Statistical Estimation of Flow Components and Diffusivity from Ocean Drifter Velocity Observations

Sarah Oscroft, MMath(Hons.), M.Res

Submitted for the degree of Doctor of Philosophy at Lancaster University.

June 2022

STOR-i
excellence with impact

Abstract

Position and velocity measurements from freely-drifting surface buoys, known as drifters, provide a unique observational dataset for measuring ocean surface flow. The most immediate value of such observations is to aggregate data across drifters to determine the currents and the mean flow at given spatial locations. However, with more sophisticated models and statistical estimation techniques we can capture additional flow components from drifter data, as we shall demonstrate in this thesis. First, we shall demonstrate how data from closely-deployed collections of drifters can be jointly modelled to extract and identify mesoscale flow components—such as strain, vorticity, and divergence—as well as submesoscale components such as diffusivity, and background components such as inertial oscillations. We apply our methods to the LatMix deployment of drifters in the Sargasso Sea in 2011. Identification of so many flow components is made possible by considering the relative motions of the drifters with respect to each other, rather than in isolation. In the proceeding two chapters we shall build on these findings and provide evidence, both from simulation results and from analytically derived statistical properties, to demonstrate how the identification and estimation of flow components is expected to statistically behave

as both the sampling features (e.g. number of drifters and length of deployment) and the underlying flow field changes. Finally, we perform a separate analysis on Global Drifter Program data, where drifters are too far apart to be modelled relatively, and we instead perform a mean-flow and diffusivity separation using a novel estimator for estimating large-scale diffusivity which we propose from the spectral analysis of time series. A central goal of this thesis is to quantify the statistical error of parameter estimates of flow components, both in terms of bias and variance, and to then tune estimation methods to reduce these errors as much as possible.

Acknowledgements

First of all I'd like to thank my supervisors Adam Sykulski and Jeffrey Early for all their support over the last few years. Their guidance and enthusiasm has been invaluable. I would also like to thank Pascale Lelong who was initially a part of my supervisory team and helped guide me in my early work. I would like to thank STOR-i for all the opportunities that they have provided throughout this PhD. I also thank EPSRC for funding the majority of my time with STOR-i. Finally, I would like to thank my family who have supported me throughout my entire academic career, and my partner Alex for his encouragement and patience, in particular over the last couple of years.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Sarah Oscroft

Contents

List of Symbols

t	Time.
t_0	Initial time.
t_n	Time at observation n.
T	Simulation length (time).
x	Position of a drifter in the x-axis.
y	Position of a drifter in the y-axis.
z	Position of a drifter in the complex plane such that $z = x + iy$.
x_k	Position of drifter k in the x-axis.
y_k	Position of drifter k in the y-axis.
$\tilde{x}$	Position in x with centre of mass removed.
$\tilde{x}'$	Position in x after removing centre of mass and initial positions.
x_k^{sm}	Submesoscale position in x, calculated from residual velocities, u_k^{sm}.
y_k^{sm}	Submesoscale position in y, calculated from residual velocities, v_k^{sm}.
u	Velocity of a drifter in the x-axis.
v	Velocity of a drifter in the y-axis.
w	Velocity of a drifter in the complex plane such that $w = u_i v$.

$v_{\{j\}}$	Velocity in the j dimension.
$\tilde{v}_{\{j\}}$	Difference between velocity in the j dimension and its mean.
u_k	Velocity in the x direction for drifter k.
v_k	Velocity in the y direction for drifter k.
u^{bg}	Background flow in the u direction.
v^{bg}	Background flow in the v direction.
u_k^{sm}	Residual (submesoscale) velocity in u for drifter k.
v_k^{sm}	Residual (submesoscale) velocity in v for drifter k.
u_x	Partial derivative $\partial u/\partial x$.
u_y	Partial derivative $\partial u/\partial y$.
v_x	Partial derivative $\partial v/\partial x$.
v_y	Partial derivative $\partial v/\partial y$.
K	Number of drifters.
$d_{\{k\}}$	Displacement in the k direction.
$\tilde{d}_{\{k\}}$	Difference between displacement in the k dimension and its mean.
$s_{\{j,k\}}$	Single particle dispersion in j and k dimensions.
$Q(x,t)$	Probability that a particle is in position $x(t)$ at time t.
$E[\cdot]$	Expectation of $\cdot$.
$< \cdot >$	Average over multiple particles.
$\kappa(t)$	Absolute diffusivity at time t.
κ	Isopycnal diffusivity.
$\kappa_{\{xy\}}$	Diffusivity tensor between x and y directions.

$\kappa_{k,x}$ Diffusivity x direction for drifter k.

$\kappa_{k,x}^{\text{sm}}$ Submesoscale diffusivity in x for drifter k.

$\hat{\kappa}_{\text{obs}}$ Diffusivity estimate using observed positions with centre of mass removed.

$\hat{\kappa}_{\text{re-centred}}$ Diffusivity estimate using positions with centre of mass and initial positions removed.

$\hat{\kappa}_{\text{unbiased}}$ Unbiased diffusivity estimate using observed positions with centre of mass removed.

κ_{∞} Long-term diffusivity.

$\hat{\kappa}^{(P)}$ Periodogram diffusivity estimator.

$\hat{\kappa}_m^{(CP)}$ Chunked periodogram diffusivity estimator, where a time series has been chunked into m new time series.

δ Divergence.

ζ Vorticity.

σ_n Normal strain.

σ_s Shear strain.

σ Strain rate.

θ Strain angle.

m_x First moment of x.

m_y First moment of y.

$m_{\tilde{x}\tilde{x}}$ Second moment of $\tilde{x}$.

$m_{\tilde{y}\tilde{y}}$ Second moment of $\tilde{y}$.

$m_{\tilde{x}\tilde{y}}$ Cross second moment of $\tilde{x}$ and $\tilde{y}$.

$m_{\cdot\cdot}^r$ Unrotated second moment of $\cdot$.

Δ	Sampling interval of drifter observations in seconds.
σ_I	Isotropic standard deviation.
dW	Weiner increment.
W	Width of time window.
u_0	Mesoscale model parameter allowing the redefinition of the expansion point x_0 without changing the other model parameters.
v_0	Mesoscale model parameter allowing the redefinition of the expansion point x_0 without changing the other model parameters.
a_1	Mesoscale model parameter allowing a small linear increase in the mesoscale component.
b_1	Mesoscale model parameter allowing a small linear increase in the mesoscale component.
x_0	Expansion location for Taylor series in x-axis.
y_0	Expansion location for Taylor series in y-axis.
s^2	Okubo-Weiss parameter.
$\bar{s}$	Complementary Okubo-Weiss parameter.
B^m	B-splines.
$\hat{\sigma}^m$	Spline coefficients.
D	Degree of splines.
f_0	Inertial frequency.
e	Eccentricity.
σ_x	Standard deviation of simulated drifters in x.
σ_y	Standard deviation of simulated drifters in y.

σ_I	Isotropic standard deviation of simulated drifters.
d_0	Mean squared distance of drifters to the origin.
$\bar{t}$	Dimensionless timescale.
$\bar{x}$	Dimensionless lengthscale.
$\bar{y}$	Dimensionless lengthscale.
$\mu_{\tilde{x}}$	Expectation of $m_{\tilde{x}\tilde{x}}$.
$\mu_{\tilde{y}}$	Expectation of $m_{\tilde{y}\tilde{y}}$.
$\sigma^2_{\tilde{x}}$	Variance of $m_{\tilde{x}\tilde{x}}$.
$\sigma^2_{\tilde{y}}$	Variance of $m_{\tilde{y}\tilde{y}}$.
f	Frequency.
τ	Lag period.
$R(\tau)$	Autocorrelation.
γ_τ	Autocovariance.
$S(\omega)$	Lagrangian frequency spectrum.
$\hat{S}^{(P)}$	Periodogram estimator for spectral density.
ϕ	Autoregressive damping parameter.
ϵ_t	White noise signal.
m_{opt}	The approximated optimal number of chunks for the ACP diffusivity estimator.
X	Realisation from a unit normal distribution.
Y	Realisation from a unit normal distribution.
Z	Realisation from a unit normal distribution.

$z_{\tilde{x}}$ Realisation from a noncentral chi-squared distribution with $K-1$ degrees of freedom and noncentrality parameter $\lambda_{\tilde{x}}$.

$z_{\tilde{y}}$ Realisation from a noncentral chi-squared distribution with $K-1$ degrees of freedom and noncentrality parameter $\lambda_{\tilde{y}}$.

z_m^2 Realisation from a central chi-squared distribution with one degree of freedom.

z_{nm}^2 Realisation from a central chi-squared distribution with M degrees of freedom.

$\lambda_{\tilde{x}}$ Noncentrality parameter.

$\lambda_{\tilde{y}}$ Noncentrality parameter.

Chapter 1

Introduction

Knowledge of the ocean and its circulation has a number of practical applications. The ocean's currents move heat across the globe, regulating the climate and weather patterns (Rhines and Häkkinen, 2003). One quantity that is of particular importance and interest to oceanographers is the diffusivity (Taylor, 1922), this is the rate at which particles spread out over time in a fluid. Observing and measuring horizontal or lateral diffusivity at the surface can be performed using position observations from free-floating buoys known as drifters, which are designed to mimic the motion of a fluid parcel or particle at the surface. When multiple drifters are deployed in a cluster, they can be further used to obtain estimates of the mesoscale parts of the flow, such as strain and vorticity, as we shall show in this thesis. Strain causes particles to stretch along the axis aligned with the strain angle, and compress against the axis that is orthogonal to the strain angle, whereas vorticity causes particles to rotate.

Splitting the flow into a mean flow, as well as mesoscale components and diffusivity creates a more coherent picture of ocean flow at the surface. Knowledge of diffusivity,

strain and vorticity helps us to understand, for example, the impact area of an oil spill, or aid in search and rescue missions (Lumpkin et al., 2017). We can also gain insight into the dispersion of ocean life (Lumpkin and Pazos, 2007), or discover how plastics dropped into the ocean will spread (La Daana et al., 2017; Hardesty et al., 2017).

Diffusivity, strain and vorticity in the ocean can be separated and estimated using observations from "Lagrangian" instruments such as drifters (Koszalka et al., 2011). Lagrangian instruments freely drift in the ocean and are hence usually modelled using a moving or "Lagrangian" frame of reference. Each individual particle trajectory is in large part unpredictable, usually following a complex and meandering path (LaCasce, 2008). Therefore multiple instruments and observations are usually required to understand a flow field. Commonly used Lagrangian instruments are floats and drifters; they are designed to move with the currents and cover large regions of the ocean. Drifters track the surface currents, they consist of a transmitter which sits on the ocean surface and is tracked by satellites, and often have a subsurface drogue (Niiler et al., 1995) which causes the transmitter to drift with the currents under the surface, at the depth of the drogue which is typically of the order of 5–50 metres. Floats track the currents below the surface, meaning they cannot be tracked by satellite while at depth, but they either triangulate their position under water using sound sources or, more commonly, periodically resurface to measure position and transmit data. This thesis is focused on statistical methodology for surface drifter data, but many of our techniques and findings could be applied to float data too after some basic modifications.

The Global Drifter Program (GDP) is a global array of drifters which collect information about the surface currents of the ocean. It is part of the National Oceanic and Atmospheric Administration's (NOAA) Global Ocean Observing System and Global Climate Observing System. The GDP aims to provide data on surface velocities, sea surface temperatures, and pressure, amongst other things, that is used for weather forecasting, research, and calibration of satellite observations. The GDP maintains over 1,000 drifters globally at any given point in time, which has all together collected over 100 million individual velocity observations over 40 years covering most of the World's oceans (Elipot et al., 2016).

Diffusivity cannot be measured directly by individual Lagrangian instruments, it must instead be inferred from multiple position or velocity observations. There are a number of ways that diffusivity can be estimated from Lagrangian instruments, and there have been a number of different studies into measuring the diffusivity in different parts of the ocean (e.g. Zhurbas and Oh, 2004; Buehler and Holmes-Cerfon, 2009; Cronin et al., 2015; Sallée et al., 2008; LaCasce and Bower, 2000), however different estimators yield markedly different values for the diffusivity. In this thesis we focus on understanding the *error* of diffusivity estimates (as well as the error of other flow parameter estimates), which motivates us to propose new diffusivity estimators which may have lower error under certain assumptions and conditions. By "error" we mean both contributions from bias and variance, and in this thesis we shall generally work with the notion of mean squared error of parameter estimates and trying to minimise this.

Mesoscale parts of a flow can be estimated from clusters of drifters that are de-

ployed together, for example as part of smaller experiments than the full scope of the Global Drifter Program. An example of one such experiment is the LatMix (or Lateral Mixing) Campaign which was carried out in the Sargasso Sea (part of the North Atlantic Ocean) in Summer 2011 (Sundermeyer et al., 2020). There were deployments in two sites, one exhibiting weak straining, and the other with moderate straining (Shcherbina et al., 2015). Nine drifters were deployed at each site and each experiment was run for six days, with position observations recorded every 30 minutes. The LatMix experiment had the aim to determine what drives isopycnal (horizontal) mixing and stirring at the submesoscale (smaller than mesoscale). In this thesis we use the drifter trajectories from the LatMix experiment to split the flow into background, mesoscale, and submesoscale components, providing estimates of strain, vorticity and submesoscale diffusivity. We use the LatMix drifter trajectories as the foundation for our analysis of the errors of parameter estimates, which enables us to build a more general model and comment on the prospective optimal initial configuration of drifters based on the anticipated error of parameter estimates, to help guide future drifter deployments and statistical analyses.

This thesis will take a statistical approach to approximate the error of estimating mesoscale flow components and diffusivity from drifter position observations. Throughout this thesis we develop estimators for diffusivity and mesoscale components, and use simulation and analytical techniques to analyse the errors. We apply these estimators on drifter observations from the LatMix experiment and the Global Drifter Program, as well as synthetically generated trajectories using stochastic Lagrangian models. Our main aim throughout this thesis is to quantify and reduce the

errors of our estimates, to ensure correct scientific conclusions can be drawn from observed surface drifter positions and velocities.

1.1 Structure of Thesis

This thesis is structured as follows. Chapter 2 introduces some of the oceanographic concepts that will be used throughout this thesis. These include providing an overview of different instruments that are used to collect data, as well as definitions of diffusivity and mesoscale flow features (e.g. strain and vorticity) which specify the modelling components and parameters that will feature repeatedly throughout this thesis. We also introduce some methods to estimate these parameters from existing literature. We highlight some weaknesses of the different methods, and use these weaknesses to provide motivation for the following chapters in the thesis.

Drifters deployed in close proximity collectively provide a unique observational data set with which to separate mesoscale and submesoscale flows. In Chapter 3 we provide a principled approach for doing so by fitting observed velocities to a local Taylor expansion of the velocity flow field. We demonstrate how to estimate mesoscale and submesoscale quantities that evolve slowly over time, as well as their associated statistical uncertainty. We show that in practice the mesoscale component of our model can explain much first and second-moment variability in drifter velocities, especially at low frequencies. This results in much lower and more meaningful measures of submesoscale diffusivity, which would otherwise be contaminated by unresolved mesoscale flow. We quantify these effects theoretically via computing Lagrangian

frequency spectra, and demonstrate the usefulness of our methodology through simulations as well as with real observations from the LatMix deployment of drifters. The outcome of this method is a full Lagrangian decomposition of each drifter trajectory into three components that represent the background, mesoscale, and submesoscale flow.

In Chapter 4 we generalise the findings of Chapter 3, and provide an insight into how different drifter deployment configurations impact upon the errors of mesoscale and submesoscale parameter estimates. This allows us to make recommendations on how many, for how long, and in what configuration drifters should be deployed, depending on what we wish to learn from an observational experiment. This chapter uses simulations of particle trajectories to obtain error estimates, by comparing parameter estimates with the true value that was used to simulate the particle trajectories. We show that the optimal deployment "morphology" is dependent upon the number and combination of mesoscale parameters (strain, vorticity, and/or divergence) that we wish to estimate. If only one parameter is being estimated, then we will show that it is optimal to deploy drifters along a straight line, which in the case of strain is aligned with the strain angle. However, if more than one parameter is being estimated, or it is not known *a priori* which mesoscale features are present, then it is usually (but not always) optimal to deploy drifters with an isotropic configuration in two dimensions at the surface. We compare our results and recommendations with others made in the literature to establish general principles.

In Chapter 5 we derive the error of parameter estimates analytically in a strain-only setting and use these errors to comment on the scaling behaviour as different

parameters change, such as the number and temporal length of the drifter observations. This enables us to reinforce the simulation results found in Chapter 4, as well as quantify the error and approximate the full distribution of various parameter estimates. In this chapter we use method-of-moments to derive estimators for strain rate, strain angle and diffusivity, and by fitting distributions to these estimators we are able to calculate the root mean square error (RMSE) of different estimators. We will show that increasing the distance from the centre of mass of the initial position, in the axis aligned with the strain angle, reduces the error of strain rate estimates. However, changing this distance in the orthogonal axis has no effect on the error. This is not the case for diffusivity estimates, where we show that there is a more complex relationship between the initial positions and the error.

In Chapter 6 we focus on drifters deployed in isolation, and as such develop a new diffusivity estimator for single drifters, and apply this estimator to trajectories from the Global Drifter Program. We compare this estimator to the standard periodogram-based estimator to show that our new estimator has a lower variance, and a reduced error overall. We prove this by fitting chi-squared distributions to the different diffusivity estimators, by assuming drifter velocities are normally-distributed. The new estimator works by splitting a time series into multiple segments, and then smoothing over the segments, with the number of segments chosen as a trade-off between bias and variance to optimally reduce the error.

Finally, in Chapter 7 we make concluding remarks and point to interesting areas of outstanding further work and future directions that are motivated by the research performed in this thesis.

Chapter 2

Background and Motivation

2.1 Introduction

The ocean plays a major role in regulating the weather and climate across the globe (Rhines and Häkkinen, 2003). Its circulation transports heat between the tropics and the poles, balancing the temperatures around the world. Ocean currents impact the weather patterns worldwide, while transporting organisms and sediments around the water. These currents are generated from a number of forces which are acting on the water such as the rotation of the Earth, the wind and the gravitation of the moon (Seager, 2006). Studies of the ocean have a number of practical applications. For example, knowledge of the currents allows ships to take the most fuel efficient path across the ocean, track pollution such as an oil or sewage spill, or aid in search and rescue operations (Lumpkin et al., 2017). These studies can help with building models of the climate and weather which can be used in predicting severe weather events such as hurricanes.

Lumpkin and Pazos (2007) provide a summary of some early methods of measuring ocean quantities. In the early 17th century drift bottles were used to map surface currents. The bottles would usually carry a note containing the location and time of their launch (Sverdrup et al., 1942), and were weighed down so that they were submerged almost completely. They have been used to map currents in the North Sea (Fulton, 1897; Tait, 1930) and the northwest Pacific Ocean. Another early method used to study ocean surface circulation uses measurements taken from the drift of ships (Lumpkin and Pazos, 2007). The current can be found by finding the difference between the absolute and relative motion of a ship with no sailing force. However, a ship has forces acting upon it from both the currents and wind, and it is difficult to separate the two forces.

Methods for tracking currents have advanced in more recent years beyond sending a message in a bottle. Modern oceanography has allowed us to develop an increasingly accurate picture of the properties of the ocean. However knowledge of ocean circulation is not complete and depends on assumptions about its internal dynamics which are used to estimate ocean currents (Rossby, 2007).

There are a number of instruments used to gather data that is used to model the ocean. These can be either Lagrangian or Eulerian; Lagrangian instruments freely drift in the ocean and are hence usually modelled using a moving frame of reference, whereas Eulerian instruments are spatially fixed in position and so have a fixed reference frame. Throughout this thesis we consider data from Lagrangian instruments, and not from Eulerian sources. Lagrangian instruments are used to gather a large variety of data about the ocean (including e.g. temperature), but this

thesis will only consider the recorded positions and velocities of the instruments.

Drifter positions can be used to build knowledge of the ocean's circulation. Throughout this thesis, we are interested in the estimation of mean flow, as well as mesoscale and submesoscale parts of the flow. Mesoscale flow refers to features that spread from ten to hundreds of kilometres across the ocean, whereas the submesoscale part of the flow is typically between 100m-10km. Large scale winds and tides drive ocean flows at $\mathcal{O}$(1000km) which through turbulent interactions drive mesoscale and submesoscale eventually dissipating at the smallest scales. The combination of flows at all these scales simultaneously acting on a particle placed in the ocean will determine where it will end up. It isn't possible to perfectly estimate all these different contributions to the flow, and so a deterministic model would not be able to accurately predict the path this particle could take through the ocean. We will therefore use concepts from the field of statistics to build simplified stochastic models for the different parts of the flow.

This background chapter will introduce some of the major concepts that will be explored in more depth throughout this thesis. Specifically, Section 2.2 will introduce the type of data that will be used throughout the rest of the thesis—particle positions from Lagrangian instruments known as drifters. We also give an overview of a few drifter deployments. Section 2.3 introduces the concepts of mesoscale flow and diffusivity, which will be investigated in depth throughout the thesis. In Section 2.4 we will give an overview of methodologies in the literature for estimating diffusivity, and discuss some of their limitations. Section 2.5 reviews some existing methodology for separating mesoscale and submesoscale flow components as well as introducing the

Lateral Mixing and Coherent Turbulence (LatMix) experiment, from which we will use drifter velocity observations through this thesis. Finally, Section 2.6 provides the motivation for introducing our own estimators for flow component parameters that we shall introduce in the proceeding chapters.

2.2 Oceanographic Instruments

There are two ways to model the flow of a fluid, the first of these is the Eulerian method, where we have a fixed frame of reference. This method describes the velocity field as a function of both location and time, and is the preferred method for many applications in fluid dynamics (Rossby, 2007). In practice, Eulerian data from the ocean are collected using a number of spatially fixed in situ instruments (such as a network of moored buoys) or from remotely sensed data such as satellite images. The spatiotemporal resolution of the data is determined both by the rate of temporal sampling of the instruments, and by the spatial distance between measurement points (e.g. the distance between buoys or the resolution of satellite imagery).

The other way to model a fluid is the Lagrangian method, where we use a moving frame of reference. A fluid particle or parcel is followed by the observer, which can be specified by a velocity field as a function of time. Here the spatiotemporal information comes from the recorded position of each particle over time. In practice, the data comes from floats or drifters (see Section 2.2.1 for more details) and the spatiotemporal resolution of the data comes from the number of instruments that are being tracked, as well as their temporal sampling rate. This method tells us about how particles move

in a fluid, and is particularly useful in the problem of determining the horizontal structure of ocean currents, as we shall discuss. Lagrangian data can map out the structure of a fluid flow with a high level of detail, with even a single trajectory being able to provide a good insight into the underlying fluid dynamics.

Fluid flows are often turbulent, and therefore non deterministic. This means each individual particle trajectory is unpredictable, following a complex path, and so statistical methods can often work better to provide a description of the path than a physical deterministic system. We now introduce a couple of Lagrangian instruments that are used in oceanography, which will be relevant to the particle trajectories we will explore throughout this thesis.

2.2.1 Lagrangian Instruments

Floats and drifters are the two main classifications of Lagrangian instruments which are used to study the ocean. Over their lifetime they can cover large distances, and so these free-drifting instruments provide a more cost effective method of providing coverage of the ocean than Eulerian instruments which have fixed locations (Sykulski et al., 2016).

Floats

Floats are one type of Lagrangian instrument typically used in ocean studies. They move with the currents and can often travel large distances, providing data from just a single instrument across many spatial locations. Floats are used to track the currents below the surface. Argo floats are the most commonly used type of float. They move

horizontally through the ocean at a set depth of around 1000 metres, and periodically return to the ocean surface to transmit data to a satellite. This includes the position of the float as well as depth profiles of other quantities of interest such as temperature and conductivity, from which salinity and density can be computed. This resurfacing occurs at regular intervals (often between 7-10 days), or the float can remain below the surface for years collecting data. Therefore the velocities that are obtained have a low spatiotemporal resolution as position data is only obtained for each float at a minimum of every 7 days when the float surfaces. Another type of float is a RAFOS float, which is able to measure underwater velocities while submerged using acoustic tracking, therefore providing a much better temporal sampling rate.

Throughout this thesis we will not directly work with particle trajectories from floats, however the methods described in the thesis could be applied to float data with some modification. We instead focus our analysis to trajectories that are obtained from drifters.

Drifters

Drifters are the modern equivalent of a message in a bottle. They move freely over the surface of the ocean, tracking the surface currents. Drifters consist of a transmitter which sits on the ocean surface and is tracked by satellites. Drifters often have a subsurface drogue (Sybrandy and Niiler, 1991; Niiler et al., 1995) which causes the transmitter to drift with the currents at a distance between 5-50m under the surface, at the depth of the drogue. The drogue helps the drifter to move with the currents as it is less influenced by winds. Figure 2.2.1 (Lumpkin and Pazos, 2007) shows what

a drifter looks like. One type of drogue is known as a 'holey-sock' as they are made from nylon cloth and consist of between 4-7 pairs of holes, with each consecutive pair rotated 90 degrees from the previous opposing pair.

There are two main sizes of drifters, the original Surface Velocity Program (SVP) drifter which has a subsurface float, and a larger drogue which has 7 pairs of holes. The newer drifters are smaller and have no subsurface float and have only 4 or 5 pairs of holes in the drogue. SVP drifters can transmit position data for approximately 400 days in the ocean before they cease transmission, but some have managed to keep transmitting positions for up to 10 years.

Drifters were previously tracked using Doppler ranging by the Argos satellite system (Niller et al., 1987), and provided locations up to 14 times each day (Ohlmann et al., 2005) with an accuracy of the order of a few hundred metres. Modern drifters can be tracked by the global positioning system (GPS), and have a higher spatial and temporal resolution, giving the location up to every 30 minutes with accuracy of the order of a few meters (Elipot et al., 2016). The data are recorded internally, and are transmitted more infrequently.

Drifters are a quasi-Lagrangian instrument, they are drogued to a certain depth, and hence are only able to move in two dimensions. The potential density of a particle in the ocean is not constant at a given depth, and therefore the potential density each drifter observes will change as the drifter moves through the ocean. Throughout this thesis, we make no attempt to quantify the effect of the differing potential densities on the different parameters that we will estimate, and the results should be interpreted accordingly.

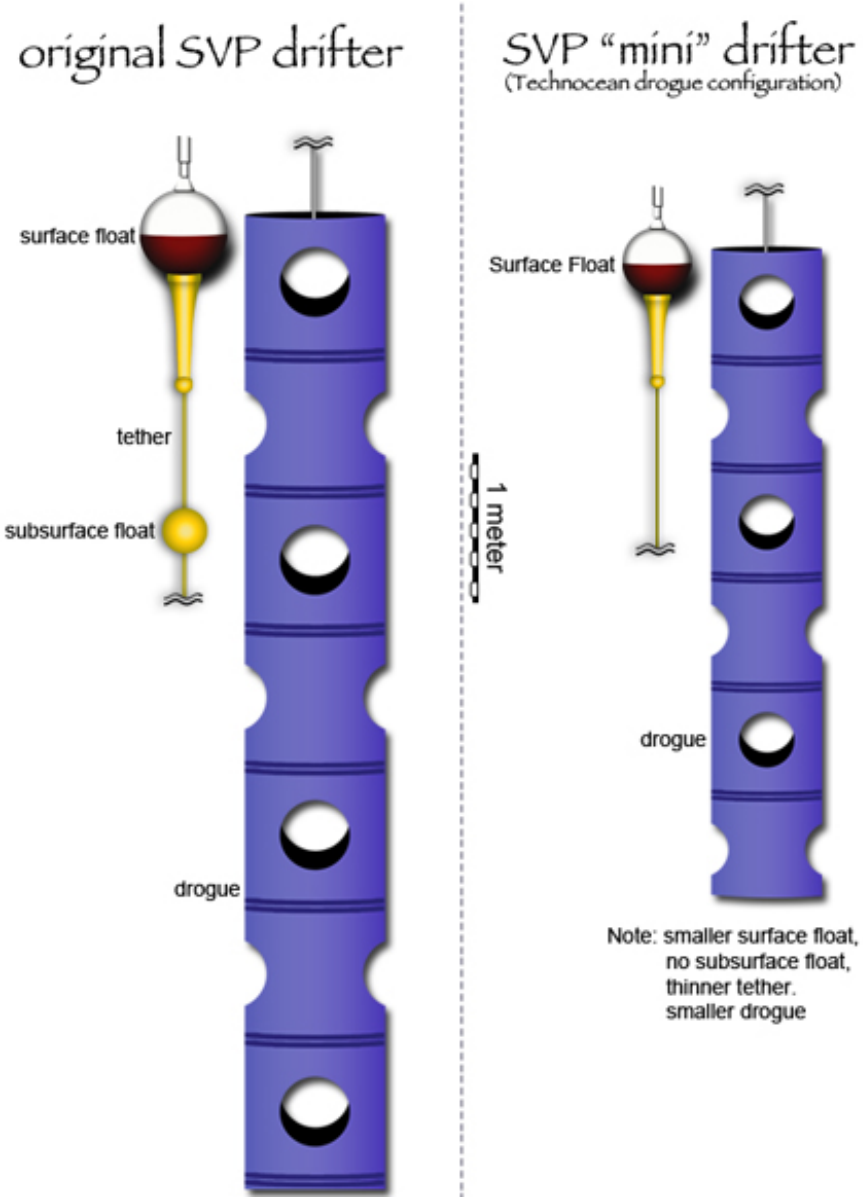

Figure 2.2.1: An illustration of a SVP drifter. Taken from `https://www.adp.noaa.gov/Students/LetsDissectaDrifter.aspx` (Lumpkin and Pazos, 2007).

2.2.2 Drifter Deployments

There have been a large number of drifter deployments, each with their own goals. The largest drifter array is the Global Drifter Program which provides data across the globe. Many smaller deployments have also taken place, each focusing on a smaller region, usually providing higher resolution data in space and time. We are only interested in the spatial and temporal position and velocity data through this thesis, but oceanographers use other drifter data (e.g. sea surface temperature, pressure, etc) and apply the information to a range of applications. Drifters can be used to learn about the circulation of the ocean, mapping out its currents, to its biology and learning about ocean life.

In this thesis we will explore results from both large-scale deployments such as the Global Drifter Program, and also from smaller-scale deployments where multiple drifters are deployed in the same region of the ocean approximately at the same time. These smaller-scale deployments provide insight into smaller features and regions of the ocean, known as the submesoscale and mesoscale, whereas global deployments provide understanding of large scale ocean currents and circulation. We will introduce submesoscale and mesoscale flow in more detail in Section 2.3. Examples of smaller deployments include the Grand Lagrangian Deployment (GLAD, Gulf of Mexico), Lateral Mixing and Coherent Turbulence Campaign (LatMix, Sargasso Sea), Lagrangian Submesoscale Experiment (LASER, Gulf of Mexico), and Coherent Lagrangian Pathways from the Surface Ocean to Interior (CALYPSO, Alboran Sea). Through this thesis we will focus on data from the Global Drifter Program and from

the LatMix campaign, but note that the methodology discussed can be applied to other drifter and float data sets.

2.2.3 The Global Drifter Program

The World Climate Research Program recognised in 1982 that a global array of drifters would be extremely useful for research in oceanography and the climate. The first Surface Velocity Program (SVP) drifters were deployed in 1979, and the modern set of SVP drifters includes all drifters deployed between 1979-1993. Large-scale deployments of the first modern SVP drifters took place in 1988 with the aim of mapping the surface circulation of the tropical Pacific Ocean.

