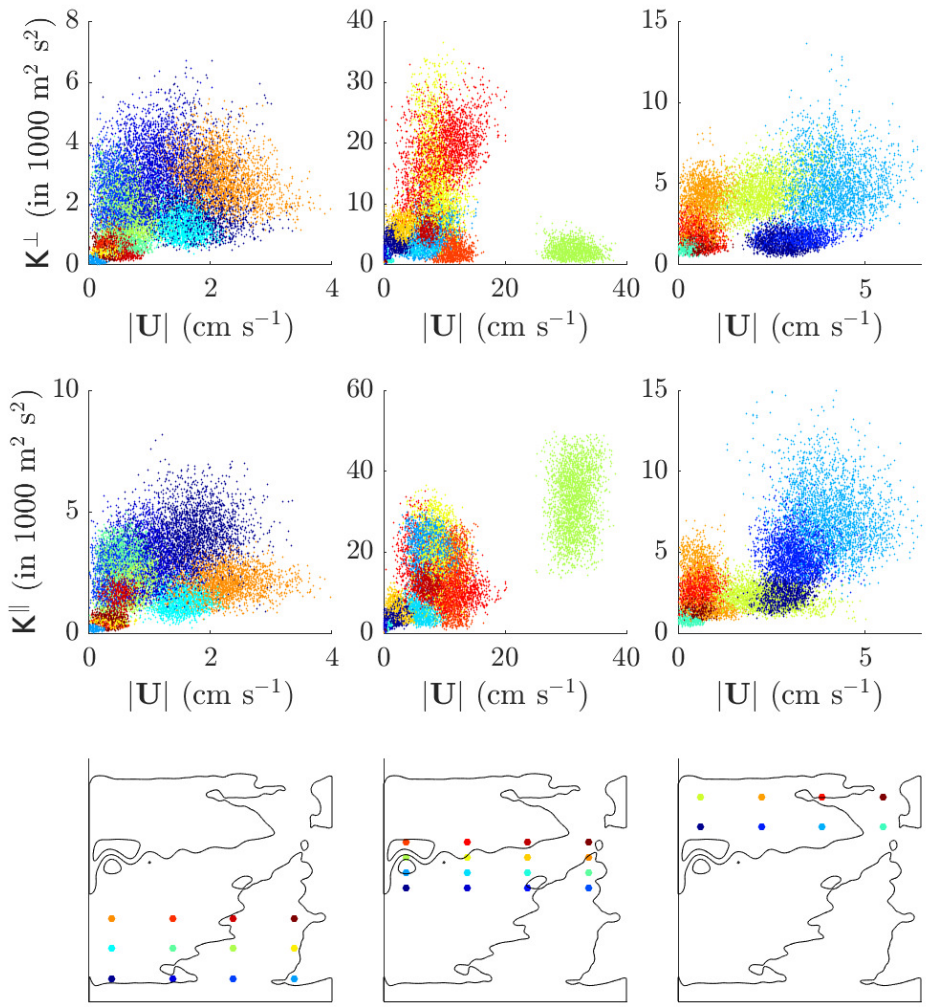

Figure 5.20: Posterior sample distributions of the cross-stream $\mathsf{K}^{\perp}$ (top row) and along-stream diffusivity (middle row) $\mathsf{K}^{\parallel}$ against the magnitude of mean velocity $|\mathbf{U}|$ at different geographical locations, inferred from the 10-year trajectory of 676 particles sampled at a 96-day interval in the middle layer quasigeostrophic model. Every coloured point in the scatter plots in the top two rows corresponds to a posterior sample. Points are coloured according to geographic location as indicated in the bottom row.

.

sian profile. This is repeated for 200 times and to give 200 samples of day-96 tracer distribution, which are then averaged to determine the benchmark.

Each tracer independently is advected under a streamfunction interpolated, linearly in time and bilinearly in space, from the discrete streamfunction computed at every time step of the quasigeostrophic calculation. The tracer advection is solved on a grid of doubled spatial resolution and quadrupled temporal resolution to the quasigeostrophic calculations, using the same finite volume method, with diffusion set to zero, for the test tracers in the data coarse-grained likelihood.

Upon normalisation, any instantaneous, non-negative and normalisable tracer profile can be interpreted as a probability density for particle positions. The benchmark serves as an empirical measure for the probability density on day 96. Should the inference perform well, this probability density will also be predicted by solving the Fokker–Planck equation with the same initial Gaussian distribution under the inferred velocity and diffusivity fields. The discrepancy between the benchmark $\Pi(\mathbf{x})$ and the predicted probability density $\pi(\mathbf{x})$ is measured by a performance metric to assess the inference quality.

An example performance metric is the L^2- distance between the prediction $\pi(\mathbf{x})$ and the benchmark $\Pi(\mathbf{x})$, divided by the L^2-norm of the benchmark

$$d\left(\Pi(\mathbf{x}), \pi(\mathbf{x})\right) = \frac{\|\Pi(\mathbf{x}) - \pi(\mathbf{x})\|}{\|\Pi(\mathbf{x})\|}. \tag{5.32}$$

This measures the relative error of $\pi(\mathbf{x})$ from $\Pi(\mathbf{x})$. The L^2 distance quantifies the average relative error of the prediction over the complete domain.

Another example of a performance metric is the Kullback-Leibler divergence of a prediction $\pi(\mathbf{x})$ from the benchmark $\Pi(\mathbf{x})$

$$D_{\mathrm{KL}}\left(\Pi(\mathbf{x})||\pi(\mathbf{x})\right) = \int \Pi(\mathbf{x}) \log\left(\frac{\Pi(\mathbf{x})}{\pi(\mathbf{x})}\right) \mathrm{d}\mathbf{x}. \tag{5.33}$$

This measures the information loss when the benchmark $\Pi(\mathbf{x})$ is approximated by prediction $\pi(\mathbf{x})$. Note that the Kullback-Leibler divergence is non-negative, and becomes zero only when the benchmark is identical to the prediction. The Kullback-Leibler divergence provides a formal quantification of the differences between two probability distribution.

5.5.2 Configurations

Two classes of predictions are considered – 1) 'MAP': obtained using the mean flow and diffusivity at the MAP estimates of the global Bayesian approach described in section 5.4.4. The fields are defined piecewise bilinearly on a 16-by-16 grid and are derived from 10-year trajectories of 676 Lagrangian particles sampled at an interval $s = 96$ days; 2) 'Davis': obtained using a prescribed Eulerian mean flow and the Davis (1987) diffusivity estimated in section 5.4.4. For a fair comparison, the Eulerian mean velocity field is L^2-optimally projected onto a piecewise bilinear function space on a 16-by-16 grid. Since the numerical solver is designed for symmetric positive-definite tensors, we post-process the Davis (1987) diffusivity by symmetricising and taking absolute values artificially on the eigenvalues. The latter step alters the diffusivity in 9 cells (see the right panel of figure 4.7 and Appendix A.1 for the post-processing). These define a piecewise constant Davis (1987) diffusivity field on the same 16-by-16 grid.

The prediction of tracer profile at $T = 96$ days is conducted by solving the Fokker–Planck equations on a fine 256 grid. We independently consider the prediction from the

MAP estimates of the velocity and diffusivity fields inferred from sampling intervals $s = 32$, 64 and 96 days. The same solver as in the likelihood evaluation (see section 5.3.3 and Appendix C.1) is applied.

5.5.3 Results

Figure 5.21 and 5.22 show several examples of the tracer distribution at time $T = 96$ days. Tracers on the same row are initialised identically, and their initial profile is shown on the leftmost column. For reference we show contour lines of the Eulerian mean streamfunction in white at the background of each panel. The tracer distributions predicted by both the MAP estimates and Davis (1987) diffusivity share common large-scale feature with the benchmark, in the eye-ball metric. Specifically Figure 5.21 and Figure 5.22 corresponds to initial conditions for which the 'MAP' predictions and 'Davis' predictions respectively fit better with the benchmark.

Both the L^2 relative error and Kullback-Leibler divergence are computed to evaluate the proximity between the benchmark and the two predictions from 'MAP' and 'Davis'. By subtracting the metric value for the 'MAP' prediction from the 'Davis' prediction, we can diagnose the relative performance between the 'MAP' prediction and the 'Davis' prediction – here a positive value indicates the 'MAP' prediction a better fit than the 'Davis' one.

We diagnose the relative performance for each initial Gaussian profile. The initial profiles, labelled zonally from 1 to 6 and meridionally from A to F, are shown in the left panel in figure 5.23. The middle and right panel of figure 5.23 show the diagnostics in L^2 relative error and Kullback-Leibler divergence respectively for each initial profiles. Both metrics indicate the superior performance of the 'MAP' estimates near the northern, southern and western boundaries (row A and F, column 1), at the head of the jet ($C1$ to $E2$). Near the eastern boundary (column 6, see for example the bottom row of figure 5.22) the 'MAP' prediction performed badly due to the poorly estimated zonal-zonal diffusivity in the Eastern boundary (discussed in 5.4.4, the zonal-zonal diffusivity is not well constrained on the eastern boundary). Another area where the data coarse-grained inference appears to under-perform is at the tail of the jet ($D2$ to $E5$). However, we emphasise that the 'MAP' prediction is still in a reasonably good agreement with the benchmark (see first to third row in figure 5.22).

5.6 Discussion and summary

In this chapter, we developed a data coarse-grained approach to infer globally-defined velocity and eddy diffusivity fields from Lagrangian trajectories. Similar to the localised approach, the data coarse-grained approach divides the domain of interest into cells and assigns each transition data point to its starting cell. The precise starting positions in each cell are then discarded, and are instead modelled by the invariant distribution. This defines a new likelihood function for the data, so that only one Fokker–Planck equation needs to be numerically solved per cell. This is in practice a drastic reduction in computational cost compared with the primitive inference, where every transition data point demands an independent solution to a Fokker–Planck equation.

The difference between the primitive and data coarse-grained inference is analysed in section 5.2, where we show that the maximum likelihood estimates in the data coarse-grained and primitive inference are the same. We proceed to show also that coarse-graining the data does not introduce extra certainty to the estimates, relative

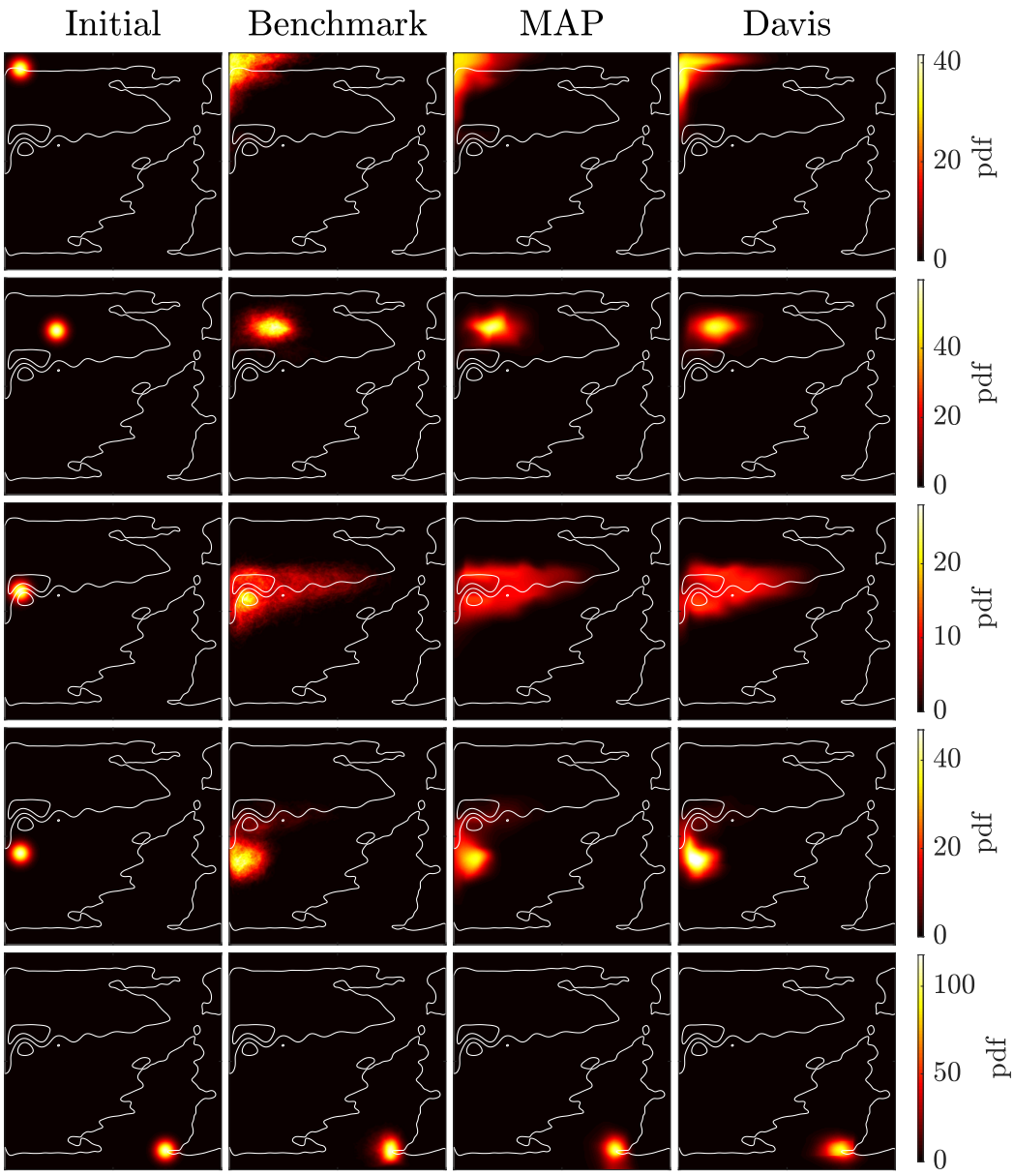

Figure 5.21: Examples of initial tracer distributions (leftmost column) and on day 96 in the middle layer (second to fourth columns) where the 'MAP' predictions fit better with the benchmark than the 'Davis' predictions, by comparing both performance metrics of Kullback-Leibler divergence and L^2 distance. The second to fourth columns (C2 to C4) from left to right are: C2) 'Benchmark': the ensemble average of tracer distribution on day 96; C3) 'MAP': the tracer distribution on day 96 predicted using the MAP estimates of velocity and diffusivity in the global approach, derived from trajectory data sampled at interval 96 days; 4) 'Davis': the tracer distribution on day 96 predicted using the projected Eulerian mean velocity and a post-processed Davis (1987) diffusivity, derived from the trajectory data evaluated at time lag 96 days.

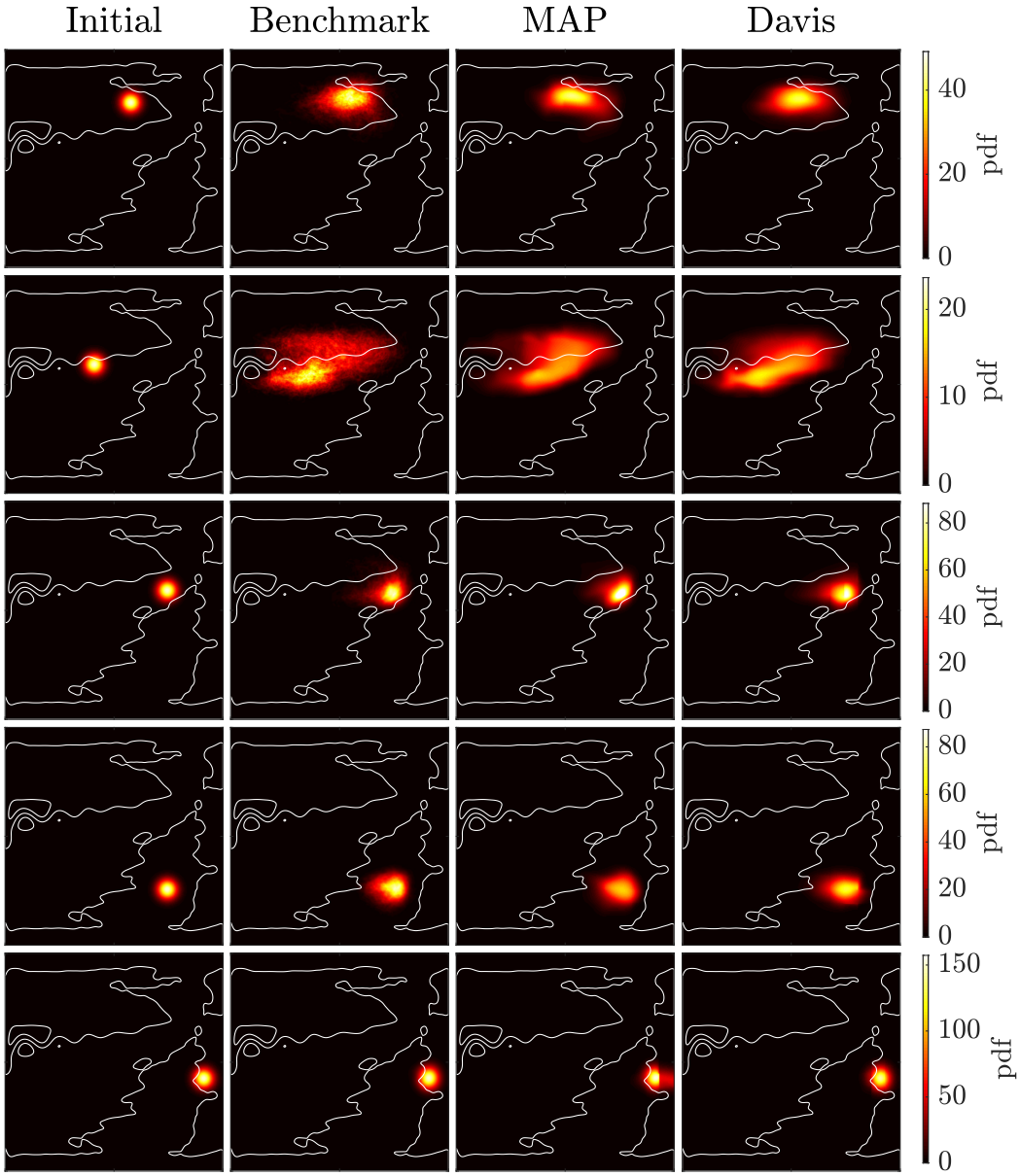

Figure 5.22: Same as figure 5.21 but for initial conditions where the 'MAP' predictions fit worse with the benchmark than the 'Davis' predictions, by comparing both performance metrics of Kullback-Leibler divergence and $L-2$ distance.

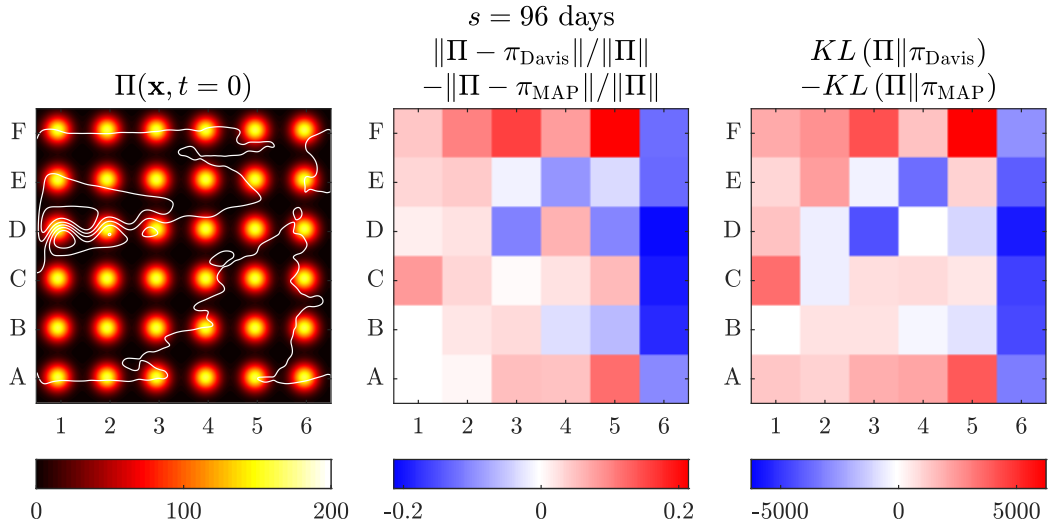

Figure 5.23: Left panel: the 6-by-6 distinct initial Gaussian profiles visualised simultaneously on one domain. Middle panel: Relative performance of the 'MAP' prediction over 'Davis', measured using the derivation of tracer distribution from the benchmark on day 96 in L^2-norm, for each of the initial profiles. Right panel: Relative performance between the 'MAP' prediction and 'Davis', measured using the Kullback-Leibler divergence of tracer distribution from the benchmark on day 96, for each of the initial profile.

to the primitive approach.

The data coarse-grained inference is tested on two sets of Lagrangian trajectory data, one simulated from an SDE and another from the quasigeostrophic double gyre calculation used also in Chapter 4. In both tests, piecewise linear streamfunction and diffusivity fields are inferred.

In the test with synthetic SDE data, the global inference demonstrates its capability to recover a discrete representation of the underlying velocity and diffusivity fields as data size increases. The global inference is also robust to the sampling intervals, showing little variability in the estimates as particles straddle multiple cells. Nonetheless, the data coarse-grained inference struggles in estimating the diffusivity on the boundaries, particular at the corners. We conjecture this is an artefact of the data coarse-grained approach and discuss a pathological case when diffusivity become non-indentifiable.

In the test with synthetic trajectories in the double gyre, the global inference is able to capture the expected spatial features in both the mean velocity and diffusivity. The inference again demonstrates its robustness with varying sampling interval. This is particularly important for estimating eddy diffusivity, as particles should be sampled at at an interval beyond the decorrelation time scale. The inference appears to encounter a pathological case near the eastern boundaries, where the zonal-zonal diffusivity on the boundary cannot be identified as the zonal transport is too weak.

To validate the inference for the double gyre flow, we devise a tracer release experiment. A benchmark is derived from ensembles of sample tracers successively dropped at different locations to follow the double gyre flow. The dispersed tracers are compared with the predictions from a coarse-grained advection–diffusion model, using the inferred MAP estimates and Davis diffusivity. We note a generally good agreement between the benchmark and the prediction from MAP estimates.

Chapter 6

Overcoming non-locality – a large-deviation approach

6.1 Introduction

To infer the velocity and diffusivity fields from trajectory data, the localised inference presented in chapter 4 relies on the locality of the likelihood function, assuming that the particle displacements over the sampling time are small relative to the variability of velocity and diffusivity, to infer cell-wise and independently defined fields. The strength of the localised approach is that the likelihood formulated is given in a closed form, avoiding a numerical solve for the Fokker–Planck equations involved in the likelihood evaluation.

The locality assumption is however often invalid in an oceanographic context. Particles could straddle multiple cells over the considered time interval, yet the displacements are attributed only to the cells they start from in the localised approach. In Chapter 5 a global inference, based on a data coarse-grained method, is proposed to overcome the locality limitation and constrain the number of Fokker–Planck equations that need to be solved per likelihood evaluation. However, it may be practically challenging to scale up the global inference to infer fields at a fine resolution and ultra-long sampling interval, as each numerical evaluation for the Fokker–Planck equations will become increasingly computational costly.

This chapter presents an analytic approach to derive an explicit likelihood function for particles displaced over two neighbouring cells. This method applies large deviation theory to construct an approximate transition density for particles in advection-dominated flows.

In section 6.2 we present the framework and analytic approximations of the large deviation principle for stochastic differential equations in the weak noise limit. Section 6.3 concerns the application of the large deviation theory to derive the transition density with a piecewise constant velocity and diffusivity field defined on two half-planes. We then apply the developed transition density in section 6.4 to perform inference from synthetic trajectories with a piecewise constant velocity and diffusivity field. Section 6.5 summarises the findings in this chapter.

6.2 The large deviation principle

6.2.1 Stochastic differential equations with weak noise

Consider the stochastic differential equation

$$\mathrm{d}\mathbf{X} = \mathbf{u}\left(\mathbf{X}\right)\mathrm{d}t + \sqrt{\epsilon}\sqrt{2\mathsf{K}\left(\mathbf{X}\right)}\,\mathrm{d}\mathbf{W}, \tag{6.1}$$

here $\mathbf{u}\left(\mathbf{x}\right)$ and $\mathsf{K}\left(\mathbf{x}\right)$ are the spatially-varying velocity and diffusivity tensor fields respectively, and ϵ is a dimensionless bookkeeping parameter that controls the relative strength of diffusion to advection. Note that the considered SDE is different from equation (1.9) and does not contain the drift due to the diffusivity gradient. The transition probability for the SDE (6.1) can be evaluated by solving the Fokker–Planck equation

$$\frac{\partial \pi}{\partial t} + \nabla \cdot (\mathbf{u}\pi) = \epsilon \mathsf{K} : \nabla^2 \pi, \tag{6.2}$$

where : is the Frobenius inner product and ∇^2 is the Hessian matrix.

In the limit of $\epsilon \to 0$, we expect the stochastic path $\mathbf{X}(t)$ to become more alike to the deterministic path $\mathbf{X}^*(t)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}^*(t) = \mathbf{U}(\mathbf{X}^*(t)), \quad \mathbf{X}^*(0) = \mathbf{y}.$$

The transition density converges in the small limit of ϵ

$$\lim_{\epsilon \to 0} \pi(\mathbf{x}, t|\mathbf{y}) = \delta(\mathbf{x} - \mathbf{X}^*(t)).$$

One may expect the rate of convergence can be determined by considering an ansatz to the Fokker–Planck equation (6.2)

$$\pi(\mathbf{x}, t|\mathbf{y}) \sim \exp\left(-\frac{1}{\epsilon} J(\mathbf{x}, t|\mathbf{y})\right),$$

for some rate function $J(\mathbf{x}, t|\mathbf{y})$. Solving the Hamilton–Jacobi equation resulting from the collection of the leading order terms leads to a solution (see e.g. E et al., 2019)

$$J(\mathbf{x}, t|\mathbf{y}) = \inf_{\mathbf{X} \in \mathcal{C}([0,t])} \left\{ \frac{1}{2} \int_0^t \left\| \dot{\mathbf{X}}(s) - \mathbf{u}(\mathbf{X}(s)) \right\|^2_{2\mathsf{K}(\mathbf{X}(s))} \mathrm{d}s : (\mathbf{X}(0), \mathbf{X}(t)) = (\mathbf{y}, \mathbf{x}) \right\}, \tag{6.3}$$

where $\|\mathbf{v}\|_{\mathsf{M}} = \sqrt{\mathbf{v}^{\mathrm{T}}\mathsf{M}^{-1}\mathbf{v}}$ is the M^{-1}-norm induced by a symmetric positive-definite matrix M. We define the *rate functional* $J(\mathbf{X})$ for a sample path $\mathbf{X} \in C([0, t])$ with deterministic starting and end positions $\mathbf{X}(0) = \mathbf{y}$ and $\mathbf{X}(t) = \mathbf{x}$,

$$J(\mathbf{X}) = \frac{1}{2} \int_0^t \|\dot{\mathbf{X}}(s) - \mathbf{u}(\mathbf{X}(s))\|^2_{2\mathsf{K}(\mathbf{X}(s))} \mathrm{d}s, \tag{6.4}$$

which is non-negative for any $t > 0$. This leads to a *large deviation principle* (LDP) on the transition probability

$$\pi(\mathbf{x}, t|\mathbf{y}) \asymp \exp\left(-\frac{1}{\epsilon} \inf_{\mathbf{X} \in \mathcal{A}(\mathbf{x}|\mathbf{y})} J(\mathbf{X})\right), \tag{6.5}$$

here the log-equality sign '$\asymp$' is defined

$$a(\epsilon) \asymp b(\epsilon) \iff \lim_{\epsilon \to 0} \log a(\epsilon) = \lim_{\epsilon \to 0} \log b(\epsilon), \tag{6.6}$$

and the infimum is over the set $\mathcal{A}(\mathbf{x}|\mathbf{y})$ of continuous paths constrained by the start and end positions $(\mathbf{X}(0), \mathbf{X}(t)) = (\mathbf{y}, \mathbf{x})$.

Instead of showing the detailed derivation of this approach, which is largely mechanical manipulations of partial differential equations, we outline an alternative probabilistic argument to explain the appearance of the path integral. The detailed derivation can be found in the textbook of E et al. (2019).

Consider the sample paths of an SDE over the space of continuous function $\mathcal{C}([0,t])$, defined on a fixed time interval $[0,t]$, which have a fixed initial position $\mathbf{X}_0 = \mathbf{y}$. The joint probability of observing a sample path $\mathbf{X}(t)$ at uniform sample time $t = t_j$, where $0 = t_0 < t_1 < t_2 < \cdots < t_{N+1} = t$ and $\Delta t = t_{j+1} - t_j$, is given by the product of transition densities $\pi(\cdot, \cdot|\cdot)$

$$p(\{\mathbf{X}(t_j) : \ j = 0, 1, \dots N\} = \prod_{j=0}^{N} \pi(\mathbf{X}(t_{j+1}), \Delta t|\mathbf{X}(t_j). \tag{6.7}$$

In the limit of small time increment $\Delta t \to 0$, each transition density can be approximated by a Gaussian distribution

$$\pi(\mathbf{X}(t_{j+1}), \Delta t|\mathbf{X}(t_j) = \frac{1}{Z_j(\Delta t)} \exp\left(-\frac{1}{2\epsilon\Delta t}\|\mathbf{X}(t_{j+1}) - \mathbf{X}(t_j) - \mathbf{u}(\mathbf{X}(t_j))\Delta t\|^2_{2\mathsf{K}(\mathbf{X}(t_j))}\right),$$

where $Z_j(\Delta t) = 4\pi\epsilon\Delta t\sqrt{|\mathsf{K}(\mathbf{X}(t_j)|}$ is the normalisation constant. Introducing the Gaussian density into the joint density (6.7) and simplifying lead to

$$p(\{\mathbf{X}(t_j) : \ j = 0, 1, \dots N\}) \propto \exp\left(-\frac{\Delta t}{2\epsilon}\sum_{j=0}^{N}\left\|\frac{\mathbf{X}(t_{j+1}) - \mathbf{X}(t_j)}{\Delta t} - \mathbf{u}(\mathbf{X}(t_j))\right\|^2_{2\mathsf{K}(\mathbf{X}(t_j))}\right). \tag{6.8}$$

In the limit of $\Delta t \to 0$, define $\dot{\mathbf{X}} = (\mathbf{X}(t_{j+1}) - \mathbf{X}(t_j))/\Delta t$, the finite sum converges to an integral. The 'probability' of a path $\mathbf{X}$, which formally should be expressed as a measure $\mathrm{d}\mu(\cdot)$ on the path is defined by

$$\mathrm{d}\mu[\mathbf{X}] \propto \exp\left(-\frac{1}{2\epsilon}\int_0^t \left\|\dot{\mathbf{X}}(s) - \mathbf{u}(\mathbf{X}(s))\right\|^2_{2\mathsf{K}(\mathbf{X}(s))} \mathrm{d}s\right) \prod_{s\in[0,t]} \mathrm{d}\mathbf{X}(s). \tag{6.9}$$

Note that this is a measure defined on infinite-dimensional space. Now consider the subset of paths A which terminates at $\mathbf{X}(t) = \mathbf{x}$. The transition probability is given by 'integrating' the path measure $\mathrm{d}\mu(\cdot)$ over the set $\mathcal{A}(\mathbf{x}|\mathbf{y})$ of paths with start and terminal positions $\mathbf{y}$ and $\mathbf{x}$,

$$\begin{aligned}\pi(\mathbf{x}, t|\mathbf{y}) &\propto \int_{\mathcal{A}(\mathbf{x}|\mathbf{y})} \mathrm{d}\mu[\mathbf{X}] \\ &= \int_{\mathcal{A}(\mathbf{x}|\mathbf{y})} \exp\left(-\frac{1}{2\epsilon}\int_0^t \left\|\dot{\mathbf{X}}(s) - \mathbf{u}(\mathbf{X}(s))\right\|^2_{2\mathsf{K}(\mathbf{X}(s))} \mathrm{d}s\right) \prod_{s\in[0,t]} \mathrm{d}\mathbf{X}(s).\end{aligned}$$

The contribution of $J(\mathbf{X})$ for any path to the transition density $\pi(\mathbf{x}, t|\mathbf{y})$ decays exponentially as ϵ decreases. Ultimately as $\epsilon \to 0$, the transition density is dominated by the path in set A that maximises the exponent $-J(\mathbf{X})/\epsilon$. This leads to the formula

$$\pi(\mathbf{x}, t|\mathbf{y}) \asymp \exp\left(-\frac{1}{\epsilon}\inf_{\mathbf{X}\in\mathcal{A}(\mathbf{x}|\mathbf{y})} J(\mathbf{X})\right), \tag{6.10}$$

coinciding with the formula (6.5) derived from the approach of asymptotic methods.

A highly technical proof can be found in Section 5.6 of Dembo and Zeitouni 1998 and section 5.3 of Freidlin et al. 2012, where the authors rigorously show that a large deviation principle exists for the probability of any closed set $\mathcal{B}$ of sample paths

$$\lim_{\epsilon\to 0} \epsilon \log(P(\mathbf{X}\in\mathcal{B})) = -\inf_{\mathbf{X}\in\mathcal{B}} J(\mathbf{X}),$$

whereas we only consider a special set of paths with fixed endpoints in this chapter.

6.2.2 Transition density and normalisation constant

The notion of log-equivalence in the transition density (6.5) only outlines the exponential component of $\pi(\mathbf{x},t|\mathbf{y})$. In order for $\pi(\mathbf{x},t|\mathbf{y})$ to define a probability density function over $\mathbf{x}$, a prefactor $1/C(\epsilon)$ needs to be determined in

$$\pi(\mathbf{x},t|\mathbf{y}) = \frac{1}{C(\epsilon)} \exp\left(-\frac{1}{\epsilon}\inf_{\mathbf{X}\in A} J(\mathbf{X})\right), \tag{6.11}$$

so that $\int \pi(\mathbf{x},t|\mathbf{y})\mathrm{d}\mathbf{x} = 1$. Define a non-negative auxiliary function

$$g(\mathbf{x}) := \inf_{\mathbf{X}\in\mathcal{A}(\mathbf{x}|\mathbf{y})} \frac{1}{2}\int_0^t \|\dot{\mathbf{X}}(s) - \mathbf{u}(\mathbf{X}(s))\|^2_{2\mathsf{K}(\mathbf{X}(s))}\mathrm{d}s. \tag{6.12}$$

The normalisation constant $C(\epsilon)$ is then an exponential integral in the form of

$$C(\epsilon) = \int \exp\left(-g(\mathbf{x})/\epsilon\right)\mathrm{d}\mathbf{x}. \tag{6.13}$$

In the limit of $\epsilon\to 0$, the integral can be approximated using Laplace's method. The method assumes the existence and uniqueness of a global minimiser $\mathbf{x}^*$ to $g(\mathbf{x})$. By a Taylor expansion on $g(\mathbf{x})$ around $\mathbf{x}^*$ up to second order (note first order derivative vanishes), the integrand term

$$\exp\left(-g(\mathbf{x})/\epsilon\right) \approx \exp\left(-g(\mathbf{x}^*)/\epsilon - (\mathbf{x}-\mathbf{x}^*)^{\mathrm{T}}\mathsf{H}(g)(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)/2\epsilon\right)$$

is locally Gaussian, here $\mathsf{H}(g)(\mathbf{x}^*)$ is the Hessian matrix of g evaluated at $\mathbf{x}^*$. A direct integration yields

$$C(\epsilon) \sim 2\pi\epsilon \frac{\exp\left(-g(\mathbf{x}^*)/\epsilon\right)}{\sqrt{|\mathsf{H}(g)(\mathbf{x}^*)|}},$$

with $|\cdot|$ the determinant of a tensor.