The array of SVP drifters is today known as the Global Drifter Program (GDP), which is part of the National Oceanic and Atmospheric Administration's (NOAA) Global Ocean Observing System and Global Climate Observing System. It is a scientific project of the Data Buoy Cooperation Panel of the World Meteorological Organisational and International Oceanographic Commission. The GDP aims to provide data on operations, surface velocities, sea surface temperatures, and pressure that is used for weather forecasting, research and calibration of satellite observations. The GDP maintains over 1,000 drifters globally at any point in time and almost 20,000 drifters have been deployed in total over its history. There are two interpolated products (one hour resolution and six hour resolution) which together cumulatively yield several hundred million position and velocity measurements. The positions of all active drifters on 23$^{\text{rd}}$ May 2022 is displayed in Figure 2.2.2.

In March 2014 the GDP announced that they were transitioning their array of

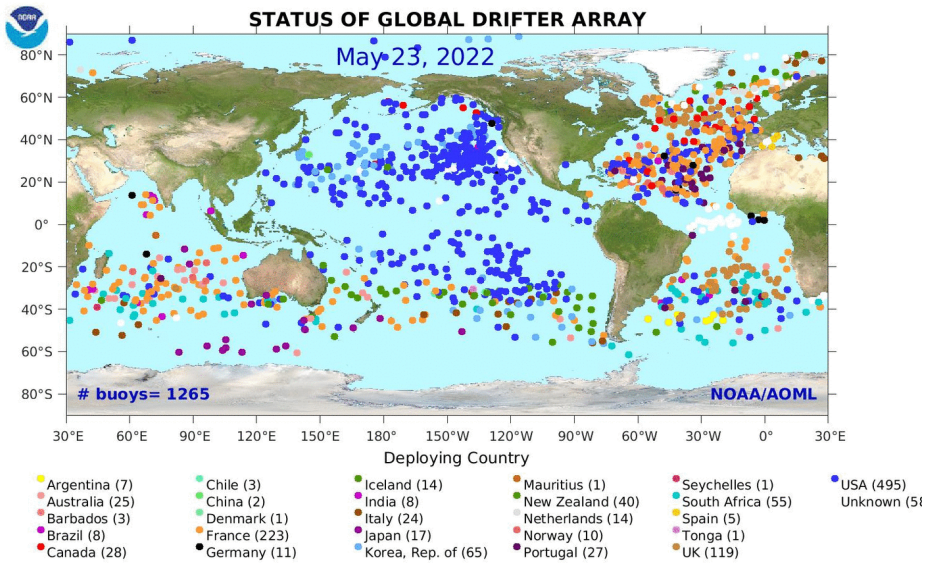

Figure 2.2.2: Position of drifters on 23rd May 2022 as taken from `https://www.aoml.noaa.gov/phod/gdp/`

drifters from Argos to Iridium transmissions (Lumpkin et al., 2017) in order to reduce costs. Iridium drifters include GPS and provide their positions hourly. The transition to Iridium is now complete (Elipot et al., 2022).

2.2.4 The LatMix Campaign

The LatMix campaign included two separate field experiments. We are interested in the first campaign which took place in the summer of 2011 in the Sargasso Sea, a region of the Atlantic Ocean with no land boundaries—it is instead bounded by currents. The campaign aimed to help build understanding of mixing at scales of 0.1–10km, also known as the sumbesoscale. A combination of drifters, floats, dye releases and ships were used as part of the campaign (Shcherbina et al., 2015; Sundermeyer

et al., 2020).

Observations were made at two overlapping sites, which were characterised as being weak and moderate straining. The experiment took place from the 1st–20th June, in an area of good water clarity which was relatively close to the shore, chosen for its relatively weak mesoscale field. The study in the region of weak strain (known as site 1) was carried out between the 2nd–10th June, with the dye and nine drifters being released on the 4th June. The study of the moderate straining site (known as site 2) took place between the 12–20th June, with the dye and nine drifters being released on the 13th June.

2.3 Ocean Circulation and Diffusivity

The primary application of floats and drifters is to measure the currents of the ocean. The large-scale mean currents move drifters across the globe, and have a large impact on the path a drifter will take. While these instruments generally follow the large-scale currents, many different physical effects will act upon a drifter as it moves through the ocean, such as wind, tides, eddies and fronts, amongst others. Through this thesis we are primarily interested in estimating the effects smaller than the large-scale currents, such as mesoscale features which we will decompose into strain, vorticity and divergence. Drifter motion is generally considered to be stochastic, meaning that the particles will also spread randomly from a given location, and the rate of this spread is known as the diffusivity, which we also study.

Knowledge of how particles move through the ocean allows us to gain a better

understanding of, for example, how an oil spill will spread in the water (Lumpkin et al., 2017), and the impact that it will cause. Knowledge of the ocean's currents and circulation can also be used to gain insight into the spreading of radioactive materials that are released into the water, or how ocean life such as fish larvae and plankton (Lumpkin and Pazos, 2007) will disperse. Another application is in aeroplane crashes in the ocean, as knowledge of the currents and circulation can be used to predict where debris came from or can aid recovery missions (e.g. Trinanes et al., 2016).

2.3.1 Mesoscale Flow

Movement of particles through the ocean is caused by processes at various scales. At the largest scale, the background flow encompasses large scale currents and inertial oscillations. At the other extreme is the microscale $\mathcal{O}$(0.01-1m), where stirring is at a molecular scale (Shcherbina et al., 2015), and at the submesoscale $\mathcal{O}$(0.1-10km) mixing is due to such diffusivity.

When drifters are deployed close together, their positions can be used to estimate mesoscale parts of the flow. Mesoscale flow $\mathcal{O}$(10-100km) is smaller than background flow which encompasses large scale currents and inertial oscillations, but is larger than submesoscale flow, the effects of which are often parameterised as a submesoscale diffusivity. The mesoscale describes effects between ten to a few hundred kilometers, for which it is useful to characterise in terms of strain, divergence and vorticity.

Strain in two dimensions refers to the stretching of the flow along one axis, and the compression along the other axis. It is broken down to normal and shear strain which refer to strain in different directions. A positive normal strain refers to stretching in

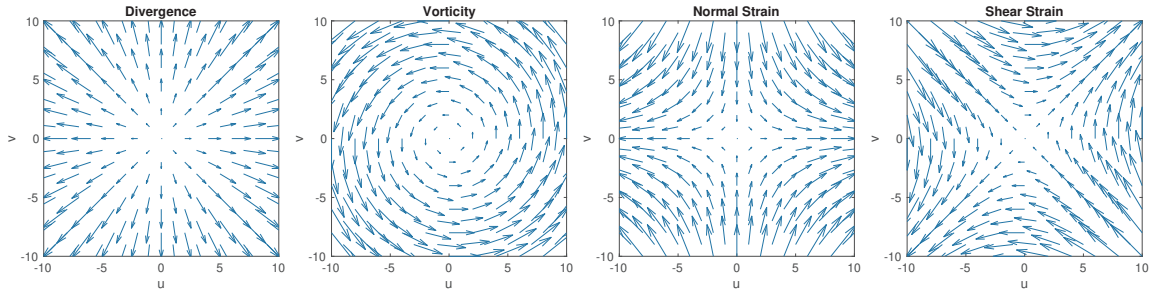

Figure 2.3.1: Velocity fields of divergence, vorticity, normal strain and shear strain (in order left to right).

the x-axis whereas a negative normal strain refers to stretching in the y-axis. Shear strain is equivalent to rotating the normal strain anticlockwise by 45 degrees. Strain is often characterised by strain rate and strain angle instead of normal and shear strain, which are based on the concept of radial and angular coordinates in polar coordinates.

Vorticity refers to the rotation of the flow, which can be either clockwise or anti-clockwise. Finally, divergence describes a flow that separates from a point (or 'diverges'), the opposite of which is convergence where different flows come together. Divergence could be considered to be similar to diffusivity as it causes particles to spread further apart from one another, however divergence causes correlated and deterministic spreading of particles whereas diffusivity refers to the uncorrelated and stochastic spread. Figure 2.3.1 displays simulated velocity fields for each of divergence, vorticity, normal and shear strain.

2.3.2 Diffusivity

Drifters remain at a constant depth near the surface of the water, and so they work well in measuring how particles at the surface disperse. The drifters are constrained to

move only in 2-dimensions, and so provide horizontal measurements of the dispersion, which can be thought of as the mean square separation from either a fixed starting point (absolute dispersion) or another particle (relative dispersion) (LaCasce and Bower, 2000). Oceanographers are interested in finding the diffusivity in different parts of the ocean, which is found by taking the time derivative of the dispersion—concepts which we shall explain mathematically shortly. Therefore the dispersion provides information about the separation of particles in space, and diffusivity is the *rate* in time at which this separation occurs.

Diffusivity is typically measured in units of metres squared per second. The size of diffusivity estimates from drifters can vary from $\mathcal{O}(0.1)\text{m}^2/\text{s}$ up to $\mathcal{O}(10^4)\text{m}^2/\text{s}$ depending on the region of the ocean, and how exactly diffusivity is estimated, and therefore different diffusivity estimates should be interpreted carefully. When diffusivity is estimated from clustered drifters, then background flow such as currents and mesoscale flow such as strain and vorticity can be estimated and removed from the drifter positions and velocities. Diffusivity estimates will therefore be much lower, as more of the flow has been explained by these other effects. When diffusivity is estimated from single drifters however, it might only be possible to remove the mean flow (or even this might not be known), as mesoscale flow cannot be easily estimated from single drifters. The size of the diffusivity estimates therefore depend on whether mean flow and/or mesoscale features have been estimated and removed from drifter data, with the diffusivity estimates generally being lower as more features are estimated and removed. We therefore can think of diffusivity as a way to describe the part of the flow that can't otherwise be explained.

We work with isotropic diffusivities, which are diffusivities at a set depth (at the depth of the drifter drogue) that are assumed to be the same in every direction (where the depth z is constant). Diffusivity can also be measured at a certain potential density, called an isopycnal diffusivity, which is often the case when diffusivity is estimated from dye release experiments. Because we are working with drifters that follow a consistent depth, we are unable to estimate the isopycnal nor diapycnal diffusivity, the latter of which is the diffusivity in the z axis.

Throughout this thesis we focus primarily on two definitions of diffusivity, submesoscale diffusivity and large-scale diffusivity. Submesoscale diffusivity can be measured from clustered drifters, and predicts the stochastic behaviour of the flow after we have removed background and mesoscale parts of the flow. An experiment designed to estimate submesoscale diffusivities will typically last only a few days and cover a small region of the ocean. Large-scale diffusivities are much larger than submesoscale diffusivities as they are caused by larger eddies and turbulent flow. They can be estimated from single drifters, and so frequently only the background mean flow is removed before estimation. To estimate large-scale diffusivities, a drifter experiment must cover larger regions over longer timescales in order to capture these larger eddies. The size of diffusivity estimates will depend both on which and how many other parts of the flow have been successfully resolved before estimating the diffusivity, and on the size of the eddies that cause the diffusivity. Because diffusivity is estimated from the residual after estimating other parts of the flow, our diffusivity estimates might also contain other unresolved effects such as internal waves. When providing estimates for diffusivity, it is important to state clearly what assumptions

are made and which parameters have been estimated prior to estimating diffusivity, as this affects the interpretation of the diffusivity estimate.

2.4 Existing Diffusivity Estimation Methodology

In this section we will introduce some existing methodology for estimating diffusivity, which we use to provide motivation for developing our own estimators in this thesis (both in Chapter 3 for submesoscale diffusivity and Chapter 6 for large-scale diffusivity). Current estimation methods provide vastly varying diffusivity estimates, which is in part due to the differing ways that diffusivity is defined.

The estimation of ocean circulation using Lagrangian observations has been studied increasingly in recent years. One of the earliest applications of diffusion to ocean dynamics was by Taylor (1922), with their equation for diffusivity forming the basis for many studies into diffusivity, including our work in Chapter 6. It defines diffusivity from a single particle to be proportional to the time derivative of the square of the displacement of a particle from its initial position. This definition has been summarised in review papers by LaCasce (2008) and Van Sebille et al. (2018), as well as in application papers such as Sallée et al. (2008) who estimate the diffusivity in the Southern Ocean.

To now define diffusivity more formally using the Taylor (1922) definition, consider the probability density function (pdf) $Q(x,t)$ which is the probability that a particle is in position $x(t)$ at time t. This pdf is defined independently of the starting position and time and so holds for a flow which is both spatially homogeneous and stationary,

meaning that the probability is invariant to shifts in position or time. The first moment of Q gives the mean displacement $E[\tilde{x}(t)] = \int \tilde{x}(t)Q(\tilde{x},t)d\tilde{x}$, and the second moment gives the dispersion $E[\tilde{x}(t)^2] = \int \tilde{x}(t)^2 Q(\tilde{x},t)d\tilde{x}$, where $\tilde{x}(t) = x(t) - E[x(t)]$ is the position with the centre of mass removed at time t. The time derivative of the dispersion gives the absolute diffusivity

$$\kappa(t) = \frac{1}{2}\frac{d}{dt}E[\tilde{x}(t)^2]. \tag{2.4.1}$$

This definition is equivalent to the integral of the velocity covariance (LaCasce, 2008; Zhurbas and Oh, 2003; Davis, 1991), which is frequently used to estimate diffusivity from single particles (e.g. Koszalka and LaCasce, 2010). The equivalence is shown in LaCasce (2008, Equation 10).

In an inhomogeneous (spatially-dependent) setting, diffusivity can be defined in terms of its dependence on position x as well as t, following the definition of Davis (1991) who defines the single particle diffusivity tensor as

$$\kappa_{\{jk\}}(\mathbf{x},t) = -\left\langle \tilde{v}_{\{j\}}(t_0|\mathbf{x},t_0)\tilde{d}_{\{k\}}(t_0 - t|\mathbf{x},t_0)\right\rangle, \tag{2.4.2}$$

where $\tilde{v}_{\{j\}} = v_{\{j\}} - E(v_{\{j\}})$, $\tilde{d}_{\{k\}} = d_{\{k\}} - E(d_{\{k\}})$ are the differences of the velocity and displacement from their means respectively, t_0 is some fixed time, and the notation $a(t|\mathbf{x},t_0)$ is the value of a at time t of a particle which passed through $\mathbf{x}$ at time t_0. The angle brackets refer to calculating an average over multiple particles, and j and k refer to dimensions such that diffusivity can vary in different directions and we can also compute a "cross-diffusivity" between dimensions when diffusivity is not expected to be independent in each dimension. Note that the definition in Equation

(2.4.2) computes backwards in time from t_0 such that diffusivity is computed from velocities and displacements prior to arriving at x at $t = t_0$.

In the rest of this section we will introduce different methods of estimating diffusivity from the literature which derive from Equations (2.4.1) and (2.4.2).

2.4.1 Clustering Drifter Trajectories

Diffusivity is not constant across the ocean, so in practice the ocean is typically split into spatial bins in which it is assumed that diffusivity is constant. The size of these bins is an area of interest to oceanographers, as the bins need to be large enough that the estimates are statistically significant and larger than the eddies containing the most energy, but small enough that estimates can be reasonably assumed to be constant across the bins. When diffusivity is estimated from residual velocities after mean flow has been removed, the bins also need to be small enough to resolve the mean flow. Mean velocity is often estimated by simply taking the average of all velocity observations in a spatial bin. Diffusivity, on the other hand, is estimated (whether directly or indirectly) as an autocorrelation between multiple observations (see Equation (2.4.1) and surrounding discussion), and so the same velocity observations yield fewer diffusivity observations which can then be averaged to obtain a diffusivity estimate for the bin. It could therefore be beneficial to estimate mean flow and diffusivity using different bin sizes. A variety of different bin sizes have been explored in the literature, for example Zhurbas and Oh (2004) and Swenson and Niiler (1996) use bins of size $5^\circ \times 5^\circ$ to estimate diffusivity, but use bins that are $2^\circ \times 2^\circ$ to estimate the average velocities.

Different regions of the ocean have different numbers of velocity observations from drifters. For example drifters near (but not on) the equator are faster moving and are more densely populated, resulting in a high number of velocity observations. Therefore, splitting the ocean into equally sized bins will result in some bins having more observations than others. Koszalka and LaCasce (2010); Koszalka et al. (2011) explore the use of different bin sizes in estimating the mean flow of simulated particle trajectories, and observe than bins that are $1^\circ \times 0.5^\circ$ are able to recover the major structures in the surface current. Bins of $2^\circ \times 1^\circ$ are able to show where currents are weaker or stronger, but lose the higher resolution structures, and when bins of $4^\circ \times 2^\circ$ are used the currents are difficult to observe.

Koszalka and LaCasce (2010); Koszalka et al. (2011) propose a method to choose the bin size where instead of the bin size being fixed, they fix the number of drifters in a bin. The drifters are organised into bins via clustering. This ensures that the standard error of the estimates are determined just by the velocities and physical and measurement uncertainty and not by the number of drifters in a bin. They also apply clustering to diffusivity estimation. They take time series of fixed lengths and calculate the velocity autocorrelation for each time series, allocating the position to be the midpoint of the trajectory. They then cluster the positions, and average the autocorrelation in each cluster. Diffusivity is then estimated as the integral of the average autocorrelation.

The clustering algorithm used by Koszalka and LaCasce (2010); Koszalka et al. (2011) is k-means clustering, which assigns all velocity observations into k clusters, $C = C_1, C_2, \cdots, C_k$, such that the sum of the squared distance between cluster mem-

bers, x_j, and the centre of the cluster, c_i is minimised,

$$\min \sum_{i=1}^{k} \sum_{x_j \in C_i} |x_j - c_i|^2. \tag{2.4.3}$$

The clustering is an iterative process as whenever the clusters change, so do the centres. In each iteration, each data point is assigned to the nearest cluster centre, and then the cluster centres are recalculated using the new cluster members. This process is repeated until the cluster assignments do not change. For the clustering, only the number of clusters, k, needs to be pre-specified. If there are a desired number of observations in a cluster, then k is chosen as the total number of observations divided by the desired number of observations in a cluster.

The clustering algorithm is applied to simulated drifter trajectories in Koszalka and LaCasce (2010), and to trajectories in the Nordic seas in Koszalka et al. (2011).

2.4.2 Principal Component Analysis

An alternative method to estimate diffusivity is described in Zhurbas and Oh (2003). It uses principal component analysis, which in two-dimensions can be thought of as drawing an ellipse around the data where each orthogonal axis of the ellipse represents a principal component (such that there are just two components in total for two-dimensional data). These axes are found by calculating the covariance matrix and then finding the eigenvalues and eigenvectors of the covariance matrix. The eigenvectors are normalised to become unit vectors, representing the principal (major) and secondary (minor) axes of the ellipse.

Zhurbas and Oh (2003) use the minor principal component of both the diffusivity tensor by Davis (1991), as given in Equation (2.4.2), and half the growth rate of the single particle dispersion, as given in Equation (2.4.1) to estimate diffusivity. In more detail, they first estimate diffusivity along each of the principal components from the diffusivity tensor defined by Davis (1991). They then separate the diffusivity tensor into the symmetric and anti-symmetric parts, denoted $\kappa^S_{\{jk\}}$ and $\kappa^A_{\{jk\}}$ respectively, by

$$\kappa^S_{\{xy\}} = \kappa^S_{\{yx\}} = \frac{\kappa_{\{xy\}} + \kappa_{\{yx\}}}{2}, \qquad \kappa^S_{\{xx\}} = \kappa_{\{xx\}}, \qquad \kappa^S_{\{yy\}} = \kappa_{\{yy\}}, \tag{2.4.4}$$

$$\kappa^A_{\{xy\}} = -\kappa^A_{\{yx\}} = \frac{\kappa_{\{xy\}} - \kappa_{\{yx\}}}{2}, \qquad \kappa^A_{\{xx\}} = \kappa^A_{\{yy\}} = 0. \tag{2.4.5}$$

The symmetric part, $\kappa^S_{\{xy\}}$ is decomposed into its principal axis, with major principal component κ_1, and minor principal component κ_2, such that $\kappa_1 > \kappa_2 > 0$.

Zhurbas and Oh (2003) also define diffusivity to be half the growth rate of the single particle dispersion $s_{\{jk\}}$ which is defined as

$$s_{\{jk\}}(t|x) = \left\langle d'_{\{j\}}(t_0 + t|x, t_0) d'_{\{k\}}(t_0 + t|x, t_0) \right\rangle. \tag{2.4.6}$$

They decompose $s_{\{jk\}}$ into the principal axis components, $s_1 \geq s_2 \geq 0$, and obtain diffusivity estimates by taking time derivatives,

$$\kappa^*_1 = \frac{1}{2}\frac{\partial s_1}{\partial t}, \qquad \kappa^*_2 = \frac{1}{2}\frac{\partial s_2}{\partial t}, \tag{2.4.7}$$

where κ^*_1 and κ^*_2 are the diffusivity estimates along the principal components, and the * refers to the estimates coming from the growth rate of the dispersion.

Diffusivity will usually increase for short timescales $t > 0$ and asymptote to a constant value, usually before 30 days (Oh et al., 2000). The long term diffusivities

for the minor principal components are therefore estimated via

$$\kappa_2(\infty) = \max_{0<t<30\text{days}}[\kappa_2(t)], \qquad \kappa_2^*(\infty) = \max_{0<t<30\text{days}}[\kappa_2^*(t)]. \tag{2.4.8}$$

Finally, they estimate diffusivity as the average between long-term diffusivity of the minor principal component estimated by each method,

$$\kappa = \frac{\kappa_2(\infty) + \kappa_2^*(\infty)}{2}. \tag{2.4.9}$$

Therefore, Zhurbas and Oh (2003) estimate the diffusivity using just the minor principal component. They apply their method to observed drifter velocities in the Pacific and Atlantic Oceans to produce maps of diffusivity estimates.

2.4.3 Fitting a Parametric Model to Drifter Velocities

Estimating diffusivity from either the half growth rate of the dispersion as in Equation (2.4.1) or from the diffusivity tensor given in Equation (2.4.2) require observations over long times for the diffusivity to converge to the long-term diffusivity. This isn't always possible as a drifter will move to another part of the ocean, where the diffusivity and mean flow are likely to be different.

Griffa et al. (1995) propose a method of estimating diffusivity which doesn't necessarily require a long time series. They fit a parametric model to the velocity observations, and therefore they assume that the velocity autocovariance will have a known shape, which can then be used to estimate diffusivity. Drifter velocities are often broken down into a mean flow estimate and a diffusivity estimate, and so it is intuitive to model a drifters position using the mean flow $E[u]$ and turbulent velocity

u''. Griffa et al. (1995) define their parametric model as

$$dx = udt = (E[u] + u'')dt \tag{2.4.10}$$

$$du'' = \Theta u'' dt + \sigma_I \sqrt{2\Theta} dW, \tag{2.4.11}$$

where dx is the displacement of a particle during time dt, u is the velocity, $\Theta = 1/t_{\text{turbulent}}$ where $t_{\text{turbulent}}$ is the turbulent time scale, σ_I is the standard deviation of the turbulent velocity, and dW is an increment from a Wiener process (also known as a Brownian motion). The stochastic differential equation of Equation (2.4.11) is known as an Ornstein-Uhlenbeck process, and has also been used by other authors to model Lagrangian velocities (e.g. Lilly et al., 2017). The timescale T refers to the time that this process is able to remember its initial turbulent velocity, as at each time step the process will gradually lose memory of this initial velocity.

Diffusivity is often estimated as the integral of the autocovariance, and the autocorrelation corresponding to Equation (2.4.11) is

$$R(\tau) = \frac{1}{\sigma_I^2} \langle u(t)u(t+\tau) \rangle = e^{-\Theta\tau}. \tag{2.4.12}$$

From the autocorrelation, the autocovariance can be calculated as $\sigma_I^2 R(\tau)$, and the diffusivity can be estimated.

The diffusivity of an Ornstein-Uhlenbeck process is therefore

$$\kappa(t) = \int_0^t \sigma_I^2 e^{-\Theta\tau} d\tau = \frac{\sigma_I^2}{\Theta} \left(1 - e^{-\Theta t}\right). \tag{2.4.13}$$

2.4.4 Discussion

The three diffusivity estimators introduced above are all based on either Taylor (1922) or Davis (1991), as are many other Lagrangian diffusivity estimators used in the

literature. Each of the three estimators will yield different estimates of diffusivity, as they are all built on different assumptions and try to minimise the error in a different way. Koszalka and LaCasce (2010) aim to control the error by restricting the number of drifters in each spatial bin to be spatially constant. Zhurbas and Oh (2003) propose an estimator which assumes spatially anisotropic flow, and estimate an isotropic diffusivity from the minor principal component. Finally, Griffa et al. (1995) estimate diffusivity using a parametric approach, and so the error comes from the error from fitting the model, as well as from any model misspecification in the assumptions made. Each of the different estimators therefore has its own strengths and weaknesses.

Clustering the drifters can lead to a higher spatial resolution of estimates for mean flow and diffusivity on average, compared with binning the observations geographically. It also ensures statistical significance of their estimates (under assumptions of stationarity within each bin), as the number of drifters in a cluster could be chosen. However, the amount of data available is not consistent across the ocean, and so the spatial size of a cluster could vary dramatically across the globe. In regions of the ocean where drifter observations are sparse, clustering observations will result in a poor spatial resolution as the clusters in these areas must cover larger spatial areas before they reach the required number of cluster members. In addition it may not be reasonable to assume stationarity and constant diffusivity levels for such large bins.

Zhurbas and Oh (2003) assume that any discrepancies between κ_x and κ_y are due to the mean flow, and that the diffusivity itself is isotropic. This assumption suggests that any differences between the minor and major principal components of diffusivity

are due to mean flow, and not differences in the diffusivity in different directions. In the case where this assumption is correct, and diffusivity is isotropic, then estimating diffusivity from the minor principal component will smooth out the variance from estimating the mean flow. However, in the case where the assumption is untrue, and diffusivity varies in the different axes, then estimating diffusivity from just the minor component will result in an underestimate of the diffusivity. The diffusivity may also be underestimated if the estimates have not reached the long-term diffusivity within 30 days, as this is the maximum number of days the long-term diffusivity is estimated from.

Estimating diffusivity from a parametric model means that diffusivity is estimated from a theoretical form of the autocovariance, rather than an estimated empirical autocovariance from observations or simulations. This means, regardless of the length of the time series, the diffusivity estimates will always represent the long-term diffusivity associated with the model. With infinite length data, the diffusivity estimate will have zero error for trajectories exactly fitting the parametric model, however no model is able to perfectly describe the complex nature of ocean circulation, and statistical error will always remain with finite data. There will therefore be error associated with fitting a model to the drifters, which will result in the diffusivity estimates having an error determined by how good a fit the parametric model is and how much data there is and how noisily it is observed. This approach to estimating diffusivity requires the model parameters to be estimated, as well as a suitable model be chosen to fit the drifter velocities. If the model were not to be a good fit to the trajectories, then this would result in diffusivity estimates with high errors.

The different estimators will all give different diffusivity estimates, even if they were applied to drifter trajectories in the same part of the ocean. This is due to the differences in the way that each method estimates diffusivity and the assumptions they are built on. Diffusivity will vary both spatially and temporally, but is generally assumed to be consistent at a given scale within a given spatiotemporal region. This highlights the need for more robust ways to estimate diffusivity, and in Chapters 3 and 6 we propose new diffusivity estimators for submesoscale and large-scale flow respectively, with the aim of trying to reduce and quantify sources of error from data sparsity and other flow components that are present.

2.5 Mesoscale and Submesoscale Flow Estimation

In this section we will introduce some existing methodologies for estimating mesoscale flow and submesoscale diffusivity. We also look at the LatMix experiment in more depth, including some of the estimates of strain rate and diffusivity from this study found in the literature. We use this to provide motivation for developing our own methodology, and applying it to the LatMix drifter observations later in this thesis.

Mesoscale flow is often estimated using tracer release experiments, as well as observations from Lagrangian instruments such as drifters. Drifters are drogued to a certain depth, and can therefore only move in two dimensions. Dye is not constrained to a set depth and will move in three dimensions, usually following the potential density. Therefore, drifters estimate statistics at a given depth, and tracers estimate statistics along an isopycnal. This will result in differing estimates from Lagrangian

instruments and dye releases.

2.5.1 Mesoscale Flow Estimation

Okubo and Ebbesmeyer (1976) provide a method to estimate the mesoscale flow components which involves taking a Taylor series of the drifter velocities around the "centre of mass" (i.e. the average location of the drifter positions), and then using least squares regression to estimate strain, vorticity and divergence.

Specifically, they approximate the drifter velocities to be decomposed as

$$\begin{bmatrix} u_k(t) \\ v_k(t) \end{bmatrix} = \begin{bmatrix} E[u(t)] \\ E[v(t)] \end{bmatrix} + \frac{1}{2} \begin{bmatrix} u_x(t) & u_y(t) \\ v_x(t) & v_y(t) \end{bmatrix} \begin{bmatrix} \tilde{x}_k(t) \\ \tilde{y}_k(t) \end{bmatrix} + \begin{bmatrix} u_k^{\mathrm{sm}}(t) \\ v_k^{\mathrm{sm}}(t) \end{bmatrix}, \quad (2.5.1)$$

where $\{u_k(t), v_k(t)\}$ are the drifter velocities at time t for drifter $k \in \{1, 2, \cdots, K\}$, $\{E[u(t)], E[v(t)]\}$ are the drifter velocities at the location of the centre of mass at time t, $\{u_x(t), u_y(t), v_x(t), v_y(t)\}$ are velocity gradients found by taking the partial derivative in x or y, and $\{u_k^{\mathrm{sm}}(t), v_k^{\mathrm{sm}}(t)\}$ are the residual velocities. Finally, $\{\tilde{x}_k(t), \tilde{y}_k(t)\}$ are the positions of drifter k at time t after removing the "centre of mass", which is the the average position of the drifters at time t. The Taylor series can be written in matrix notation at each time point t as

$$U = RA + E, \quad V = RB + F, \quad (2.5.2)$$

where the matrices are defined as follows:

$$U = \begin{bmatrix} u_1(t) \\ u_2(t) \\ . \\ . \\ u_K(t) \end{bmatrix}, \; V = \begin{bmatrix} v_1(t) \\ v_2(t) \\ . \\ . \\ v_K(t) \end{bmatrix}, \; R = \begin{bmatrix} 1 & \tilde{x}_1(t) & \tilde{y}_1(t) \\ 1 & \tilde{x}_2(t) & \tilde{y}_2(t)) \\ . & . & . \\ . & . & . \\ 1 & \tilde{x}_K(t) & \tilde{y}_K(t) \end{bmatrix},$$

$$A = \begin{bmatrix} E[u(t)] \\ u_x(t) \\ u_y(t) \end{bmatrix}, \quad B = \begin{bmatrix} E[v(t)] \\ v_x(t) \\ v_y(t) \end{bmatrix}, \quad E = \begin{bmatrix} u_1^{\mathrm{sm}}(t) \\ u_2^{\mathrm{sm}}(t) \\ . \\ . \\ u_K^{\mathrm{sm}}(t) \end{bmatrix}, \; F = \begin{bmatrix} v_1^{\mathrm{sm}}(t) \\ v_2^{\mathrm{sm}}(t) \\ . \\ . \\ v_K^{\mathrm{sm}}(t) \end{bmatrix}. \quad (2.5.3)$$

Okubo and Ebbesmeyer (1976) use least squares regression to approximate A and B as

$$A = (R'R)^{-1}R'U,$$

$$B = (R'R)^{-1}R'V, \quad (2.5.4)$$

and then estimate the residual velocities as

$$E = [1 - R(R'R)^{-1}R']U = U - RA,$$

$$F = [1 - R(R'R)^{-1}R']V = V - RV. \quad (2.5.5)$$

The mesoscale flow components can be estimated from the velocity gradients as

Divergence: $\delta(t) = u_x(t) + v_y(t)$, Vorticity: $\zeta(t) = v_x(t) - u_y(t)$,

Normal strain: $\sigma_n(t) = u_x(t) - v_y(t)$, Shear strain: $\sigma_s(t) = v_x(t) + u_y(t)$. (2.5.6)

2.5.2 Submesoscale Diffusivity Estimation

Isopycnal diffusivity is often estimated from a tracer using the advection-diffusion equation (e.g. Sundermeyer et al., 2020)

$$\frac{\partial \gamma}{\partial t} + \sigma x \frac{\partial \gamma}{\partial x} - \sigma y \frac{\partial \gamma}{\partial y} = \kappa \nabla^2 \gamma, \tag{2.5.7}$$

where γ is the dye concentration, σ is the non-divergent strain rate, and κ is the isopycnal diffusivity which is isotropic. It is assumed that the axes are aligned with the principal axis of the strain tensor, with x corresponding to the principal axis and y corresponding to the minor axis. To obtain second moments of the tracer in x and y, they multiply the advection-diffusion equation from Equation (2.5.7) by x^2 and y^2 respectively to obtain

$$\frac{dm_{xx}^2}{dt} - 2\sigma m_{xx}^2 = 2\kappa, \tag{2.5.8}$$

$$\frac{dm_{yy}^2}{dt} + 2\sigma m_{yy}^2 = 2\kappa, \tag{2.5.9}$$

where m_{xx}^2 and m_{yy}^2 are the second moments in x and y respectively. The second term of each equation has a different sign corresponding to stretching or convergence of the tracer patch in the major and minor axis respectively. The strain causes the second moment in the major (x) axis to grow exponentially, while diffusion grows linearly, whereas in the minor (y) axis the strain causes the tracer patch to converge, counteracting the diffusion causing the patch to grow.