Since the deterministic path satisfying $\dot{\mathbf{X}}(s) = \mathbf{u}(\mathbf{X}(s))$ keeps the path integral zero throughout the time-interval $s\in[0,t]$, the terminal position $\mathbf{x}^*$ of the deterministic path, given by

$$\mathbf{x}^* = \mathbf{X}(0) + \int_0^t \mathbf{u}(\mathbf{X}(s))\mathrm{d}s, \qquad \mathbf{X}(0) = \mathbf{y}, \tag{6.14}$$

is the global minimiser of $g(\mathbf{x})$ where $g(\mathbf{x}^*) = 0$. These imply the normalisation constant $C(\epsilon)$ to be

$$C(\epsilon) \approx 2\pi\epsilon|\mathsf{H}(g)(\mathbf{x}^*)|^{-1/2}. \tag{6.15}$$

6.3 The transition density of a piecewise constant flow

In this section we apply the large deviation theory presented in the last section to derive an analytic transition density for piecewise constant velocity and diffusivity field defined on two half-planes. This may serve as an adaptation for the localised approach in Chapter 4, where cell-wise fields are inferred independently, for particles transported between two adjacent cells. In an advection-dominated flow, the majority of particles would be transported in the along stream direction. If the interval considered is short enough to confine particles transport in only the adjacent cells, one may expect that the transition density is largely controlled by the velocity and diffusivity in the cells in the along stream direction. We utilise this property in the large deviation theory to define a transition density that depends simultaneously on fields in both cells.

It is worthwhile noting that strong solutions to stochastic differential equations with discontinuous velocity and diffusivity do not always exist, and thus, the large deviation results may not be applicable. We proceed anyway with the large deviation framework, hoping the analysis in Chen et al. (2013) which proves the validity of the large deviation theory in a one-dimensional piecewise constant field can be extrapolated to the two-dimensional space. We are also reassured of the utility by later numerical verification in section 6.4.

6.3.1 The piecewise constant setting

Consider piecewise constant velocity $\mathbf{u}(\mathbf{x})$ and diffusivity $\mathsf{K}(\mathbf{x})$ defined on a pair of half-planes sharing an interface at $x = a$:

$$\text{(left half-plane) } D_l := \{(x, y) \in \mathbb{R}^2 : x \leq a\}, \tag{6.16}$$

$$\text{(right half-plane) } D_r := \{(x, y) \in \mathbb{R}^2 : x > a\}, \tag{6.17}$$

with

$$\mathbf{u}(\mathbf{x}) = \begin{cases} \mathbf{u}_l = (u_l, v_l)^{\mathrm{T}}, & \forall \mathbf{x} \in D_l, \\ \mathbf{u}_r = (u_r, v_r)^{\mathrm{T}}, & \forall \mathbf{x} \in D_r, \end{cases} \tag{6.18}$$

and

$$\mathsf{K}(\mathbf{x}) = \begin{cases} \mathsf{K}_l, & \forall \mathbf{x} \in D_l, \\ \mathsf{K}_r, & \forall \mathbf{x} \in D_r. \end{cases} \tag{6.19}$$

On a time interval $s \in [0, t]$, define the sub-time interval for a path $\mathbf{X}(s)$ to found in the left and right half-plane:

$$\mathcal{T}_l := \{s \in [0, t] : \mathbf{X}(s) \in D_l\}, \tag{6.20a}$$

$$\mathcal{T}_r := \{s \in [0, t] : \mathbf{X}(s) \in D_r\}. \tag{6.20b}$$

The rate functional (6.4) for the path takes the form of

$$J(\mathbf{X}) = \frac{1}{2} \left(\int_{\mathcal{T}_l} \|\dot{\mathbf{X}}(s) - \mathbf{u}_l\|^2_{2\mathsf{K}_l} \mathrm{d}s + \int_{\mathcal{T}_r} \|\dot{\mathbf{X}}(s) - \mathbf{u}_r\|^2_{2\mathsf{K}_r} \mathrm{d}s \right). \tag{6.21}$$

Consider the continuous paths that transit from one half-plane to another at most once over the time-interval $[0, t]$ such that the two sub-intervals $\mathcal{T}_l$ and $\mathcal{T}_r$ are connected.

The optimal path to maximise each of the functionals on the right-hand side of (6.21) respectively satisfies the Euler–Lagrange equation

$$\left(\nabla_{\mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\nabla_{\dot{\mathbf{x}}}\right)\left(\|\dot{\mathbf{X}}(t) - \mathbf{u}_l\|^2_{2\mathsf{K}_l}\right) = 0, \tag{6.22}$$

$$\left(\nabla_{\mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\nabla_{\dot{\mathbf{x}}}\right)\left(\|\dot{\mathbf{X}}(t) - \mathbf{u}_r\|^2_{2\mathsf{K}_r}\right) = 0, \tag{6.23}$$

here $\nabla_{\mathbf{x}}$ and $\nabla_{\dot{\mathbf{x}}}$ are the gradients with respect to the position and velocity. By noting that the velocity and diffusivity are constant in both equations, a direct evaluation gives

$$\ddot{\mathbf{X}}(s) = 0, \quad \forall s : x(s) \neq a. \tag{6.24}$$

This means the optimal velocity $\dot{\mathbf{X}}(s)$ remains constant on each of the half-planes D_l and D_r, which geometrically leads to a trajectory of piecewise straight lines.

When diffusion is absent, paths of particles are deterministically controlled by the velocity field. With small diffusivity, the optimal velocity $\dot{\mathbf{X}}$ for a particle trajectory with fixed end-points is a small perturbation from the background velocity. This motivates a perturbative expression of the optimal velocity on the background velocity

$$\dot{\mathbf{X}}(s) = \begin{cases} \mathbf{u}_l + \delta\mathbf{u}_l = (u_l + \delta u_l, v_l + \delta v_l)^{\mathrm{T}}, & \text{when } \mathbf{X}(s) \in D_l, \\ \mathbf{u}_r + \delta\mathbf{u}_r = (u_r + \delta u_r, v_r + \delta u_r)^{\mathrm{T}}, & \text{when } \mathbf{X}(s) \in D_r. \end{cases} \tag{6.25}$$

This simplifies the rate functional (6.21) into

$$J(\mathbf{X}) = \frac{1}{2}\left(\int_{\mathcal{T}_l} \|\delta\mathbf{u}_l\|^2_{2\mathsf{K}_l}\mathrm{d}s + \int_{\mathcal{T}_r} \|\delta\mathbf{u}_r\|^2_{2\mathsf{K}_r}\mathrm{d}s\right), \tag{6.26}$$

which suggests that the search of an optimal path for the rate functional (6.21) is equivalent to identifying the perturbing velocities $\delta\mathbf{u}_l$ and $\delta\mathbf{u}_r$ of the optimal path, and determining the time intervals $\mathcal{T}_l$ and $\mathcal{T}_r$ the optimal path spent in each half-plane.

In the following sections, we consider the transition density $\pi(\mathbf{x}, t|\mathbf{y})$ with an initial position in the left half-plane $\mathbf{y} \in D_l$ without loss of generality.

6.3.2 Derivation for transition probability in the left half-plane

This section concerns the evaluation of $\pi(\mathbf{x}, t|\mathbf{y})$ for $\mathbf{x}$ in the left half-plane D_l.

Knowing that the optimal path is piecewise straight, the only feasible path that connects $\mathbf{y}$ and $\mathbf{x}$ on the same half-plane is the straight line through the two points. This can be formally derived by noting that the optimal path never crosses the half-plane interface so that $\mathcal{T}_r = \{s : \mathbf{X}(s) \in D_r\} = \emptyset$. Minimising the rate functional (6.26) is therefore equivalent to minimising a finite-dimensional cost function

$$J^d_l(\delta\mathbf{u}_l) = \frac{1}{2}t\|\delta\mathbf{u}_l\|^2_{2\mathsf{K}_l}, \tag{6.27}$$

subject to the constraints $\mathbf{x} = \mathbf{y} + (\mathbf{u}_l + \delta\mathbf{u}_l)\,t$. The constraint uniquely determines $\delta\mathbf{u}_l$ and gives the minimum of $J^d_l(\delta\mathbf{u}_l)$:

$$\min_{\mathbf{x}=\mathbf{y}+(\mathbf{u}_l+\delta\mathbf{u}_l)t} J^d_l(\delta\mathbf{u}_l) = \frac{1}{2}\|\mathbf{x} - \mathbf{y} - t\mathbf{u}_l\|^2_{2\mathsf{K}_l t}.$$

The transition density is therefore

$$\pi(\mathbf{x}, t|\mathbf{y}) \asymp \exp\left(-\frac{1}{2}\|\mathbf{x} - (\mathbf{y} + t\mathbf{u}_l)\|^2_{2\epsilon\mathsf{K}_l t},\right), \quad \text{for } \mathbf{x} \in D_l, \tag{6.28}$$

which is a multivariate Gaussian distribution with mean $\mathbf{y} + t\mathbf{u}$ and covariance $2\epsilon\mathsf{K}_l t$, coinciding with the transition density for a uniform constant velocity and diffusivity. Note that (6.28) is defined up to a normalisation constant, whose evaluation is discussed in section 6.3.4.

6.3.3 Transition probability in the right half-plane

This section concerns the evaluation of $\pi(\mathbf{x}, t|\mathbf{y})$ for $\mathbf{x}$ in the right half-plane D_r.

The optimal piecewise straight paths $\mathbf{X}(s)$ that made a transition from the left to right half-plane is characterised by the time τ and location (a, b) at which the transition occurs.

To systematically obtain an analytic approximation to the transition probability, we first consider its peak of the transition probability profile. This is given by following the path without stochastic fluctuations, which minimises the constraint-free rate function (6.21). Two cases for the deterministic path are considered – **Case 1**: throughout the time interval $[0, t]$, the deterministic path remains in the left half-plane; **Case 2**: At some time $\tau^* \in [0, t]$, the deterministic trajectory makes its way to the interface at (a, b^*) before continuing its journey in the right half-plane. Examples of two cases are respectively shown the top two left panels and top two right panels in figure 6.2, with details of the figure given in the upcoming section 6.4.

Case 1 In this case, of interest is the transition density at a point in the right half-plane, even that the peak of the transition density lies in the left half-plane. Suppose the optimal path from $\mathbf{y}$ to $\mathbf{x}$ crosses the interface at point (a, b) and time τ, it follows that the perturbing velocities $\delta\mathbf{u}_l$ and $\delta\mathbf{u}_r$ in (6.25) are given by

$$\mathbf{y} + \tau\,(\mathbf{u}_l + \delta\mathbf{u}_l) = (a, b)^{\mathrm{T}} \iff \delta\mathbf{u}_l = \big((a, b)^{\mathrm{T}} - \mathbf{y}\big)/\tau - \mathbf{u}_l, \tag{6.29}$$

and

$$\mathbf{x} = (a, b)^{\mathrm{T}} + (t - \tau)\,(\mathbf{u}_r + \delta\mathbf{u}_r) \iff \delta\mathbf{u}_r = \big(\mathbf{y} - (a, b)^{\mathrm{T}}\big)/(t - \tau) - \mathbf{u}_r. \tag{6.30}$$

With the time intervals for the path to remain in each half-plane determined, the rate functional (6.26) becomes

$$J(\mathbf{X}) = \frac{1}{2}\left(\tau\|\delta\mathbf{u}_l\|^2_{2\mathsf{K}_l} + (t - \tau)\|\delta\mathbf{u}_r\|^2_{2\mathsf{K}_r}\right) \tag{6.31}$$

Rewriting (6.31) in terms of (6.29) and (6.30) shows the constrained minimisation of the rate functional $J(\mathbf{X})$ is equivalent to minimisation of a finite-dimensional cost function

$$J_r^{(1)}(b, \tau) = \frac{1}{2}\left(\frac{\|(a, b)^{\mathrm{T}} - \mathbf{y} - \tau\mathbf{u}_l\|^2_{2\mathsf{K}_l}}{\tau} + \frac{\|\mathbf{x} - (a, b)^{\mathrm{T}} - (t - \tau)\mathbf{u}_r\|^2_{2\mathsf{K}_r}}{t - \tau}\right), \tag{6.32}$$

with a free parameter b and a constrained τ satisfying an inequality $0 \le \tau \le t$. A numerical method is generally needed to minimise the cost function $J_r^{(1)}(b, \tau)$ and evaluate the transition density.

Case 2 In this case, both the point of interest $\mathbf{x}$ and the peak of the transition density lie in the right half-plane. We first examine the deterministic path, which is characterised by the crossing time τ^* and y-coordinate b^* of the point of crossing. This reads

$$\mathbf{y} + \tau^* \mathbf{u}_l = (a, b^*)^{\mathrm{T}} \iff \begin{cases} \tau^* = (a - x_i)/u_l, \\ b^* = y_i + v_l \tau^*. \end{cases} \tag{6.33}$$

The endpoint $\mathbf{x}^* = (x^*, y^*)$ of the deterministic path at time t is therefore given by

$$\mathbf{x}^* = (a, b^*)^{\mathrm{T}} + (t - \tau^*)\mathbf{u}_r \iff \begin{cases} x^* = a + (t - \tau^*)u_r, \\ y^* = b^* + (t - \tau^*)v_r. \end{cases} \tag{6.34}$$

For a position $\mathbf{x}$ in the neighbourhood of $\mathbf{x}^*$, the optimal piecewise straight path $\mathbf{X}$ is a small deviation from the deterministic path. This gives rises to a perturbative expression of the crossing time τ and crossing y-coordinate b of the optimal path

$$\begin{aligned} \tau &= \tau^* + \delta\tau \\ b &= b^* + \delta b \end{aligned}$$

in addition to perturbative velocities in (6.25). The rate functional (6.26) is thus expressed by

$$J(\mathbf{X}) = \frac{1}{2}\left(\int_0^{\tau^* + \delta\tau} \|\delta \mathbf{u}_l\|^2_{2\mathsf{K}_l} \mathrm{d}s + \int_{\tau^* + \delta\tau}^{t} \|\delta \mathbf{u}_r\|^2_{2\mathsf{K}_r} \mathrm{d}s \right), \tag{6.35}$$

subject to the constraints

$$\mathbf{y} + (\tau^* + \delta\tau)(\mathbf{u}_l + \delta\mathbf{u}_l) = (a, b^* + \delta b)^{\mathrm{T}}, \tag{6.36a}$$

$$(a, b^* + \delta b)^{\mathrm{T}} + (t - \tau^* - \delta\tau)(\mathbf{u}_r + \delta\mathbf{u}_r) = \mathbf{x}, \tag{6.36b}$$

$$0 \leq \tau^* + \delta\tau \leq t. \tag{6.36c}$$

Introducing the conditions (6.33) and (6.30) into the equality constraints in (6.36), they simplify into

$$\delta\mathbf{u}_l = \frac{(0, \delta b)^{\mathrm{T}} - \delta\tau \mathbf{u}_l}{\tau^* + \delta\tau}, \tag{6.37a}$$

$$\delta\mathbf{u}_r = \frac{\mathbf{x} - \mathbf{x}^* - (0, \delta b)^{\mathrm{T}} + \delta\tau \mathbf{u}_r}{t - \tau^* - \delta\tau}. \tag{6.37b}$$

Introducing (6.37) into (6.35), the constrained minimisation of the rate functional $J(\mathbf{X})$ is equivalent to minimising the two-dimensional cost function

$$J_r^{(2)}(\delta b, \delta\tau) = \frac{1}{2}\left(\frac{\|(0, \delta b)^{\mathrm{T}} - \delta\tau \mathbf{u}_l\|^2_{2\mathsf{K}_l}}{\tau^* + \delta\tau} + \frac{\|\mathbf{x} - \mathbf{x}^* - (0, \delta b)^{\mathrm{T}} + \delta\tau \mathbf{u}_r\|^2_{2\mathsf{K}_r}}{t - \tau^* - \delta\tau} \right), \tag{6.38}$$

subject to a sole inequality constraint

$$0 \leq \tau + \delta\tau \leq t. \tag{6.39}$$

Similar to the cost function $J_r^{(2)}(\delta b, \delta\tau)$ in **case 1**, the cost function $J_r^{(2)}(\delta b, \delta\tau)$ in this case, to the knowledge of the author, does not have a closed-form minimiser.

Analytic approximations to simplify the analysis Our analysis reveals that the transition density in the right half-plane can be evaluated by minimising the finite-dimensional cost functions (6.32) or (6.38). However, the absence of an analytic minimiser makes the cost functions unappealing to be used in Bayesian inference. We therefore attempt to obtain a closed-form approximation to the transition density, by firstly discarding the case for (6.32) with justifications, and secondly make an asymptotic approximation to (6.38).

The first claim is that the cost function (6.32), corresponding to a point in the right half-plane even though the deterministic path never reaches it, is practically negligible. In the vanishing diffusivity limit $\epsilon \to 0$, the transition probability mass would increasingly concentrate around the peak where deterministic path terminates at. Only when the peak is close to the intersection will a particle have a significant chance to be found in the right half-plane. With only a small fraction of particles expected to fall into this category, the values of transition density on the right half-plane becomes irrelevant in an inference. To preserve the full interpretability as a probability density function, we extend the transition density defined in (6.28) to the right half-plane

$$\pi(\mathbf{x}, t|\mathbf{y}) \asymp \exp\left(-\frac{1}{2}\|\mathbf{x}-\mathbf{y}-t\mathbf{u}_l\|^2_{2\epsilon\mathsf{K}_l t},\right), \quad \text{for } (x > a \text{ and } x^* < a), \tag{6.40}$$

acknowledging the crudeness of this definition. This approach unavoidably incurs an error, expected to be controlled jointly by the scale of diffusivity and time t. A detailed analysis can be carried out to identify the erroneous area and its dependence on different parameters.

The second claim is that the cost function (6.38) can be approximated without significantly altering the transition density. In this case the deterministic and optimal paths cross the half-plane interface at different crossing times (and positions). If crossing time difference $\delta\tau$ between the optimal path and the deterministic path is short compared with the time the optimal path spent in each half-plane, such that

$$|\delta\tau| \ll \tau \quad \text{and} \quad |\delta\tau| \ll t-\tau, \tag{6.41}$$

the fractions $1/(\tau+\delta\tau)$ and $1/(t-\tau-\delta\tau)$ in (6.38) can be approximated by

$$\frac{1}{\tau+\delta\tau} = \frac{1}{\tau} + \mathcal{O}\left(\delta\tau\right), \tag{6.42}$$

$$\frac{1}{t-\tau-\delta\tau} = \frac{1}{t-\tau} + \mathcal{O}\left(\delta\tau\right), \tag{6.43}$$

to formulate an approximating cost function

$$\tilde{J}^{(2)}_r(\delta b, \delta\tau) = \frac{1}{2}\left(\frac{\|(0,\delta b)^{\mathrm{T}} - \delta\tau\mathbf{u}_l\|^2_{2\mathsf{K}_l}}{\tau} + \frac{\|\mathbf{x}-\mathbf{x}^*-(0,\delta b)^{\mathrm{T}}+\delta\tau\mathbf{u}_r\|^2_{2\mathsf{K}_r}}{t-\tau}\right). \tag{6.44}$$

The approximating cost function $\tilde{J}^{(2)}_r$ approximates the exact cost function $J^{(2)}_r$ to second order in $\delta\tau$

$$J^{(2)}_r(\delta b, \delta\tau) = \tilde{J}^{(2)}_r(\delta b, \delta\tau) + \mathcal{O}\left(\delta\tau^3\right). \tag{6.45}$$

The approximating cost function $\tilde{J}^{(2)}_r(\delta b, \delta\tau)$ is quadratic with respect to $(\delta b, \delta\tau)$, whose minimiser has a closed-form expression. Upon algebraic manipulations, it turns

out that the minimum of the approximate $\tilde{J}_r^{(2)}(\delta b, \delta\tau)$ is quadratic with respect to $\mathbf{x} - \mathbf{x}^*$, that is,

$$\min_{\delta b, \delta\tau} \tilde{J}_r^{(2)}(\delta b, \delta\tau) = \frac{1}{4}(\mathbf{x} - \mathbf{x}^*)^{\mathrm{T}} \mathsf{\Gamma}^{-1} (\mathbf{x} - \mathbf{x}^*), \tag{6.46}$$

where $\mathsf{\Gamma} = \mathsf{\Gamma}(t)$ is a symmetric tensor with components given by

$$\mathsf{\Gamma}^{11}(t) = \frac{u_r^2}{u_l^2}\mathsf{K}_l^{11}\tau^* + \mathsf{K}_r^{11}(t - \tau^*), \tag{6.47a}$$

$$\mathsf{\Gamma}^{12}(t) = \left(\frac{u_r}{u_l}\mathsf{K}_l^{12} - \frac{u_r(v_l - v_r)}{u_l^2}\mathsf{K}_l^{11}\right)\tau^* + \mathsf{K}_r^{12}(t - \tau^*), \tag{6.47b}$$

$$\mathsf{\Gamma}^{22}(t) = \left(\mathsf{K}_l^{22} - 2\frac{v_l - v_r}{u_l}\mathsf{K}_l^{12} + \frac{(v_l - v_r)^2}{u_l^2}\mathsf{K}_l^{11}\right)\tau^* + \mathsf{K}_r^{22}(t - \tau^*), \tag{6.47c}$$

where the superscript ij presents the ij-th component of a tensor.

With the approximating cost function $J_r^{(2)}(\delta b, \delta\tau)$ replacing the exact (6.38), the transition density is therefore

$$\pi(\mathbf{x}, t|\mathbf{y}) \asymp \exp\left(-\frac{1}{\epsilon}\min_{\delta b, \delta\tau} \tilde{J}_r^{(2)}(\delta b, \delta\tau)\right) = \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}^*\|^2_{2\epsilon\mathsf{\Gamma}(t)}\right), \tag{6.48}$$

which is a Gaussian distribution with mean $\mathbf{x}^*$ and covariance $2\epsilon\mathsf{\Gamma}(\mathsf{t})$.

Note that the assumption (6.41) for the asymptotic expansion involves a large time-interval $t - \tau^*$ spent in the right half-plane relative to $\delta\tau$. This equivalently implies the deterministic position $\mathbf{x}^*$ at time t should be sufficiently far from the half-plane interface. Hence (6.48) is not expected to be valid at positions close to the interface. However, in the limit of small diffusivity, we expect this area of invalidity shrinks as the probability mass increasingly concentrates around the deterministic path.

For the same reason suggested in the first claim of the analytic approximations, we expect only a small number of particle samples would be found around the interface. We therefore extend the formula (6.48) to cover the complete right half-plane, resulting in

$$\pi(\mathbf{x}, t|\mathbf{y}) \asymp \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}^*\|^2_{2\epsilon\mathsf{\Gamma}(t)}\right), \quad \text{for } (x > a \text{ and } x^* > a). \tag{6.49}$$

Note that (6.28) is defined up to a normalisation constant, which is discussed in section 6.3.4.

6.3.4 Normalisation constant of the transition density

Section 6.3.2 shows that the transition density is Gaussian on the left half-plane. Section 6.3.3 shows that transition density can be approximated by a Gaussian distribution on the right half-plane. Combining the results yields

$$\pi(\mathbf{x}, t|\mathbf{y}) \asymp \begin{cases} \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{y} - t\mathbf{u}_l\|^2_{2\epsilon\mathsf{K}_l t},\right), & \text{for } (x < a) \text{ or } (x > a \text{ and } x^* < a), \\ \exp\left(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}^*\|^2_{2\epsilon\mathsf{\Gamma}(t)}\right), & \text{for } (x > a \text{ and } x^* > a). \end{cases} \tag{6.50}$$

Note that the transition density in (6.50) is defined up to a normalisation constant, which cannot be obtained by a brute-force integration of (6.50) over the whole domain due to the multi-sourced approximations made in deriving (6.50).

However, the normalisation constant can be approximated using the local Hessian of the transition density profile around its peak, as suggested in 6.2.2. The physical

interpretation is that in the small diffusivity limit, the probability mass of the transition density concentrates around the peak $\mathbf{x}^*$ so that the transition density can be locally approximated by a Gaussian profile with the local Hessian matrix being the covariance. Applying the formula (6.15), the normalisation constant $C(\epsilon)$ reads

$$C(\epsilon) \approx \begin{cases} 2\pi\epsilon\sqrt{|2\mathsf{K}_l t|}, & \text{for } x^* < a, \\ 2\pi\epsilon\sqrt{|2\mathsf{\Gamma}(t)|}, & \text{for } x^* > a. \end{cases} \tag{6.51}$$

This shows the normalisation constant depends only on the half-plane where the terminal position of the deterministic path is found.

6.4 Idealised example

6.4.1 Configuration

Consider the non-dimensionalised stochastic differential equation (6.1) with mean velocity

$$\mathbf{u}(\mathbf{x}) = \begin{cases} \mathbf{u}_l = (1/2, 1/5)^{\mathrm{T}}, & \text{for } x < 0, \\ \mathbf{u}_r = (1, -3/10)^{\mathrm{T}}, & \text{for } x > 0, \end{cases} \tag{6.52}$$

and

$$\mathsf{K}(\mathbf{x}) = \begin{cases} \mathsf{K}_l = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, & \text{for } x < 0, \\ \mathsf{K}_r = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}, & \text{for } x > 0. \end{cases} \tag{6.53}$$

A range of diffusivity strength ϵ over $\log_{10}\epsilon = -5, -4.75, \cdots 0$ is considered.

To generate particle trajectory data, for each ϵ, a total of 1024 particles are uniformly placed in the square domain $(x, y) \in [-1, 0] \times [-0.5, 0.5]$ with a regular spacing 10^{-5} in each direction. The Euler–Maruyama method with a small timestep 2^{-12} is applied to simulate the trajectories up to time $t = 4$.

With the prescribed velocity and diffusivity fields, particles are expected to drift from the left half-plane to the right over time. Following the deterministic path, the majority of particles are expected to make the transition at $t \approx 1$. Figure 6.1 shows the spatial distribution of the particles, obtained by binning the position records, at different times t and different values of ϵ.

A universal pattern for all ϵ is that more particles are found in the right half-plane as t increases, as expected by the advection. However, the detailed spatial distribution of particles depends heavily on the diffusivity strength ϵ. At $t = 4$, a large majority of particles fall into the right half-plane for $\epsilon = 10^{-2}$ and 10^{-4}, while a significant number of particles remain dispersed in the left half-plane for $\epsilon = 10^{0}$. The advection of particles is identical for all ϵ, and at time $t = 4$ all particles would be found in the right half-plane should they followed the deterministic path, yet this is not the case for strong diffusion.

To verify the formulas of the transition density (6.50) and normalisation constants (6.51), derived from the large deviation theory, we independently simulate trajectories of another 4096 particles, all deployed at $\mathbf{y} = (-0.5, 0)^{\mathrm{T}}$, using the Euler-Maruyama method with the same timestep 2^{-12}. In this configuration, ϵ is taken to be 10^{-2} and the particle will all cross the interface exactly at $t = 1$ if they follow the deterministic drift. The position of particles at different times $t = 0.25$, 0.875, 1.125 and 4 are binned

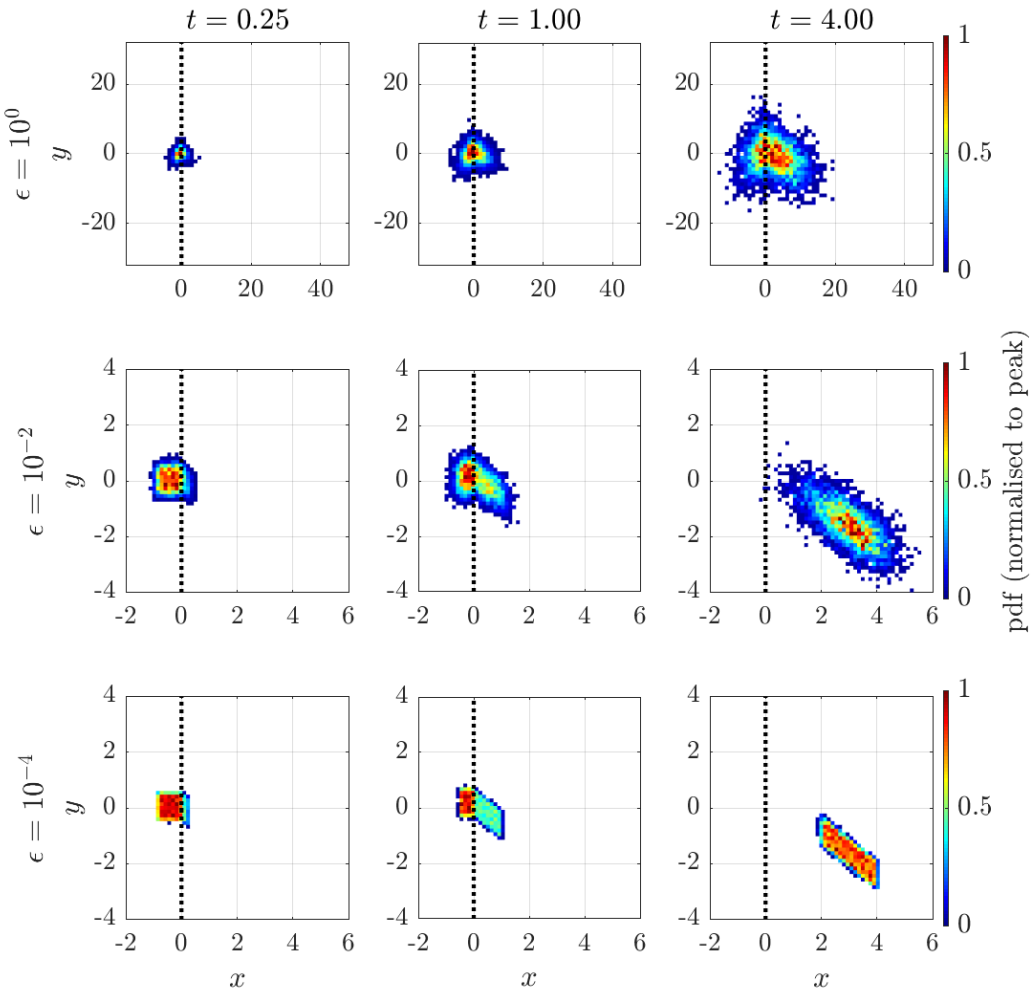

Figure 6.1: Probability density function estimated using 4096 particle samples at time $t = 0.25$, 1 and 4 (from left to right columns). The three rows correspond to three different values of diffusivity strength ϵ, from top to bottom respectively for 10^0, 10^{-2} and 10^{-4}, which are increasingly advection-dominated. The colour scales are normalised with respect to the corresponding peaks. The black dashed lines mark the interface of the half-planes.

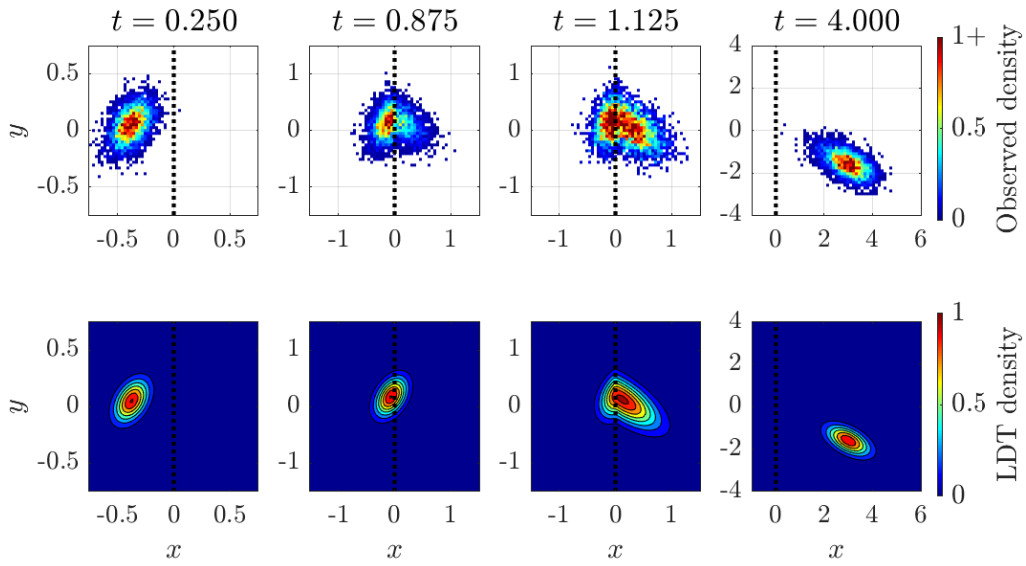

Figure 6.2: The instantaneous probability density functions at time $t = 0.25, 0.875, 1.125$ and 4 (from left to right columns) for particles started at $(-0.5, 0)$ and $\epsilon = 10^{-2}$. The top panels are the empirical probability density from binning the positions of 4096 particles. The bottom panels are the transition density derived from the large deviation principle (LDT) (6.50) equipped with the normalisation constant (6.51). The colour scale is normalised over the peak values of the LDT densities at each time. The vertical black dashed lines mark the interface of the half-planes.

to estimate a benchmark transition density profile, which is shown in the top row of figure 6.2.

The transition densities under the large deviation approach, shown in the bottom row of figure 6.2, are compared with the benchmark. For small and large times $t = 0.25$ and 4, when the deterministic path is far enough from the half-plane interface ($x = 0$), the majority of particles are found respectively on the left and right half-plane. The derived formula (6.50) agrees well with the benchmark in these cases. For the intermediate time $t = 0.875$ and 1.125 when the deterministic path is not too far from the half-plane interface, the derived formula (6.50) fails to capture the density profiles on the side opposite of the deterministic position. The mismatch between the formula (6.50) and the benchmark is expected, and are precisely the neglected cases in the analytic approximations discussed in section 6.3.3.

In this inference we are primarily interested in whether the maximum a posteriori (MAP) estimates can recover the exact fields where the particles are generated.