To estimate diffusivity, the variances m_{xx}^2 and m_{yy}^2 and their rate of change with time can be calculated from the tracer patch, and all that is left is two equations and two unknowns, which the equations can be solved to find.

An alternative way to estimate diffusivity and strain rate jointly is to assume that after a time of $1/\sigma$, the strain rate dominates the diffusion in the x axis, but in the y axis, the strain rate and diffusion balance out to a steady streak width, simplifying Equations (2.5.8) and (2.5.9) to

$$\frac{dm_{xx}^2}{dt} - 2\gamma m_{xx}^2 = 0, \tag{2.5.10}$$

$$\gamma m_{yy}^2 = \kappa. \tag{2.5.11}$$

Therefore the strain rate can be estimated from the growth rate of the second moment in x, and diffusivity can then be estimate using the estimated strain rate as well as the second moment in y.

2.5.3 Earlier Work Separating Mesoscale and Submesoscale Components

Sundermeyer and Price (1998) compare statistics estimated from floats and tracer releases from the North Atlantic Tracer Release Experiment (NATRE). The aim of Sundermeyer and Price (1998) is to gain a better understanding of what causes mixing and stirring in the ocean. It is thought (Sundermeyer et al., 2020) that submesoscale processes transfer energy between the mesoscale and the microscale, which are the smallest scales. Sundermeyer and Price (1998) define mixing processes to be those which can be modelled by diffusion, being small-scale processes which cannot be resolved, and they define stirring to be larger scale events that can be resolved, such as streaking and folding of a tracer. Estimating the dispersion of a purely diffusive process has been more thoroughly studied, whereas estimating dispersion in the pres-

ence of strain is more complex and is therefore less well understood. Throughout this thesis we aim to develop a better understanding of techniques to estimate mesoscale and submesoscale flow, and the properties of the resulting estimates.

When a tracer is released it can be modelled in three distinct phases (Garrett, 1983). In the first phase, the tracer patch is much smaller than the straining eddies, and so the growth is modelled as a diffusive process. In the second phase, the tracer patch is large enough to be acted upon by the strain and is advected into long thin streaks which grow in length exponentially with time. When the patch enters its final stage, it is larger than the mesoscale eddies, with the streaks wrapping around each other due to stirring from the eddies. The patch becomes more homogeneous, and therefore can again be modelled as a diffusive process. Therefore, in order to estimate the mesoscale effects, the tracer must be in its second phase, where it is approximately the size of the mesoscale processes that are being estimated. The single particle diffusivity estimators that were discussed in Section 2.4 would be suitable for a tracer in the third phase of evolution.

Sundermeyer and Price (1998) used both tracer patch release experiments and float trajectories from NATRE to compare the statistics estimated from each. They also aimed to determine whether the three stage tracer model is consistent with observed dispersion rates from the NATRE experiment, and understanding the biases of following Garrett (1983) to estimate the strain rate, streak width, and small scale diffusivity.

They found that estimates from floats and the tracer closely agreed. The dispersion rates they found were consistent with an exponential growth phase, then a reduced

growth phase, and finally a linear regime, which was consistent with the three phase model by Garrett (1983). Estimating the small-scale diffusivity as in Garrett (1983), lead to the requirement of dividing estimates by a factor of 2 to obtain unbiased estimates. They believe that this bias is due to unidentified horizontal mixing at scales of 1-10km. This motivated the study of submesoscale processes, which was addressed in the LatMix Field Campaign.

2.5.4 Findings from the LatMix Experiment

The Lateral Mixing (LatMix) Experiment (Sundermeyer, 2017; Sundermeyer et al., 2020; Shcherbina et al., 2015) was carried out to gain a better understanding of stirring and mixing at the submesoscale. Shcherbina et al. (2015) provide an overview of the campaign, which included floats and tracer releases, and a suite of other observations, as well as providing an overview of ongoing research by many different authors using data from the many different sources.

It was hypothesised (Shcherbina et al., 2015) that stirring and mixing at the submesoscale could be due to any (or none) of the following:

- internal wave shear dispersion,
- stirring from vortical modes caused in internal waves breaking,
- straining of tracer fields to smaller scales by mesoscale processes,
- high Rossby number subinertial submesoscale motions.

Testing the above hypotheses required observations from a tracer with different levels

of mesoscale straining, and hence there were two major field campaigns with instruments deployed in areas with different straining levels. Throughout this thesis, we only look at the first of these two campaigns, which took place in the Sargasso Sea in summer 2011 with two separate experiments, one at a region with low straining [$\sim 10^{-6}$/s or $\sim 0.01 f_c$, where $f_c = 7.7 \times 10^{-5}$rad/s is the Coriolis frequency at 32°N], and the other at a region with moderate straining flow [$\sim 10^{-5}$/s or $\sim 0.1 f_c$]. In each site, three research vessels carried out an initial survey, and then dye was released, accompanied by a Lagrangian float within the patch (Shcherbina et al., 2015). Drifters and Electromagnetic Autonomous Profiling Explorer (EM-APEX) floats were also deployed to surround the dye patch. The vessels carried out surveys in the area at the centre of the dye, with four Slocum gliders and one SeaGlider supplementing these surveys. There were also eight smaller dye releases which were each tracked for about a day by towed instruments, and four of these were tracked for the first 6 hours by airborne lidar. Throughout this thesis, we are interested in estimation techniques for position observations obtained from the drifters, although some of our techniques might be applied to some of the other data as well.

The study at the site of weak straining (Site 1) took place between 2$^{\text{nd}}$ June until the 10$^{\text{th}}$ June 2011. On the 4$^{\text{th}}$ June the dye was released along a 1.4km streak with nine drifters drogued at 30m also released with their initial positions forming a cross shape centred on the dye streak. Over a period of six days, the instruments travelled 30km in the direction south-southeast, with the drifters being stretched along the northwest-southeast axis, and relatively little rotation. The dye stretched into an ellipse that was 12km ×5km. Shcherbina et al. (2015) provide estimates for the

diffusivity, with an isopycnal diffusivity of $1\text{m}^2/\text{s}$ estimated from the dye patch and a lateral diffusivity of $0.2\text{m}^2/\text{s}$ estimated from the drifters. They note that the difference between these diffusivities is likely to be due to the difference between dye following an isopycnal and drifters being drogued at a set height. Therefore the drifters cannot accurately estimate diffusivity at a consistent potential density.

Sundermeyer et al. (2020) estimated the strain rate to be $3 \times 10^{-6}/\text{s}$ from the dye and $0.1 - 2 \times 10^{-5}/\text{s}$ from the drifters. They produced three different estimates for the isopycnal diffusivity at Site 1 as they found that the streak width did not approach steady state during the 6 days of the experiment. They therefore estimated diffusivity from both Equations (2.5.8) and (2.5.9), as well as Equation (2.5.11). Their estimates ranged from $0.4\text{m}^2/\text{s}$ up to $0.7\text{m}^2/\text{s}$ from dye, and $2.8\text{m}^2/\text{s}$ from drifters. They note that deciding which estimate is the most correct depends on which assumptions are assumed to be true.

The moderate straining study (Site 2) took place between 12^{th} June until the 20^{th} June 2011. Dye was released on the 13^{th} June, this time along a 2km line. Drifters were deployed with the dye, and again had initial positions in the shape of a cross. This site showed much stronger stretching, elongating to over 50km long, with width approximately 3-5km (Sundermeyer et al., 2020). Shcherbina et al. (2015) state that despite the stronger straining of the dye, the isopycnal diffusivity estimate was approximately the same as that in the weak-straining site, with estimates of $1\text{m}^2/\text{s}$.

Sundermeyer et al. (2020) estimate the strain rate to be $1.7 \times 10^{-5}/\text{s}$ for days 0-2, decreasing to be $2.5 \times 10^{-6}/\text{s}$ in days 2-4 from dye, and estimates ranging from $3 \times 10^{-5}/\text{s}$ early in the evolution and $5 \times 10^{-5}/\text{s}$ in later days from drifters. They

again found that the second moment in the minor axis did not approach steady state, however they note that this is unsurprising due to the changing strain rate estimates. They again calculate different estimate for diffusivity, in this case ranging from $1.9\text{m}^2/\text{s}$ to $4\text{m}^2/\text{s}$.

Despite the difference in the estimates for strain rate at each site, the isopycnal diffusivity estimates are remarkably similar between the two sites. Neither Shcherbina et al. (2015) nor Sundermeyer et al. (2020) were able to rule out any of the four hypotheses of what lead to stirring and mixing at the submesoscale, and suggested that all four factors could possibly contribute, and no one process currently stands out to be driving dispersion. Therefore, understanding mixing and stirring at the submesoscale is still an area for ongoing research.

2.5.5 Discussion

Mesoscale and submesoscale flow can be estimated from many different Lagrangian instruments. In this section we have focused on estimation from dye releases and from drifters. The two methods are expected to give differing estimates due to differences in the way the data is collected. A tracer is free to move in all three axes, and will typically move through regions of constant potential density, known as following an isopycnal. On the other hand, drifters are drogued to a certain depth, and can only collect measurements at this depth. This means that if the density changes as a drifter moves through the ocean then the density at which the drifter is measuring will change. Sundermeyer et al. (2020) found a slight variation in their estimates of strain rate and diffusivity depending on whether they were estimating from drifter or

dye positions. Therefore we note that regardless of the instrument used to obtain an estimate, any parameter estimates are expected to vary at a different depth/density to that followed by the drifters/dye during the experiment.

The method for estimating strain, divergence and vorticity by Okubo and Ebbesmeyer (1976) provides an estimate for each parameter at every sampled time point, with the spatial gradients also being allowed to vary at every point in time. This could potentially result in estimates which are noisy and not statistically significant, as the model doesn't allow any smooth changes to ensure the error is sufficiently small to ensure significance. Furthermore, Okubo and Ebbesmeyer (1976) take a Taylor series around the centre of mass, meaning that information is lost from the mean velocities which have been removed, this could also increase the error of the estimates. We note however, that the method uses least squares regression which minimises the residual velocities, and so this method attempts to describe as much of the flow as possible as being either mean flow or mesoscale flow, resulting in the submesoscale diffusivities describing the unexplained part of the flow.

An alternative method to estimate strain rate is described in Sundermeyer et al. (2020), where strain rate and diffusivity are estimated simultaneously. This method draws similarities to the principal component analysis for estimating diffusivity discussed in Section 2.4.2, as both methods require the principal axis to be found, and diffusivity is then found primarily from the behaviour of the minor axis. This method is defined for a tracer, however drifters can be expected to behave similarly in their spreading, albeit with a less clear ellipse to determine the axes. This method is only able to estimate strain rate and diffusivity, and in the presence of vorticity and/or

divergence, the estimates are likely to be less accurate as the effects of vorticity/-divergence will be captured in the estimates of strain rate and diffusivity. Vorticity is unlikely to make significant changes to the second moments, however the ellipse will be rotated and the principal axis will move with the vorticity. Therefore, if the principal axis is able to follow the vorticity then the estimates should be largely unaffected, but if a single fixed principal axis is used then the estimates will inevitably be biased. Divergence could alias as either strain rate and/or diffusivity, and is likely to result in biased estimates if ignored. Therefore, the Sundermeyer et al. (2020) method should be used for strain only fields, or must be carefully modified to account for other mesoscale effects.

Shcherbina et al. (2015) and Sundermeyer et al. (2020) quote varying estimates for strain rate and divergence in each site of the LatMix experiment. It is not known which of these estimates more accurately describes the complex nature of the ocean dynamics in the Sargasso Sea at this point in time, and so the mesoscale and submesoscale processes driving the drifters in the LatMix experiment are still unknown. This highlights the requirement for more research into the estimation of both mesoscale and submesoscale flow, as well as more understanding of the properties and errors associated with each parameter estimate.

2.6 Conclusion

Throughout this chapter we have introduced some of the basic definitions from oceanography which will feature throughout this thesis. We introduced Lagrangian instru-

ments, with an emphasis on drifters which will be used to provide positions that will be used to gain an understanding of ocean dynamics. We are particularly interested in the estimation of mesoscale and submesoscale flow, as well as submesoscale and large-scale diffusivities. We have introduced some methodology used in the literature to estimate these processes, as well as discussing their limitations. We now will further discuss some of those methods and their limitations to motivate the work carried out in this thesis.

Okubo and Ebbesmeyer (1976) provide a relatively straightforward method for splitting drifter velocities into a background flow, as well as mesoscale and submesoscale components. In Chapter 3 we will develop a new way to estimate strain, vorticity and divergence, based on the method by Okubo and Ebbesmeyer (1976), but with some important modifications. Our model addresses the shortfalls of their method which we discussed in Section 2.5.5. Specifically, we will assume that the mesoscale processes are slowly changing in time, and we therefore will estimate each parameter over a given time window, instead of independently at each sampled time point. We will show how to choose this smoothing window to ensure that the estimates are less noisy and statistically significant. The window length can also be changed to reduce the variance of the estimates if the true parameters are known to remain constant over the experiment, for example in LatMix site 1, we saw from Sundermeyer et al. (2020) that the strain rate is approximately constant, and so our method is able to provide a single estimate for the strain rate across the entire time period. We will show that our estimator produces similar estimates to Sundermeyer et al. (2020) when applied to LatMix drifter trajectories, including the decreasing

strain rate in Site 2.

Fitting a parametric model to drifter trajectories, as in Griffa et al. (1995), may in practice result in large biases and errors for any resulting parameter estimates. This is because the ocean is complex, and no model is able to capture all the different ocean dynamics. However, parametric models can be used to provide useful summaries and a general idea of the magnitude and associated error of parameter estimates. This is because a parametric model allows us to develop estimators with known theoretical properties which are precise assuming the assumptions of the model are satisfied by the data. Such theoretical properties can be used to aid decision making in the design of drifter release experiments; for example in Chapter 5 we will fit an Ornstein-Uhlenbeck model to derive estimators for strain and diffusivity. We can use these estimators and their properties to draw conclusions about how drifters should be best deployed (in what number and in what configuration) to reduce the anticipated error of different resulting parameter estimates from the study. In Chapter 6 we use a parametric model to fine-tune our estimator to each individual drifter track to reduce the expected error of diffusivity estimates. By using a parametric model we obtain a closed form expression for the error (both bias and variance) allowing us to tune the tuning parameters of the estimator, however we do not explicitly use the model in the final estimation process, and therefore we do not directly encounter the error associated with a poorly fitting parametric model.

The diffusivity definition of Taylor (1922) will form the basis of our diffusivity estimation in Chapter 6, where we provide a new diffusivity estimator which attempts to reduce the error when compared with the standard estimator using the integral of

the autocovariance. Estimating diffusivity across the ocean in practice requires us to make decisions on how to split up the ocean into regions of approximately constant diffusivity. Koszalka and LaCasce (2010) use clustering to increase the resolution of their estimates, as well as controlling statistical significance from sample size. However, in areas with poor spatial coverage, some of their clusters could be large and in these regions they cannot predict how diffusivity changes with space which is a disadvantage. In Chapter 6 we introduce a method to estimate diffusivity that attempts to reduce the variance of diffusivity estimates from single drifters, and hence can produce meaningful estimates when aggregated across all drifters. This means that we can attempt to improve diffusivity estimates with a constant resolution across the globe, with a goal of ensuring reliable estimates even in regions with low data volumes.

This thesis builds upon some of the methods we have introduced in this chapter, with an emphasis on estimating parameters of flow components and their associated error. The goal of the subsequent chapters is to improve these methods and reduce statistical and physical sources of uncertainty and error to obtain flow estimates that are less biased and noisy and more statistically significant.

Chapter 3

Separating Mesoscale and Submesoscale Flows from Clustered Drifter Trajectories

The contents of this chapter is published in Oscroft S, Sykulski AM, Early JJ. Separating Mesoscale and Submesoscale Flows from Clustered Drifter Trajectories. Fluids. 2021; 6(1):14.

3.1 Introduction

Recent field experiments targeting submesoscale motions (100m–10km) include the deployment of dozens to hundreds of GPS tracked surface drifters in close proximity, e.g., 'LatMix' (Shcherbina et al., 2015), 'GLAD' (Poje et al., 2014), 'LASER' (Gonçalves et al., 2019) and 'CALYPSO' (Mahadevan et al., 2020). These deploy-

ments are designed to sample a narrow spatiotemporal window, but with high enough data density to resolve submesoscale motions. However, even when submesoscale motions are resolved, separating those motions from the larger, often more energetic mesoscale motions remains a significant challenge.

One approach to disentangling the submesoscales from the mesoscales with high resolution drifter data is to use the results from turbulence theory. For example, Poje et al. (2014) showed results using two-particle statistics consistent with local dispersion at submesoscales. Beron-Vera and LaCasce (2016) found ambiguous results until inertial oscillations were filtered from the trajectories. This suggests, not surprisingly, that realistic flow fields contain a combination of flow features that can be linearly separated in some contexts. In a detailed modelling study, Pearson et al. (2019) showed that, even with some filtering, these Lagrangian statistics are far more sensitive than similar Eulerian measures, and called into question the interpretation of previous studies that use variations of two-particle statistics.

An alternative approach is to parameterise the energetic mesoscale flow features from the Lagrangian trajectories, in order to disentangle them from the unparameterised, possibly submesoscale, flows. The notion of accounting for, or parameterising, the mesoscale strain in order to measure the submesoscale diffusivity, appears to originate with tracer release experiments (Sundermeyer and Price, 1998; Sundermeyer and Ledwell, 2001), and is based on ideas introduced in Garrett (1983). The basic idea is that one axis of the tracer grows exponentially with a rate proportional to the strain rate, σ, while the other axis reaches a steady state balanced by the compressing effect of σ and the elongating effect of diffusivity, κ. In the dye experiments, the

mesoscale strain rate is determined by measuring the rate of elongation of the patch, which is then used to deduce the diffusivity. The key idea to this approach is that the mesoscale strain rate is parameterised, in order to separate its effect from the submesoscale motions.

This manuscript extends the idea of parameterising mesoscale features, in order to disentangle submesoscale flow, to a more principled and robust framework appropriate for Lagrangian particles. Our work is complementary to, but distinct from, the recent works of Gonçalves et al. (2019); Lodise et al. (2020) who developed a method for projecting clustered drifter trajectories to reconstruct local *Eulerian* velocity fields using Gaussian Process regression. The goal of our work is to disentangle the trajectories in a *Lagrangian* sense, and explicitly separate each drifter trajectory into background, mesoscale and submesoscale components—where each decomposed drifter can then be analysed further within the Lagrangian framework. A key benefit is that our Lagrangian separation allows for the explicit estimation of submesoscale *diffusivity*, as we shall show.

The structure of this Chapter is as follows. In Section 3.2 we first introduce a conceptual Lagrangian flow model, and then show how this can be parameterised using a local Taylor expansion. Then in Section 3.3 we show how these parameters can be estimated from clustered drifter deployments. We pay particular focus to building a hierarchy of models, where each layer in the hierarchy adds extra parameters (e.g. strain/vorticity/divergence) that represent additional flow features. We provide novel methodology for selecting between hierarchies based on the evidence from the data. In Section 3.4 we go further and incorporate nonstationary flow features, by

allowing mesoscale parameters to slowly evolve over time. We provide methodology for estimating this evolution using splines, and then we provide techniques for quantifying the uncertainty of all parameter estimates using the bootstrap. We detail how this quantification of uncertainty provides the ideal mechanism from which to select the key parameter of the temporal window length. Throughout Sections 3.3 and 3.4 we perform detailed simulation analyses to provide further insight and motivation. Then in Section 3.5 we test and perform our novel methodologies on data collected from drifters in the LatMix deployment, which reveals new insights and discovers previously hidden mesoscale and submesoscale structures. Discussion and conclusions can be found in Section 3.6. We also perform a sensitivity analysis against the number of drifters, as well as the configuration of the initial deployment, in Chapter 4. Code to replicate all results and figures in this Chapter is available at `https://github.com/JeffreyEarly/GLOceanKit`.

Overall, the principle contribution of this Chapter is a general framework for analysing Lagrangian data from clustered drifter deployments. Specifically, this methodology provides a tool to detect for the presence of various mesoscale flow features and separate those features from the submesoscale flow—while allowing such features to evolve over time—together with providing quantified statistical uncertainty of output.

3.2 Modelling Framework

The primary conceptual model used throughout this manuscript is that the total velocity of a Lagrangian particle $\mathbf{u}^{\text{total}}$ can be decomposed into three components,

$$\mathbf{u}^{\text{total}} = \mathbf{u}^{\text{bg}} + \mathbf{u}^{\text{meso}} + \mathbf{u}^{\text{sm}}, \tag{3.2.1}$$

where $\mathbf{u}^{\text{bg}}$ is a large scale background flow, $\mathbf{u}^{\text{meso}}$ is the mesoscale flow (> 10 km, > 10 days) and $\mathbf{u}^{\text{sm}}$ is the submesoscale flow (100m–10km, 1 hr–10 days). The background flow is assumed to be spatially homogeneous in some local region around the drifters, and thus includes motions such as inertial oscillations and large scale currents. The terminology used here is appropriate for a range of oceanographic contexts, but arguably the separation into mesoscale and submesoscale are more precisely related to *non-local* and *local* dynamics, respectively. We thus use the term mesoscale to describe structures that behave *non-locally* across the drifters, and are therefore the smoothly varying fluid structures that will be parameterised, such as the constant strain rate used in the tracer release experiments (Sundermeyer and Ledwell, 2001). The submesoscale currents are simply the residual motion, not captured by the background or mesoscale flow. If any statistically significant submesoscale signal remains, its energy spectrum will likely be shallower than the mesoscale portion and therefore be consistent with local dynamics. In practice, the scales captured by these three types of motion will vary depending on the deployment details and the limitations of the data, as much as the actual physical processes themselves, as we shall show. The proposed methodology therefore ultimately remains agnostic to the scales and physical processes governing the motions, but instead focuses on the statistical significance of the model.

Surface drifter motion is constrained to a fixed depth near the ocean surface, where

the two-dimensional positions are measured in geographic coordinates longitude and latitude. For the work here it is necessary to use map coordinates $\{x(t), y(t)\}$ with a projection that locally preserves area and shape. Following Early and Sykulski (2020) we use the transverse Mercator projection with central meridian placed between the minimum and maximum longitude of the drifter experiment and add a false northing and easting to shift the origin to the southwest corner. The total velocity $\mathbf{u}^{\text{total}}$ of a drifter is then two-dimensional and assumed to represent the velocity at the depth of the drifter drogue. The work here will also be generally applicable to clustered deployments of RAFOS floats with minor modification, but we will use the terminology of drifters throughout the manuscript.

3.2.1 A local Taylor expansion

One of the simplest models for separating flow components is to perform a local Taylor expansion of the velocity field. Suppose we have observations from K clustered drifters at time t, where the position of drifter k ($1 \leq k \leq K$) in x and y orthogonal directions is given by $\{x_k(t), y_k(t)\}$, measured in metres, and the corresponding velocity is given by $\frac{d}{dt}\{x_k(t), y_k(t)\}$, measured in metres per second. We then take a Taylor series expansion of the velocity field evaluated at the position of drifter k, such that we model its velocity as

$$\underbrace{\frac{d}{dt}\begin{bmatrix} x_k(t) \\ y_k(t) \end{bmatrix}}_{\mathbf{u}^{\text{total}}} = \underbrace{\begin{bmatrix} u^{\text{bg}}(t) \\ v^{\text{bg}}(t) \end{bmatrix}}_{\mathbf{u}^{\text{bg}}} + \underbrace{\begin{bmatrix} u_0 + a_1 t \\ v_0 + b_1 t \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \sigma_n + \delta & \sigma_s - \zeta \\ \sigma_s + \zeta & \delta - \sigma_n \end{bmatrix}\begin{bmatrix} x_k(t) - x_0 \\ y_k(t) - y_0 \end{bmatrix}}_{\mathbf{u}^{\text{meso}}} + \underbrace{\begin{bmatrix} u_k^{\text{sm}}(t) \\ v_k^{\text{sm}}(t) \end{bmatrix}}_{\mathbf{u}^{\text{sm}}}, \tag{3.2.2}$$

where

- $\{x_k(t), y_k(t)\}$ are observations from drifter k at time t,
- $\{u^{\text{bg}}(t), v^{\text{bg}}(t)\}$ is the spatially homogeneous time-varying background flow,
- $\{u_0, v_0, a_1, b_1, \sigma_n, \sigma_s, \zeta, \delta\}$ are the model parameters for the mesoscale flow,
- $\{x_0, y_0\}$ is the expansion location and has no consequence to the model, other than redefining $\{u_0, v_0\}$,
- $\{u_k^{\text{sm}}(t),\ v_k^{\text{sm}}(t)\}$ are the residual 'submesoscale' velocities for each drifter, assumed to be zero-mean in time, but also zero-mean in space across drifters.

Equation (3.2.2) therefore separates background, mesoscale, and submesoscale features in the data, following the conceptual model of Equation (3.2.1). We will show in Section 3.3.1 how the eight mesoscale parameters can be simultaneously estimated, and in Section 3.3.2 how the flown can be decomposed into background, mesoscale and submesoscale components.

We write Equation (3.2.2) in terms of velocities instead of positions, even though the raw data usually measure positions of drifters. This is because we will be using time series techniques to fit the drifter trajectories to the model. Time series analysis

techniques often assume that the data are stationary, hence the use of velocities is desirable as they are more stationary than positions.

Equation (3.2.2) is similar to that given by Okubo and Ebbesmeyer (1976) which we gave in Equation (2.5.1), except with the addition of the $(u_0 + a_1 t, v_0 + b_1 t)^T$ term which comes from taking a more complete Taylor expansion. The terms u_0 and v_0 allow us to redefine the expansion point without changing the other model parameters, and a_1 and b_1 allow a small linear increase in the mesoscale component. In Equation (3.2.2) we rewrite the strain tensor of Equation (2.5.1) in terms of our mesoscale parameters.

The mesoscale parameters are simply re-definitions of the standard spatial gradients: the divergence is $\delta = u_x + v_y$, the vorticity is $\zeta = v_x - u_y$, the normal strain rate is $\sigma_n = u_x - v_y$, and the shear strain rate is $\sigma_s = v_x + u_y$. The normal and shear strain rates can be combined to a scalar value for the strain rate $\sigma = \sqrt{\sigma_n^2 + \sigma_s^2}$ and rotation angle $\theta = \arctan\left[\sigma_s/\sigma_n\right]/2$, where $\sigma_n = \sigma\cos(2\theta)$, $\sigma_s = \sigma\sin(2\theta)$.

These mesoscale parameters can be rewritten to make the spatial gradients the subject and then directly plugged into Equation (2.5.1). Therefore, for $u_0 = a_1 = v_0 = b_1 = 0$, Equations (2.5.1) and (3.2.2) are equivalent. We choose to write Equation (3.2.2) in terms of the mesoscale parameters as these are what we wish to use the model to estimate. The spatial gradients have no physical meaning in our model, except for their relationship to the mesoscale parameters.

For the moment, the eight mesoscale parameters are assumed to be sufficiently slowly varying that we can treat them as constant over some time window, although we will relax this restriction later. In practice, the mesoscale component of the model will

capture *any* coherent feature that has constant spatial gradient across the cluster of drifters, whether that is a large scale more permanent feature like a Western boundary current or a transient mesoscale eddy—or nothing at all. The spatially homogeneous time-varying 'background' flow will capture inertial and tidal oscillations, but may also erroneously include parts of a time or spatially varying mesoscale flow. Finally, the residual 'submesoscale' velocity will include any velocity contributions not captured by the other components.

The model of Equation (3.2.2) was applied to drifter observations in Okubo and Ebbesmeyer (1976) to obtain estimates of the spatial gradient parameters, but with two key differences from the approach taken here. First, the spatial gradients were allowed to vary at each observational time point, without any constraints on the rate of fluctuation. Second, the expansion point $\{x_0, y_0\}$ was chosen to be the time-varying centre-of-mass of the cluster of drifters. The consequence of this choice is quite significant and is worth considering in more detail. Defining the centre-of-mass (or first moment) as $m_x(t) \equiv \frac{1}{K}\sum_{k=1}^{K} x_k(t)$ and $m_y(t) \equiv \frac{1}{K}\sum_{k=1}^{K} y_k(t)$, it follows from Equation (3.2.2) that the centre-of-mass velocity includes contributions from both the homogeneous background as well as the spatial gradients such that

$$\frac{d}{dt}\begin{bmatrix} m_x(t) \\ m_y(t) \end{bmatrix} = \begin{bmatrix} u^{\mathrm{bg}}(t) \\ v^{\mathrm{bg}}(t) \end{bmatrix} + \begin{bmatrix} u_0 + a_1 t \\ v_0 + b_1 t \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \sigma_n + \delta & \sigma_s - \zeta \\ \sigma_s + \zeta & \delta - \sigma_n \end{bmatrix} \begin{bmatrix} m_x(t) - x_0 \\ m_y(t) - y_0 \end{bmatrix}, \quad (3.2.3)$$

where no submesoscale is assumed to be present as we have defined $\frac{1}{K}\sum_{k=1}^{K} u_k^{\mathrm{sm}}(t) = 0$. That the mesoscale spatial gradients have a (potentially) significant impact on the velocity of the centre-of-mass is evident in the top row of simulated drifter trajectories

shown in Figure 3.2.1, where the entire cluster of drifters is advected by the linear flow. Now if the expansion point is taken to be the centre-of-mass, $\{x_0(t), y_0(t)\} = \{m_x(t), m_y(t)\}$, then Equation (3.2.3) reduces the background velocity to the sample mean velocity, such that $u^{\mathrm{bg}}(t) \approx \frac{d}{dt} m_x(t)$. As a result, after subtracting Equation (3.2.3) from (3.2.2), the velocities of the individual particles in the centre-of-mass frame,

$$\frac{d}{dt}\begin{bmatrix} x_k(t) - m_x(t) \\ y_k(t) - m_y(t) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sigma_n + \delta & \sigma_s - \zeta \\ \sigma_s + \zeta & \delta - \sigma_n \end{bmatrix}\begin{bmatrix} x_k(t) - m_x(t) \\ y_k(t) - m_y(t) \end{bmatrix} + \begin{bmatrix} u_k^{\mathrm{sm}}(t) \\ v_k^{\mathrm{sm}}(t) \end{bmatrix}, \quad (3.2.4)$$

only depend on the spatial gradients and submesoscale flow. In some sense, the difference between Equation (3.2.2) and Equation (3.2.4) is quite remarkable: simply by changing to centre-of-mass coordinates, the potentially complicated form of the background flow, $\{u^{\mathrm{bg}}, v^{\mathrm{bg}}\}$, is eliminated, along with all the velocity variance associated with mesoscale advection of the centre-of-mass from Equation (3.2.3). With this choice of reference frame, the spatial gradients in the model now only characterise the *spreading* of particles, i.e. the second moment, as shown in the second row of Figure 3.2.1, along with any spreading caused by the submesoscale process.