6.4.2 Parameterisation

We apply Bayesian inference to this problem, with the uniform velocity and diffusivity in each half-plane. The constant velocities in each half-plane are parameterised by the magnitude $U_\cdot$ and orientation $\Theta_\cdot$. The constant diffusivity tensors are parameterised by the eigenvalues $\Gamma_{1,\cdot}, \Gamma_{2,\cdot}$ and the orientations $\Phi_\cdot$ of the leading eigenvector

$$\mathbf{U}(\mathbf{x};\boldsymbol{\theta}) = \begin{cases} \mathbf{u}_l(\boldsymbol{\theta}) = U_l \begin{pmatrix} \cos\Theta_l \\ \sin\Theta_l \end{pmatrix}, & \text{for } x < 0, \\ \mathbf{u}_r(\boldsymbol{\theta}) = U_r \begin{pmatrix} \cos\Theta_r \\ \sin\Theta_r \end{pmatrix}, & \text{for } x > 0, \end{cases} \tag{6.54}$$

and

$$\mathsf{K}(\mathbf{x};\boldsymbol{\theta}) = \begin{cases} \mathsf{K}_l(\boldsymbol{\theta}) = \mathsf{R}(\Phi_l) \begin{pmatrix} \Gamma_{1,l} & 0 \\ 0 & \Gamma_{2,l} \end{pmatrix} \mathsf{R}(\Phi_l)^{\mathrm{T}}, & \text{for } x < 0, \\ \mathsf{K}_r(\boldsymbol{\theta}) = \mathsf{R}(\Phi_r) \begin{pmatrix} \Gamma_{1,r} & 0 \\ 0 & \Gamma_{2,r} \end{pmatrix} \mathsf{R}(\Phi_r)^{\mathrm{T}}, & \text{for } x > 0, \end{cases} \tag{6.55}$$

where

$$\mathsf{R}(\Phi) = \begin{pmatrix} \cos\Phi & -\sin\Phi \\ \sin\Phi & \cos\Phi \end{pmatrix}$$

is a rotation matrix identical to (3.7). Thus the parameters to infer are

$$\boldsymbol{\theta} = (U_l, \Theta_l, \Gamma_{1,l}, \Gamma_{2,l}, \Phi_l, U_r, \Theta_r, \Gamma_{1,r}, \Gamma_{2,r}, \Phi_r)^{\mathrm{T}}. \tag{6.56}$$

6.4.3 Posterior evaluation

Particle positions at different time t are used to independently infer the piecewise constant velocity and diffusivity fields for each of the ϵ with $\log_{10}\epsilon = -5, -4.75, \cdots 0$.

The likelihood function (2.20)

$$p(R|\boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{j=1}^{P-1} \pi\left(\mathbf{y}(t_{j+1}), t_{j+1} - t_j | \mathbf{y}(t_j)\right),$$

Parameter $\boldsymbol{\theta}$	Proposal s.d. $\sqrt{V_j}$	Prior range
U_l and U_r	1/100	[0, 100]
Φ_l and Φ_r	1	$[-100\pi, 100\pi]$
$\Gamma_{1,l}$ and $\Gamma_{1,r}$	ϵ	[0, 100]
$\Gamma_{2,l}$ and $\Gamma_{2,r}$	ϵ	[0, 100]
Φ_l and Φ_r	1	$[-100\pi, 100\pi]$

Table 6.1: Parameters used to for the Metropolis–Hastings Algorithm.

is defined by the transition density (6.50) with normalisation constant (6.51)

$$\pi(\mathbf{x}, t|\mathbf{y}) = \frac{1}{C(\epsilon)} \begin{cases} \exp\left(-\frac{1}{2}\|\mathbf{x}-\mathbf{y}-t\mathbf{u}_l\|^2_{2\epsilon\mathsf{K}_l t},\right), & \text{for } (x<a) \text{ or } (x>a \text{ and } x^*<a), \\ \exp\left(-\frac{1}{2}\|\mathbf{x}-\mathbf{x}^*\|^2_{2\epsilon\Gamma(t)}\right), & \text{for } (x>a \text{ and } x^*>a). \end{cases},$$

$$\text{where } C(\epsilon) = \begin{cases} 2\pi\epsilon\sqrt{|2\mathsf{K}_l t|}, & \text{for } x^*<a, \\ 2\pi\epsilon\sqrt{|2\Gamma(t)|}, & \text{for } x^*>a. \end{cases}$$

A uniform prior on a bounded range is imposed on all parameters, with the range chosen to be $[0, 100]$ for the magnitude of velocities and the eigenvalues of diffusivity tensors, and $[-100\pi, 100\pi]$ for the remaining angular parameters.

The random–walk Metropolis–Hastings algorithm (see section 2.2.2) is applied to sample the posterior distribution. The algorithm is started using the exact values under which the particle trajectories are simulated. This allows a handy validation of the inference, without worrying on the convergence of the Markov chains. In essence we are using the sampling algorithm to seek the MAP estimates. One set of $200,000$ samples are drawn using the sampler. The standard deviation in the proposal distribution is listed in Table 6.1.

6.4.4 Results

The MAP estimates of the velocity and diffusivity fields, approximated using the sample with (unnormalised) maximum posterior value, are compared against the exact fields. The relative error of the inferred velocities and diffusivity tensors from the exact ones are evaluated in terms of 2-norms. The left column of figure 6.3 shows the errors against ϵ at different time t (solid lines with marker 'x' in different colours). We note an improving capability to recover the exact fields when ϵ diminishes. Specifically for $\epsilon \lesssim 10^{-2}$, the relative discrepancy rarely exceeds 20%. The figure also shows that as the interval t increases so that more particles are found in the right half-plane, the estimations in the right half-plane gain accuracy.

To investigate the causes of the relative discrepancy from the MAP estimates to the exact values, we explore two parameters in the inference - data size and MCMC steps. The right column of figure 6.3 corresponds to the inference using 4-times as many particles, which shows qualitatively insignificant difference to results using fewer particles. Having doubled the MCMC steps, we again find the minimal difference between the results and those reported in 6.3.

As the large deviation approach is devised as an adaption to the localised inference in chapter 4, we compare the results with those from the localised one. The relative discrepancies of the velocities and diffusivities inferred using the localised approach are shown in the top and third row of figure 6.3. The relative error is smallest for the

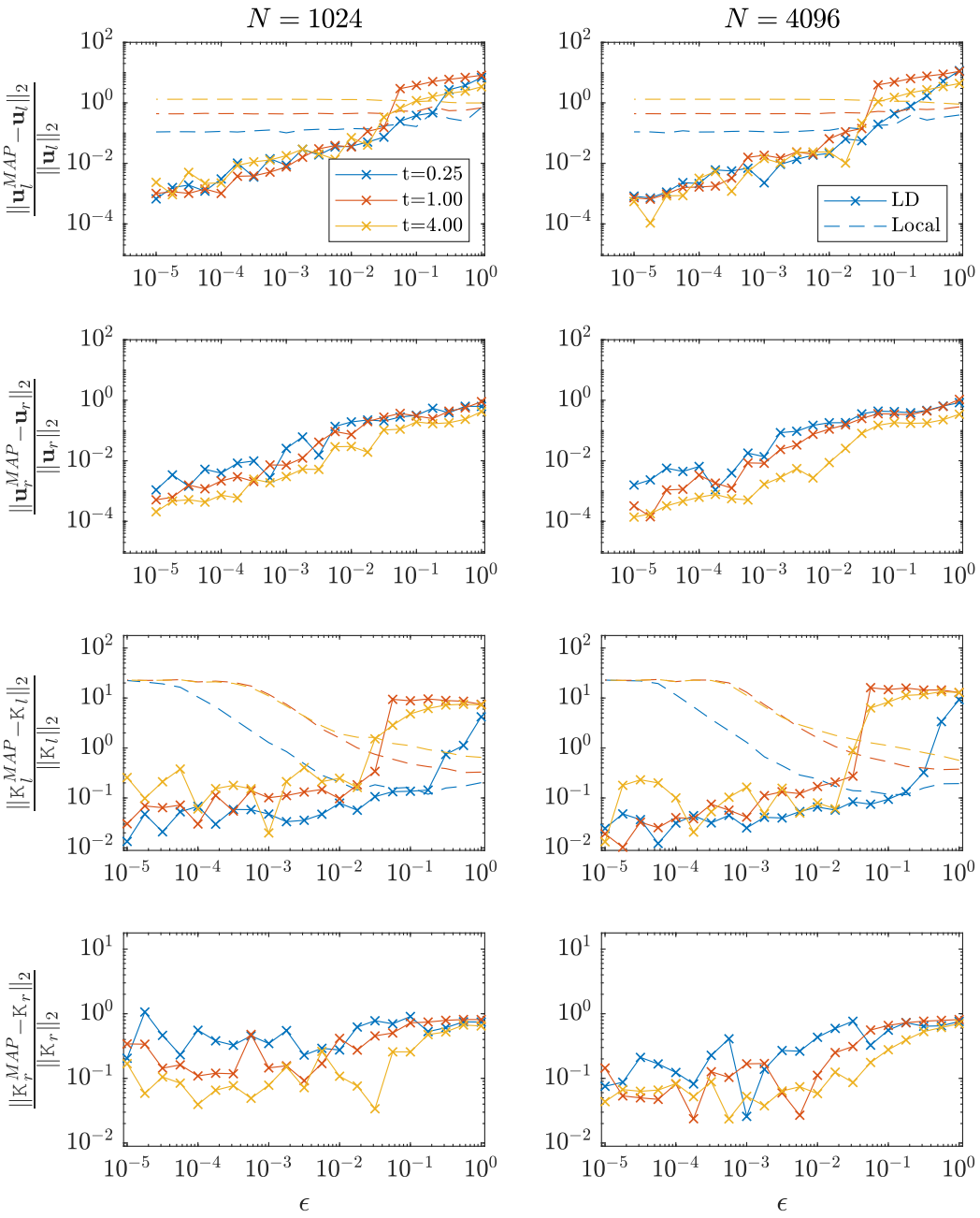

Figure 6.3: The relative discrepancy, measured in 2-norms, of the MAP estimates from the exact fields against the diffusivity strength ϵ at sampling time $t = 0.25, 1$ and 4 on a log-log scale. The top two rows correspond to the velocities in the left and right half-plane, while the bottom two rows correspond to the diffusivities. The left columns and right columns differ by the number of particles used, with 1024 on the left and 4096 on the right. The line colours represent the considered sampling time and are labelled in the legend of the top left panel. The broken lines on the first and third row corresponds to the MAP estimates of cellwise velocity and diffusivity under the local approach.

smallest considered time $t = 0.25$, in which the locality assumption is better satisfied. For large diffusivity strength $\epsilon \gtrsim 10^{-1}$, we note that the localised approach is superior to the large deviation approach. This can be explained by reviewing the top panel of figure 6.1 – a majority of particles are near the plane interface on both sides, relative to the dispersion length scale. This is precisely the neglected scenario, that particles concentrate on the same half-plane as the peak of transition density, in the analytic approximations in section 6.3.3 to simplify the transition density formula. As diffusivity strength ϵ falls in the range of 10^{-1} to $10^{-3/2}$, the localised approach and the large deviation approach provide comparable estimates for the case of $t = 0.25$ as they share the same transition density formula. As the diffusivity strength ϵ further reduces to $\lesssim 10^{-2}$, the relative error in the localised approach gradually becomes several orders of magnitude larger than that in large deviation approach for all considered time t.

6.5 Discussion and summary

In this chapter we applied the large deviation theory to derive analytic expressions for the transition density of a stochastic differential equation with weak diffusivity. Specifically the method is applied to a piecewise-constant velocity and diffusivity field defined on two half-planes.

Analytic arguments are supplied in section 6.3.3 to simplify the transition density formulas for particle displacements under different conditions. While there is an over-simplification in that not all scenarios are considered, we argue that in a practical inference it may well be sufficient to rely on the considered cases.

The inference is applied to an idealised example in section 6.4. The results show that, in the limit of small diffusivity, the large deviation approach is capable of tracking the spatial variation in the velocity and diffusivity. In the same limit the large deviation approach is also shown to outperform the localised approach when the majority of particles are transported across half-planes, violating the locality assumption.

An unsurprising limitation of the large deviation approach is that it requires the transition probability mass to concentrate on one-side of the half-planes to justify the simplifying assumptions. We did not pursue a detailed error analysis but we emphasised this is viable. The simplification assumptions can also be lifted by solving the constrained optimisation problems (6.32) and (6.38).

It is worth examining if the approach is applicable to infer ocean diffusivity in the advection-dominated area. However, a considerable amount of adaption is needed to generalise from a domain of two half-planes to a domain of cells with multiple faces. It is noted in the quasigeostrophic double gyre considered in previous chapters that on the eastward jet where the advection is the strongest, the particles are rarely adhered to the streamlines of the Eulerian mean flow, but are driven by eddies to straddle multiple cells at $\mathcal{O}(10 \text{ days})$ sampling intervals. This raises concerns on the applicability of the large deviation approach.

Chapter 7

Conclusions and future work

7.1 Conclusions

This thesis developed a novel Bayesian approach to diagnose eddy diffusivities from Lagrangian trajectory data. In Chapter 1 we introduced the concept of eddy diffusivity resulting from coarse-graining the turbulent dynamics from both Lagrangian and Eulerian perspectives. We also reviewed existing oceanographic diffusivity diagnostics and outlined their strengths and limitations.

Chapter 2 continued with an introduction to the Bayesian framework and the computational tools for practical inference. The central outcome of the Bayesian paradigm is the posterior distribution, which combines prior information and data to estimate and quantify uncertainty in a parameter of interest. In this thesis, we modelled the Lagrangian dynamics in a turbulent fluid by a stochastic differential equation. This leads to a posterior distribution, built by solutions of Fokker–Planck equations, on the parameters of the mean velocity and eddy diffusivity fields using discretely observed particle trajectory data. We emphasise here that the posterior distribution is a probability density for the parameters, assuming the parameterised stochastic differential equation can perfect generate the trajectory data. The spread of the posterior cannot account for errors in the underlying model.

In Chapter 3 we applied Bayesian inference to a multiscale cellular flow and verified that a sufficiently long sampling interval is essential to infer the eddy diffusivity. Since in practice the sufficient sampling interval is rarely known in advance, the inference should be repeated at different downsampling rates, and its convergence property must be validated to conclude on an eddy diffusivity estimate.

We showed in Chapter 4 the first attempt to apply the Bayesian inference to a more complex and oceanographically-relevant example. With a modest amount of particle trajectory data in a three-layer quasigeostrophic double gyre system, the mean velocity and diffusivity field are inferred using a localised approach. This approach divides the domain into cells and infers locally-defined fields independently. By restricting to locally-defined linear stochastic differential equations, this approach allows the inference to be carried out using a likelihood with explicit formulas at an affordable computational cost. In the middle layer the inference is capable of constraining the eddy diffusivity in most the domain within a factor of about 2, using 676 trajectories over a 10-year period.

The localised likelihood function assumes that particles observed from a given cell experience the same linear velocity and constant diffusivity field over the entire sampling interval. This however places a strain on the principle of sampling at long intervals,

which is the necessary condition for the particle motion to be well modelled by the SDE with an eddy diffusivity. It is therefore essential to overcome the locality limitation in the inference.

We proceeded in two directions, one is an adaptation to the localised approach and the other is a novel global approach. In Chapter 5, the latter direction is taken. In principle it is possible to overcome the locality limitation by inferring simultaneously all degrees of freedom in globally-defined mean velocity and diffusivity fields, and account for the particle displacements with the field consistently in the likelihood. This brute-force approach turns out to be computationally infeasible, as a tremendous number of Fokker–Planck equations needs to be solved.

To overcome the computational challenge we devised a data coarse-grained method and reformulate the inference problem. Modelling the precise starting positions of the transition data point by a probability distribution, we arrived at with a coarse-grained likelihood whose computational cost is independent of the data size. We further showed that, in the limit of large data, the data coarse-grained inference shares the same maximum a posteriori (MAP) estimates as primitive inference, despite an additional uncertainty.

The data coarse-grained global inference is applied to the quasigeostrophic double gyre system. We showed that both the inferred velocity and diffusivity become less sensitive to sampling interval than in the localised inference. We also validated the inference by comparing the tracer distributions predicted by the inferred fields against empirical tracer profiles, and note a generally good agreement.

In Chapter 6 we explored an adaptation to the localised approach, by accounting for particles displaced across two cells. This keeps the computational cost for likelihood evaluation affordable. We considered a simplified setting where piecewise constant velocity and diffusivity fields are defined on two half-planes. Assuming that the diffusivity is weak relative to velocity, the large deviation theory is applied to derive an analytical expression of the likelihood. The inference based on this approach improves on the localised approach, but only when advection is sufficiently strong relative to the diffusion.

7.2 Future work

This thesis opens up avenues for further research.

The most pressing issue is to apply the inference, especially the global inference, to real drifters in the ocean. We look forward to providing answers to a few key questions, 1) how the mean velocity and diffusivity vary in the global ocean, 2) how many drifters are needed to confidently estimate the diffusivity, 3) how the estimates from Bayesian inference compare with existing diagnostics and 4) whether the diagnosed diffusivity is capable of representing the unresolved turbulent mixing in a coarse-grained model. These questions have significant implications for climate model parameterisations and deserve a definite answer.

We can however foresee several bottlenecks when applying the inference to real data. The first issue is the highly uneven geographical distribution of drifters in the ocean. It is not only about a wider spread of the posterior at locations with few drifter data, but the more concerning issue that the clustering of drifters, especially the surface ones, are spatially correlated. We expect to the inference, which assumes independence in the data, may over-fit with data and provide a sharp estimate to a nonphysical flow at locations with clustered data. The clustering of drifters also indicates that the mean

velocity of drifters may be compressible. Introducing new degrees of freedom to account for the compressibility unavoidably increases the already high computational cost.

Another challenge is on the choice of boundary conditions. In the quasigeostrophic simulation a square domain with zero-flux boundary condition is considered. In the open ocean, boundary conditions must be supplied to the Fokker–Planck equation to account for the net in and outflow of particles in the transition density. An optimal strategy to tune the boundary condition is currently out of sight but a pressing answer is needed to apply the inference in real data.

In the global inference, we note that the data coarse-grained method may struggle in areas with particles rarely transported cross-cells. A handy fix to this is to restore to a localised or a large derivation approach in such cells, and enable the individual trajectory information to estimate a weak velocity or diffusivity. This hybrid approach may be the most favourable in inference for the global ocean, where flow properties could vary by orders of magnitude in different geographical locations.

To perform the hybrid inference, the large deviation based inference should also be generalised. The first adaptation needed is a coherent likelihood formula for particles transported over cells in all directions. The second amendment required is to account for fields beyond the piecewise constant ones, which are compatible with ones used in the global and localised inference.

Beyond the scope of this thesis is the assessment of the suitability of the chosen stochastic model for Lagrangian trajectories. The Bayesian inference can in principle be extended for this task. This can be conducted by incorporating a list of model classes and defining a posterior over the space of models. The relative fitness of each model to the data can be identified through the posterior analysis.

Another direction for further work is to upgrade the sampling technology. In this thesis we only apply the Metropolis–Hasting algorithm, which produces highly correlated samples in successive steps. This means a large number of samples, and hence also long computation time, is needed for a convergence of sample distribution to the target posterior. It is worthwhile to examine if any speed-up is achieved by switching to the sophisticated sampling algorithms discussed in section 2.2.5.

Appendix A

Velocity autocorrelation calculations

This chapter is modified from the published article Ying et al. (2019).

A.1 Calculating the Davis diffusivity

The Davis (1987) diffusivities (1.32) are calculated with respect to the 10-year Eulerian mean flow. We use particle locations observed every 3 hours over 10 years. We adopt the method of Griesel et al. (2010) to evaluate the two diffusivities in each of the 16×16 cells partitioning the domain. The position of each particle every 3 hours is treated as a new independent starting point, to generate a set of particle trajectories each with time lag τ. The conditional averaging operator $\langle \cdot \rangle_{\{\mathbf{X}_i(t)=\mathbf{x}\}}$ in (1.32) is then modified to include all particle trajectories that end in a given cell, and the time integral is computed using the trapezoidal rule.

The along-stream and cross-stream Davis (1987) diffusivities are calculated using the 10-year Eulerian mean flow at the centre of each cell to define the mean velocity $\bar{\mathbf{u}}(\mathbf{x})$ appearing in equation (1.32). Note that the (1.32) diffusivity tensor need not be symmetric positive definite and hence corresponding diffusivity ellipses cannot be shown without further processing.

In the model validation experiments in section 5.5, the estimated diffusivity tensor (1.32) is post-processed to become symmetric positive definite for usage in the numerical partial differential equation solver. This is conducted by symmetrising the tensor $(\mathsf{K}_{\text{Davis}} + \mathsf{K}_{\text{Davis}}{}^{\text{T}})/2$ which is guaranteed diagonalisable. The symmetrised tensor is eigendecomposed. If any eigenvalue is negative, its sign is flipped in the diagonal matrix and composed with the original eigenvector to form a symmetric positive definite tensor.

A.2 Estimating an decorrelation timescale

Decorrelation scales of Lagrangian trajectories can be defined using the eddy velocity autocorrelation function

$$R(\mathbf{x};\tau) = \left\langle \left[\dot{\mathbf{X}}_i(t) - \overline{\mathbf{u}}(\mathbf{x})\right] \cdot \left[\dot{\mathbf{X}}_i(t+\tau) - \overline{\mathbf{u}}(\mathbf{X}_i(t+\tau))\right] \right\rangle_{\{\mathbf{X}_i(t)=\mathbf{x}\}}, \tag{A.1}$$

where $\overline{\mathbf{u}}(\mathbf{x})$ is the time-averaged Eulerian mean flow, τ is the time lag and $\langle \cdot \rangle_{\{\mathbf{X}_i(t)=\mathbf{x}\}}$ is the conditional average over particle trajectories that leave position $\mathbf{x}$ at time t.

Garraffo et al. (2001) (also appendix of Lumpkin et al., 2002) propose the empirical form

$$R(\mathbf{x};\tau) = R(\mathbf{x};0)\cos\left(\frac{\pi\tau}{2T_z}\right)e^{-(\tau/T_e)^2}, \tag{A.2}$$

where T_z is the first zero-crossing time and T_e is an e-folding scale, which we interpret as a decorrelation time – note that the envelope decays more rapidly than exponential. The implementation requires local binning of $\mathbf{X}_i(t)$ and results in an eddy velocity autocorrelation function in each cell. We estimate both parameters T_z and T_e through a least-squares fitting.

Appendix B

Linear stochastic differential equations

This chapter is modified from the published article Ying et al. (2019).

B.1 Linear velocity parameterisation

A divergence-free linear velocity field $\mathbf{U}(\mathbf{x})$ in Cartesian coordinates can be expressed as

$$\mathbf{U}(\mathbf{x}) = \mathsf{A}(\mathbf{x} - \mathbf{x}_0) + \mathbf{U}_0 \tag{B.1}$$

where $\mathsf{A} = \begin{pmatrix} A_{xx} & A_{xy} \\ A_{yx} & -A_{xx} \end{pmatrix}$ is a constant trace-free velocity-gradient tensor and $\mathbf{U_0}$ is a constant vector.

The velocity gradient tensor A can be re-expressed in terms of rotationally invariant quantities Υ_1 and Υ_2 via

$$\begin{aligned} \mathsf{A} &= \mathsf{R}(\Phi_\mathsf{A}) \begin{pmatrix} 0 & \Upsilon_2 + \Upsilon_1 \\ \Upsilon_2 - \Upsilon_1 & 0 \end{pmatrix} \mathsf{R}(\Phi_\mathsf{A})^\mathrm{T} \\ &= \Upsilon_1 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \Upsilon_2 \begin{pmatrix} -\sin 2\Phi_A & \cos 2\Phi_A \\ \cos 2\Phi_A & \sin 2\Phi_A \end{pmatrix}, \end{aligned}$$

with rotation matrix

$$\mathsf{R}(\Phi_\mathsf{A}) = \begin{pmatrix} \cos\Phi_\mathsf{A} & -\sin\Phi_\mathsf{A} \\ \sin\Phi_\mathsf{A} & \cos\Phi_\mathsf{A} \end{pmatrix}.$$

Υ_1 sets the magnitude of the anti-symmetric part of the velocity gradient tensor and is related to the vorticity via

$$\Upsilon_1 = -\frac{1}{2} \begin{pmatrix} -\partial_y \\ \partial_x \end{pmatrix} \cdot \mathbf{U}(\mathbf{x}) = -\frac{1}{2}\omega.$$

Υ_2 sets the magnitude of the symmetric part of the velocity gradient tensor. If the angle Φ_A is chosen such that Υ_2 is non-negative,

$$\Upsilon_2 = \frac{1}{\sqrt{2}} \left\| \dot{\varepsilon} \right\|_F,$$

where $\dot{\varepsilon} = \left(\nabla\mathbf{U} + (\nabla\mathbf{U})^\mathrm{T}\right)/2$ is the strain rate tensor and $\|\cdot\|_F$ denotes the Frobenius norm.

B.2 Constant diffusivity parameterisation

A constant symmetric diffusivity tensor K in Cartesian coordinates can be expressed as

$$\mathsf{K} = \begin{pmatrix} K_{xx} & K_{xy} \\ K_{xy} & K_{yy} \end{pmatrix}.$$

The diffusivity tensor K can be re-expressed in terms of rotationally invariant eigenvalues Γ_1 and Γ_2 via

$$\begin{aligned} \mathsf{K} &= \mathsf{R}\left(\Phi_{\mathsf{K}}\right) \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix} \mathsf{R}\left(\Phi_{\mathsf{K}}\right)^{\mathrm{T}} \\ &= \frac{\Gamma_1 + \Gamma_2}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{\Gamma_1 - \Gamma_2}{2} \begin{pmatrix} \cos 2\Phi_{\mathsf{K}} & \sin 2\Phi_{\mathsf{K}} \\ \sin 2\Phi_{\mathsf{K}} & -\cos 2\Phi_{\mathsf{K}} \end{pmatrix}, \end{aligned}$$

with rotation matrix

$$\mathsf{R}\left(\Phi_{\mathsf{K}}\right) = \begin{pmatrix} \cos\Phi_{\mathsf{K}} & -\sin\Phi_{\mathsf{K}} \\ \sin\Phi_{\mathsf{K}} & \cos\Phi_{\mathsf{K}} \end{pmatrix}.$$

The level set satisfying the quadratic equation $\mathbf{x}^{\mathrm{T}}\mathsf{K}^{-1}\mathbf{x} = 1$ is the 'diffusivity ellipse' shown (normalised) in figure 4.5. Geometrically, $\sqrt{\Gamma_1}, \sqrt{\Gamma_2}$ are the two radii of this ellipse and ϕ_{K} is the angle the semi-major axis makes with the x-axis.

B.3 Linear stochastic differential equations

The transition density $\pi(\mathbf{x}, t|\mathbf{y})$ associated with the stochastic differential equation

$$\mathrm{d}\mathbf{X}(t) = (\mathsf{A}\mathbf{X}(t) + \mathbf{b})\mathrm{d}t + \sqrt{2\mathsf{K}}\mathrm{d}\mathbf{W}, \quad \mathbf{X}(0) = \mathbf{y}, \tag{B.2}$$

is a Gaussian, because of the linearity of (B.2), of the form

$$\pi(\mathbf{x}, t|\mathbf{y}) = \frac{1}{2\pi\sqrt{\det\boldsymbol{\Sigma}(t)}} \exp\left(-\frac{1}{2}\left(\mathbf{x} - \mathbf{m}(t)\right)^{\mathrm{T}} \boldsymbol{\Sigma}(t)^{-1} \left(\mathbf{x} - \mathbf{m}(t)\right)\right), \tag{B.3}$$

with mean and covariance

$$\mathbf{m}(t) = \mathbb{E}\mathbf{X}(t) \quad \text{and} \quad \boldsymbol{\Sigma}(t) = \mathbb{E}(\mathbf{X}(t) - \mathbf{m}(t)) \otimes (\mathbf{X}(t) - \mathbf{m}(t)), \tag{B.4}$$

where $\mathbb{E}$ denotes expectation over the Brownian motion $\mathbf{W}$. Differentiating (B.4) with respect to time and using (B.2) yields

$$\frac{\mathrm{d}\mathbf{m}(t)}{\mathrm{d}t} = \mathsf{A}\mathbf{m}(t) + \mathbf{b}, \tag{B.5a}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \mathsf{A}\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\mathsf{A}^{\mathrm{T}} + 2\mathsf{K}, \tag{B.5b}$$

using Ito's formula (e.g. Pavliotis, 2014). Solving and taking the initial conditions $\mathbf{m}(0) = \mathbf{y}$ and $\boldsymbol{\Sigma}(0) = \mathsf{0}$ into account gives

$$\mathbf{m}(t) = e^{\mathsf{A}t}\mathbf{y} + \int_0^t e^{\mathsf{A}(t-s)}\mathbf{b}\,ds, \tag{B.6a}$$

$$\boldsymbol{\Sigma}(t) = 2\int_0^t e^{\mathsf{A}(t-s)}\mathsf{K}e^{\mathsf{A}^{\mathrm{T}}(t-s)}\,ds. \tag{B.6b}$$

Equation (4.4) is recovered by letting $\mathbf{b} = \mathbf{U}_0 - \mathsf{A}\mathbf{x}_0$ in (B.6).

Appendix C

Numerical methods for partial differential equations

C.1 Advection and diffusion operator

Consider the advection–diffusion equation with an incompressible velocity field $\mathbf{U}$ and diffusivity tensor field K with no flux boundary condition

$$\frac{\partial q}{\partial t} + \nabla \cdot (\mathbf{U}q) = \nabla \cdot (\mathsf{K}\nabla q)\,, \tag{C.1a}$$

$$(\mathbf{U}q + \mathsf{K}\nabla q) \cdot \hat{n} = 0, \quad \text{on boundary } \partial\Omega, \tag{C.1b}$$

where $\partial\Omega$ is the boundary of the domain and $\hat{n}$ is its outward normal. In this thesis we only consider a velocity field where $\mathbf{U} \cdot \hat{n} = 0$ with zero net flow across the boundary.

By considering the advection operator $\mathcal{A}q = -\nabla \cdot (\mathbf{U}q)$ and the diffusion operator $\mathcal{D}q = \nabla \cdot (\mathsf{K}\nabla q)$, the advection–diffusion equation and its solution can be written in

$$\frac{\partial q}{\partial t} = (\mathcal{A} + \mathcal{D})q,$$
$$Q(t) = \mathrm{e}^{(\mathcal{A}+\mathcal{D})t}Q(0),$$

where $Q(t)$ is a shorthand for the solution $q(x, y, t)$

We apply a cell-based finite volume method to discretise the equation on a unit square domain. We also apply Strang splitting scheme for the advection and diffusion operators is applied to march the solution in time, that is, over each time step

$$Q(t + \Delta t) = \mathrm{e}^{\mathcal{D}\Delta t/2}\mathrm{e}^{\mathcal{A}\Delta t}\mathrm{e}^{\mathcal{D}\Delta t/2}Q(t),$$

which is known to be second-order accurate in time (if the approximation for $\mathcal{A}$ and D are at least second order). We verify this using the method of manufactured solutions on the implemented solver.

C.1.1 Advective step – $\mathrm{e}^{\mathcal{A}\Delta t}$

We apply the *corner-transport upwind* presented in section 20.2 LeVeque (2002) to discretise the advection operator.

Denote Δx and Δy the spacing of the Arakawa C-grid. The integer indices (i, j) corresponds to quantities defined at cell centres and the half indices $(i + 1/2, j + 1/2)$ refers to quantities defined on edges. Applying the divergence theorem over the cell

C_{ij} centred at (x_i, y_j) and dividing the equation by the area $\Delta x \Delta y$ to the integral form of advection equation $\partial_t q = \mathcal{A}q$ leads to a fully discrete updating scheme for $Q_{ij}^n = \iint_{C_{ij}} q(x, y, t_n)\mathrm{d}x\mathrm{d}y/(\Delta x \Delta y)$

$$Q_{ij}^{n+1} = Q_{ij}^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2,j}^n - F_{i-1/2,j}^n\right) - \frac{\Delta t}{\Delta y}\left(G_{i,j+1/2}^n - G_{i,j-1/2}^n\right), \tag{C.2}$$

where $F_{i-1/2,j}^n$ and $G_{i,j-1/2}^n$ are the time-averaged fluxes across the left and bottom interface of cell C_{ij}

$$F_{i-1/2,j}^n = \frac{1}{\Delta t \Delta y}\int_{t_n}^{t_{n+1}}\int_{y_{j-1/2}}^{y_{j+1/2}} u(x_{i-1/2}, y)q(x_{i-1/2}, y, t)\mathrm{d}y\mathrm{d}t, \tag{C.3a}$$

$$G_{i,j-1/2}^n = \frac{1}{\Delta t \Delta x}\int_{t_n}^{t_{n+1}}\int_{x_{i-1/2}}^{x_{i+1/2}} v(x, y_{j-1/2})q(x, y_{j-1/2}, t)\mathrm{d}x\mathrm{d}t. \tag{C.3b}$$

The flux $F_{i-1/2,j}$ in the x-direction is numerically approximated by a hybrid of a one-dimensional first-order upwind flux $\bar{F}_{i-1/2,j}$, a second-order correction term $F'_{i-1/2,j}$ (with the superscript time index n dropped for clarity, see also equation (6.32) of LeVeque (2002)) and a cross-derivative correction $\tilde{F}_{i-1/2,j}$

$$\bar{F}_{i-1/2,j} = u^-_{i-1/2,j}Q_{ij} + u^+_{i-1/2,j}Q_{i-1,j}, \tag{C.4a}$$

$$F'_{i-1/2,j} = \frac{1}{2}|u_{i-1/2,j}|\left(1 - \left|\frac{u_{i-1/2,j}\Delta t}{\Delta x}\right|\right)(Q_{ij} - Q_{i-1,j}), \tag{C.4b}$$

$$\begin{aligned}\tilde{F}_{i-1/2,j} = \frac{1}{2}\frac{\Delta t}{\Delta y}\big(&u^-_{i-1/2,j}v^-_{i,j-1/2}(Q_{ij} - Q_{i,j-1}) + u^+_{i-1/2,j}v^-_{i-1,j-1/2}(Q_{i-1,j} - Q_{i-1,j-1})\\ &+u^-_{i-1/2,j-1}v^+_{i,j-3/2}(Q_{i,j-1} - Q_{i,j-2}) + u^+_{i-1/2,j-1}v^+_{i-1,j-3/2}(Q_{i-1,j-1} - Q_{i-1,j-2})\big),\end{aligned} \tag{C.4c}$$

via a limiter $\phi(\cdot)$

$$F_{i-1/2,j} = \bar{F}_{i-1/2,j} + \phi(\theta_{i-1/2,j})F'_{i-1/2,j} + \tilde{F}_{i-1/2,j}, \tag{C.5}$$

here $u^\pm_{i-1/2,j} = \pm\max\left(0, \pm u_{i-1/2,j}\right)$ is the velocity from the upwind side and $\theta_{i-1/2,j} = (Q_{Ij} - Q_{I-1,j})/(Q_{ij} - Q_{i-1,j})$ is a smoothness measure in the upwind side $I = i \mp \mathrm{sgn}\left(u_{i-1/2,j}\right)$. The limiter enables weighting between a higher-order scheme for a smooth Q and a first-order yet stable upwind scheme for a non-smooth Q so as to eliminate numerical oscillation without hampering accuracy. A common limiter choice is the *monotonized central–difference (MC)* $\phi(\theta) = \max\left(0, \min\left((1+\theta)/2, 2, 2\theta\right)\right)$ (Van Leer, 1977). Note also that the second-order Lax–Wendroff method can be recovered by setting $\phi(\theta) = 1$ (Lax and Wendroff, 1960).