3.2.2 Diffusivity

A key measure with which we evaluate our techniques is to measure the diffusivity of observed and modelled velocities. We define the submesoscale diffusivity for each drifter k as in Equation (21) of LaCasce (2008), such that

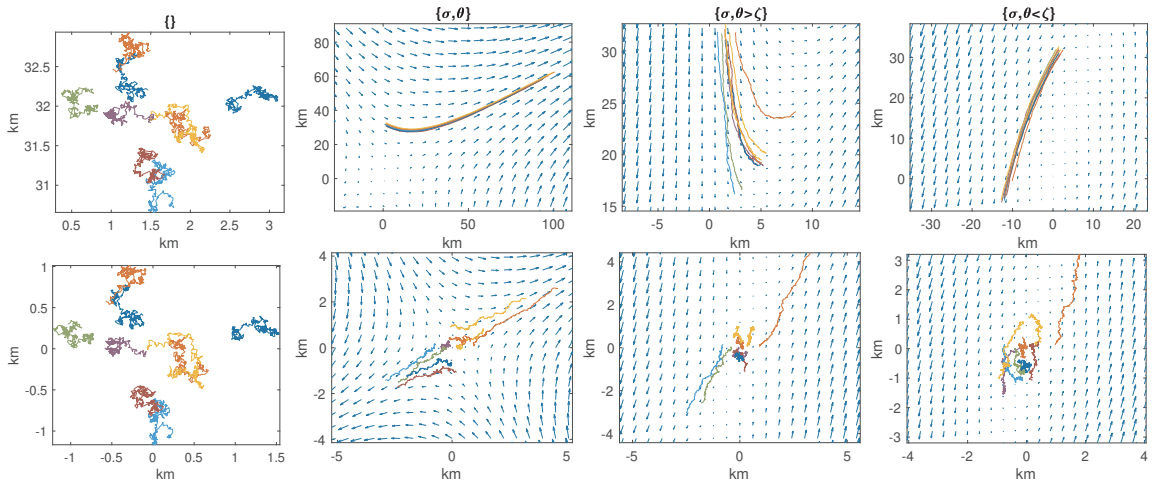

Figure 3.2.1: Simulation of 9 drifters from Equation (3.2.2) over 6.25 days, with starting positions, number of drifters, and experiment length taken to match LatMix Site 1. In each panel the submesoscale velocities $\{u_k^{\mathrm{sm}}(t), v_k^{\mathrm{sm}}(t)\}$ follow a Wiener increment process with diffusivity equal to 0.1 m^2/s. The top row shows drifter positions, and the bottom row shows positions with respect to centre-of-mass at each time step. From left to right we include the following model components. Left: diffusivity only. Centre left: strain and diffusivity. Centre right: strain, vorticity, and diffusivity (strain dominated). Right: strain, vorticity, and diffusivity (vorticity dominated). In each plot where a parameter is present, it has been set as $\sigma = 7 \times 10^{-6}/\mathrm{s}$, $\theta = 30°$, $\zeta = 6 \times 10^{-6}/\mathrm{s}$ (centre right), and $\zeta = 8 \times 10^{-6}/\mathrm{s}$ (right). We have set $u_0 = v_0 = a_1 = b_1 = u^{\mathrm{bg}} = v^{\mathrm{bg}} = 0$. The trajectories are simulated using the Euler-Maruyama scheme (Kloeden and Platen, 2013) and we include quivers in all plots representing the underlying velocity field.

$$\kappa_{k,x}^{\mathrm{sm}}(t) = \frac{1}{2}\frac{d}{dt}x_k^{\mathrm{sm}}(t)^2 = \int_0^t u_k^{\mathrm{sm}}(t)u_k^{\mathrm{sm}}(\tau)d\tau, \tag{3.2.5a}$$

$$\kappa_{k,y}^{\mathrm{sm}}(t) = \frac{1}{2}\frac{d}{dt}y_k^{\mathrm{sm}}(t)^2 = \int_0^t v_k^{\mathrm{sm}}(t)v_k^{\mathrm{sm}}(\tau)d\tau, \tag{3.2.5b}$$

where $x_k^{\mathrm{sm}}(t)$ is calculated from the residual velocities, $u_k^{\mathrm{sm}}(t)$, such that $x_k^{\mathrm{sm}}(t) = \int_0^t u_k^{\mathrm{sm}}(t)dt$, and similarly for $y_k^{\mathrm{sm}}(t)$. As in Equation (10) of LaCasce (2008), a joint diffusivity measure across all drifters could be defined by averaging the positions/velocities before applying the derivatives/integrals in Equations (3.2.5a) and (3.2.5b); however, we initially choose to calculate diffusivity separately for each drifter k to reflect the fact that drifters are spatially spread in a clustered deployment, and hence their diffusivity values may depend on spatial scale within a spatially inhomogeneous flow field.

In general, it is also useful to consider the isotropic diffusivity as this is rotationally invariant, and as such, does not depend on our choice of coordinate system. The isotropic submesoscale diffusivity for drifter k is defined as

$$\kappa_{k,z}^{\mathrm{sm}}(t) = \frac{1}{4}\frac{d}{dt}|z_k^{\mathrm{sm}}(t)|^2 = \frac{1}{2}\int_0^t w_k^{\mathrm{sm}}(t)w_k^{\mathrm{sm}}(\tau)d\tau, \tag{3.2.6}$$

where $z_k^{\mathrm{sm}}(t) = x_k^{\mathrm{sm}}(t) + \mathrm{i}y_k^{\mathrm{sm}}(t)$, $w_k^{\mathrm{sm}}(t) = u_k^{\mathrm{sm}}(t) + \mathrm{i}v_k^{\mathrm{sm}}(t)$, and $\mathrm{i} \equiv \sqrt{-1}$. The isotropic diffusivity is the average of $\kappa_{k,x}^{\mathrm{sm}}(t)$ and $\kappa_{k,y}^{\mathrm{sm}}(t)$ such that $\kappa_{k,z}^{\mathrm{sm}}(t) = \frac{1}{2}\{\kappa_{k,x}^{\mathrm{sm}}(t) + \kappa_{k,y}^{\mathrm{sm}}(t)\}$.

The diffusivity is also related to the power spectral density of complex velocity $w_k(t)$ where

$$S(\omega) \equiv \frac{1}{T}\left|\int_0^T w_k(t)e^{-\mathrm{i}\omega t}dt\right|^2. \tag{3.2.7}$$

$S(\omega)$ is known as the Lagrangian frequency spectrum and is related the isotropic diffusivity in Equation (3.2.6) with

$$\kappa_{k,z}(T) = \frac{1}{4}S(0), \tag{3.2.8}$$

as shown in Lilly et al. (2017). Formally, diffusivity requires the process to be stationary and is defined in the limit as $T \to \infty$, but in practice we are always limited to finite observation times. The total variance of a complex particle velocity is conserved with the Lagrangian frequency spectrum, $\frac{1}{T}\int w_k(t)^2 dt = \int S(\omega)d\omega$, and in this sense it will be useful to think of how the model components in Equation (3.2.2) each describe the distribution of variance in the frequency spectrum.

Equations (3.2.5), (3.2.6) and (3.2.7) are theoretical constructs as they require submesoscale velocities to be observed continuously in time. In practice, drifter observations are only observed at discrete time points. In Section 3.3 we will discuss how to *estimate* submesoscale diffusivity from clustered drifter data using our modelling and estimation approach.

We note that diffusivities could also be calculated directly from raw velocities $\{\frac{d}{dt}x_k(t), \frac{d}{dt}y_k(t)\}$, or from centre-of-mass velocities that have only had the background removed and still contain mesoscale flow contribution (as in Equation (3.2.4)), and such values of diffusivity will in general be much larger than the submesoscale diffusivities. This highlights the scale-dependent nature of diffusivity, as well as the challenges in comparing different measurements of diffusivity.

3.2.3 Model solutions

The mesoscale component of Equation (3.2.2) is a linear ordinary differential equation with tractable analytical solutions (e.g. Haynes, 2001; Lilly, 2018). However, the submesoscale component of Equation (3.2.2) is assumed unknown, and may represent a range of different phenomena. Thus, for our simulation analyses that follow in this chapter we generate the submesoscale process stochastically using trajectory paths defined by

$$\frac{d}{dt}\begin{bmatrix} x_k(t) \\ y_k(t) \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \sigma_n + \delta & \sigma_s - \zeta \\ \sigma_s + \zeta & \delta - \sigma_n \end{bmatrix}\begin{bmatrix} x_k(t) - x_0 \\ y_k(t) - y_0 \end{bmatrix} + \sqrt{2\kappa}\mathbf{dW}_t, \tag{3.2.9}$$

where the function **dW** represents an increment of a two-dimensional Wiener process (a random walk in the discrete-time limit) that forms the submesoscale component. The Lagrangian frequency spectrum of the submesoscale process is therefore simply that of a white noise process:

$$S(\omega) = 4\kappa. \tag{3.2.10}$$

The frequency spectrum of internal waves (perhaps the best known submesoscale process) will have either more or less contribution to the total variance, depending on the frequency. We thus consider a white noise velocity process to be a reasonably agnostic choice. Notably absent from Equation (3.2.9) is the spatially homogeneous background flow. In practice this contains a significant amount of power from inertial and tidal oscillations, but does not significantly impact the estimation of mesoscale quantities as we shall show. The particle trajectories shown in Figure 3.2.1 are sampled

from Equation (3.2.9), where each column contains different choices for the mesoscale parameters, but the submesoscale diffusivity κ is held constant (the first column has no mesoscale and hence the particles follow a random walk).

In the absence of the stochastic submesoscale white noise process, the Lagrangian trajectories from Equation (3.2.9) are purely deterministic and thus their Lagrangian frequency spectra can be computed exactly, as we shall now show. For the following analytical solutions we set $\delta = 0$, but make no such assumption in the estimation procedure that follows. To integrate Equation (3.2.9) with $\kappa = 0$, note that simply re-positioning a particle's initial location can be used to redefine $\{u_0, v_0\}$. Specifically, if the initial position of the particle is given by $\{x(0), y(0)\}$ with nonzero $\{u_0, v_0\}$, the $\{u_0, v_0\}$ can be set to zero, so long as the initial position is set to $\{x(0) - x_u, y(0) - y_u\}$ where

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \frac{2}{s^2} \begin{bmatrix} \sigma_n & \sigma_s - \zeta \\ \sigma_s + \zeta & -\sigma_n \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}, \tag{3.2.11}$$

and the Okubo-Weiss parameter is defined by $s^2 \equiv \sigma^2 - \zeta^2$. Thus, without loss of generality, we can simply take $\{u_0, v_0\}$ and the expansion point to be zero. The complex path $z(t) = x(t) + \mathrm{i}y(t)$ with initial position given by $\{x(0), y(0)\} = \{r\cos\alpha, r\sin\alpha\}$ is therefore

$$z(t) = \begin{cases} \frac{r}{s} e^{\mathrm{i}\alpha} \left(s \cosh\left(\frac{st}{2}\right) + \left(\sigma e^{\mathrm{i}2(\theta-\alpha)} + \mathrm{i}\zeta\right) \sinh\left(\frac{st}{2}\right)\right) & \text{if } \sigma^2 > \zeta^2 \\ \frac{r}{\bar{s}} e^{\mathrm{i}\alpha} \left(\bar{s} \cos\left(\frac{\bar{s}t}{2}\right) + \left(\sigma e^{\mathrm{i}2(\theta-\alpha)} + \mathrm{i}\zeta\right) \sin\left(\frac{\bar{s}t}{2}\right)\right) & \text{if } \sigma^2 < \zeta^2 \end{cases} \tag{3.2.12}$$

and the associated velocity $w(t) = u(t) + \mathrm{i}v(t)$ is given by

$$w(t) = \begin{cases} \frac{r}{2}e^{\mathrm{i}\alpha}\left(s\sinh\left(\frac{st}{2}\right) + \left(\sigma e^{\mathrm{i}2(\theta-\alpha)} + \mathrm{i}\zeta\right)\cosh\left(\frac{st}{2}\right)\right) & \text{if } \sigma^2 > \zeta^2 \\ \frac{r}{2}e^{\mathrm{i}\alpha}\left(-\bar{s}\sin\left(\frac{\bar{s}t}{2}\right) + \left(\sigma e^{\mathrm{i}2(\theta-\alpha)} + \mathrm{i}\zeta\right)\cos\left(\frac{\bar{s}t}{2}\right)\right) & \text{if } \sigma^2 < \zeta^2 \end{cases} \tag{3.2.13}$$

where we have defined the complementary Okubo-Weiss parameter by $\bar{s}^2 \equiv \zeta^2 - \sigma^2$. The mean-square distance of a particle from the origin is given by

$$\frac{1}{T}\int_0^T |z(t)|^2\, dt = \begin{cases} \frac{2r^2}{Ts^3}\sinh\left(\frac{sT}{2}\right)\left[\sigma A\cosh\left(\frac{sT}{2}\right) + sB\sinh\left(\frac{sT}{2}\right)\right] - \frac{r^2}{s^2}\zeta C & \text{if } \sigma^2 > \zeta^2 \\ \frac{2r^2}{T\bar{s}^3}\sin\left(\frac{\bar{s}T}{2}\right)\left[-\sigma A\cos\left(\frac{\bar{s}T}{2}\right) + \bar{s}B\sin\left(\frac{\bar{s}T}{2}\right)\right] + \frac{r^2}{T\bar{s}^2}\zeta C & \text{if } \sigma^2 < \zeta^2 \end{cases} \tag{3.2.14}$$

and total velocity variance,

$$\frac{1}{T}\int_0^T |w(t)|^2\, dt = \begin{cases} \frac{r^2}{2sT}\sinh\left(\frac{s}{2}T\right)\left[\sigma A\cosh\left(\frac{s}{2}T\right) + sB\sinh\left(\frac{s}{2}T\right)\right] + \frac{r^2\zeta C}{4} & \text{if } \sigma^2 > \zeta^2 \\ \frac{r^2}{2\bar{s}T}\sin\left(\frac{\bar{s}}{2}T\right)\left[\sigma A\cos\left(\frac{\bar{s}}{2}T\right) - \bar{s}B\sin\left(\frac{\bar{s}}{2}T\right)\right] + \frac{r^2\zeta C}{4} & \text{if } \sigma^2 < \zeta^2 \end{cases} \tag{3.2.15}$$

where

$$A = \sigma + \zeta\sin 2\left(\theta-\alpha\right),\ B = \sigma\cos 2\left(\theta-\alpha\right),\ C = \zeta + \sigma\sin 2(\theta-\alpha), \tag{3.2.16}$$

and T is the length of time that has passed since the particle has moved from its initial position.

The Lagrangian frequency spectrum of a particle in a linear velocity field can now be computed using Equations (3.2.13) and (3.2.7) which yields

$$S(\omega) = \begin{cases} \frac{r^2}{T}\sinh^2\left(\frac{sT}{4}\right)\left[\frac{\sigma A\cosh\left(\frac{s}{2}T\right)+sB\sinh\left(\frac{s}{2}T\right)-\zeta C}{\omega^2+\frac{s^2}{4}}+\frac{s^2C(\omega+\zeta/2)}{\left(\omega^2+\frac{s^2}{4}\right)^2}\right] & \text{if } \sigma^2>\zeta^2 \\ \frac{r^2}{T}\sin^2\left(\frac{\bar{s}T}{4}\right)\left[\frac{-\sigma A\cos\frac{\bar{s}T}{2}+\bar{s}B\sin\frac{\bar{s}T}{2}+\zeta C}{\omega^2-\frac{\bar{s}^2}{4}}+\frac{\bar{s}^2C(\omega+\zeta/2)}{\left(\omega^2-\frac{\bar{s}^2}{4}\right)^2}\right] & \text{if } \sigma^2<\zeta^2 \end{cases} \tag{3.2.17}$$

where the Lagrangian frequency spectra of complex-valued velocities are permitted to be asymmetric in ω (see Sykulski et al. (2016)), which will occur in Equation (3.2.17) when $\zeta \neq 0$. Asymmetric spectra arise when the rotary spectra are unequal and there is a preferred direction of spin (Sykulski et al., 2017b). With no strain and after sufficiently long observation time ($T >> 1/\zeta$), Equation (3.2.17) becomes a single frequency delta function, reflecting the rotation of a particle from the vorticity. However, for the cases considered here, observation times are at most $O(1/s, 1/\bar{s})$, and often much less. The result is a spectrum that is generally very red ($S(\omega) \sim \omega^{-2}$), with total power increasing in observation time T.

The Lagrangian frequency spectrum in Equation (3.2.17) would appear to indicate that particles advected by a linear velocity field have a non-zero diffusivity, following the definition of Equation (3.2.8). However, while it is true that the linear velocity field causes particles to disperse, increasing their second moment with T, this spreading is entirely deterministic with correlations between particles spatially and across time, and thus does not formally meet the requirement that diffusivity results from a stationary random velocity process. From the perspective of trying to isolate and estimate the diffusivity of submesoscale processes (which may be stationary at these scales), the linear velocity field may be viewed as contaminating the lowest frequencies in the spectrum, providing erroneously high values of diffusivity if not removed

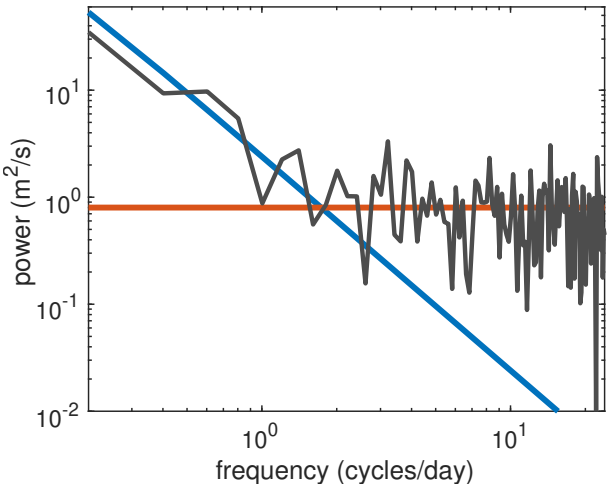

Figure 3.2.2: The one-sided frequency spectrum for a particle integrated with Equation (3.2.9) is shown in black. The particle is initially placed at $\{x(0), y(0)\} = \{1 \text{ km}, 1 \text{ km}\}$ and integrated for 5 days in a strain-only model with simulation parameters set to $\kappa = 0.1$ m^2/s and $\sigma = 1 \times 10^{-5}$/s. The theoretical spectrum of the mesoscale process, Equation (3.2.17), is shown in blue, and the theoretical spectrum of the white noise process, Equation (3.2.10), is shown in red.

correctly.

Figure 3.2.2 shows the one-sided Lagrangian frequency spectrum of a single particle simulated using Equation (3.2.9). The Lagrangian frequency spectrum thus has two distinguishing parts: the white noise submesoscale process given by Equation (3.2.10) and the deterministic red process given by Equation (3.2.17). In Figure 3.2.2 the observed particle spectrum is very nearly the linear addition of the theoretical Lagrangian frequency spectra of the mesoscale and submesoscale models of Equations (3.2.10) and (3.2.17) respectively. In terms of Figure 3.2.2, the objective of the methodology is to remove the deterministic contribution of the mesoscale flow (in

blue), in order to study the submesoscale process that remains.

3.3 Estimation and Hierarchical Modelling

The spreading of particles in the ocean can be categorised into three distinct stages of diffusivity according to the size of the drifter separation (or the tracer patch) relative to the size of mesoscale features (Sundermeyer and Price, 1998). At the smallest spatial scales, the mesoscale features may be so weak that the submesoscale processes dominate across all resolved scales and therefore completely control the spreading (e.g., when the mesoscale spectrum in Figure 3.2.2 is below the submesoscale spectrum). At the other extreme, where drifters are separated by distances that exceed the size of mesoscale features such as with the Global Drifter Program, the motions between any two drifters are uncorrelated and there are no common features to parameterise. We are interested in the middle stage, where the spread of the drifters is within the size of the mesoscale features. The upper bound of separation is dictated by the requirement that the spatial gradients in Equation (3.2.2) must be similar between drifters, while the lower bound is simply determined by lack of statistical significance of the mesoscale parameters. We place no upper bound on the number of drifters required, however there should be at least two drifters to remove the background part of the flow. The drifters should be sampled frequently enough that there is enough data to obtain estimates which are statistically significant whilst keeping the spread of the drifters within the mesoscale. Further discussion of how to ensure significance of results will be given in Section 3.4 and Chapter 4.

3.3.1 Parameter estimation

Estimates for the mesoscale parameters in Equation (3.2.2) from observations will be obtained using least squares regression, by minimising the sum of the squared residuals representing the non-mesoscale flow, as we shall now show. This approach therefore fits as much of the data to the mesoscale part of the model as possible. To perform the fits, we make the important step of decomposing the K drifter velocities into K drifter velocities relative to the centre-of-mass, plus a centre-of-mass velocity, as represented in Equations (3.2.3) and (3.2.4) respectively. In other words the summation of Equations (3.2.3) and (3.2.4) recovers Equation (3.2.2). When put into matrix-vector notation for observations these models can be jointly written as

$$U = XA + \epsilon, \tag{3.3.1}$$

where we have defined

$$U = \underbrace{\frac{d}{dt}\begin{bmatrix} \tilde{x}_k(t_n) \\ \tilde{y}_k(t_n) \\ m_x(t_n) \\ m_y(t_n) \end{bmatrix}}_{2(K+1)N\times 1}, \quad \epsilon = \underbrace{\begin{bmatrix} u_k^{\mathrm{sm}}(t_n) \\ v_k^{\mathrm{sm}}(t_n) \\ u^{\mathrm{bg}}(t_n) \\ v^{\mathrm{bg}}(t_n) \end{bmatrix}}_{2(K+1)N\times 1}, \tag{3.3.2}$$

and

$$X = \underbrace{\frac{1}{2}\begin{bmatrix} \mathbf{0}_{KN} & \mathbf{0}_{KN} & \mathbf{0}_{KN} & \mathbf{0}_{KN} & \tilde{x}_k(t_n) & \tilde{y}_k(t_n) & -\tilde{y}_k(t_n) & \tilde{x}_k(t_n) \\ \mathbf{0}_{KN} & \mathbf{0}_{KN} & \mathbf{0}_{KN} & \mathbf{0}_{KN} & -\tilde{y}_k(t_n) & \tilde{x}_k(t_n) & \tilde{x}_k(t_n) & \tilde{y}_k(t_n) \\ 2\cdot\mathbf{1}_N & \mathbf{0}_N & 2t_n & \mathbf{0}_N & \bar{m}_x(t_n) & \bar{m}_y(t_n) & -\bar{m}_y(t_n) & \bar{m}_x(t_n) \\ \mathbf{0}_N & 2\cdot\mathbf{1}_N & \mathbf{0}_N & 2t_n & -\bar{m}_y(t_n) & \bar{m}_x(t_n) & \bar{m}_x(t_n) & \bar{m}_y(t_n) \end{bmatrix}}_{2(K+1)N\times p}, \; A = \underbrace{\begin{bmatrix} u_0 \\ v_0 \\ a_1 \\ b_1 \\ \sigma_n \\ \sigma_s \\ \zeta \\ \delta \end{bmatrix}}_{p\times 1}. \tag{3.3.3}$$

In this notation, $\tilde{x}_k(t_n) \equiv x_k(t_n) - m_x(t_n)$, $\tilde{y}_k(t_n) \equiv y_k(t_n) - m_y(t_n)$ are length KN column vectors of the N observations at times $t_1 \leq t_n \leq t_N$ from each of the K drifters in a chosen time window of width $W = t_N - t_1$. Similarly $\bar{m}_x(t_n) \equiv m_x(t_n) - x_0$, $\bar{m}_y(t_n) \equiv m_y(t_n) - y_0$ are length N column vectors of the moving centre-of-mass at times $t_1 \leq t_n \leq t_N$. The particular ordering of the observations within each vector in Equations (3.3.2) and (3.3.3) does not matter, so long as it is consistent, and in fact, there is no restriction that the drifter observations occur at the same time, despite our choice of notation. We have defined $\mathbf{0}_{KN}$ and $\mathbf{1}_{KN}$ to be $KN \times 1$ column vectors of zeros and ones, respectively. Under each matrix we have given its size, where p is the number of parameters, and in this case $p = 8$. The vector A contains model parameters which are estimated using the least squares solution

$$A = (X'X)^{-1}X'U. \tag{3.3.4}$$

By combining Equations (3.3.1) and (3.3.4) the residual submesoscale and background velocities can be estimated by taking

$$\epsilon = [1 - X(X'X)^{-1}X']U. \tag{3.3.5}$$

The least-squares solution is equivalent to the optimal maximum likelihood solution when the residuals are Gaussian and independent and identically distributed. In general, weighted least squares solutions should be used if residuals are correlated or have unequal variance, and although this will likely be the case here, weighted-least squares requires prior knowledge of the distributional structure of the residuals which we do not wish to assume is known. Overall, we found the (non-weighted) least squares solution of Equations (3.3.4)–(3.3.5) to be robust in simulation experiments and real data analysis, and to perform better than performing least squares directly on the representation of Equation (3.2.2) on raw velocities for each drifter without removing centre-of-mass. This is due to the fact that the K drifter velocities in centre-of-mass coordinates, with the addition of the centre-of-mass velocity, can be thought of as a collection of $K + 1$ drifters that are more independent of each other than the K drifters in fixed-reference frame coordinates. This leads to errors that are more uncorrelated over drifters yielding better least squares parameter fits.

3.3.2 Flow decomposition

Once the parameters have been estimated using Equation (3.3.4), the constituent parts of the conceptual model of Equation (3.2.1) can be computed. The mesoscale contribution to each drifter is computed using

$$\begin{bmatrix} u_k^{\text{meso}}(t_n) \\ v_k^{\text{meso}}(t_n) \end{bmatrix} \equiv \begin{bmatrix} u_0 + a_1 t \\ v_0 + b_1 t \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \sigma_n + \delta & \sigma_s - \zeta \\ \sigma_s + \zeta & \delta - \sigma_n \end{bmatrix} \begin{bmatrix} x_k(t_n) - x_0 \\ y_k(t_n) - y_0 \end{bmatrix}. \tag{3.3.6}$$

The background is assumed to be spatially homogeneous, and thus can be recovered from the residuals by averaging across drifters at each time,

$$\begin{bmatrix} u^{\text{bg}}(t_n) \\ v^{\text{bg}}(t_n) \end{bmatrix} \equiv \frac{1}{K} \sum_{k=1}^{K} \left(\frac{d}{dt} \begin{bmatrix} x_k(t_n) \\ y_k(t_n) \end{bmatrix} - \begin{bmatrix} u_k^{\text{meso}}(t_n) \\ v_k^{\text{meso}}(t_n) \end{bmatrix} \right). \tag{3.3.7}$$

Finally, the submesoscale contribution to each drifter is all that remains,

$$\begin{bmatrix} u_k^{\text{sm}}(t_n) \\ v_k^{\text{sm}}(t_n) \end{bmatrix} \equiv \frac{d}{dt} \begin{bmatrix} x_k(t_n) \\ y_k(t_n) \end{bmatrix} - \begin{bmatrix} u_k^{\text{meso}}(t_n) \\ v_k^{\text{meso}}(t_n) \end{bmatrix} - \begin{bmatrix} u^{\text{bg}}(t_n) \\ v^{\text{bg}}(t_n) \end{bmatrix}. \tag{3.3.8}$$

This accomplishes the conceptual decomposition of velocities proposed in Equation (3.2.1). We emphasise that the fits of Equations (3.3.1)–(3.3.5) could be performed without the centre-of-mass velocity by removing the bottom 2 rows of U, ϵ and X in Equations (3.3.2) and (3.3.3). This is in effect only fitting observations to the second-moment model of Equation (3.2.4), as also proposed in Okubo and Ebbesmeyer (1976). While this fit still obtains estimates of mesoscale quantities $\{\sigma, \theta, \zeta, \delta\}$, and disentangles the submesoscale $\{u^{\text{sm}}(t), v^{\text{sm}}(t)\}$, the first-moment

mesoscale parameters $\{u_0, a_1, v_0, b_1\}$ and the background $\{u^{\text{bg}}, v^{\text{bg}}\}$ can no longer be estimated directly (unless fitted *a posteriori*). This means a full decomposition of the flow as performed in Equations (3.3.6)–(3.3.8) is not directly accomplished using the K drifters in centre-of-mass frame only. We shall refer to this reduced technique as the **second-moment fitting method**. In contrast, we refer to the full estimation technique from Equations (3.3.1)–(3.3.8) as the **first and second-moment fitting method**.

Regardless of the fitting method, we estimate the isotropic submesoscale diffusivity $\kappa^{\text{sm}}_{k,z}(t)$, defined in Equation (3.2.6), by measuring the *implied square displacement* of the submesoscale velocities within the window. This yields

$$\widehat{\kappa}^{\text{sm}}_{k,z}(t_n) = \frac{\Delta}{4N}\left|\sum_{t=t_1}^{t_N} u^{\text{sm}}_k(t) + \mathrm{i}v^{\text{sm}}_k(t)\right|^2, \tag{3.3.9}$$

where Δ is the sampling interval of drifter observations measured in seconds. Equation (3.3.9) is equivalent to taking 1/4 of the periodogram of the velocities—or the absolute square of the Fourier Transform—at frequency zero. This is consistent with the fact that the theoretical diffusivity of a stationary complex-valued process is determined by 1/4 of the zero-frequency of the Lagrangian frequency spectrum as per Equation (3.2.8).

The above equations produce estimates of the background, mesoscale and submesoscale parts of the flow over *some* choice of temporal window length $W = t_N - t_1$. A small value of W results in a reduced number of data points in the regression causing potentially noisy parameter estimates. Conversely, a large value of W incorporates

more distant observations in time and smooths over this noise, but may lead to poor estimates if the underlying mesoscale parameters are evolving over time. This is the classic bias-variance trade-off in statistical estimation. In Section 3.4 we address the issue of choosing an appropriate window length, and we introduce a principled estimation method using splines that allow parameters to evolve slowly over time, resulting in smoother less-variable estimates.

3.3.3 Hierarchical modelling

The Taylor series model of Equation (3.2.2) specifies 8 mesoscale parameters which are collectively denoted $\{u_0, v_0, a_1, b_1, \sigma, \theta, \zeta, \delta\}$, and these can be estimated from clustered drifter data using the machinery of Section 3.3.1. However, not every clustered set of drifters will necessarily experience all of these effects (as we illustrated in Figure 3.2.1), or the data might not give statistically significant estimates of some of the parameters even if they are truly present. Alternatively, we might already know the true values of some of the parameters and so we do not wish to estimate these. Motivated by this, we now introduce a simple method of removing certain parameters from the model, by either setting them to be zero or a pre-specified fixed value, and then estimating only the remaining unspecified parameters. If we were to instead set parameters to zero (or fixed values) after estimation, we would sub-optimally lose part of the data contained in the removed estimate.

To remove a parameter from the model, one simply removes the parameter from the vector A in Equation (3.3.3) and the corresponding column from the matrix X. In a similar vein, multiple parameters can be removed by repeating this procedure.

Ultimately, depending on the number of parameters removed, the matrix X will be sized $2(K+1)N \times p$, and the column vector A, will be sized $p \times 1$, where p is the number of free parameters that remain in the model. If $p = 8$, as presented in Equation (3.3.2), then this represents the full mesoscale solution. If any parameter values are known *a priori* then they should be inserted as fixed values into A and then multiplied by the corresponding respective columns from X and then subtracted from the vector U, before proceeding with the least squares minimisation of Equation (3.3.4) to estimate remaining parameters.

We now consider the special case of only estimating the mesoscale quantities $\{\sigma, \theta, \zeta, \delta\}$ using the second-moment fitting method discussed in Section 3.3.2. If we estimate all quantities in $\{\sigma, \theta, \zeta, \delta\}$ then $p = 4$. In contrast, if we remove all mesoscale parameters such that $\{\sigma, \theta, \zeta, \delta\} = \{0, 0, 0, 0\}$, then $p = 0$, and only submesoscale velocities remain in the centre-of-mass frame of Equation (3.2.4). If $0 < p < 4$, this represents scenarios where some mesoscale components from $\{\sigma, \theta, \zeta, \delta\}$ are present, and some are not, and we display this schematically in Figure 3.3.1. We consider strain rate and strain angle (or equivalently shear and normal strain rates) to be either jointly present or both missing. Overall, there are therefore eight possible models we might consider, shown explicitly in Figure 3.3.1. Regardless of the choice of model, the remaining non-zero parameters are estimated using Equation (3.3.4) as before.

Figure 3.3.1 also shows that the eight models exist in a *hierarchy*. The simplest model, the null hypothesis shown at the top of Figure 3.3.1, corresponds to velocities in a centre-of-mass frame that are submesoscale only. There are three direct descendants of this model in the hierarchy, the addition of vorticity or divergence, each of which

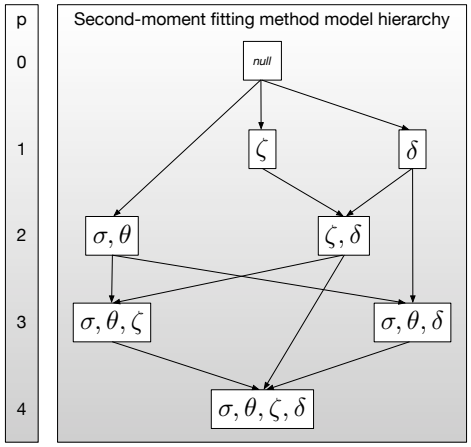

Figure 3.3.1: Hierarchy of mesoscale models using the second-moment fitting method where p indicates the number of parameters. A model with increased complexity is used only if it explains significantly more variance than the lower complexity model. Models with fewer parameters are favoured when a choice must be made.

requires one more parameter, or strain, which requires two additional parameters. The central philosophy is that **a descendent in the hierarchy should only be used if it shows meaningful improvement in some relevant error metric**, essentially disproving the null hypothesis. Because adding parameters will *always* produce at most the same residual (which may itself be the error metric), this approach avoids using too many degrees-of-freedom and producing meaningless or noisy parameter estimates.

It is worth noting that estimating all four mesoscale parameters $\{\sigma, \theta, \zeta, \delta\}$ at each time point (as is often done in the literature) would benefit from this conceptual approach. With K drifters there are $2K$ position observations at a given time point, from which 4 parameters must be estimated at each time point. For modestly sized drifter deployments, this computation runs the risk of producing estimates with no

statistical significance.

In general, when selecting between the model hierarchies for all 8 mesoscale parameters $\{u_0, v_0, a_1, b_1, \sigma, \theta, \zeta, \delta\}$ then we are faced with an increased complexity of selecting between reduced permutations of the full specification. Motivated by this, in Section 3.4.3 we will introduce methodology for estimating time-varying parameters using splines, which allows for a natural mechanism from which to build a full hierarchy of first and second-moment candidate models, as we shall show.

3.3.4 Selecting between hierarchies

We have provided a mixed background-mesoscale-submesoscale modelling framework in Equation (3.2.2) and a corresponding estimation framework in Section 3.3.1. Then in Section 3.3.3 we discussed how to estimate parameters using different hierarchies of mesoscale components in the overall model. The appropriateness of a chosen model in the hierarchy, for a given set of observational drifter data, can be evaluated by estimating the *error* resulting from the fitted model at a given point in time. We argue there is more than one meaningful way in which *error* can be computed—and in this section we shall define two such ways that prove to be very useful in terms of model evaluation.

Fraction of Variance Unexplained (FVU)

The first method is perhaps the most intuitive. Here we calculate how much variance remains in the 'unexplained' residual submesoscale velocities found in Equation (3.3.8). This value in itself, however, is not a meaningful quantity unless it is

presented in reference to some other quantity. Therefore, to provide a normalised and meaningful metric we introduce the notion of *the Fraction of Variance Unexplained* (FVU), which is defined as

$$\text{FVU} = \frac{\sum_{t_n=t_1}^{t_N}\sum_{k=1}^{K}\left\{u_k^{\text{sm}}(t_n)^2 + v_k^{\text{sm}}(t_n)^2\right\}}{\sum_{t_n=t_1}^{t_N}\sum_{k=1}^{K}\left\{\left[\frac{d}{dt}\left(x_k(t_n) - m_x(t_n)\right)\right]^2 + \left[\frac{d}{dt}\left(y_k(t_n) - m_y(t_n)\right)\right]^2\right\}}, \quad (3.3.10)$$

and hence quantifies the *proportion* of the variability remaining in the submesoscale model, as compared to velocities that have only had the centre-of-mass removed (and will hence still contain second-moment mesoscale effects present in Equation (3.2.4)). The FVU will therefore in general be some value between zero and one. An FVU value close to one occurs when there is little to no mesoscale component estimated from the data. In contrast, an FVU value equal to zero means the mesoscale model successfully explains all variability in the data after the background is removed, and there is no residual submesoscale process left behind. For mixed mesoscale and submesoscale flow the FVU will be somewhere between zero and one, and this will vary dependent on the magnitude and number of mesoscale components present in the model fit.

In Figure 3.3.2, in the left column we display FVU values obtained from our simulation setup shown in Figure 3.2.1. Specifically, we generate 100 replicated simulations of each of the four model scenarios shown in Figure 3.2.1—diffusivity only, strain+diffusivity, strain+vorticity+diffusivity (strain dominated), strain+vorticity+ diffusivity (vorticity dominated)—where the stochasticity between replicates occurs from simulating submesoscale velocities from a Gaussian white noise process as in Equation (3.2.9). Again, as in LatMix Site 1, we simulate nine drifters within each

simulation with matching initial positions, but this time we just simulate half-hourly records for one day. We use the procedures described in Section 3.3.3 to fit four hierarchies of models to each simulation within each scenario. Note that we perform a global fit by setting the window length W to be the full length of the observations (one day). The FVU values are calculated from Equation (3.3.10) and the resulting spread of values across simulations are shown by box and whisker plots in Figure 3.3.2. We also provide the spread of observed FVU values in an oracle case where the true mesoscale parameters are known.

In the figure we have also indicated the estimated theoretical FVU value obtained by combining the mesoscale variance obtained from Equation (3.2.15) for each drifter k (let us denote this $\sigma^2_{w^{\text{meso}}}(k)$) with the submesoscale variance of a white noise process given from the spectral form of Equation (3.2.10) yielding $\sigma^2_{w^{\text{sm}}} = 4\kappa(1-1/K)$, which is the same for each drifter, where the $(1-1/K)$ rescaling is required to account for moving to a centre-of-mass reference frame. We can then obtain an estimated theoretical FVU value, which we denote $\widetilde{\text{FVU}}$, by taking

$$\widetilde{\text{FVU}} = \frac{\sigma^2_{w^{\text{sm}}}}{\left\{\frac{1}{K}\sum_{k=1}^{K}\sigma^2_{w^{\text{meso}}}(k)\right\} + \sigma^2_{w^{\text{sm}}}}. \tag{3.3.11}$$

This an *estimated* theoretical FVU, rather than an exact solution, because we have ignored the co-dependence between the mesoscale and submesoscale processes and assumed these variances aggregate separately. The results however indicate remarkable agreement between theoretical and observed quantities for FVU over all scenarios (except when insufficient mesoscale parameters are proposed in the candidate model),

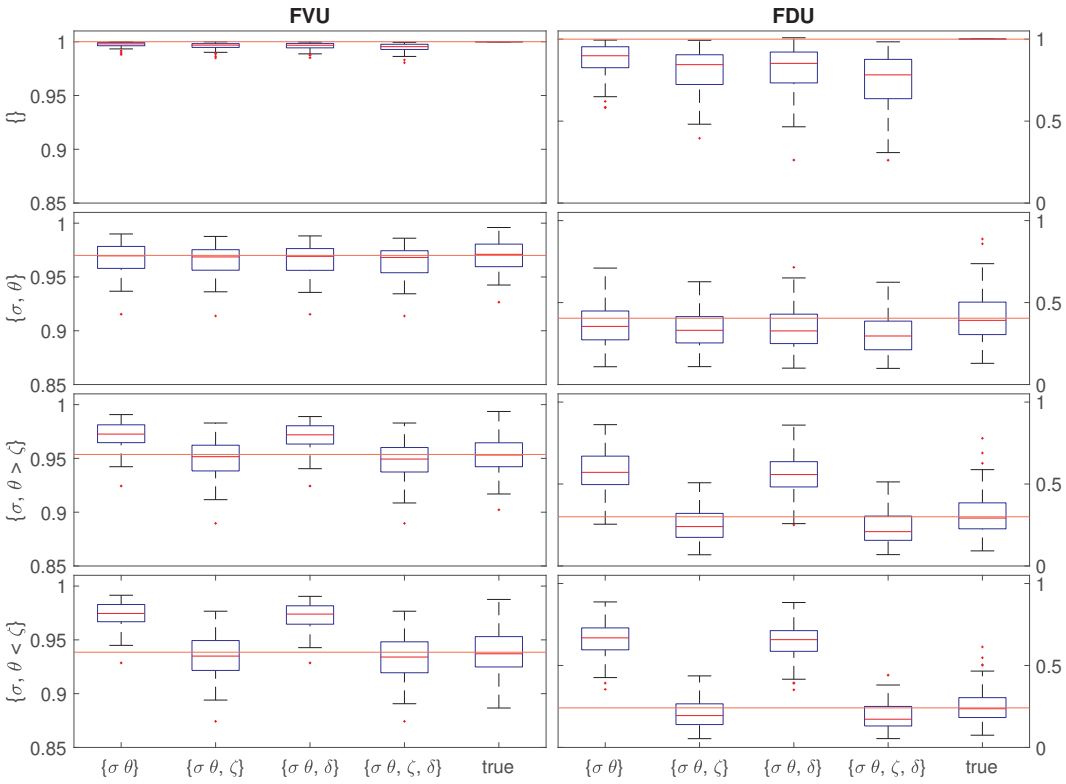

Figure 3.3.2: FVU (left column) and FDU (right column) for candidate models fitted to trajectories generated from the four model scenarios from Figure 3.2.1. Each subplot here is for a different true model scenario (the y-axis), and each box and whisker within a subplot provides the spread of FVU/FDU values from a fitted candidate model (the x-axis). The final box and whisker in each subplot is using the true mesoscale parameter values. The spread of results is over 100 repeated simulations using 9 drifters sampled every 30 minutes for one day. The estimated theoretical FVU, obtained from Equation (3.3.11), and the estimated theoretical FDU, obtained from Equation (3.3.13), are overlaid by a red horizontal line in each subplot. Parameters are estimated using the second-moment fitting method, where results using the first and second-moment fitting method yield near identical results as $u_0 = v_0 = a_1 = b_1 = u^{\text{bg}} = v^{\text{bg}} = 0$ in these simulations.

suggesting Equation (3.3.11) is an accurate approximation for the spatial and temporal scale of the simulation performed.

Overall, the key finding of Figure 3.3.2 (left column) is that the FVU helps identify the correct model in all true model scenarios considered, and correctly estimates how much of the variance is explained by the mesoscale and submesoscale components in agreement with the theory. The addition of a mesoscale parameter which is truly present significantly reduces the FVU, but adding further unnecessary mesoscale parameters (such as the divergence which is not present in any of the scenarios) does not significantly reduce FVU. This diagnostic tool therefore shows utility as a method for detecting the presence of mesoscale effects on drifter velocities, and for selecting between mesoscale model hierarchies. We shall scrutinise this further when we apply our procedures to LatMix data in Section 3.5.

Fraction of Diffusivity Unexplained (FDU)

The FVU is a measure of how much of the variability of the data remains in the submesoscale residuals. However, we argue this is not the only metric with which to ultimately select from a model hierarchy. First of all, as the residual velocities are being directly minimised (along with the background) in the least squares fits of Equations (3.3.1)–(3.3.5), the more complex models will generally have a lower FVU than nested simpler models with fewer or no mesoscale components (as seen in Figure 3.3.2). This may lead to over-fitting models unless parameter penalisation methods are introduced. Secondly, mesoscale processes are primarily low frequency processes with decaying Lagrangian velocity frequency spectra, as we showed in Fig-

ure 3.2.2. Submesoscale processes, on the other hand, will likely have Lagrangian velocity frequency spectra that are spread across frequencies and concentrated away from frequency zero. For example, white noise submesoscale residuals will have a flat spectrum, and an internal wave process, represented by the Garrett-Munk spectrum for instance, will have significant energy at the inertial frequency f_0, but very small energy at frequency zero.

For these reasons, we now motivate a second metric with which to evaluate different model hierarchies. Specifically, we measure the *diffusivity* of the residual process for each drifter, and compare this with the implied total diffusivity of each drifter when no mesoscale is removed. In other words, we compare the variability of the aggregated and submesoscale-only components in terms of their respective diffusivities, with a view that submesoscale diffusivity should be much lower than total diffusivity when even a mild mesoscale component is present (as mesoscale energy is dominant at low frequencies in the velocity spectra). To quantify this effect we introduce the notion of *the Fraction of Diffusivity Unexplained* (FDU), which we define by

$$\text{FDU} = \frac{\sum_{t_n=t_1}^{t_N} \sum_{k=1}^{K} \hat{\kappa}_{k,z}^{\text{sm}}(t_n)}{\sum_{t_n=t_1}^{t_N} \sum_{k=1}^{K} \hat{\kappa}_{k,z}^{\text{c.o.m.}}(t_n)}, \tag{3.3.12}$$

where $\hat{\kappa}_{k,z}^{\text{sm}}(t_n)$ has already been defined in Equation (3.3.9). $\hat{\kappa}_{k,z}^{\text{c.o.m.}}(t_n)$ is the diffusivity for drifter k with only centre-of-mass removed, which is defined by replacing $u_k^{\text{sm}}(t)$ with $\frac{d}{dt}\left(x_k(t_n) - m_x(t_n)\right)$ and $v_k^{\text{sm}}(t)$ with $\frac{d}{dt}\left(y_k(t_n) - m_y(t_n)\right)$ in Equation (3.3.9). The FDU measures how much diffusivity is present in the submesoscale residual after removing the mesoscale, as compared to the diffusivity that is observed relative to

the centre-of-mass when no mesoscale has been explicitly removed. An FDU value of zero means that the submesoscale process has no observed diffusivity, and an FDU of one will occur when either no mesoscale is present, or the mesoscale does not create any diffusive-type behaviour on the particles.

We display observed FDU values across our simulations in the right column of Figure 3.3.2, mirroring the simulation setup used for FVU described in Section 3.3.4. The *estimated* theoretical FDU values are overlaid by a red horizontal value from computing

$$\widetilde{\mathrm{FDU}} = \frac{\kappa_z^{\mathrm{sm}}}{\left\{\frac{1}{K}\sum_{k=1}^{K}\kappa_{k,z}^{\mathrm{meso}}\right\} + \kappa_z^{\mathrm{sm}}}, \tag{3.3.13}$$

where the expected submesoscale diffusivity for all drifters is $\kappa_z^{\mathrm{sm}} = \kappa(1-1/K)$ where again the $(1-1/K)$ rescaling is required to account for moving to a centre-of-mass reference frame. We obtain $\kappa_{k,z}^{\mathrm{meso}}$ by taking 1/4 of the zero-frequency value from Equation (3.2.17) (as per the definition of Equation (3.2.8)). Similarly to Equation (3.3.11), Equation (3.3.13) is an *estimated* theoretical FDU because we are assuming independent dispersion caused by the mesoscale and submesoscale. Nevertheless, Figure 3.3.2 indicates consistent agreement between observed and theoretical FDU values (when the correct model is fitted), highlighting the accuracy of this approximation.

The main finding of the FDU analysis in Figure 3.3.2 is that the mesoscale explains significantly more of the total diffusivity than the total variance. This is as expected because of the low-frequency nature of mesoscale processes (see Figure 3.2.2) and highlights the usefulness of computing FDU values to test for mesoscale presence. In

all cases we can see that FDU analysis reveals the correct generating mesoscale model even better than FVU does. We shall further use this diagnostic method of assessing model fits with LatMix data in Section 3.5.

3.4 Uncertainty Quantification and Capturing Temporal Evolution

3.4.1 Uncertainty Quantification

We now provide a method for estimating the uncertainty of parameter estimates when applied to observational datasets. In a simulation setting, uncertainty estimates can be obtained by repeating experiments several times stochastically or with different initial conditions, but this cannot be done in the real world where clustered drifter deployments are scarcely repeated in the same region of the ocean, and will likely be measuring different mesoscale and submesoscale features each time.

Instead, we resort to the bootstrap, which resamples the observed data in such a way as to provide a population of different datasets with which to measure uncertainty. Specifically, the bootstrap is implemented by taking a random sample of K trajectories from the K drifters *with* replacement, such that the same trajectories may be selected multiple times as if they were different drifters. Then the mesoscale parameters are estimated for this random sample of trajectories. Let us denote any one of these parameter estimates as $\hat{p}_b$. The process is then repeated B times, every time randomly resampling a set of K trajectories with replacement, such that we obtain B parameter

estimates $\{\hat{p}_1, \ldots, \hat{p}_B\}$. These replicated bootstraps can be used to form quantiles which then provide confidence intervals for the parameter of interest, often set to values such as 90% or 95%. Alternatively, we can also estimate the *standard error* of $\hat{p}$, the parameter estimate for p, by measuring the sample standard deviation of $\hat{p}_b$ given by

$$\mathrm{SE}_B(\hat{p}) = \left[\frac{1}{B-1} \sum_{i=1}^{B} \left\{ \hat{p}(i) - \hat{p}_{(\cdot)} \right\}^2 \right]^{1/2}, \tag{3.4.1}$$

where $\hat{p}_{(\cdot)} = \frac{1}{B} \sum_{i=1}^{B} \hat{p}(i)$.

In Figure 3.4.1 we show a histogram of bootstrap parameter estimates for $\{\sigma, \theta, \zeta\}$, with a red vertical line at the true value, and a blue vertical line showing the average bootstrap estimate. The purpose of this simulation is simply to show that bootstrap parameter estimates are *centred* at their true values and symmetrically distributed, despite the fact that drifter trajectories are sampled with replacement. We found this to be a consistent feature across different true parameter values and simulation settings.

Next we establish that the bootstrap estimate for the standard error of parameter estimates, given in Equation (3.4.1), agrees with standard errors of parameter estimates observed from repeated simulations. In Table 3.4.1 we compare *simulated* and *bootstrap* standard errors for two experiments: the strain-only and the strain-dominated simulations of Figure 3.2.1. The standard errors from simulations are across 100 repeated simulations, but the bootstrap standard error approximation is just from 1 simulation of drifters each time (as we would have with real data). Despite

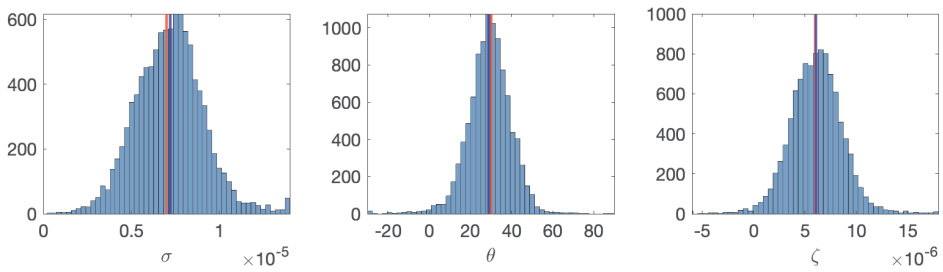

Figure 3.4.1: Histogram of bootstrap parameter estimates for strain rate, strain angle, and vorticity, over 100 repeated simulations where $B = 100$ for each simulation, thus obtaining 10,000 total bootstrapped parameter values. The trajectories are generated as in Figure 3.2.1 in the strain-dominated model for 1 day, and the parameters are estimated using the second-moment fitting method. Any bootstrap estimates outside the range of the x-axis are placed in the limiting visible bar in the histogram on each side. The red vertical line is the true parameter value, and the blue vertical line is the average bootstrap estimate.

	$\sigma\ (s^{-1}) \times 10^6$	$\theta\ (^\circ)$	$\zeta(s^{-1}) \times 10^6$
Strain-only Simulation			
Simulated	1.17	6.68	N/A
Bootstrap	1.32 ± 0.365	6.29 ± 2.47	N/A
Strain-dominated Simulation			
Simulated	1.22	6.78	1.61
Bootstrap	1.55 ± 0.459	8.08 ± 3.43	1.94 ± 0.572

Table 3.4.1: Observed standard errors from simulation, and average bootstrap standard error estimates from Equation (3.4.1) (where $B = 100$), over 100 repeated simulations, for both the strain-only and strain-dominated simulations of Figure 3.2.1 over 1 day. We also provide the standard deviation of bootstrap standard error estimates across the 100 simulations, as indicated after the $\pm$ symbol.

this, the average bootstrap standard error estimate is very close to the standard error from repeated simulations (with the standard deviation of the bootstrap standard error accounting for any difference). Notice also that the bootstrap standard error estimates are usually conservative, which is better than the converse, and correctly increase when more parameters need to be estimated. This demonstrates the accuracy of Equation (3.4.1) in estimating the standard error of parameter estimates obtained from Equation (3.3.4). We will make use of the bootstrap in the analysis of LatMix data in Section 3.5.

3.4.2 Time-evolving parameters using rolling windows

To estimate the temporal evolution of mesoscale features across a drifter deployment we allow the mesoscale parameters to evolve over time. In this section we first introduce a simple method for doing so where we use a rolling time window of width W and estimate the parameters $\{u_0(t_n), v_0(t_n), a_1(t_n), b_1(t_n), \delta(t_n), \zeta(t_n), \sigma_n(t_n), \sigma_s(t_n)\}$ in Equation (3.3.3) over time using velocity observations contained in the interval $[\frac{d}{dt}x_k(t_n - \frac{W}{2}), \frac{d}{dt}x_k(t_n + \frac{W}{2})]$ and $[\frac{d}{dt}y_k(t_n - \frac{W}{2}), \frac{d}{dt}y_k(t_n + \frac{W}{2})]$ using the same approach outlined in Section 3.3.1, repeated at every observation time-step t_n in the experiment.

In general, the window width parameter W should be chosen to be large enough to ensure we have reduced variance and statistically significant estimates of each mesoscale parameter, but not so large that resolution is lost from over-smoothing. To examine this effect we display simulated trajectories in Figure 3.4.2 which exactly follows the strain-only simulation from Figure 3.2.1, except that the strain rate parameter now decreases linearly by a factor of 10 across the length of the 6.25 day simulation, and we have increased κ to $0.5\text{m}^2/\text{s}$. We then use the second-moment fitting method with the strain-only model over rolling windows with 3 choices of W (6-hours, 1-day, or 3-days). In Figure 3.4.3 we display the time-varying strain rate estimate over time from the data in Figure 3.4.2, alongside the standard error of this estimate over time (obtained over 100 repeated simulations). With this increased diffusivity, the inherent trade-off with the rolling-window method becomes apparent. Long window lengths provide low uncertainty, but the parameter estimates are only provided in the temporal centre of the experiment (and would be biased if extended

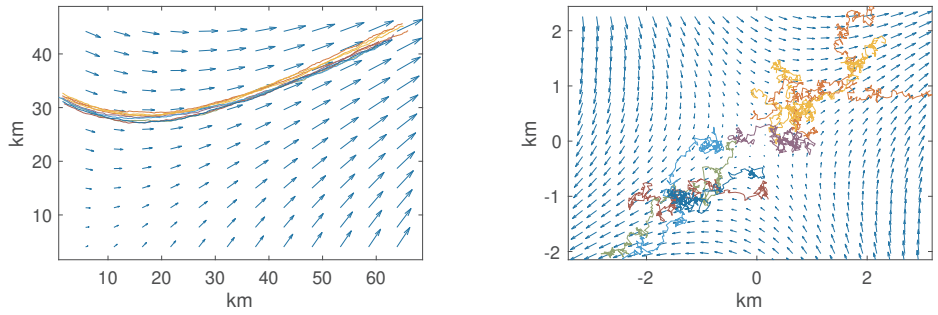

Figure 3.4.2: Simulation of 9 drifters using the identical configuration of Figure 3.2.1 (strain only) except that the strain rate changes linearly across time from $\sigma = 1 \times 10^{-5}$/s to $\sigma = 1 \times 10^{-6}$/s and $\kappa = 0.5\text{m}^2/\text{s}$. The left panel displays drifter positions. The right panel displays drifter positions with respect to their centre-of-mass. The quiver arrows indicate the velocity field at the beginning of the simulation.

outwards). Short windows, on the other hand, provide variable estimates with large standard errors that exceed half the parameter value, as we see on the right panel—meaning such estimates cannot be statistically distinguished from zero in a " two sigma" sense. A daily window length is perhaps the most appropriate balance here.

Motivated by these challenges, we shall shortly provide a more principled approach to generating smoothly-evolving parameter estimates using splines in Section 3.4.3. Before doing so, we present results of a large simulation analysis which we will use to guide our window length selection choices in the LatMix experiment. Specifically, in Figure 3.4.4 we plot a heatmap of standard errors in strain rate estimation, over a grid of values of true constant strain rate, σ, and estimation window length, W. We repeat the analysis for a low diffusivity $\kappa = 0.1$ m^2/s and high diffusivity setting $\kappa = 1$ m^2/s. Otherwise the settings are the LatMix-type settings used in Figure 3.2.1, using 9 drifter trajectories with matching starting locations. The standard errors are

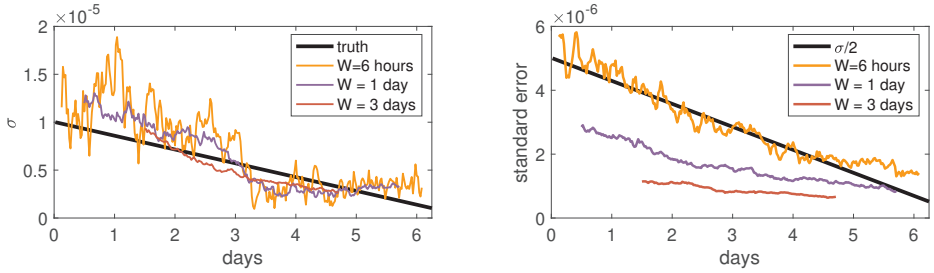

Figure 3.4.3: The left panel shows rolling-time window estimates of the varying strain rate from the data presented in Figure 3.4.2 over 3 choices of window lengths using the second-moment fitting method. The right panel shows the standard error of these time-varying estimates over 100 repeated simulations, plotted against the true value of $\sigma/2$.

in units of the true strain rate, and we have marked with a red line the point at which the standard error is approximately equal to half the true strain rate. The way in which this plot should be interpreted is that for a given strain rate (and diffusivity), the window length should be *at least* as long as the red line marking the point at which estimates become statistically significant. For example, higher diffusivities, or lower strain rates, will require longer windows with which to estimate the parameters significantly. We focus on strain in these simulations, as this was found to be the most pronounced mesoscale effect in the LatMix analysis that follows, but this analysis could be repeated with other mesoscale parameters to inform window length selection for other drifter deployments. In Chapter 4 we perform a sensitivity analysis of these results for varying numbers of drifters and initial deployment configurations, to help generalise our findings to wider settings.

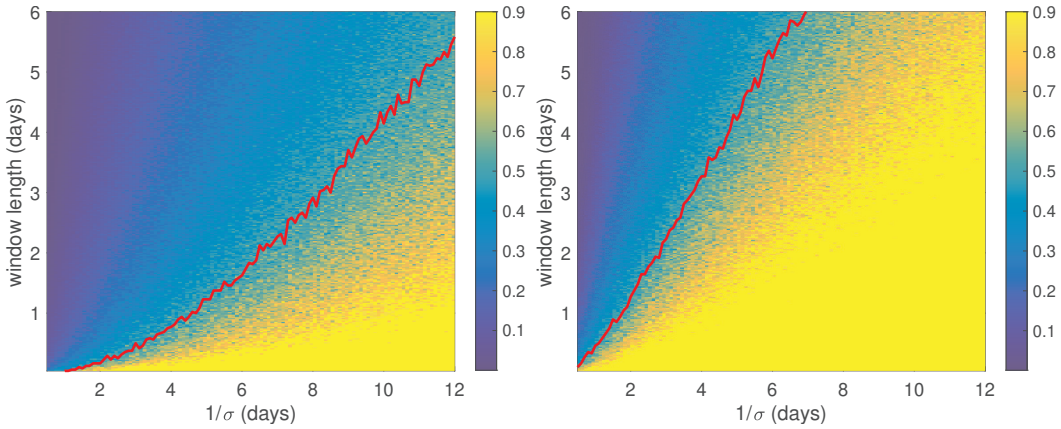

Figure 3.4.4: Estimated standard errors for the strain rate (in the units of the true strain rate) across a dense grid of fixed strain rate values σ and window lengths W in a strain-only simulation mirroring the setup in Figure 3.2.1. In the left panel we have set $\kappa = 0.1$ m^2/s and in the right $\kappa = 1$ m^2/s. The strain rate estimates are obtained using the second-moment fitting method of a strain-only model, and the standard errors are obtained over 100 repeated simulations. The standard errors in the heatmap are upper-bounded by 0.9 for representation purposes. We draw a red line where the standard error is approximated to be half the true parameter value for each value of the strain rate.

3.4.3 Slowly-evolving parameters using splines

To generalise the idea of time windowing to estimate the mesoscale parameters, we represent the parameters as coefficients as a finite sum of B-splines,

$$\sigma(t) = \sum_{m=1}^{M} \hat{\sigma}^m B^m(t), \tag{3.4.2}$$

where M is the total number splines over the experiment window and $\hat{\sigma}^m$ are the M coefficients. A B-spline (or basis spline) of degree D is a local piecewise polynomial that maintains nonzero continuity across D knot points placed at times τ_i. These knot points define the extent of the B-splines, and therefore let us choose an effective window length for parameter fluctuations. The lowest degree ($D = 0$) splines are boxcar functions between the knot points, and are thus identical to non-overlapping windows in Section 3.4.2. At degree $D = 1$, B-splines are triangle functions that span two knot points, thus providing continuity in time as well as a piecewise first derivative. This generalises to higher degrees, where a B-spline of degree D has D non-zero derivatives, as reviewed in Early and Sykulski (2020). The key benefit to this approach is that we can allow for time variation in the parameters while simultaneously choosing an effective window length—all while adding only a few coefficients to the model.