The flux $G_{i,j-1/2}$ in the y-direction is numerically approximated in a similar fashion. We refer readers to section 20.2 to 20.8 of LeVeque (2002) for the details of the schemes.

C.1.2 Diffusion step – $\mathrm{e}^{\mathcal{D}\Delta t}$

Denote Δx and Δy the spacing of the Arakawa C-grid. A finite volume discretisation results in

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_{ij} = -\frac{1}{\Delta x}\left(\mathcal{F}_{i+1/2,j} - \mathcal{F}_{i-1/2,j}\right) - \frac{1}{\Delta y}\left(\mathcal{G}_{i,j+1/2} - \mathcal{G}_{i,j-1/2}\right), \tag{C.6}$$

where

$$\mathcal{F}_{i-1/2,j} = \frac{1}{\Delta y}\int_{y_j}^{y_{j+1}} \hat{x}^{\mathrm{T}}\mathsf{K}(x_{i-1/2}, y)\nabla q(x_{i-1/2}, y, t)\mathrm{d}y, \tag{C.7a}$$

$$\mathcal{G}_{i,j-1/2} = \frac{1}{\Delta x}\int_{x_i}^{x_{i+1}} \hat{y}^{\mathrm{T}}\mathsf{K}(x, y_{j-1/2})\nabla q(x, y_{j-1/2}, t)\mathrm{d}x, \tag{C.7b}$$

are the fluxes across cell C_{ij} at its left and bottom edge, with $\hat{x}$ and $\hat{y}$ being the unit vector in x and y-direction. The instantaneous gradient $\nabla q(x_{i-1/2}, y_j, \cdot)$ and $\nabla q(x_i, y_{j-1/2}, \cdot)$ are approximated by a second-order accurate central difference of discrete Q_{ij}, which leads to the numerical flux $\mathcal{F}_{i-1/2,j}$ and $\mathcal{G}_{i,j-1/2}$

$$\begin{aligned}
\mathcal{F}_{i-1/2,j} =& K_{11}(x_{i-1/2}, y_j)\left(\frac{Q_{ij} - Q_{i-1,j}}{\Delta x}\right) \\
&+ K_{12}(x_{i-1/2}, y_j)\left(\frac{Q_{i,j+1} - Q_{i,j-1} + Q_{i-1,j+1} - Q_{i-1,j-1}}{4\Delta y}\right) \\
\mathcal{G}_{i,j-1/2} =& K_{21}(x_i, y_{j-1/2})\left(\frac{Q_{i+1,j} - Q_{i-1,j} + Q_{i+1,j-1} - Q_{i-1,j-1}}{4\Delta x}\right) \\
&+ K_{22}(x_i, y_{j-1/2})\left(\frac{Q_{ij} - Q_{i,j-1}}{\Delta y}\right)
\end{aligned}$$

for $\mathsf{K} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$. The resulting system of linear ordinary differential equation (C.6) is solved by the time-stepping via the second-order accurate Heun's method (also known as explicit trapezoidal method, see Ascher and Petzold, 1998).

C.2 Quasigeostrophic calculations

The three-layer quasigeostrophic equations (1.34) are discretised using finite differencing, with a mesh with 513×513 nodes uniformly spaced on a square grid, in a 3840 km $\times$ 3840 km square horizontal domain. The resolution allows mesoscale eddies to be resolved. The advection term in the quasigeostrophic potential vorticity equation is discretised using the Arakawa (1966) Jacobian, and Laplace operators are discretised using second order centered differencing. The prognostic equation is marched in time using a third-order Adams–Bashforth scheme with uniform timestep Δt_{QG}, started with a second order Runge–Kutta step followed by a second order Adams–Bashforth step.

Over each time step of prognostic equation (1.34), the instantaneous streamfunction and relative vorticity is inverted from the instantaneous potential vorticity. The elliptic diagnostic problem (1.34) for potential vorticity inversion is solved via projection onto discrete baroclinic modes, and the resulting Poisson or modified Helmholtz problems are solved using a Fast Poisson Solver (e.g. Strang, 1986, section 5.5), with the decoupled tri-diagonal systems arrived at using a Discrete Sine Transform using FFTPACK 5.1.

Appendix D

Numerical solvers for particle trajectories

D.1 Two-dimensional exact solver

D.1.1 Formulations

Motions of Lagrangian particles are simulated on the fly when running the idealised quasigeostrophic model. At every time step t_n, the quasigeostrophic model returns a layer-wise two-dimensional discrete streamfunction defined on the vertices of the 513×513 grid denoted ψ_{ij}^n with spatial indices i and j and temporal index n.

A continuous streamfunction $\psi(\mathbf{x}, t)$ is constructed by linear interpolation both spatially and temporally. To perform the spatial interpolation, each square cell on the grid is meshed into four isosceles triangular elements along the diagonals ('criss-cross' mesh). The value of streamfunction at the centre of the square is assigned by bilinearly interpolating from the corner values. This enables a nodal linear interpolation within each triangular element. To perform the temporal interpolation, the spatially interpolated streamfunction $\psi(\mathbf{x}, t_n)$ at successive time steps are temporally linear interpolated.

An element-wise velocity field is defined by the taking (weak) skew-gradient of streamfunction $\mathbf{u}(\mathbf{x}, t) := \nabla^\perp \psi(\mathbf{x}, t)$, where $\nabla^\perp := (-\partial_y, \partial_x)^\mathrm{T}$. At any given time, the velocity is piecewise constant across elements, equipped with an additional constraint of continuous normal components across element edges.

The velocities at the vertices of the elements have no (strong) pointwise definition. However as the vertices constitute only a zero-measure subset of the domain, the particles trajectories governed by the ordinary differential equation

$$\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} = \mathbf{u}(\mathbf{X}(t), t), \tag{D.1}$$

can still be defined almost everywhere. In the time-independent case, as long as the initial position does not lie on a vertex, the trajectory remains well-posed over time.

Suppose a particle $\mathbf{X}(s)$ lies within an element e_k at time $s \in [t_n, t_{n+1})$. Denote $\mathbf{u}_k^n$ and $\mathbf{u}_k^{n+1}$ the velocity in element at time t_n and t_{n+1}. So long as the particle remains in the element, its position $\mathbf{X}(t)$ is governed by

$$\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} = \mathbf{u}_k(s) + \frac{t-s}{t_{n+1}-s}\left(\mathbf{u}_k^{n+1} - \mathbf{u}_k(s)\right) \tag{D.2}$$

where $\mathbf{u}_k(s) := \mathbf{u}_k^n + \left(\mathbf{u}_k^{n+1} - \mathbf{u}_k^n\right)(s - t_n)/(t_{n+1} - t_n)$ is the velocity in element e_k at

time s. Solving the differential equation gives

$$\mathbf{X}(t) = \mathbf{X}(s) + (t-s)\mathbf{u}_k(s) + \frac{(t-s)^2}{2(t_{n+1}-s)}\left(\mathbf{u}_k^{n+1} - \mathbf{u}_k(s)\right), \tag{D.3}$$

cautioning that its validity is up to a time $t \leq \min(t_{n+1}, s+\tau)$, where τ is the collision time for a particle to hit any edge of the element. The computation of the full trajectory can hence be divided into element-wise sub-problems of *collision detection* - identifying the time and location of each collision - and advecting particle positions using (D.3) until the designated time t is reached.

D.1.2 Collision detection

To determine the collision time τ in an element, it suffices to compare the time needed for a particle to hit each of the edges. For an edge with endpoints coordinates $\mathbf{N}_0$ and $\mathbf{N}_1$ (ordered anti-clockwise), its unit normal $\hat{n}$ (pointing outward from the element) is given by rotating and normalising the vector $\mathbf{l}_{01} := \mathbf{N}_1 - \mathbf{N}_0$ by $\pi/2$ clockwise, that is,

$$\hat{n} := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \frac{\mathbf{l}_{01}}{\|\mathbf{l}_{01}\|}.$$

The distance $S(t)$ between the particle $\mathbf{X}(t)$ and the edge is hence given by projecting $\mathbf{X}(t) - \mathbf{N}_0$ along the normal vector $\hat{n}$, which gives

$$\begin{aligned} S(t) :=& (\mathbf{X}(t) - \mathbf{N}_0)\cdot\hat{n}, \\ =& S(s) + (t-s)\mathbf{u}_k(s)\cdot\hat{n} + \frac{(t-s)^2}{2(t_{n+1}-s)}\left(\mathbf{u}_k^{n+1} - \mathbf{u}_k(s)\right)\cdot\hat{n}, \end{aligned}$$

with the second line followed from (D.3). The condition for the collision time $S(s+\tau) = 0$ leads to a quadratic equation

$$0 = \left(\frac{\left(\mathbf{u}_k^{n+1} - \mathbf{u}_k(s)\right)\cdot\hat{n}}{2(t_{n+1}-s)}\right)\tau^2 + \left(\mathbf{u}_k(s)\cdot\hat{n}\right)\tau + S(s), \tag{D.4}$$

Only the smaller positive root of the quadratic equation can be a candidate of collision time. The collision time with three edges are evaluated, and the unique minimal one would correspond to the particle advection.

D.2 Properties

The particle solver forms the full trajectory by bridging pieces of exact particle trajectories within each element. This preserves the exactness of the solution to equation (D.1) and results in several desired properties.

1. *Symplecticity.* The time-evolution of the two coordinates $\mathbf{X}(t) = (x(t), y(t))^{\mathrm{T}}$ is exactly governed by a set equations

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\partial\psi}{\partial y}, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = +\frac{\partial\psi}{\partial x},$$

which resembles the Hamiltonian equation with an area-preserving property.

2. *Quantity conservation.* Given a flow prescribed by a streamfunction $\psi(\mathbf{x}, t)$, a Lagrangian trajectory follows the streamlines and preserve the quantity $\psi(\mathbf{X}(t), t) = \psi(\mathbf{X}(0), 0)$.

3. *Reversibility.* A Hamiltonian system is time-reversible and so is this numerical scheme.

4. *Disburdening boundary conditions.* No extra boundary conditions need to be imposed, provided the boundary is a streamline, since the particle is adhered on the streamlines of the interpolated $\psi(\mathbf{x}, t)$.

However, the element-wise constant velocity field construction also exposes a number of undesired properties.

1. *Incomplete domain coverage.* The time-evolution of particles lying on an element vertex or edge risk singularity. Particle velocities at element vertices are not well-defined, and a zero normal velocity on a boundary edge would result in ambiguity of which element the particle should be attributed to.

2. *Numerical Instability.* Any small numerical error in particle positions $\mathbf{X}(t)$ may mis-attribute the particle to a wrong element, which could make it travel on a substantially different streamline with unbounded deviations from the exact trajectory.

3. *Potential stalls.* A particle stalls if it lies at a local maximum or minimum point of the streamfunction. This is not problematic as the spatially linear interpolation guarantees that local peak or trough of streamfunction are only found on the element vertices.

To prevent undesirable outcomes, several measures are taken when implementing the numerical scheme.

1. *Avoid initialisation on element vertices.* In practice any small numerical deviations will prevent a particle from being advected exactly to any element vertex. As long as particles are not initialised on the vertices, they will not likely visit these vertices either.

2. *Extended precision floating-points representation.* All floating-points quantities are implemented with extended precision to avoid overflow when encountering division by small values.

3. *Stable algorithms.* The order of floating-points operations shall be designed with caution to suppress numerical artefacts. For example, the stable algorithm in Forsythe (1966) is implemented to solve the quadratic equation (D.4) and handle different corners cases. The magnitude of normal velocity at an edge is implemented to be the average of the two adjacent elements, despite their theoretical identity, to ensure a unique normal velocity across elements. The initial identification of elements particles lie are implemented using code from libsupermesh (Panourgias and Maddison, 2016) for a quad-tree based search (Samet, 1984).

Appendix E

Uniformity of invariant distribution

Consider the Fokker–Planck (advection–diffusion) equation (5.2)

$$\frac{\partial \pi}{\partial t} + \nabla \cdot (\mathbf{U}\pi) = \nabla \cdot (\mathsf{K}\nabla\pi)\,, \tag{E.1a}$$

$$\pi\left(\mathbf{x}, 0|\mathbf{x}_0\right) = \delta\left(\mathbf{x} - \mathbf{x}_0\right). \tag{E.1b}$$

on a closed domain Ω with an incompressible mean flow $\mathbf{u}$ with no net flow on the boundary: $\mathbf{u}\cdot\hat{n} = 0$, where $\hat{n}$ is the outward normal on the boundary $\partial\Omega$. Imposing a no-flux boundary condition on the transition density $\pi = \pi(\mathbf{x}, t|\mathbf{y})$

$$(\mathbf{u}\pi - \mathsf{K}\nabla\pi)\cdot\hat{n} = 0, \quad \forall \mathbf{x} \in \partial\Omega,$$

ensures Lagrangian particles governed by the associated SDE never escape the closed domain Ω. Note that the no-flux boundary condition is implied by

$$\hat{n}^{\mathrm{T}}\mathsf{K}\nabla\pi = 0 \tag{E.2a}$$

$$\mathbf{u}\cdot\hat{n} = 0 \tag{E.2b}$$

thanks to the boundary condition on mean flow $\mathbf{u}\cdot\hat{n} = 0$. In the special case where the unit tangent and normal to the domain boundary are colinear with the eigenvectors of K on the boundary $\partial\Omega$, the only admissible $\nabla\pi$ satisfies $\nabla\pi\cdot\hat{n} = 0$ where particles reflected elastically along the normal direction on the boundary.

Consider the second moment of π by multiplying the Fokker–Planck equation by π and applying vector identities, which yields

$$\begin{aligned} & \pi\frac{\partial\pi}{\partial t} + \pi\mathbf{u}\cdot\nabla\pi - \pi\nabla\cdot(\mathsf{K}\nabla\pi) = 0 \\ \Rightarrow \qquad & \frac{\partial\pi^2}{\partial t} + \nabla\cdot\left(\mathbf{u}\pi^2 - 2\pi\mathsf{K}\nabla\pi\right) = -2\nabla\pi\cdot\mathsf{K}\nabla\pi, \end{aligned}$$

where the final line followed from the incompressibility $\nabla\cdot\mathbf{u} = 0$. Integrating over the domain and applying divergence theorem gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \pi^2 \mathrm{d}\mathbf{x} + \int_{\partial\Omega}\left(\mathbf{u}\pi^2 - 2\pi\mathsf{K}\nabla\pi\right)\cdot\hat{n}\mathrm{d}l = -2\int_\Omega \nabla\pi\cdot\mathsf{K}\nabla\pi\mathrm{d}\mathbf{x}.$$

With the simplified no-flux boundary conditions (E.2) the above surface integral vanishes. Note that also $\nabla\pi \cdot \mathsf{K}\nabla\pi \geq 0$ due to the positive-definiteness of the tensor K. Hence the integrated second moment is bound to be non-increasing in time

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\pi^2\mathrm{d}\mathbf{x} \leq 0 \tag{E.3}$$

with equality holds only when $\nabla\pi = \mathbf{0}$ almost everywhere, which suggests the invariant distribution has no variation over the domain and is given by $\rho(\mathbf{x}) = 1/\int_{\Omega}\mathrm{d}\mathbf{x}$.

Bibliography

Abernathey, R., Ferreira, D., Klocker, A., 2013. Diagnostics of isopycnal mixing in a circumpolar channel. Ocean Modelling 72, 1–16.

Abernathey, R. P., Marshall, J., 2013. Global surface eddy diffusivities derived from satellite altimetry. Journal of Geophysical Research: Oceans 118, 901–916.

Alnæs, M. S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., Wells, G. N., 2015. The FEniCS Project Version 1.5. Archive of Numerical Software 3 (100).

Andrews, D. G., Holton, J. R., Leovy, C. B., 1987. Middle atmosphere dynamics, 1st Edition.

Arakawa, A., 1966. Computational design for long-term numerical integration of theequations of fluid motion: Two-dimensional incompressible flow. Part I. Journal of Computational Physics 1 (1), 119–143.

Argo, 2000. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE.

Ascher, U. M., Petzold, L. R., 1998. Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations, 1st Edition. Society for Industrial and Applied Mathematics.

Asmussen, S., Glynn, P. W., 2007. Stochastic simulation: algorithms and analysis. Springer.

Bachman, S., Fox-Kemper, B., Bryan, F. O., 02 2015. A tracer-based inversion method for diagnosing eddy-induced diffusivity and advection. Ocean Modelling 86, 1–14.

Bachman, S., Fox-Kemper, B., Bryan, F. O., 2020. A Diagnosis of Anisotropic Eddy Diffusion From a High-Resolution Global Ocean Model. Journal of Advances in Modeling Earth Systems 12 (2), e2019MS001904.

Bauer, S., Swenson, M. S., Griffa, A., 2002. Eddy mean flow decomposition and eddy diffusivity estimates in the tropical Pacific Ocean: 2. Results. Journal of Geophysical Research: Oceans 107 (C10), 18–1–18–18.

Bennett, A. F., 1984. Relative Dispersion: Local and Nonlocal Dynamics. Journal of the Atmospheric Sciences 41 (11), 1881–1886.

Berloff, P., Hogg, A. M. C., Dewar, W., 2007. The turbulent oscillator: A mechanism of low-frequency variability of the wind-driven ocean gyres. Journal of Physical Oceanography 37 (9), 2363–2386.

Berloff, P. S., McWilliams, J. C., 2002. Material Transport in Oceanic Gyres. Part II: Hierarchy of Stochastic Models. Journal of Physical Oceanography 32 (3), 797–830.