To extend the estimation method presented in Section 3.3.1, we now require M coefficients for each of the p parameters, resulting in pM total coefficients to estimate. Rewriting vector A from Equation (3.3.3) we have that

$$A = \underbrace{\begin{bmatrix} u_0^m \\ v_0^m \\ \sigma_n^m \\ \sigma_s^m \\ \zeta^m \\ \delta^m \end{bmatrix}}_{pM\times 1}, \tag{3.4.3}$$

where each coefficient, e.g. u_0^m, is a column vector of the M B-spline coefficients (we will shortly discuss why a_1 and b_1 can be dropped here). The data matrix X correspondingly expands from p to pM columns,

$$X = \underbrace{\frac{1}{2}\begin{bmatrix} \mathbf{0}_{KN} & \mathbf{0}_{KN} & \tilde{x}_k(t_n)B^m(t_n) & \tilde{y}_k(t_n)B^m(t_n) & -\tilde{y}_k(t_n)B^m(t_n) & \tilde{x}_k(t_n)B^m(t_n) \\ \mathbf{0}_{KN} & \mathbf{0}_{KN} & -\tilde{y}_k(t_n)B^m(t_n) & \tilde{x}_k(t_n)B^m(t_n) & \tilde{x}_k(t_n)B^m(t_n) & \tilde{y}_k(t_n)B^m(t_n) \\ 2B^m(t_n) & \mathbf{0}_N & \bar{m}_x(t_n)B^m(t_n) & \bar{m}_y(t_n)B^m(t_n) & -\bar{m}_y(t_n)B^m(t_n) & \bar{m}_x(t_n)B^m(t_n) \\ \mathbf{0}_N & 2B^m(t_n) & -\bar{m}_y(t_n)B^m(t_n) & \bar{m}_x(t_n)B^m(t_n) & \bar{m}_x(t_n)B^m(t_n) & \bar{m}_y(t_n)B^m(t_n) \end{bmatrix}}_{2(K+1)N\times pM}, \tag{3.4.4}$$

where each column is repeated for each of the M B-splines. Note that, because the B-splines are local functions, the resulting matrix may be relatively sparse.

Parameter estimation is as before, but Equation (3.3.6) for the mesoscale flow is replaced by,

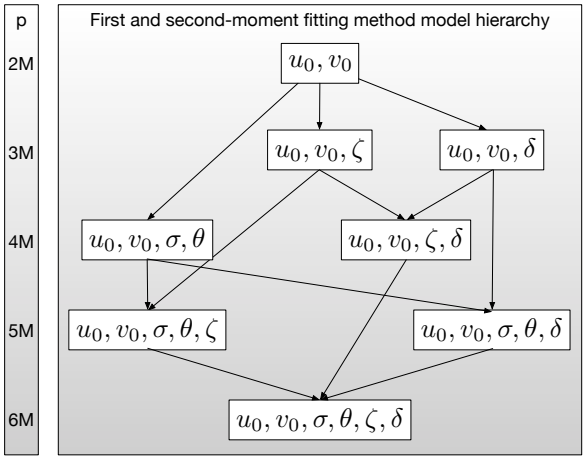

Figure 3.4.5: Hierarchy of first and second-moment mesoscale models where p indicates the number of parameters. A model with increased complexity is used only if it explains significantly more variance than the lower complexity model. Models with fewer parameters are favoured when a choice must be made.

$$\begin{bmatrix} u_k^{\text{meso}}(t_n) \\ v_k^{\text{meso}}(t_n) \end{bmatrix} \equiv \sum_{m=1}^{M} \begin{bmatrix} u_0^m B^m(t_n) \\ v_0^m B^m(t_n) \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \sigma_n^m + \delta^m & \sigma_s^m - \zeta^m \\ \sigma_s^m + \zeta^m & \delta^m - \sigma_n^m \end{bmatrix} \begin{bmatrix} (x_k(t_n) - x_0)\, B^m(t_n) \\ (y_k(t_n) - y_0)\, B^m(t_n) \end{bmatrix}. \tag{3.4.5}$$

The background flow and submesoscale flow are still recovered using Equations (3.3.7) and (3.3.8), respectively.

One of the advantages of using B-splines is that the model hierarchy is simplified. Figure 3.4.5 shows the complete model hierarchy that includes the first and second-moment fitting method, unlike Figure 3.3.1 which only showed the hierarchy for the second-moment fitting method. The key simplification is that with B-splines we can drop (a_1, b_1) from X when going from Equation (3.3.3) to (3.4.4), since time dependence is encoded in the B-spline estimates for (u_0, v_0). Choosing the appropriate model

from Figure 3.4.5 proceeds exactly as in Section 3.3.3, but with the additional caveat that one must choose the spline degree D and the number of splines M. With the restriction that the spline degree $D < M$, a reasonable upper bound is $D = 3$, the cubic spline. The number of splines M can be chosen by assuming a minimum window length (as discussed in Section 3.4.2), treating the centre of each window as a data point, and then applying the formula for the canonical interpolating spline in Early and Sykulski (2020). To compute this explicitly, assume a time series of length T, with minimum window length W, then this results in a total of $M = \max(\lfloor T/W \rfloor, 1)$ evenly sized windows of minimum length. Now apply Equations (7) and (8) in Early and Sykulski (2020) using pseudo points at $\{t_1, t_1 + T/M(j - 1/2), t_N\}$ where $j = 2, \ldots, M - 1$. When the drifters are evenly sampled in time, this will result in M splines that each have support from the same number of data points, and each data point will intersect $D + 1$ splines. As a result, there is really only one parameter to adjust: the effective window length or, alternatively, the number of splines M. Because setting $M = 1$ exactly reproduces the approach in Section 3.3.1 using fixed parameters, the freedom for parameters to vary over time can be systematically increased by increasing M.

Quantifying uncertainty with spline solutions requires a modification to the approach in Section 3.4.1. This is because the resulting bootstrapped parameter estimates are no longer pointwise estimates of each parameter, but rather time-varying global solutions. This means that computing the mean of each mesoscale parameter at each instant in time will not, in general, result in a valid solution since each solution is a global fit to the data. As a result, rather than considering a *mean* value from bootstrap solutions, as in Figure 3.4.1, we must establish the *most likely* bootstrap

solution. Applying the bootstrap B times results in B continuous time varying model solutions of the parameters. Thus, we compute the most likely solution (of the B solutions) from an estimated joint probability distribution function (PDF). Specifically, for each estimated parameter in the model, we use a kernel density estimator to estimate a PDF from the bootstrap replicates for each parameter at each point in time using the methodology in Botev et al. (2010). For example, at time t_n we estimate a one-dimensional PDF $\hat{P}_\zeta(t_n, \hat{\zeta})$ using the B bootstrap parameter estimates for ζ and a two-dimensional PDF $\hat{P}_{\sigma_n,\sigma_s}(t_n, \hat{\sigma_n}^b(t_n), \hat{\sigma_s}^b(t_n))$ for σ_n, σ_s. The likelihood of each path is then found with

$$L(\hat{\sigma_n}^b, \hat{\sigma_s}^b, \hat{\zeta}^b) = \prod_{n=1}^{N} \hat{P}_{\sigma_n,\sigma_s}(t_n, \hat{\sigma_n}^b(t_n), \hat{\sigma_s}^b(t_n)) \cdot \hat{P}_\zeta(t_n, \hat{\zeta}^b(t_n)), \tag{3.4.6}$$

where, in practice, we include probabilities from all estimated parameters. The most likely solution is that with maximum L, where confidence intervals are similarly calculated by including the Y percent of the B most likely solutions.

3.5 Application to the LatMix Experiment

The lateral mixing (LatMix) field campaign of 2011 (Shcherbina et al., 2015; Sundermeyer et al., 2020) deployed drifters and dye with the aim of understanding what causes mixing at the submesoscale, and how this varies both spatially and temporally. The experiment consisted of two drifter deployments in the Sargasso Sea, where the drifters were deployed in a cluster. The first deployment, which we refer to as 'Site 1', consisted of 9 drifters tracked for 6.1 days in an area of low strain, and the second

deployment, 'Site 2', consisted of 8 drifters tracked for 6.3 days in an area of moderate strain. There has been a large amount of interest and research from the experiment, e.g. Shcherbina et al. (2013).

In Figure 3.5.1 we plot the drifter trajectories for each site both in terms of their $\{x, y\}$ positions, but also with respect to the time-varying centre-of-mass across drifters. The effect of the mesoscale, especially strain, can already be seen visually by inspecting this plot, both in the absolute and centre-of-mass reference frames. There are also possible signs of divergence in Site 1 (the drifters spreading in a non-random way), and vorticity in Site 2. We will now inspect this in more rigorous statistical detail using the methodology of this Chapter.

3.5.1 Fixed mesoscale parameter estimates

We first fit fixed (i.e. non-time-varying) mesoscale parameters to Equation (3.2.4) at each site using the second-moment fitting method described in Section 3.3. We present the results in the top half of Table 3.5.1 using several model hierarchies. For each model hierarchy we present the estimated mesoscale quantities, and the resulting submesoscale diffusivity. We also present FVU and FDU values (Equations (3.3.10) and (3.3.12) respectively) to assess model fit, where we remind the reader that lower values correspond to model fits with reduced error. To select the best model we use the conceptual approach illustrated earlier in Figure 3.3.1.

For Site 1 we see reasonable evidence for adding the parameters $\{\sigma, \theta\}$ ahead of vorticity ζ or divergence δ, as this creates the lowest FDU values thereby creating low submesoscale diffusivities of $\kappa \approx 0.2\text{m}^2/\text{s}$, as reported in Shcherbina et al. (2015).

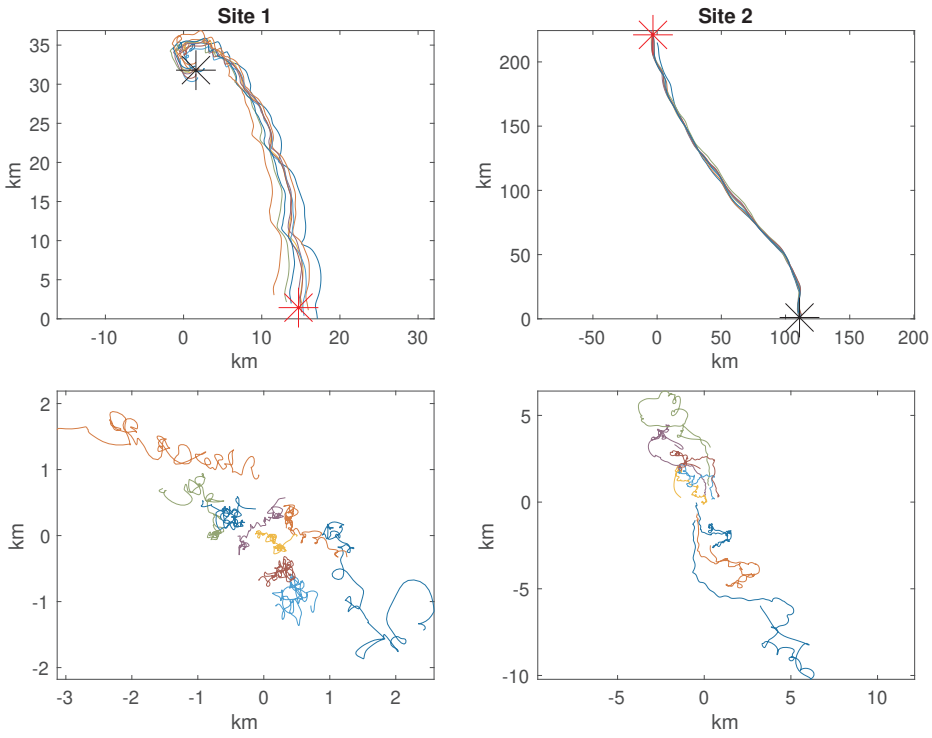

Figure 3.5.1: LatMix trajectories of Site 1 (9 drifters) and Site 2 (8 drifters). Top row are the positions in $\{x_k(t), y_k(t)\}$, bottom row are relative to centre-of-mass $\{\tilde{x}_k(t), \tilde{y}_k(t)\} = \{x_k(t) - \frac{1}{K}\sum_{k=1}^{K} x_k(t), y_k(t) - \frac{1}{K}\sum_{k=1}^{K} y_k(t)\}$. The black and red star in the top row of plots indicate the respective starting and ending centre-of-mass positions. $\{0, 0\}$ in the $\{x, y\}$ components corresponds to $\{-73.0234, 31.7424\}$ degrees longitude-latitude for Site 1 and $\{-73.6776, 32.2349\}$ degrees longitude-latitude for Site 2.

Next, we follow the hierarchy and consider adding vorticity or divergence to the strain. Here we see little evidence for vorticity, but some for divergence, with a marginal reduction in the FDU value for the latter. Finally, just for completion, we show the full hierarchy. While this full hierarchy will always yield the lowest FVU compared to all simpler models (as this is the objective function being minimised)—the FVU value does not appear to drop significantly, and the FDU value has in fact increased, suggesting this to be an overfitted choice if we are only selecting among fixed mesoscale parameters.

For Site 2 we see mixed evidence for either initially adding divergence or strain, but the vorticity-only fit performs poorly and in fact adds diffusivity as compared to raw centre-of-mass velocities. As divergence is only one parameter (vs two for strain), we would normally proceed this way down the hierarchy using Figure 3.3.1. However, as we shall see when we account for time-variation in the mesoscale parameters, there will be more evidence for a strain-only model than a divergence-only model, therefore for comparison we follow this route down the hierarchy. When considering adding vorticity or divergence, then now there is interestingly more evidence for vorticity, with reduced FVU and FDU values. Overall however, we note that diffusivity values are much larger at Site 2 using fixed parameters, with $\kappa \approx 2\text{m}^2/\text{s}$. This is likely due to the presence of time-varying mesoscale features not being account for, as we shall now explore.

Fixed estimates (Site 1)

model	σ (f_0)	θ (°)	ζ (f_0)	δ (f_0)	κ (m^2/s)	FVU	FDU
$\{\zeta\}$	0	0	-0.000137	0	0.974	1.000	1.001
$\{\delta\}$	0	0	0	0.0493	0.361	0.983	0.371
$\{\sigma, \theta\}$	0.0591	-27.8	0	0	0.188	0.976	0.193
$\{\sigma, \theta, \zeta\}$	0.0785	-15.3	-0.0443	0	0.229	0.971	0.235
$\{\sigma, \theta, \delta\}$	0.0489	-25.6	0	0.0137	0.174	0.976	0.179
$\{\sigma, \theta, \zeta, \delta\}$	0.0711	-12.2	-0.0443	0.0137	0.216	0.971	0.221

Fixed estimates (Site 2)

model	σ (f_0)	θ (°)	ζ (f_0)	δ (f_0)	κ (m^2/s)	FVU	FDU
$\{\zeta\}$	0	0	0.00613	0	4.011	0.999	1.000
$\{\delta\}$	0	0	0	0.0125	1.886	0.997	0.470
$\{\sigma, \theta\}$	0.0131	-67.0	0	0	1.906	0.996	0.475
$\{\sigma, \theta, \zeta\}$	0.0642	78.0	0.0650	0	1.950	0.985	0.486
$\{\sigma, \theta, \delta\}$	0.0107	-67.9	0	0.00258	1.874	0.996	0.467
$\{\sigma, \theta, \zeta, \delta\}$	0.0637	77.0	0.0650	0.00258	1.919	0.985	0.478

Table 3.5.1: LatMix submesoscale diffusivity estimates and associated FVU and FDU, estimated over candidate models in the hierarchy at each site using fixed parameter estimates. For fixed estimates we also show the mesoscale parameter estimates (scaled by the inertial frequency, f_0). The fixed method uses the second-moment fitting methods.

	Rolling estimates (Site 1)			Rolling estimates (Site 2)		
model	κ (m^2/s)	FVU	FDU	κ (m^2/s)	FVU	FDU
$\{\zeta\}$	0.995	0.992	1.022	2.924	0.872	0.729
$\{\delta\}$	0.325	0.974	0.334	2.341	0.838	0.584
$\{\sigma, \theta\}$	0.183	0.961	0.188	1.680	0.710	0.419
$\{\sigma, \theta, \zeta\}$	0.282	0.937	0.290	0.825	0.675	0.206
$\{\sigma, \theta, \delta\}$	0.147	0.966	0.151	1.753	0.704	0.437
$\{\sigma, \theta, \zeta, \delta\}$	0.248	0.941	0.255	0.722	0.669	0.180

	Spline estimates (Site 1)			Spline estimates (Site 2)		
model	κ (m^2/s)	FVU	FDU	κ (m^2/s)	FVU	FDU
$\{\zeta\}$	1.742	1.025	1.791	3.059	0.973	0.697
$\{\delta\}$	0.342	0.983	0.352	3.438	0.831	0.783
$\{\sigma, \theta\}$	0.178	0.976	0.183	2.118	0.837	0.483
$\{\sigma, \theta, \zeta\}$	1.433	0.997	1.473	1.041	0.808	0.237
$\{\sigma, \theta, \delta\}$	0.159	0.974	0.163	2.501	0.783	0.570
$\{\sigma, \theta, \zeta, \delta\}$	1.446	0.996	1.487	1.466	0.770	0.334

Table 3.5.2: LatMix submesoscale diffusivity estimates and associated FVU and FDU, estimated over candidate models in the hierarchy at each site using either rolling window, or spline parameter estimates. The rolling-window method uses the second-moment fitting methods, whereas the spline method uses the first and second-moment fitting method.

3.5.2 Time-evolving parameters using rolling windows

We now apply the rolling-window estimates using the second-moment fitting method, as discussed in Section 3.4.2. To pick a suitable window length W, we see from Table 3.5.1 that diffusivity scales as order $0.1 - 1$ m^2/s, and the strain rate when converted to days is approximately 1/3 days. Therefore, using Figure 3.4.4 as a guide we choose a window length of $W = 1$ day (corresponding to 49 observations over 30-minute sampling intervals for each drifter). This choice also coincides approximately with the inertial and diurnal periods meaning inertial oscillations and tides will be relatively close to zero mean within the window, thus being closer to satisfying the zero-mean assumption of the average submesoscale residuals across drifters made in Equations (3.2.2)–(3.2.4).

Within Table 3.5.2 we provide the estimated submesoscale diffusivity, and FVU and FDU error metrics, using rolling one-day windowed mesoscale parameter estimates for each hierarchy. As expected, the FVU decreases everywhere (as more parameters are being fitted) in comparison to the fixed-parameter fits. The FDU values, on the other hand, decrease in some but not all cases, providing mixed evidence for time-variation. We notice the reductions in FVU and FDU are most pronounced for Site 2, indicating this is the site most likely to have a time-evolving mesoscale. Overall, there is now evidence for a time-varying strain-vorticity model. Including divergence is now a less favourable choice than with the earlier analysis with fixed estimates.

In Figure 3.5.2 we display some examples of the time-varying parameter estimates

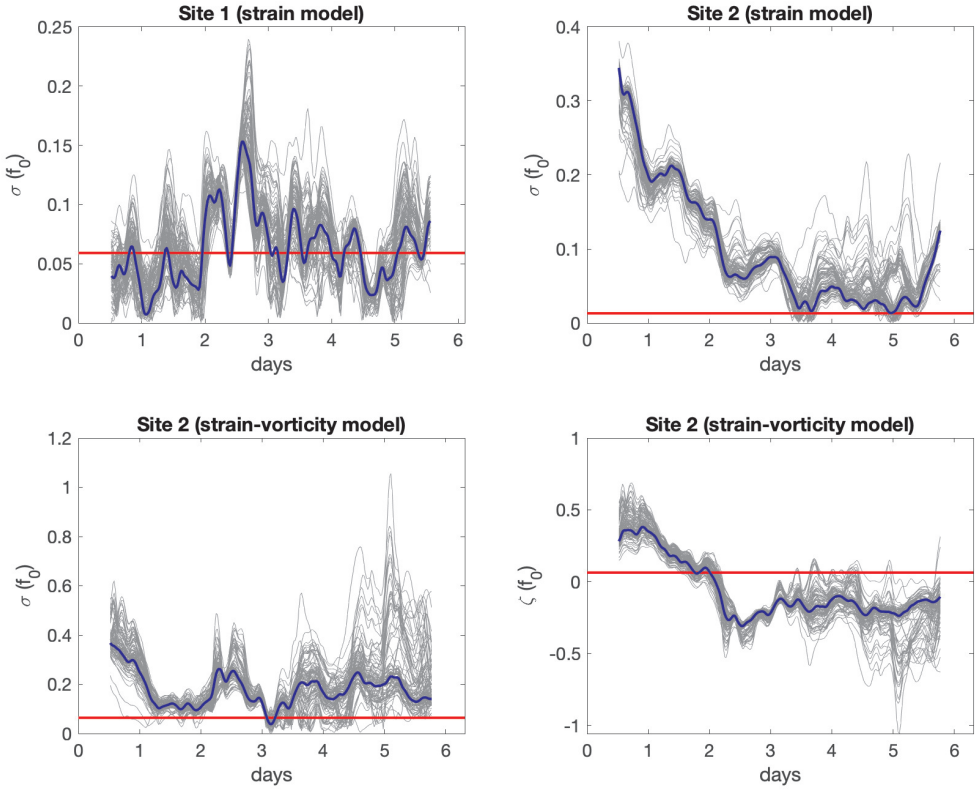

Figure 3.5.2: Fixed (red) and time-varying (blue) parameter estimates, where the latter are generated with a one-day rolling window using the second-moment fitting method. Top-Left: strain rate estimates with the strain-only model (Site 1). Top-Right: strain rate estimates with the strain-only model (Site 2). Bottom-Left: strain rate estimates with the strain-vorticity model (Site 2). Bottom-Right: vorticity estimates with the strain-vorticity model (Site 2). 100 bootstrapped time-varying trajectories are shown in grey in each subplot.

using this approach. In the top panels we show the strain rate over time at each respective site using a strain-only model, where the evidence for temporal evolution at Site 2 is clear. We overlay bootstrap trajectories of these time series (as well as the fixed parameter estimates from Table 3.5.1) which indicates this variation appears significant at Site 2, but largely not at Site 1. Also, the low values for strain rate of $\approx 0.01 f_0$ in the fixed-parameter estimate appears to be a misfit due to model misspecification from not allowing time-variation. The values for the strain rate are now larger at Site 2 than at Site 1 when allowing time evolution, as expected. In the bottom panels we show the time-varying strain rate and vorticity estimates using a strain-vorticity model. Again there is evidence for time-variation which we will explore further with spline fitting.

Although the parameter estimates obtained using rolling windows are overfitted and not slowly varying, these fits however provide an extremely useful *lower* bound, in terms of interpreting estimated submesoscale diffusivities and FVU/FDU values. This will help guide the implementation for modelling time-variation more smoothly using significantly fewer parameters in the spline methodology that follows. In contrast, the fixed parameter estimates provide a useful *upper* bound on diffusivities and FVU/FDU values, as this approach is the most parsimonious.

3.5.3 Slowly-evolving parameters using splines

We continue our analysis of the LatMix data by fitting time-evolving mesoscale parameters using the splines approach defined in Section 3.4.3. We will use the full first *and* second-moment fitting method allowing us to make a complete decomposition of

the flow at both sites into background, mesoscale, and submesoscale components.

First, in Figure 3.5.3 we compare estimates of strain rate between the second-moment and the first and second-moment fitting methods during the first two days of the LatMix Site 1 experiment. This particular window has relatively low strain rates that may not be distinguishable from zero, as seen in the top-left panel of Figure 3.5.2. Using the bootstrap estimates and a kernel density estimator, the left panel of Figure 3.5.3 shows the distribution of strain rates using the second-moment fitting method. While the peak of the distribution is consistent with the strain rate estimated over the entire six day experiment, the 90% contour of the distribution includes an enormous range of strain rates, including zero. In contrast, by including the first-moment as part of the fitting method, the right-panel of Figure 3.5.3 shows a narrower range of strain rates that do not include zero. Thus, at least in this example, the combined first and second-moment fitting method provides more robust estimation than the second-moment fitting method by including extra information in the fit.

In Figure 3.5.4 we display the time-evolving parameter estimates at Sites 1 and 2 using a strain-only and strain-vorticity model respectively. We overlay confidence intervals obtained using the bootstrap procedure outlined in Section 3.4.3. The time evolution of the strain-vorticity parameters is clear at Site 2, where all 3 mesoscale parameters $\{\sigma, \theta, \zeta\}$ are seen to change in a smooth fashion across the 6 days. In contrast, at Site 1, evidence of time variability for the strain rate is less clear, as the estimate of constant strain rate (dashed-line) fits entirely within the confidence intervals. Figure 3.5.4 also shows estimates of $\{u_0, v_0\}$, but their particular values

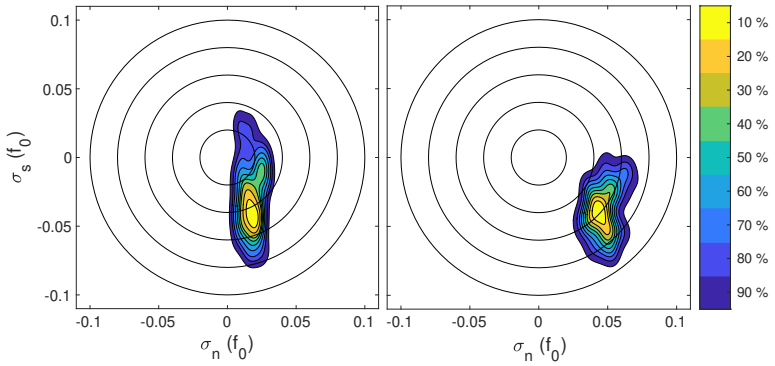

Figure 3.5.3: Distribution of strain rate parameters estimated for the first two days of the LatMix experiment at Site 1. Contours indicate the percentage of samples enclosed. The left panel shows estimated strain rate parameters using only the second-moment fitting method, where the right panel shows estimates using the first and second-moment fitting method.

are not directly interpretable, as they depend on the location of the expansion point, $\{x_0, y_0\}$. Instead, from Equation (3.2.3), it can be seen that they contribute to the mesoscale description of the flow at the location of the centre-of-mass.

We include the submesoscale diffusivity estimates, as well as FVU and FDU values, in the bottom portion of Table 3.5.2, along with comparison values from a hierarchy of models at each site. What we observe is quite remarkable: we can achieve FVU and FDU values that are very close to the rolling window estimates, despite using *significantly* fewer parameters to describe the evolution of the mesoscale velocity field. The evidence from Table 3.5.2 continues to support the choice of a strain model at Site 1 (with minor evidence for the additional presence of divergence), and a strain-vorticity model at Site 2. The estimated submesoscale diffusivities after performing the fits are around $\kappa = 0.2$ m^2/s at Site 1 and $\kappa = 1.0$ m^2/s at Site 2, nearly an

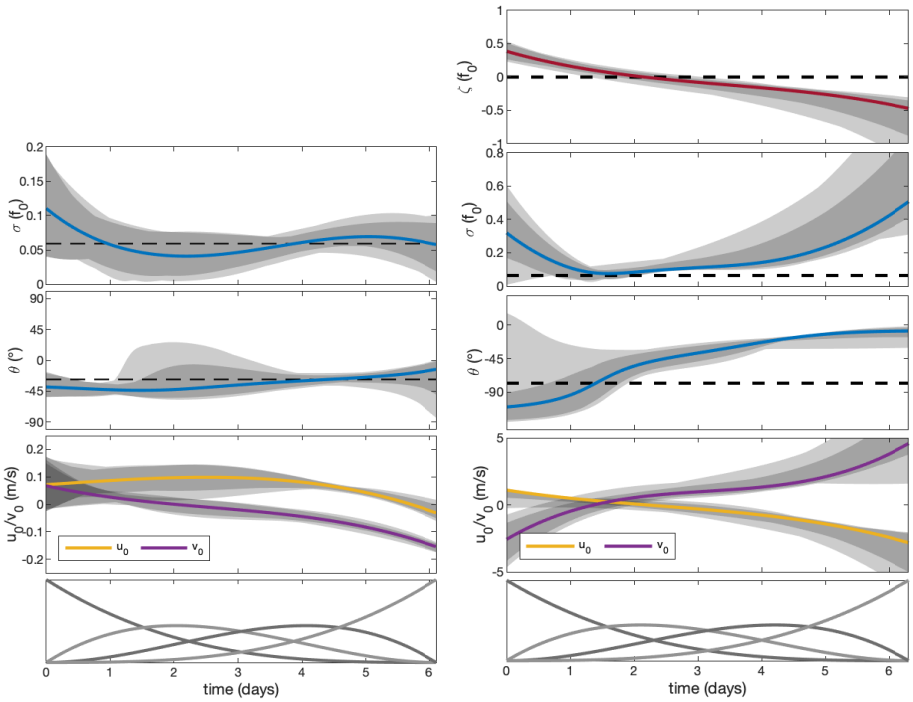

Figure 3.5.4: Parameters of the spline based strain model fits to Site 1 (left panel) and strain-vorticity model fits to Site 2 (right panel) using the first and second-moment fitting method. The most likely solution is highlighted, with 90% and 68% most likely solutions shown in grey and dark grey, respectively. The models are fit using four degrees of freedom per parameter with the splines shown in the bottom row.

order-of-magnitude difference.

Finally, we complete our analysis of the LatMix data by using the spline fits of Figure 3.5.4 to decompose the flow into the three components of our conceptual model of Equation (3.2.1)—background, mesoscale, and submesoscale—and then integrate over time to construct an implied set of drifter trajectories for each component. This is displayed in Figures 3.5.5 and 3.5.6 for Site 1 and Site 2 respectively. We have also included the mesoscale component in centre-of-mass coordinates. We observe that the mesoscale components meander in the fixed reference frame and follow the observed

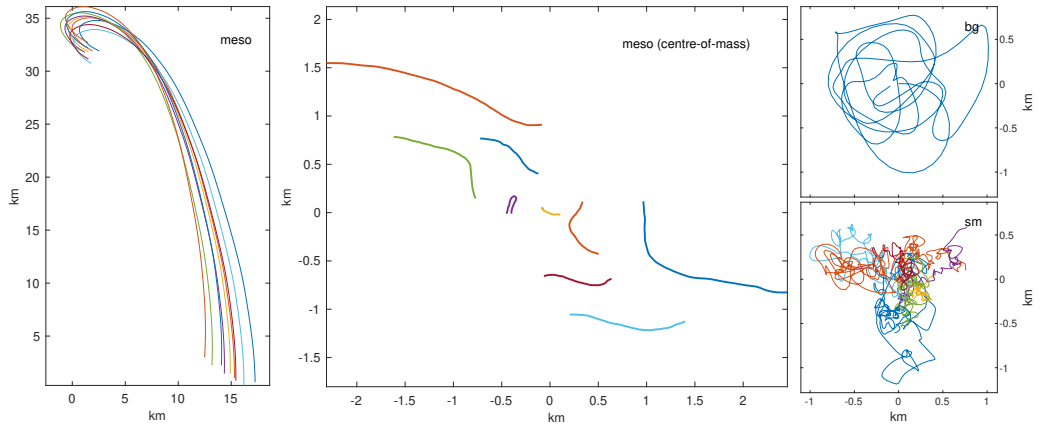

Figure 3.5.5: Decomposition of the flow at LatMix Site 1 using the strain-only model fitted with splines using the first and second-moment fitting method. The left panel shows the the mesoscale solution in the fixed coordinate reference frame (compare to the upper-left panel of Figure 3.5.1). The centre panel shows the same solution in the centre-of-mass frame (compare to the lower-left panel of Figure 3.5.1). The top-right and bottom-right panels show the path-integrated background and submesoscale flow, respectively.

particle paths explaining most of their displacement *and* explain some of the spreading in the centre-of-mass frame. This can be seen by directly comparing Figures 3.5.5 and 3.5.6 with Figure 3.5.1. The submesoscale components are random-walk like and broadly resemble a diffusive process. The background components contain inertial oscillations and tides which create looping trajectories with roughly daily periodicity.

Figure 3.5.7 shows the Lagrangian spectra of the background flow, the mean (across drifters) of the mesoscale flow, and the mean (across drifters) of the submesoscale flow, for Sites 1 and 2 respectively. A number of features standout in Figure 3.5.7. The Coriolis frequency is almost exactly the diurnal frequency at this latitude,

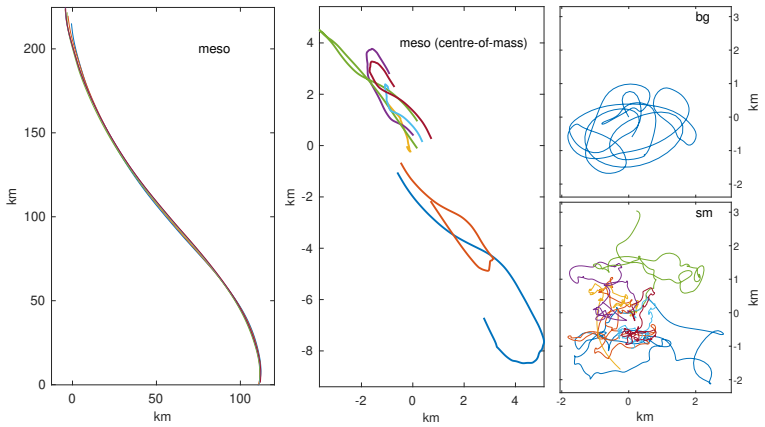

Figure 3.5.6: Same as Figure 3.5.5, but for LatMix Site 2 using the strain-vorticity model. The mesoscale solution in fixed frame can be compared to the upper-right panel of Figure 3.5.1), and the mesoscale solution in centre-of-mass frame can be compared to the lower-right panel of Figure 3.5.1

and this has the effect of creating a relatively substantial peak of energy on the anti-cyclonic side of the spectrum of the background flow at Site 1, with no corresponding peak on the cyclonic side. This means that the oscillation is anticyclonic and nearly circular. Furthermore, the semi-diurnal tide appears primarily on the cyclonic side, although with some energy on the anticyclonic side. The background flow at Site 2 shows significantly more power, especially at lower frequencies and also has a strong inertial signal. The mesoscale flow at Site 2 is much stronger than Site 1, as expected.

If the drifters were governed by the stochastic model given with Equation (3.2.9), then removing the effects of the strain in centre-of-mass coordinates would reveal a submesoscale signal given by increments of the Wiener process. The Lagrangian power spectrum would show a (flat) white noise process. However, Figure 3.5.7 shows that the submesoscale spectra from both Site 1 and 2 have significantly more structure.

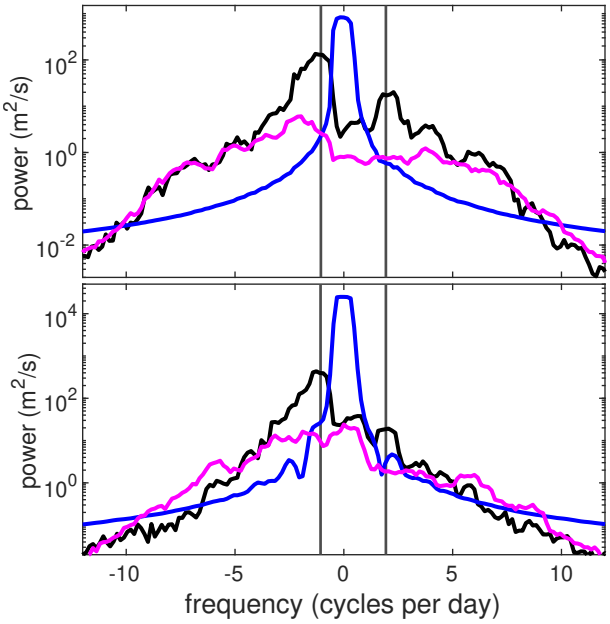

Figure 3.5.7: The top and bottom panels show the power spectra of the decomposed flow for Sites 1 and 2, respectively. The spectra shown are the spatially homogeneous background flow $\mathbf{u}^{\mathrm{bg}}$ (black), the average of the mesoscale component of the flow $\mathbf{u}^{\mathrm{meso}}$ (blue), and the average of the submesoscale component $\mathbf{u}^{\mathrm{sm}}$ (magenta). Anticyclonic oscillations are indicated by negative frequencies and cyclonic oscillations by positive frequencies. The vertical lines indicate the semi-diurnal tidal frequency and the inertial frequency on the positive and negative side, respectively.

The spectra are characterised by low power at sub-inertial frequencies, roughly an order of magnitude more power on the anticyclonic side than the cyclonic side at near inertial frequencies, and a decay of power at higher frequencies.

3.6 Discussion and Conclusion

The separation in Equation (3.2.1) is a compelling conceptual model, based on the ideas of non-local spreading in turbulence theory—but is the separation actually doing something useful in practice? This idea can be tested by considering the cross-terms in the total energy of the model, as was done in Lelong et al. (2020). Specifically, the cross terms in the kinetic energy equation,

$$\mathbf{u}^2_{\text{total}} = \mathbf{u}^2_{\text{bg}} + \mathbf{u}^2_{\text{meso}} + \mathbf{u}^2_{\text{sm}} + 2\left(\mathbf{u}_{\text{bg}}\mathbf{u}_{\text{meso}} + \mathbf{u}_{\text{meso}}\mathbf{u}_{\text{sm}} + \mathbf{u}_{\text{bg}}\mathbf{u}_{\text{sm}}\right), \tag{3.6.1}$$

should remain small if this is truly an orthogonal linear decomposition. To assess this quantity we compute the coherence between the complex submesoscale signal and the complex mesoscale signal in the centre-of-mass frame, as shown in Figure 3.6.1. The results show remarkably low coherence ($\mathcal{O}(0.1)$) at Site 1, across all frequencies, suggesting no relation between the two signals. In contrast, Site 2 does show more coherence between the two signals, likely reflecting the challenges of the separation in time-varying conditions. Despite this, the average coherence across frequency bands is ≈ 0.2, suggesting the decomposition is successfully separating two mostly distinct signals. The validity of this separation can be made precise using the methodology that unambiguously separates waves and geostrophic motions at each instant in time

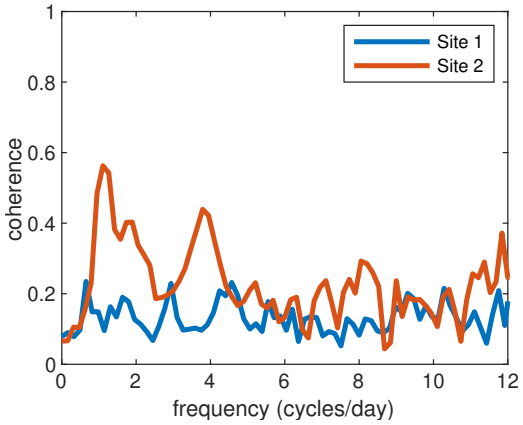

Figure 3.6.1: Coherence between the mesoscale signal in the centre-of-mass frame and the submesoscale signal at Site 1 and Site 2, using the disentangled velocities corresponding to the trajectories shown in Figures 3.5.5 and 3.5.6 respectively.

in an Eulerian reference frame (Early et al., 2021).

One of the key strengths of this methodology is how few parameters are needed to estimate the mesoscale parameters and perform the decomposition. For example, at Site 1 there are $N = 294$ observations of position from $K = 9$ drifters, resulting in $2NK$ degrees-of-freedom. The second-moment fitting method uses $2N$ degrees-of-freedom to remove the centre-of-mass. Using a single window across the entire time series to estimate the 2 parameters in the strain model, such as Site 1 which is well described by a single set of strain rate parameters across the entire window, leaves $2N(K-1)-2$ degrees-of-freedom to describe the submesoscale flow. In contrast, daily rolling windows with $N_W = 49$ points (corresponding to one day) that estimate strain rate parameters at each of the $N - N_W$ time points, leaves only $2N(K-2) + 2N_W$ degrees-of-freedom to describe the submesoscale flow. As is evident in Figure 3.5.2, these extra degrees of freedom capture time-variability in the parameters that may not

be appropriate. Finally, the spline fits require estimating M coefficients per mesoscale parameter, and thus the spline based time-varying fits leave $2N(K-1)-2M$ degrees-of-freedom to describe the submesoscale flow using the second-moment fits. With $M = 4$ sufficient to capture any time variability at Sites 1 and 2, this approach uses remarkably few parameters to perform this estimation. The benefit of which is that the decomposed submesoscale trajectory will contain rich statistical information with which to do further Lagrangian analysis.

As discussed in the introduction, we view this works as complementary to that of Gonçalves et al. (2019); Lodise et al. (2020) who recently developed a method for projecting clustered drifter trajectories to local Eulerian velocity fields using Gaussian Process regression. The ultimate goal of Lodise et al. (2020) was to compute horizontal velocity gradients with which to better understand vertical transport. The method was applied to the CALYPSO an LASER drifter deployments. Applying our method to these datasets is a natural avenue for further investigation. More broadly speaking, what our method provides to complement Gonçalves et al. (2019); Lodise et al. (2020), is not the *Eulerian* velocity field, but rather the *Lagrangian* decomposition of the trajectories into various components. This allows us to extract the specific submesoscale component from the trajectory for further analysis within the Lagrangian setting. This allows for the estimation of submesoscale diffusivity, which is not a topic covered in Gonçalves et al. (2019); Lodise et al. (2020). However, there is certainly scope to merge and compare our methodologies, particular because the constructed Eulerian velocity field can be directly compared with the mesoscale parameters we estimate locally over time (and hence space) using our slowly-evolving

spline fits. Again, this is certainly a topic that warrants further investigation. We also see potential for our work to naturally follow-on from the recent methodology developed in Vieira et al. (2020) who identify clusters of drifter trajectories that share coherent structures. For example, such clustering could be used to divide larger deployments into smaller clusters, after which our method can then be applied to each cluster to separate flow components within coherent structures.

Chapter 4

Sensitivity Analysis of the Drifter Deployment Configuration

4.1 Introduction

In this chapter we shall build on Chapter 3 by investigating the sensitivity of the simulation results with respect to the number of drifters in the cluster, as well as the cluster *morphology* (i.e. the spatial distribution of the initial deployment configuration). Our simulation results in Chapter 3 were using 9 drifters configured to start as at Latmix Site 1—and we used these results to motivate and help interpret our real data analysis of the Latmix data. In other drifter deployments however, the number of drifters, the number and duration of position observations, and the deployment configuration will all potentially vary, and we now investigate what impact this may have through simulation studies.

This chapter is structured as follows. In Section 4.2 we investigate how the num-

ber of drifters in the deployment impacts upon the error of the mesoscale parameter estimates. We show how the error scales inversely with the number of drifters. Section 4.3 performs a similar analysis to Section 4.2, but changing the number of observation time points in the experiment for each drifter. The number of time points can be changed by either observing the drifters more frequently, or by allowing the deployment to run for longer at the same temporal sampling rate. We investigate both these possibilities and how they affect the error of the parameter estimates. Section 4.4 looks at varying the initial configuration of the drifter deployment. There are many different ways that the initial deployment configuration could be changed, but we limit this analysis to the following: (a) changing the average initial distance of the drifters from the centre of mass, (b) rotating the initial deployment relative to the strain angle, and (c) changing the eccentricity of the deployment configuration. This final analysis (c) shall consist of deploying drifters in an ellipse, varying how eccentric this ellipse is, and then looking at the errors for each eccentricity averaged over different possible angles that the ellipse could be aligned with. In our simulations, one or more mesoscale parameters could be estimated (and the others assumed known), and so we repeat this analysis for different combinations of known and unknown parameters. In Section 4.5 we conclude our analysis from this chapter and discuss the optimal drifter deployments based on our simulation studies.

4.1.1 Simulation Set-Up

In this section, we will introduce the simulation set-up that will be used throughout this chapter, and comment on how this affects the validity of our results.

We base our simulations upon the LatMix experiment at Site 1. Therefore, if a parameter is present in the model, it is chosen to be the value that was used in Figure 3.2.1. Specifically, the parameters are set as $\sigma = 7 \times 10^{-6}$/s, $\theta = 30°$, $\zeta = 6 \times 10^{-6}$/s, and $\zeta = 8 \times 10^{-6}$/s. The simulations last for one day, with sampling rate $\Delta = 30$ minutes. The submesoscale diffusivity will follow an isotropic 2D Wiener process with the isotropic diffusivity equal to 0.1m^2/s. Standard errors of parameter estimates will be calculated by estimating parameters over 100 repeated simulations. Initial positions will be chosen iostropically (unless otherwise stated) where the expected distance to the centre of mass is chosen to match that in Site 1 of the LatMix experiment.

We assume throughout this chapter that the stochastic noise (i.e. the Wiener component) of each drifter observation is independent across time and across drifters. This would likely only be true in real drifters if the observations occur with sufficiently large length and time scales.

The simulated particle trajectories contain only the parameters that we are estimating. This would not be the case when we estimate using real drifter trajectories. For example, we commented in Chapter 3 that there was evidence of internal waves in the region covered by the LatMix experiment, and we do not include internal waves in our simulations. Therefore, any results given in this chapter can only be used as a guide, and are not fully representative of the errors from a real drifter deployment.

The scaling behaviours found in this chapter are expected to be the same as that from drifter deployments such as LatMix, as the scaling behaviours are unchanged by the inclusion of additional mesoscale parameters, or the value of each simulated parameter. Therefore the results in this chapter are designed to be useful in choosing

how to deploy clustered drifter arrays. In Chapter 5 we expand upon the results given in this chapter, where the results can be found for other values of the mesoscale and submesoscale parameters.

4.2 Varying the Number of Drifters

First we vary the number of drifters, denoted K, where we keep the same notation as Chapter 3. Throughout this section we will also use the model and inference procedure introduced in Chapter 3 to estimate mesoscale parameters, unless stated otherwise and parameters will be estimated as fixed quantities across this entire window.

In Figure 4.2.1 we report the relative standard error of mesoscale parameters (strain rate, strain angle and vorticity) in the strain-dominated simulation of Figure 3.2.1. The initial drifter positions are sampled isotropically with expected distance to centre-of-mass fixed over all experiments to be identical to Latmix Site 1. Relative standard error is computed for each parameter by dividing the observed sample standard error by the true parameter value. We have included a reference line that scales as $1/\sqrt{K}$ which is the asymptotic limit we expect to see standard errors reduce by from asymptotic theory when observations are independent. For this simulation environment we see that this scaling behaviour is approximately correct for $K > 5$. We emphasise that in practice this scaling behaviour will not apply to real deployments. Here we have simulated drifters that experience independent submesoscale errors across drifters, which is an idealised scenario. In reality an increasing number of densely packed co-located drifters will experience correlated motions thus

eventually limiting the amount of new information that can be gained by adding more drifters to a cluster. Nevertheless, the simple rule from the observed scaling behaviour is that one must have approximately four times as many drifters to reduce the standard error by a factor of two. At least three drifters are needed before the parameter estimates become statistically significant. However, the SE decreases rapidly (faster than the rate of $1/\sqrt{K}$) until we have around six drifters before the decrease in the SE begins to slow, and so one could argue that it is better to have at least around six drifters to reduce the SE significantly, at least in this simulation scenario.

4.3 Varying the Number of Observation Time Points

We now vary the number of observed time points for the drifter positions in the simulation. This can be done by (a) increasing the simulation time T whilst keeping the sampling time between observations Δ constant, or (b) decreasing Δ while keeping T constant.

In Figure 4.3.1 we display the relative standard error for strain only and vorticity only models in scenario (a), along with a reference line of $1/\sqrt{T}$. Drifters were simulated as in the strain only model of Figure 3.2.1, replacing strain with vorticity of 6×10^{-6}/s for the vorticity only model. The relative standard error follows the $1/\sqrt{T}$ relationship for the case where we only have vorticity. For the strain-only simulation this relationship is followed for shorter simulations, however if we allow the simulation to run longer then we see a faster decrease in reducing the standard error.

The difference in the two simulations is due to the nature of each mesoscale feature

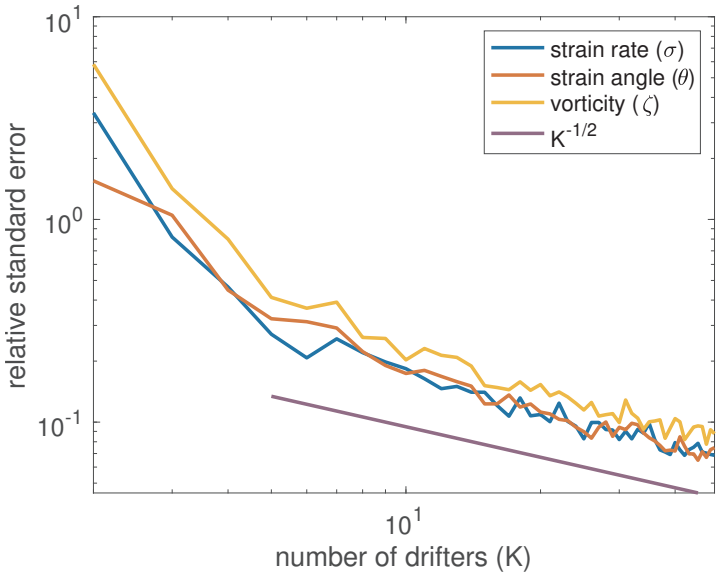

Figure 4.2.1: Relative standard error of stain rate, strain angle and vorticity for a varying number of drifters K, averaged over over 100 repeated simulations for each K. The simulation setup is as in Figure 3.2.1 in the strain-dominated model with trajectories simulated for 1 day. The initial drifter positions are sampled isotropically with expected distance to centre-of-mass fixed over all experiments to be identical to Latmix Site 1. Relative standard error is computed by dividing the observed sample standard error by the true parameter value. Therefore in this experiment we require approximately 3 drifters in the cluster before the standard errors are approximately half the true parameter value (and hence significantly non-zero).

in a linear velocity field with zero mean flow: vorticity causes drifters to orbit the centre of mass, whereas strain will move drifters further from the centre of mass with time. The strain rate is set to be constant throughout the simulation, resulting in the velocity increasing exponentially as time increases. Diffusivity is also set to be constant, and so the increased velocity results in a larger signal to noise ratio, and thus a faster reduction in standard error. In the case of vorticity only, the average speed of the drifters is constant and so the expected standard error is simply inversely proportional to the square root of the number of observation time points from asymptotic theory.

We now consider scenario (b): the effect on the relative standard error of increasing the number of time points by reducing the time between observations Δ. We require the length of simulation T to be $1/\sigma$ or $1/\zeta$ in order to reliably estimate strain or vorticity respectively, which corresponds to $T \approx 5/3$ days for strain ($\sigma = 7 \times 10^{-6}$/s) and almost two days for vorticity ($\zeta = 6 \times 10^{-6}$/s). Therefore, in Figure 4.3.2, which displays the relative standard errors, we set the length of the simulation to be constant at $T = 2$ days, and otherwise the drifters are simulated as in Figure 4.3.1. When $\Delta = 2$ days (at the left extreme) we have just two observations—one at the start of the simulation and one at the end. In the deterministic setting, these would correspond to almost the same point for vorticity-only particle trajectories (with the variation due to the cycle period being just less than two days). Both plots of Figure 4.3.2 show a significant but small decrease in the standard error as we decrease Δ, thus increasing the number of observation time points.

For the examples shown in Figures 4.3.1 and Figure 4.3.2, increasing the number

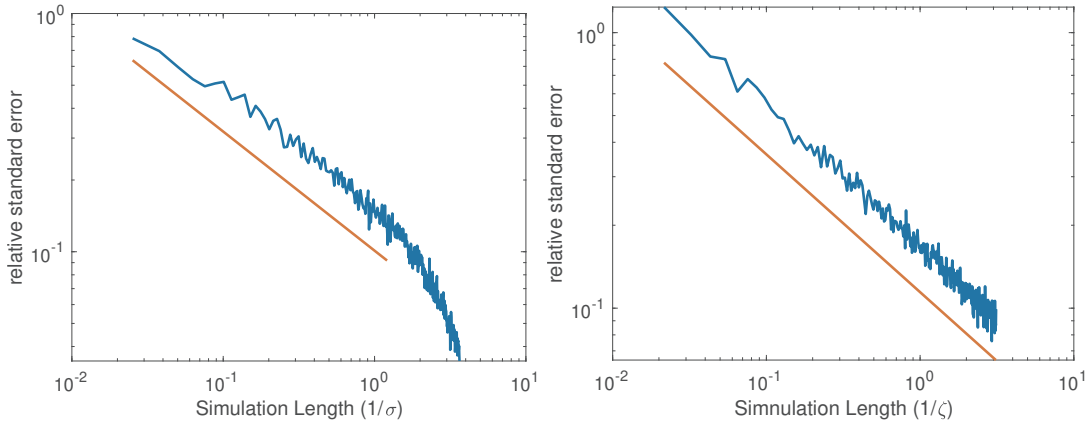

Figure 4.3.1: Relative standard error of stain rate (left) and vorticity (right) for 9 drifters and a varying simulation length T from 0 to 6 days (in increments of 30 minutes), plotted on a log scale. For each value of T the results are averaged over 100 repeated simulations. The x-axes are plotted in units of $1/\sigma$ and $1/\zeta$ respectively. The initial drifter positions are set to be identical to Latmix Site 1, and the time between observations is $\Delta = 30$ minutes. Relative standard error is computed as in Figure 4.2.1. The left and right panels show relative standard errors for strain only ($\sigma = 7 \times 10^{-6}$/s) and vorticity only ($\zeta = 6 \times 10^{-6}$/s) models respectively.

of observation time points by decreasing Δ does not appear to have as significant an impact to the standard error as adding to the total length of the simulation. This is because infilling between observations does not provide as much information about the mesoscale effects as collecting data over a longer time window. Therefore, to reduce the standard error, it is better to increase the total length of the simulation than to decrease the time between observations. This finding is of course idealised because (a) the assumption we have made of spatially and temporally fixed mesoscale parameters is of course unrealistic, and (b) the sampling rate of drifter observation cannot always be controlled or altered. In reality, the velocity field will only be *locally* stationary, therefore there may be some practical time limit for which mesoscale parameters might be stationary. For more discussion of this please see the earlier work in Chapter 3 on optimal window length selection for estimating mesoscale parameters for the LatMix drifters.

4.4 Varying the Initial Configuration

We can also vary the initial cluster morphology in our simulated environment—where by cluster morphology we refer to the initial spatial configuration of the drifter deployment at first measurement. In Figure 4.4.1 we consider two classes of configurations. First, we consider deployments where the drifters are initially configured to be either closer or further apart than in Latmix Site 1, as shown in Figure 4.4.1 (left). Specifically, the red drifters are twice as far from the centre-of-mass as Latmix Site 1 drifters (in blue), and the green drifters are half this distance. We repeat the same

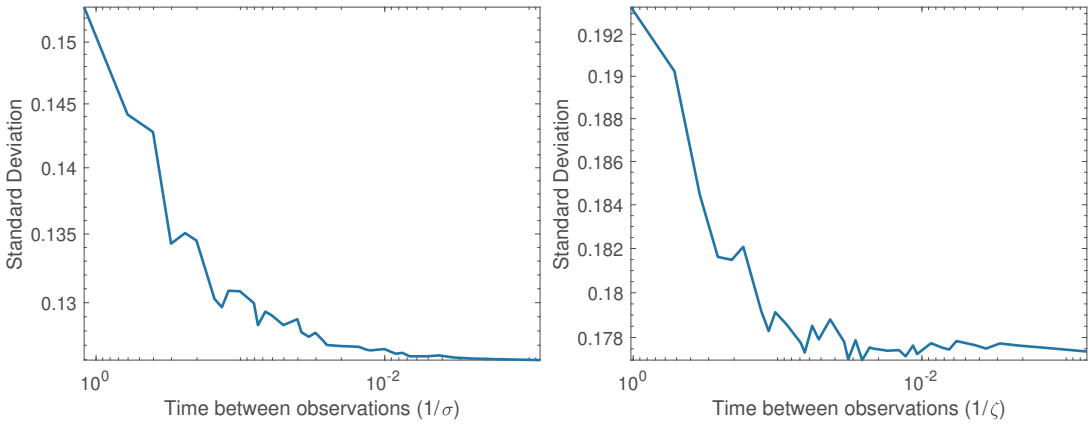

Figure 4.3.2: Relative standard error of stain rate (left) and vorticity (right) for drifters simulated as in Figure 4.3.1, plotted on a log scale, except varying the time between observations Δ on a decreasing scale (from $\Delta = 2$ days to 2 minutes in increments of 2 minutes, using only values of Δ which exactly divide into T) whilst keeping the length of the simulation fixed to $T = 2$ days. Again the x-axes are plotted in units of $1/\sigma$ and $1/\zeta$ respectively where $\sigma = 7 \times 10^{-6}$/s and $\zeta = 6 \times 10^{-6}$/s.

analysis as Figure 3.4.4 over different true strain rates to find the required window lengths (the segment of the time series over which we will estimate the parameters, see Section 3.4.2) in the left panel of Figure 4.4.2. We observe that drifters initialised further apart require shorter window lengths to obtain significant strain rate estimates, and conversely require longer window lengths when initialised closer together. This phenomenon is easily understood in this idealised simulation scenario where spacing drifters farther provides richer information on mesoscale features as distances to centre-of-mass are increased. In practice the flow field is not spatially homogeneous, so as with the number of drifters (Section 4.2) and length of simulation (Section 4.3), here there will be a practical limit as to how far apart drifters should be initially placed to ensure they are sampling the same homogeneous background flow field.

In the right panel of Figure 4.4.1 we contrast Latmix Site 1 configurations (in blue dots) versus two other deployment configurations: one that is parallel to the true strain angle (red dots), and another that is orthogonal to the true strain angle (green dots). For each colour, the average distance to centre-of-mass is held constant to 0.6km. To see how this affects parameter estimation we repeat the analysis of Figure 3.4.4 to find the required window length to get significant estimates of the strain rate over a range of true strain rate values—these are displayed for each configuration in the right panel of Figure 4.4.2. We see that the required window length is significantly reduced when the configuration is aligned parallel to the strain angle (red drifters), and conversely the required window length is increased when this is orthogonal (green drifters). The results with the Latmix configuration, which is more isotropic, are sandwiched in between.

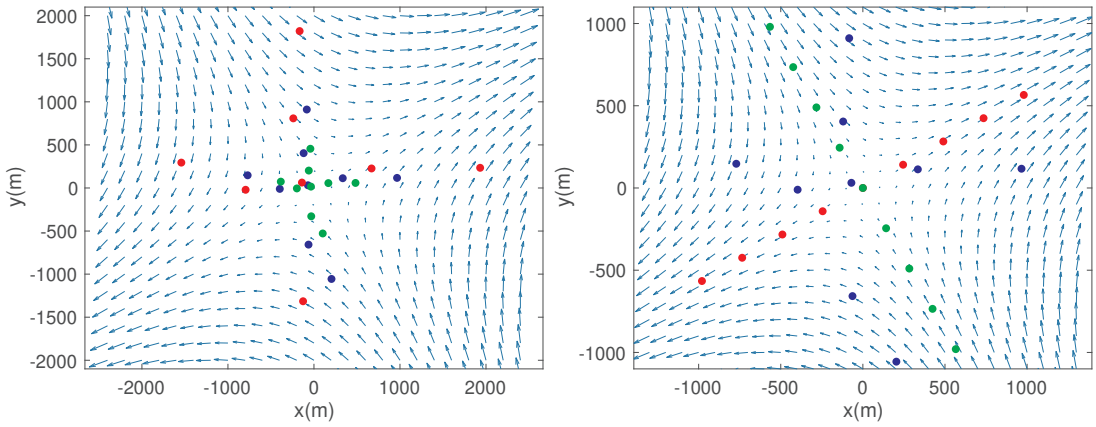

Figure 4.4.1: Different cluster morphologies (deployment configurations) we shall consider. In the left panel we again consider 9 drifters deployed as at Latmix Site 1 (blue dots), together with 9 drifters deployed with the same morphologies but with the respective distances to the centre-of-mass either doubled or halved (red and green dots respectively). In the right panel we consider 9 drifters deployed as at Latmix Site 1 (blue dots), but this time the red and green dots are drifters deployed parallel and orthogonal to the strain angle (red and green dots respectively) but with the same average distance to centre-of-mass as the blue dots. In both panels the velocity field is as in the strain-only simulation of Figure 3.2.1, and the positions are given in centre-of-mass coordinates.

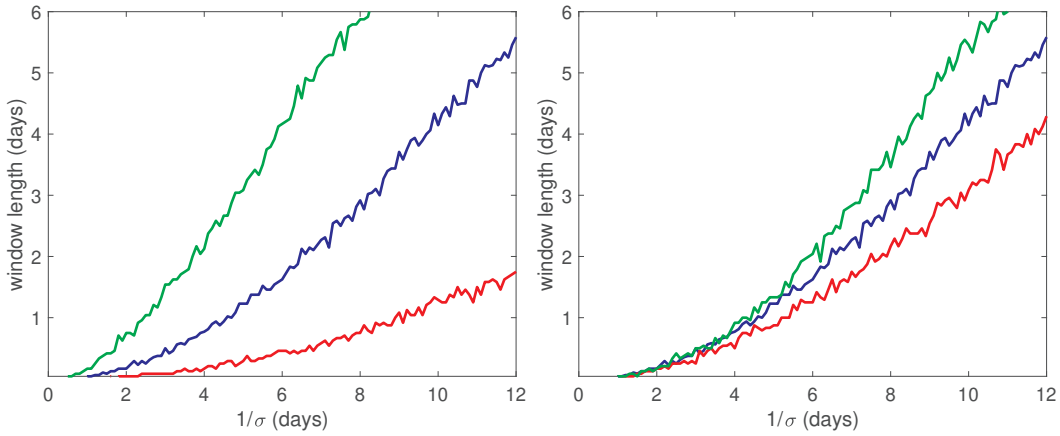

Figure 4.4.2: Required window lengths to obtain significant strain rate estimates for different drifter configurations. The lines in the left/right panels correspond to the drifter configurations considered in the left/right panels of Figure 4.4.1 respectively, with the colours matching the corresponding configurations. Each line corresponds to the level where the standard error of the strain rate estimate is approximately half the true strain rate value. These lines are found as in Figure 3.4.4 over 100 repeated simulations over a grid of true strain rates and window lengths.

This analysis shows that in a strain-only field (with no vorticity or divergence), then the optimal morphology is to align drifters along the expected strain angle—but more investigation is needed to understand how the optimal configuration may change in the presence of vorticity and/or divergence, as well as background and submesoscale effects. For example, Ohlmann et al. (2017) showed that an isotropic configuration has the lowest error for estimating divergence, whereas configurations along a straight line, such as those in Figure 4.4.1, have the largest errors. This is in contrast to our results for a strain-only field, where the LatMix configuration is the most isotropic yet has higher error than aligning drifters along the strain angle. Therefore, the optimal morphology appears to be dependent upon the mesoscale features present in the data, and unless these are known *a priori* then the best model-agnostic morphology is likely to be an isotropic cluster. Motivated by this, for the remainder of this section we will setup a new simulation environment to more thoroughly explore how the optimal morphology changes when we estimate different combinations of mesoscale parameters.

4.4.1 Simulation Set-Up

We shall consider how varying the eccentricity of the initial deployment configuration impacts upon the standard error of the resulting parameter estimates. Eccentricity, e is a measure of how circular an ellipse is, with $e = 0$ being a perfect circle and $e = 1$ a straight line. Therefore the red and green simulations of Figure 4.4.1 have deployment eccentricity equal to one.