Beskos, A., Pillai, N., Roberts, G. O., Sanz-Serna, J.-M., Stuart, A., 2013. Optimal tuning of the hybrid Monte Carlo algorithm. Bernoulli 19 (5A), 1501–1534.

Borgas, M., Flesch, T. K., Sawford, B. L., 1997. Turbulent dispersion with broken reflectional symmetry. Journal of Fluid Mechanics 332, 141—-156.

Bracco, A., LaCasce, J. H., Provenzale, A., 2000. Velocity Probability Density Functions for Oceanic Floats. Journal of Physical Oceanography 30 (3), 461–474.

Brooks, S. P., Gelman, A., 1998. General Methods for Monitoring Convergence of Iterative Simulations. Journal of Computational and Graphical Statistics 7 (4), 434–455.

Charney, J. G., 1947. The dynamics of long waves in a baroclinic westerly current. Journal of Meteorology 4, 136–162.

Chen, Y., Baule, A., Touchette, H., Just, W., Nov 2013. Weak-noise limit of a piecewise-smooth stochastic differential equation. Phys. Rev. E 88, 052103.

Chiswell, S. M., 2013. Lagrangian Time Scales and Eddy Diffusivity at 1000 m Compared to the Surface in the South Pacific and Indian Oceans. Journal of Physical Oceanography 43 (12), 2718–2732.

Cotter, C. J., Pavliotis, G. A., 2009. Estimating eddy diffusivities from noisy Lagrangian observations. Communications in Mathematical Sciences 7 (4), 805–838.

Cotter, S. L., Roberts, G. O., Stuart, A. M., White, D., 2013. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. Statistical Science 28 (3), 424–446.

Crommelin, D., Vanden-Eijnden, E., 2006. Reconstruction of diffusions using spectral data from timeseries. Communications in mathematical sciences 4, 651–668.

Dashti, M., Stuart, A. M., 2017. The bayesian approach to inverse problems. In: R., G., D., H., H., O. (Eds.), Handbook of Uncertainty Quantification. Springer, Cham.

Davis, R. E., 1987. Modeling eddy transport of passive tracers. Journal of Marine Research 45 (3), 635–666.

Davis, R. E., 1991. Observing the general circulation with floats. Deep Sea Research Part A. Oceanographic Research Papers 38, S531–S571.

Dembo, A., Zeitouni, O., 1998. Large Deviations Techniques and Applications. Springer.

E, W., Li, T., Vanden-Eijnden, E., 2019. Applied Stochastic Analysis. American Mathematical Society.

Eady, E. T., 1949. Long Waves and Cyclone Waves. Tellus 1 (3), 33–52.

Eden, C., Greatbatch, R. J., 2008. Towards a mesoscale eddy closure. Ocean Modelling 20 (3), 223–239.

Evans, L. C., 2013. An Introduction to Stochastic Differential Equations. American Mathematical Society.

Ferrari, R., Nikurashin, M., 2010. Suppression of Eddy Diffusivity across Jets in the Southern Ocean. Journal of Physical Oceanography 40 (7), 1501–1519.

Ferrari, R., Wunsch, C., 2009. Ocean Circulation Kinetic Energy: Reservoirs, Sources, and Sinks. Annual Review of Fluid Mechanics 41 (1), 253–282.

Forsythe, G. E., 1966. How Do You Solve a Quadratic Equation? Tech. rep., Stanford University, Stanford, CA, USA.

Freidlin, M., Szücs, J., Wentzell, A., 2012. Random Perturbations of Dynamical Systems. Springer.

Froyland, G., Gottwald, G. A., Hammerlindl, A., 2014. A Computational Method to Extract Macroscopic Variables and Their Dynamics in Multiscale Systems. SIAM Journal on Applied Dynamical Systems 13 (4), 1816–1846.

Gargett, A. E., 1984. Vertical eddy diffusivity in the ocean interior. Journal of Marine Research 42 (2), 359–393.

Garraffo, Z. D., Mariano, A. J., Griffa, A., Veneziani, C., Chassignet, E. P., 2001. Lagrangian data in a high-resolution numerical simulation of the North Atlantic: I. Comparison with in situ drifter data. Journal of Marine Systems 29 (1), 157–176.

Gelman, A., B., C. J., S., S. H., Dunson, D., A., V., D.B., R., 2013. Bayesian Data Analysis, Third Edition. Chapman and Hall/CRC.

Gelman, A., Rubin, D. B., 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science 7 (4), 457–472.

Green, J. S. A., 1970. Transfer properties of the large-scale eddies and the general circulation of the atmosphere. Quarterly Journal of the Royal Meteorological Society 96 (408), 157–185.

Griesel, A., Gille, S. T., Sprintall, J., McClean, J. L., LaCasce, J. H., Maltrud, M. E., 2010. Isopycnal diffusivities in the Antarctic Circumpolar Current inferred from Lagrangian floats in an eddying model. Journal of Geophysical Research: Oceans 115 (C6).

Haidvogel, D. B., McWilliams, J., Gent, P. R., 1992. Boundary current separation in a quasigeostrophic, eddy-resolving ocean circulation model. Journal of Physical Oceanography 22 (8), 882–902.

Haigh, M., Sun, L., Shevchenko, I., Berloff, P., 2020. Tracer-based estimates of eddy-induced diffusivities. Deep Sea Research Part I: Oceanographic Research Papers, 103264.

Hairer, M., Stuart, A. M., Vollmer, S. J., 2014. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. Annals of Applied Probability 24 (6), 237–574.

Hallberg, R., Gnanadesikan, A., 2006. The Role of Eddies in Determining the Structure and Response of the Wind-Driven Southern Hemisphere Overturning: Results from the Modeling Eddies in the Southern Ocean (MESO) Project. Journal of Physical Oceanography 36, 2232–2252.

Ham, D. A., 2006. On techniques for modelling coastal and ocean flow with unstructured meshes. Ph.D. thesis, Technische Universiteit Delft.

Ham, D. A., Pietrzak, J., Stelling, G. S., 2006. A streamline tracking algorithm for semi-Lagrangian advection schemes based on the analytic integration of the velocity field. Journal of Computational and Applied Mathematics 192 (1), 168–174.

Hosseini, B., 2019. Two Metropolis–Hastings Algorithms for Posterior Measures with Non-Gaussian Priors in Infinite Dimensions. SIAM/ASA Journal on Uncertainty Quantification 7 (4), 1185–1223.

Ivanov, L. M., Chu, P. C., 2019. Estimation of turbulent diffusion coefficients from decomposition of Lagrangian trajectories. Ocean Modelling 137, 114–131.

Jayne, S. R., Marotzke, J., 2002. The Oceanic Eddy Heat Transport. Journal of Physical Oceanography 32 (12), 3328–3345.

Kaipio, J., Somersalo, E., 2005. Statistical and Computational Inverse Problems. Springer.

Karabasov, S. A., Berloff, P. S., Goloviznin, V. M., 2009. CABARET in the ocean gyres. Ocean Modelling 30 (2–3), 155–168.

Klocker, A., Ferrari, R., LaCasce, J. H., 2012a. Estimating Suppression of Eddy Mixing by Mean Flows. Journal of Physical Oceanography 42 (9), 1566–1576.

Klocker, A., Ferrari, R., LaCasce, J. H., Merrifield, S. T., 2012b. Reconciling float-based and tracer-based estimates of lateral diffusivities. Journal of Marine Research 70 (4), 569–602.

Koszalka, I., LaCasce, J. H., Orvik, K. A., 2009. Relative dispersion in the Nordic Seas. Journal of Marine Research 67 (4), 411–433.

LaCasce, J. H., 2008. Statistics from Lagrangian observations. Progress in Oceanography 77 (1), 1–29.

Laurindo, L. C., Mariano, A. J., Lumpkin, R., 2017. An improved near-surface velocity climatology for the global ocean from drifter observations. Deep Sea Research Part I: Oceanographic Research Papers 124 (1), 73–92.

Lax, P., Wendroff, B., 1960. Systems of conservation laws. Communications on Pure and Applied Mathematics 13 (2), 217–237.

Lee, M.-M., Marshall, D. P., Williams, R. G., 1997. On the eddy transfer of tracers: Advective or diffusive? Journal of Marine Research 55 (3), 483–505.

Leimkuhler, B., Reich, S., 2004. Simulating Hamiltonian Dynamics, 1st Edition. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

LeVeque, R. J., 2002. Finite Volume Methods for Hyperbolic Problems. Cambridge University Press.

Lilly, J. M., Rhines, P. B., Schott, F., Lavender, K., Lazier, J., Send, U. S., D'Asaro, E., 2003. Observations of the Labrador Sea eddy field. Progress in Oceanography 59 (1), 75–176.

Lilly, J. M., Sykulski, A. M., Early, J. J., Olhede, S. C., 2017. Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. Nonlinear Processes in Geophysics 24 (3), 481–514.

Liu, J. S., 2008. Monte Carlo Strategies in Scientific Computing. Springer Publishing Company, Incorporated.

Logg, A., Mardal, K., Wells, G. N., 2012. Automated Solution of Differential Equations by the Finite Element Method. Springer.

Lumpkin, R., Elipot, S., 2010. Surface drifter pair spreading in the North Atlantic. Journal of Geophysical Research: Oceans 115 (C12).

Lumpkin, R., Treguier, A.-M., Speer, K., 2002. Lagrangian Eddy Scales in the Northern Atlantic Ocean. Journal of Physical Oceanography 32 (9), 2425–2440.

Maddison, J. R., Marshall, D. P., Shipton, J., 2015. On the dynamical influence of ocean eddy potential vorticity fluxes. Ocean Modelling 92, 169–182.

Majda, A. J., Kramer, P. R., 1999. Simplified models for turbulent diffusion: Theory, numerical modelling, and physical phenomena. Physics Reports 314 (4), 237–574.

Majda, A. J., McLaughlin, R. M., 1993. The Effect of Mean Flows on Enhanced Diffusivity in Transport by Incompressible Periodic Velocity Fields. Studies in Applied Mathematics 89 (3), 245–279.

Mak, J., Maddison, J. R., Marshall, D. P., 2016. A new gauge-invariant method for diagnosing eddy diffusivities. Ocean Modelling 104 (4), 252–268.

Marshall, D. P., Maddison, J. R., Berloff, P. S., 2012. A framework for parameterizing eddy potential vorticity fluxes. Journal of Physical Oceanography 42 (4), 539–557.

Marshall, J., Radko, T., 2003. Residual-Mean Solutions for the Antarctic Circumpolar Current and Its Associated Overturning Circulation. Journal of Physical Oceanography 33 (11), 2341–2354.

Marshall, J., Shuckburgh, E., Jones, H., Hill, C., 2006. Estimates and implications of surface eddy diffusivity in the Southern Ocean derived from tracer transport. Journal of Physical Oceanography 36 (9), 1806–1821.

Marshall, J., Shutts, G., 1981. A note on rotational and divergent eddy fluxes. Journal of Physical Oceanography 11 (12), 1677–1680.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics 21 (6), 1087–1092.

Nakamura, N., 1996. Two-dimensional mixing, edge formation, and permeability diagnosed in an area coordinate. Journal of the Atmospheric Sciences 53 (11), 1524–1537.

Oh, I. S., Zhurbas, V., Park, W., 2000. Estimating horizontal diffusivity in the East Sea (Sea of Japan) and the northwest Pacific from satellite-tracked drifter data. Journal of Geophysical Research: Oceans 105 (C3), 6483–6492.

Panourgias, I., Maddison, J. R., 2016. Parallel supermeshing for multimesh modelling. Tech. rep., doi:10.5281/zenodo.1316942.

Pasquero, C., Bracco, A., Provenzale, A., Weiss, J. B., 2007. Particle motion in a sea of eddies. In: Griffa, A., Kirwan, Jr., A. D., Mariano, A. J., Özgökmen, T., Rossby, H. T. (Eds.), Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics. Cambridge University Press, pp. 89–118.

Pavliotis, G. A., 2002. Homogenization Theory for Advection—Diffusion Equations with Mean Flow. Ph.D. thesis.

Pavliotis, G. A., 2014. Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations. Texts in Applied Mathematics. Springer-Verlag New York.

Pavliotis, G. A., Stuart, A. M., 2007. Parameter Estimation for Multiscale Diffusions. Journal of Statistical Physics 127 (4), 741–781.

Pavliotis, G. A., Stuart, A. M., 2008. Multiscale Methods: Averaging and Homogenization. Texts in Applied Mathematics. Springer-Verlag New York.

Pedlosky, J., 1987. Geophysical Fluid Dynamics, 2nd Edition. Springer-Verlag.

Peng, S., Qian, Y.-K., Lumpkin, R., Li, P., Wang, D., Du, Y., 2015. Characteristics of the Near-Surface Currents in the Indian Ocean as Deduced from Satellite-Tracked Surface Drifters. Part II: Lagrangian Statistics. Journal of Physical Oceanography 45, 459–477.

Qian, Y.-K., Peng, S., Liang, C.-X., 2019. Reconciling Lagrangian Diffusivity and Effective Diffusivity in Contour-Based Coordinates. Journal of Physical Oceanography 49 (6), 1521–1539.

Redi, M. H., 1982. Oceanic Isopycnal Mixing by Coordinate Rotation. Journal of Physical Oceanography 12 (10), 1154–1158.

Robert, C. P., Casella, G., 2005. Monte Carlo Statistical Methods (Springer Texts in Statistics). Springer-Verlag.

Roberts, G. O., Gelman, A., Gilks, W. R., 1997. Weak convergence and optimal scaling of random walk Metropolis-Hastings algorithms. The Annals of Applied Probability 7 (1), 110–120.

Roberts, G. O., Rosenthal, J. S., 1998. Optimal scaling of discrete approximations to Langevin diffusions. Journal of the Royal Statistical Society: Series B (Statistical Methodology 60 (1), 255–268.

Roberts, G. O., Rosenthal, J. S., 2006. Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. Annals of Applied Probability 16 (4), 2123–2139.

Ross, S. M., 2006. Simulation, Fourth Edition. Academic Press, Inc.

Rühs, S., Zhurbas, V., Koszalka, I. M., Durgadoo, J. V., Biastoch, A., 2018. Eddy Diffusivity Estimates from Lagrangian Trajectories Simulated with Ocean Models and Surface Drifter Data—A Case Study for the Greater Agulhas System. Journal of Physical Oceanography 48 (1), 175–196.

Rypina, I. I., Kamenkovich, I., Berloff, P., Pratt, L. J., 2012. Eddy-Induced Particle Dispersion in the Near-Surface North Atlantic. Journal of Physical Oceanography 42 (12), 2206–2228.

Sallée, J. B., Speer, K., Morrow, R., Lumpkin, R., 2008. An estimate of Lagrangian eddy statistics and diffusion in the mixed layer of the Southern Ocean. Journal of Marine Research 66 (4), 441–463.

Samet, H., 1984. The Quadtree and Related Hierarchical Data Structures. ACM Computing Surveys (CSUR) 16 (2), 187–260.

Sansón, L. Z., Pérez-Brunius, P., Sheinbaum, J., 2017. Surface Relative Dispersion in the Southwestern Gulf of Mexico. Journal of Physical Oceanography 47 (2), 387–403.

Strang, G., 1986. Introduction to applied mathematics. Wellesley-Cambridge Press.

Swenson, M. S., Niiler, P. P., 1996. Statistical analysis of the surface circulation of the California Current. Journal of Geophysical Research: Oceans 101 (C10), 22631–22645.

Sykulski, A. M., Olhede, S. C., Lilly, J. M., Danioux, E., 2016. Lagrangian time series models for ocean surface drifter trajectories. Journal of the Royal Statistical Society: Series C (Applied Statistics) 65 (1), 29–50.

Taylor, G. I., 1922. Diffusion by continuous movements. Proceedings of the London Mathematical Society s2-20 (1), 196–212.

Vallis, G. K., 2006. Atmospheric and oceanic fluid dynamics : fundamentals and large-scale circulation, 1st Edition. Cambridge University Press.

Van Leer, B., 1977. Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. Journal of Computational Physics 23 (3), 276–299.

Veneziani, M., Griffa, A., Reynolds, A. M., Mariano, A. J., 2004. Oceanic turbulence and stochastic models from subsurface Lagrangian data for the northwest Atlantic Ocean. Journal of Physical Oceanography 34, 1884–1906.

Vollmer, S., 2015. Dimension-Independent MCMC Sampling for Inverse Problems with Non-Gaussian Priors. SIAM/ASA Journal on Uncertainty Quantification 3, 535–561.

Waterman, S., Hogg, N. G., Jayne, S. R., 2011. Eddy–Mean Flow Interaction in the Kuroshio Extension Region. Journal of Physical Oceanography 41, 1182–1208.

Wolfram, P. J., Ringler, T. D., 2017. Computing eddy-driven effective diffusivity using Lagrangian particles. Ocean Modelling 118, 94–106.

Ying, Y. K., Maddison, J. R., Vanneste, J., 2019. Bayesian inference of ocean diffusivity from Lagrangian trajectory data. Ocean Modelling 140, 101401.

Zhai, X., Greatbatch, R. J., 2006. Surface eddy diffusivity for heat in a model of the northwest Atlantic Ocean. Geophysical Research Letters 33.

Zhurbas, V., Oh, I. S., 2003. Lateral diffusivity and Lagrangian scales in the Pacific Ocean as derived from drifter data. Journal of Geophysical Research: Oceans 108 (C5).

Zhurbas, V., Oh, I. S., 2004. Drifter-derived maps of lateral diffusivity in the Pacific and Atlantic Oceans in relation to surface circulation patterns. Journal of Geophysical Research: Oceans 109 (C5).