We shall simulate the initial position of K drifters from a bivariate normal dis-

tribution with some arbitrary mean, standard deviation σ_x and σ_y (in X and Y components), and zero covariance. We shall set σ_x and σ_y such that the average expected distance from the centre of mass is the same as that in the Latmix experiment at Site 1. In the isotropic case ($\sigma_I = \sigma_x = \sigma_y$) the expected distance to centre of mass is calculated from the simulated drifters as

$$\text{cmd} = \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\sqrt{X_k^2 + Y_k^2}\right] = \sigma_I \mathbb{E}\left[\sqrt{\left(\frac{X_k}{\sigma_I}\right)^2 + \left(\frac{Y_k}{\sigma_I}\right)^2}\right]. \tag{4.4.1}$$

Defining $Z = \sqrt{\left(\frac{X_k}{\sigma_I}\right)^2 + \left(\frac{Y_k}{\sigma_I}\right)^2}$ then this exactly follows a chi distribution with 2 degrees of freedom ($Z \sim \chi_2$) which has mean $\sqrt{2}\Gamma\left(\frac{3}{2}\right)$. The expected distance can therefore be written as

$$\text{cmd} = \sigma_I \sqrt{2}\Gamma\left(\frac{3}{2}\right), \tag{4.4.2}$$

and hence the standard deviation of the normal distribution can be set to be

$$\sigma_I = \frac{\text{cmd}}{\sqrt{2}\Gamma\left(\frac{3}{2}\right)}, \tag{4.4.3}$$

such that we can insert the cmd value from Latmix Site 1 to set σ_I in the isotropic case. For the general anisotropic case, we shall simulate initial particle positions from

$$X_k \overset{iid}{\sim} \mathbb{N}(0, \sigma_x^2), \tag{4.4.4}$$

$$Y_k \overset{iid}{\sim} \mathbb{N}(0, \sigma_y^2), \tag{4.4.5}$$

where $\sigma_x = \sigma_I/\rho$, $\sigma_y = \rho\sigma_I$, and ρ is a re-scaling factor such that $\rho > 1$. The re-scaling factor is related to the eccentricity of the deployment by $e = \sqrt{1 - (1/\rho)^4}$ from the properties of an ellipse. This gives randomly distributed anisotropic particles within an ellipse with the major axis aligned along the y axis.

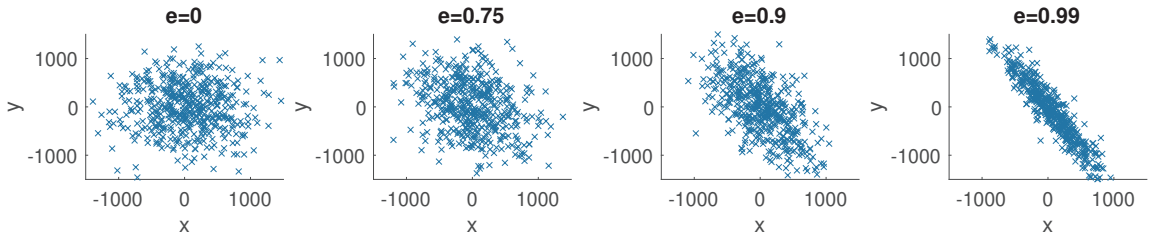

Figure 4.4.3: Initial positions of 500 drifters simulated from Equations (4.4.4) and (4.4.5). The average distance to the centre of mass is set to be the same as that in LatMix Site 1. These positions have eccentricity of (from left to right) 0, 0.75, 0.9 and 0.99 and the angle of the drifters is $-30°$ in each plot.

A rotation matrix can then be applied to the positions so that they can be aligned with or against the mesoscale features that are simulated in the data. Unless otherwise specified, the particle trajectories will be rotated randomly by sampling from a uniform distribution and multiplying the value by 360. This will result in a uniformly random angle between 0 to 360 degrees. Figure 4.4.3 displays 500 simulated particle initial positions from Equations (4.4.4) and (4.4.5) with different eccentricities.

Throughout this section, we will simulate particle trajectories with a sampling rate of 30 minutes for one day using nine drifters, motivated by the LatMix experiment. Fixed (non time-varying) mesoscale parameters will be estimated over the entire time series, and the submesoscale diffusivity is set as $0.1\text{m}^2/\text{s}$ for all simulations. Hence the particle trajectories are simulated similarly to the simulations we performed earlier in Section 4.3.

4.4.2 Single Parameter Estimation

First, we explore changing the deployment eccentricity for particle trajectories simulated using only one mesoscale effect: strain, divergence or vorticity. We display the relative standard error of mesoscale parameter estimates over 10,000 simulations in Figure 4.4.4 for varying eccentricities. The standard error for each parameter is calculated by estimating only the parameter for the mesoscale feature that is present, with all other mesoscale parameters set to zero. For the strain only model (left panel, $\sigma = 7 \times 10^{-6}$/s) we simulate particle trajectories for three different cases—the angle of the deployment ellipse aligned parallel to the strain angle, aligned orthogonal to the strain angle, and a random angle for each simulation. The corresponding standard errors are shown with the red, green, and blue curves respectively. We have also added horizontal dashed lines displaying the standard error for the deployments from Figure 4.4.1 (right). Note that the horizontal green and red dashed lines refer to the standard error of equally spaced particles, whereas the values of the solid curves at $e = 1$ correspond to randomly placed drifters along a line, we therefore do not expect these to be identical. The Latmix configuration (blue dashed line) appears to have a higher standard error than if the drifters were deployed randomly (with the same fixed distance to centre of mass), however this result does not consistently hold across all scenarios (as we shall shortly see). Furthermore, if the drifters are deployed randomly or parallel to the strain angle then on average it is better to deploy the drifters along a straight line with $e = 1$ than it is to use an isotropic deployment. In contrast, if the drifters are aligned orthogonally to the strain angle then a more

isotropic arrangement is preferable, and so the alignment of the drifters in relation to the strain angle determines the optimal configuration. Overall, aligning drifters parallel to the strain angle achieves optimal results here in the sense of minimising standard error.

For the divergence only (centre panel, $\delta = 3.8 \times 10^{-6}$/s) and vorticity only (right panel, $\zeta = 6 \times 10^{-6}$/s) simulations, we plot only the blue continuous line as there is no strain angle to align the drifters with or against. For these simulations, the standard error decreases for higher eccentricities in all cases, suggesting that it would be better to deploy drifters with $e = 1$, i.e. along a straight line. The Latmix deployment (blue dashed line) was better than an isotropic configuration where $e = 0$, however it is more optimal to position drifters along a single line for higher eccentricities when $e > 0.8$. Therefore, on average the optimal deployment morphology appears to be along a straight line for trajectories simulated with only a single mesoscale parameter, unless the trajectories are simulated using strain-only mesoscale and the initial configuration is aligned orthogonally to the strain angle.

Real-world data from drifter deployments will be much more complex than the above simulations which have assumed that there is no mean flow and only a single mesoscale parameter present. For example, Site 2 of the Latmix experiment showed that there was likely to be both strain and vorticity present, as well as other background effects. In Chapter 3 we presented parameter estimates from a strain-only model (assuming the vorticity to be zero) and from a strain-vorticity model. We therefore next consider a scenario where we simulate drifters from a strain-vorticity model, but estimate only strain rate, whilst assuming the vorticity to be zero. We now

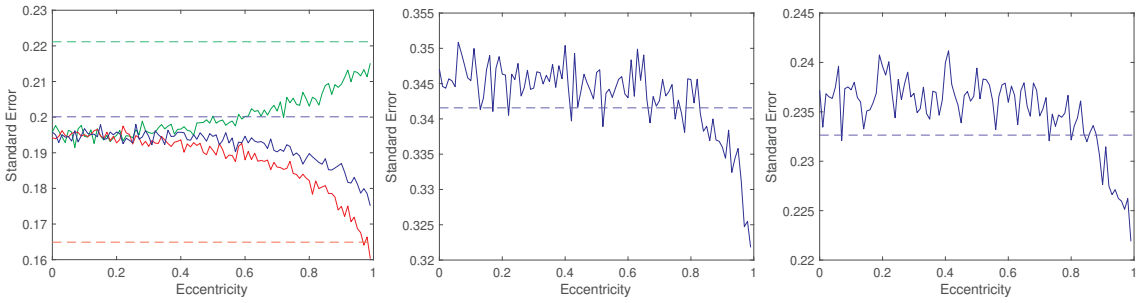

Figure 4.4.4: Standard error standardised by the parameter value over 10,000 simulations for different deployment eccentricity values for strain only (left panel, $\sigma = 7 \times 10^{-6}$/s), divergence only (centre panel, $\delta = 3.8 \times 10^{-6}$/s) and vorticity only (right panel, $\zeta = 6 \times 10^{-6}$/s) models. Particle positions are observed every 30 minutes for one day with diffusivity set at 0.1 m^2/s. Only the mesoscale parameters present in the data are estimated in each configuration. The continuous lines in each panel correspond to the average standard error for different deployment eccentricity values for drifters deployed perpendicular to the strain angle (green), aligned with the strain angle (red) and with a random angle (blue). The blue dashed lines correspond to drifters deployed with Latmix Site 1 initial positions. The green and red dashed lines correspond to equally spaced drifters deployed along a straight line that is aligned perpendicular to and parallel to the strain angle respectively. The red and green solid and dashed lines are only present in the left panel where strain is present.

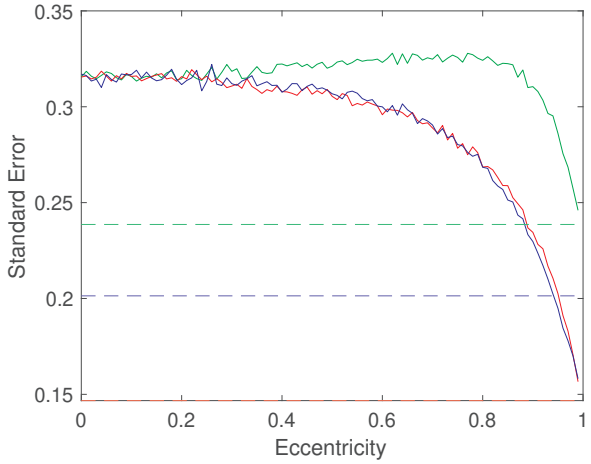

Figure 4.4.5: Standard error standardised by the parameter value of strain rate estimates over different deployment eccentricities. The drifter trajectories are simulated as in Figure 4.4.4 except using a strain-vorticity model with $\sigma = 7 \times 10^{-6}$ and $\zeta = 6 \times 10^{-6}$. The solid and dashed lines in each panel correspond to different deployment morphologies as in Figure 4.4.4.

obtain the standard errors as shown in Figure 4.4.5. It again appears to be optimal to deploy drifters along a single line, however the presence of vorticity now means that the angle of the deployment relative to the strain angle is less important as the drifters will be rotated by the vorticity. Therefore, deploying drifters in a line that is orthogonal to the strain angle (see green continuous line) results in a lower relative standard error than isotropic positions. However, the standard error is further minimised by aligning the drifters in a line that is parallel to the strain angle (red line) or simply any random angle (blue line).

So far, we have only studied the standard error when a single mesoscale parameter is estimated, and our results show the opposite to those in Ohlmann et al. (2017). Specifically, we showed that drifters deployed along a straight line have a lower stan-

dard error for most scenarios we showed, whereas Ohlmann et al. (2017) showed that drifters with an isotropic initial configuration will have lower standard errors. However, in the next section we will show that when more than one mesoscale parameter is estimated, the optimal drifter configuration will change.

4.4.3 Multiple Parameter Estimation

We now explore estimating different combinations of mesoscale parameters and how the number of estimated parameters affects the standard error of those estimates. First we estimate any two of strain, divergence and vorticity from drifter trajectories simulated using a correctly specified model with only those mesoscale features present. It is seen in Figure 4.4.6 that when more than one mesoscale parameter is estimated, $e = 0$ now minimises the standard error for some combinations of parameters — specifically when we estimate strain rate, strain angle and divergence $\{\sigma, \theta, \delta\}$ or strain rate, strain angle and vorticity $\{\sigma, \theta, \zeta\}$. However, if divergence and vorticity are estimated together $\{\delta, \zeta\}$, without any strain, then $e = 1$ minimises the standard error, although the range of the standard error across deployment eccentricities is small, meaning this result is unlikely to be practically significant.

These results show that when strain rate is estimated along with one other mesoscale parameter, then we are unable to disentangle their effects easily when the drifters are deployed in a single straight line. This can be easily understood for drifters in a strain-divergence field deployed in a straight line (aligned with the strain angle), as then it is difficult to determine whether the spreading motion along the strain angle is due to strain or divergence. Similarly, when we estimate strain and vorticity in a

strain-vorticity field, then placing the particles in a line does not exploit the spatial variability along different slices of the velocity field, making disentangling the effects all the more challenging. However, for the divergence-vorticity trajectories, the velocity field is rotationally invariant across space, as the divergence causes the drifters to spread or converge, whereas the vorticity causes the circular motion, and so these are much more easily separated regardless of spatial configuration. This explains why the standard error in the right panels is more or less similar across all eccentricities and orientations.

The standard errors when all mesoscale parameters are jointly estimated are shown in Figure 4.4.7. Each parameter was simulated in the drifter trajectories with $\sigma = 7 \times 10^{-6}$, $\zeta = 6 \times 10^{-6}$, and $\delta = 3.8 \times 10^{-6}$. Lower eccentricities have lower standard errors for each of the parameters, meaning that if all mesoscale parameters are present and being estimated then it is more optimal to deploy drifters in an isotropic configuration rather than anything eccentric. We note that the Latmix drifter configuration (blue dashed line) results in even lower standard errors than fully isotropic positions, as was also the case in Figure 4.4.6, suggesting that randomly deployed drifters might not fully minimise the error under expectation. The red and green dashed lines, corresponding to drifters deployed equally spaced along a straight line, have the highest standard errors for divergence, but not for vorticity or strain. This makes sense for strain since aligning the drifters with the strain angle (green dashed line) will decrease the standard error for strain rate (but not for the other parameters we are jointly estimating). Drifters along a single line can only estimate spatial gradients in a single dimension, and hence only a single parameter can be reliably

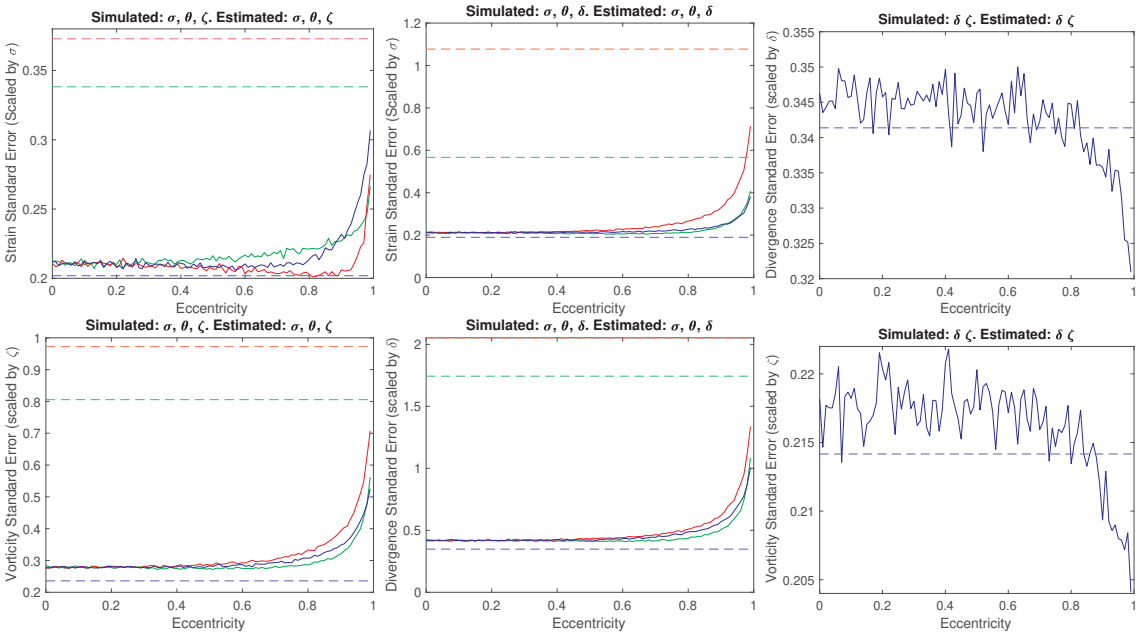

Figure 4.4.6: Standard errors standardised by the parameter value of mesoscale parameter estimates for varying eccentricities for strain-vorticity (left column), strain-divergence (centre column) and divergence-vorticity (right column) models. Within each model (or column) two parameters are estimated as indicated by the y-axis label. The drifter trajectories are simulated as in Figure 4.4.4, aside from the mesoscale parameters present in the model which vary in each column. Each solid or dashed line in each panel corresponds to a different deployment morphology as in Figure 4.4.4. The red and green solid and dashed lines are only present in the left and centre columns where strain is present.

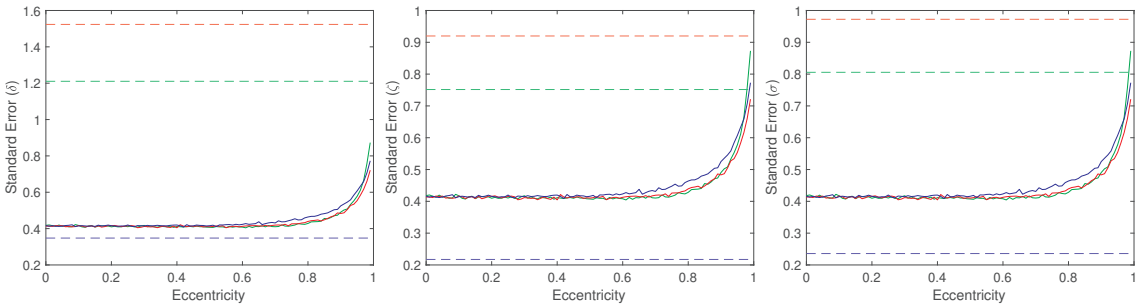

Figure 4.4.7: Standard errors standardised by parameter value of parameter estimates for varying eccentricities when all mesoscale parameters are simulated and then estimated. The drifter trajectories are simulated as in Figure 4.4.4, aside from the mesoscale parameters present in the model ($\sigma = 7\times10^{-6}, \zeta = 6\times10^{-6}, \delta = 3.8\times10^{-6}$). Each straight or dashed line in each panel corresponds to a different deployment morphology as in Figure 4.4.4.

estimated whilst minimising the standard error. To estimate multiple parameters we require drifter observations in more than one spatial dimension, and hence a lower deployment eccentricity is beneficial (i.e. a more isotropic configuration). However, note that in Figure 4.4.7 any $e < 0.6$ achieves close to the minimal standard error for all mesoscale parameter estimates, suggesting mild eccentricity at deployment does not have a significant effect and only more extreme eccentricities should be avoided.

Finally, we investigate the error of estimating a mesoscale feature parameter which isn't truly present in the model, thus allowing for a proper comparison with a similar study by Ohlmann et al. (2017). We repeat the experiment of Ohlmann et al. (2017) and simulate drifters that follow a random walk, i.e. they have no mesoscale component and zero mean flow, and then we estimate all mesoscale parameters. The standard errors of the divergence estimates (measured against a true value of 0) are

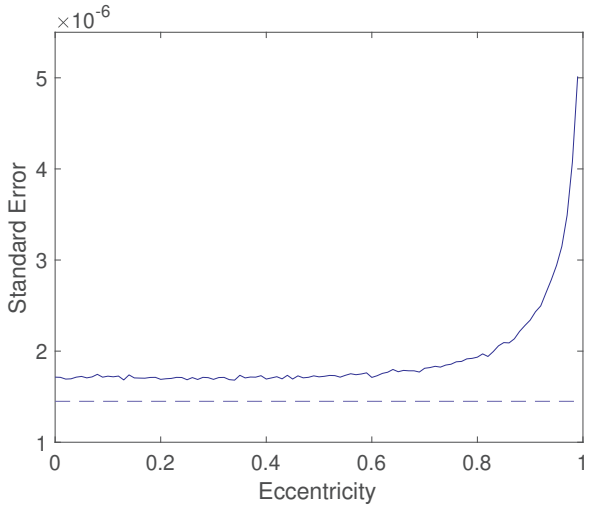

Figure 4.4.8: Standard error of divergence estimates from drifter trajectories simulated with no mesoscale component (i.e. $\sigma = \zeta = \delta = 0$) but all mesoscale parameters are estimated. Drifter trajectories are simulated as in Figure 4.4.4 and the straight and dashed lines in each panel refer to the same deployment morphologies as in this Figure.

shown in Figure 4.4.8 (note that we do not standardise the standard error by parameter value here since the mesoscale parameters are all set to zero) and are in complete agreement with Ohlmann et al. (2017): it is more optimal to deploy drifters in an isotropic configuration than in a straight line. These results are the same as when all mesoscale parameters were present in the simulation (Figure 4.4.7), meaning that it is the number and type of parameters estimated rather than the parameters truly present which affects the optimal morphology.

To conclude, this analysis suggests that the optimal morphology is dependent on the number of mesoscale parameters being estimated and which specific parameters they are. In other words, it is the mesoscale parameters being *estimated* that determines the optimal morphology, rather than the parameters that are actually present

in the drifter trajectories. Therefore we do not require prior knowledge of mesoscale features present in the ocean to choose the optimal morphology, we just need to determine what we want to learn from the experiment. If we are only interested in estimating a single mesoscale parameter, then it is generally optimal to deploy the drifters along a straight line. If we wish to estimate more than one mesoscale parameter then the optimal morphology is determined by which parameters are desired, however in most cases it is better to deploy the drifters in an isotropic initial configuration. If we have no prior knowledge whatsoever of what parameters are present, and we wish to learn which mesoscale features are present in the particle trajectories, then it would be safer to choose a morphology with low deployment eccentricity. However, reducing the deployment eccentricity below around 0.6-0.7 doesn't appear to significantly reduce the standard error when multiple mesoscale parameters are estimated, whereas for a single mesoscale parameter increasing the deployment eccentricity above 0.4 does significantly reduce the standard error. Therefore, with no prior knowledge of the data or model to be used, a deployment eccentricity between 0.4-0.7 will be relatively close to optimal most of the time.

4.5 Discussion and Conclusions

It will always be beneficial to deploy more drifters and to observe them for longer, however this additional data will have diminishing returns. Therefore, what we really want to know, is how much data is enough? In Section 4.2 we showed that we need at least three drifters before our parameter estimates become statistically significant

from zero. The relative standard error initially decreases more quickly, and after five drifters it scales as $\mathcal{O}(\frac{1}{\sqrt{K}})$. Therefore, we would recommend at least five drifters in such a setting. Similarly, in Section 4.3 we showed that in terms of the simulation time length T, then the relative standard error generally scales as $\mathcal{O}(\frac{1}{\sqrt{T}})$. In the case of strain-only data, the relative standard error then decreases more quickly when $T > 1/\sigma$, but this is not the case for the vorticity-only data, where we do not see any faster decrease in the error once $T > 1/\zeta$. For both strain-only and vorticity-only experiments, we found the experiment should run for at least $T \approx 0.1\sigma$ or $t \approx 0.1\zeta$ for the estimated parameters to be significantly different from zero. In terms of the sampling rate Δ, we found quickly diminishing returns when reducing this, and in any case this option is seldom available in practice due to the GPS technology used and the limited battery-life of drifters.

The optimal deployment configuration of drifters is determined based on what we wish to learn from an experiment. If we are only interested in estimating a single mesoscale parameter, the optimal configuration is different than if we wish to estimate a combination of mesoscale parameters. Specifically, it is generally optimal to deploy drifters along a single straight line if we only wish to estimate one mesoscale parameter. If more than one parameter is to be estimated then the optimal configuration is determined by which parameters are to be estimated, as we discussed in Section 4.4. If we do not have any prior knowledge of the expected parameter values before the experiment then a deployment configuration with low eccentricity (i.e. more circular) should be used.

The results in this chapter are simulation-derived and based on an idealised model,

and the complexity of real-world data could change the results. We have explored the scaling for only certain parameter values. Motivated by this, in Chapter 5 we will explore these results further via analytically-derived expressions for the error.

Chapter 5

Analytical Derivations of the Errors of Parameter Estimates

5.1 Introduction

Diffusivity can be estimated by taking the time derivative of the relative dispersion (LaCasce, 2008), where the dispersion is the second moment of the particle positions in centre of mass coordinates after the initial positions are removed. The second moment of particle positions grows exponentially with time for a fixed strain rate (Sundermeyer et al., 2020)—a relationship which allows us to estimate the strain rate, assuming that it is constant in time. In this chapter we will define method-of-moment estimators for both diffusivity and strain rate, and use these estimators to provide theoretical errors for these parameter estimates, which we use to advise how drifters are deployed. We will introduce three possible estimators for diffusivity in centre of mass coordinates for a diffusivity-only model, as well as a method to estimate diffusivity in a strain-

diffusivity model. We also introduce a method-of-moment estimator for the strain rate. Our estimator is similar to that by Sundermeyer et al. (2020), however they estimate strain rate using the second moment in only the major axis which is aligned with the strain angle, whereas we derive an estimator using the second moments in both the major and minor axis. Similarly, they estimate diffusivity using just the second moment in the minor direction whereas we will derive an estimator using the second moments in both axes.

The optimal deployment morphologies in the previous chapter depended on the values of the parameters chosen for our simulations. The results might not hold if any of the strain rate, strain angle, vorticity, divergence, diffusivity, number of observations, time between observations, or number of drifters are changed. These results gave a good intuition into how many, how long, and in what configuration drifters should be deployed, however repeating the simulations for different parameter values would be time consuming.

In this chapter we will instead approximate the theoretical distribution of errors for the least squares parameter estimates from our model from Chapter 3 (Equation (3.2.2)) by calculating the distribution of the second moment of the positions. This approximation will be done using method-of-moments estimation, which is simpler than the least squares estimation used in the simulations of Chapter 4, but allows (near) exact distributions to be derived as we shall show. We shall do this approximation for the simple case of a strain-only model, but note that the general method can be expanded to more complex mesoscale models. This chapter will help to mathematically reveal and explain the results of the number of drifters required (as found in

Section 4.2), the number of observations (see Section 4.3) and the initial configuration of the drifters (as in Section 4.4), and allow the results to be expanded to different parameter values to those used in the previous chapter.

Sundermeyer et al. (2020) estimate the strain rate using second moments as

$$\frac{dm_{\tilde{x}\tilde{x}}}{dt} - \sigma m_{\tilde{x}\tilde{x}} = 0, \tag{5.1.1}$$

where $\tilde{x}_k = x_k - m_x$ is the position of drifter k with the first moment removed and will be defined fully in Section 5.2. Equation (5.1.1) can be solved to find that

$$m_{\tilde{x}\tilde{x}} = m_{\tilde{x}\tilde{x}}(0)e^{\sigma t}, \tag{5.1.2}$$

and hence the strain rate is estimated as

$$\hat{\sigma} = \frac{1}{t}\left[\log(m_{\tilde{x}\tilde{x}}) - \log(m_{\tilde{x}\tilde{x}}(0))\right]. \tag{5.1.3}$$

They estimate diffusivity from second moments as

$$\hat{\kappa} = \frac{\sigma}{2} m_{\tilde{y}\tilde{y}}. \tag{5.1.4}$$

In this Chapter we will introduce a method-of-moments estimator for strain rate and diffusivity which will give estimates with lower RMSE than the estimator from Sundermeyer et al. (2020). Specifically, in Section 5.2 we derive the second moments of particles in a strain only field. We calculate the expected second moment and its variance in each of the x and y directions, and we also derive the cross second moment. Section 5.3 introduces three different ways to estimate diffusivity from second moments when there are no mesoscale components present in the particle trajectories. We calculate the error of each diffusivity estimator and show that the optimal variant

is to remove the initial positions from drifter positions before estimating diffusivity. This produces a biased diffusivity estimator, however correcting for this bias (another variant which we try) increases the error due to increased variance as we shall show. In Section 5.4 we show how to rotate the particle trajectories to align the x axis with the strain rate, and derive an estimator for strain rate in this rotated reference frame. We estimate the error of this estimator and use this to provide scaling behaviour which can be used to reinforce the results from Chapter 4 about the optimal drifter deployment. In Section 5.5 we derive an estimator for diffusivity when the strain rate is not zero, and use this estimator to show how our diffusivity estimates scale with different parameters. Finally, in Section 5.6 we summarise our results from this Chapter and provide a discussion as to the expected errors of parameter estimates with respect to mesoscale features present as well as the deployment configuration.

5.2 Second Moment Calculation

Throughout this chapter, we will consider the more simple case of a strain only field. The results we present throughout this chapter only hold when the strain rate is aligned with the x axis, which is not usually the case in real or simulated particle trajectories (unless θ has been set to be zero). Therefore, for these results to hold, we must align the axis with the strain rate by applying a rotation matrix to the positions. We will show how to carry out this rotation in Section 5.4.1.

To determine the optimal drifter deployment, we wish to derive analytical expressions for the error of strain rate and diffusivity using method-of-moments, which

can be used to calculate the errors. To derive these expressions we must first calculate the moments of the solution to the stochastic differential equation (SDE) from Equation (3.2.9).

Drifter Positions

The SDEs for $x_k(t)$ and $y_k(t)$ in a strain-only field are

$$dx_k(t) = \frac{1}{2}\sigma x_k(t)dt + \sqrt{2\kappa}dW_t, \tag{5.2.1}$$

$$dy_k(t) = -\frac{1}{2}\sigma y_k(t)dt + \sqrt{2\kappa}dW_t, \tag{5.2.2}$$

where the subscript k denotes drifter $k = 1, 2, \cdots, K$.

Both Equations 5.2.1 and (5.2.2) can be modelled using an Ornstein-Uhlenbeck (OU) process. In general, the OU process is defined as

$$dx_t = -\beta(x_t - \alpha)dt + \sigma_I^2 dW_t, \tag{5.2.3}$$

where β is the growth-rate, α is the asymptotic mean, and σ_I^2 is the noise variance. The solution to an OU process follows a normal distribution, where the distribution of x_t given initial position x_0 is

$$x_t|x_0 \sim N\left(\alpha + (x_0 - \alpha)e^{-\beta t}, \frac{\sigma_I^2(1 - e^{-2\beta t})}{2\beta}\right). \tag{5.2.4}$$

Hence, by replacing α, β, and σ_I^2 in Equation (5.2.3) with their values in Equations (5.2.1) and (5.2.2), we set $\alpha = 0$, $\beta = \pm\sigma/2$, and $\sigma_I^2 = \sqrt{2\kappa}$. The drifter positions in Equations (5.2.1) and (5.2.2) are therefore distributed as

$$x_k(t) \sim N\left(x_k(0)e^{\frac{t\sigma}{2}}, \frac{2\kappa}{\sigma}(e^{t\sigma} - 1)\right), \tag{5.2.5}$$

$$y_k(t) \sim N\left(y_k(0)e^{-\frac{t\sigma}{2}}, \frac{2\kappa}{\sigma}(1 - e^{-t\sigma})\right), \tag{5.2.6}$$

and hence can be written as

$$x_k(t) = x_k(0)e^{\frac{t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma} - 1)}X_k, \tag{5.2.7}$$

$$y_k(t) = y_k(0)e^{-\frac{t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(1 - e^{-t\sigma})}Y_k, \tag{5.2.8}$$

where X_k and Y_k each follow a unit normal distribution.

We can approximate the drifter positions for small t by taking a Taylor expansion around $t = 0$ to obtain

$$x_k(t) \approx x_k(0)\left(1 + \frac{t\sigma}{2}\right) + \sqrt{2\kappa t}X_k, \tag{5.2.9}$$

$$y_k(t) \approx y_k(0)\left(1 - \frac{t\sigma}{2}\right) + \sqrt{2\kappa t}Y_k. \tag{5.2.10}$$

At small times, the deterministic component of the drifter positions will grow linearly in time with respect to σ in the x direction whilst decreasing linearly in time with respect to σ in the y direction. On the other hand, the stochastic component will grow at a rate proportional to the *square root* of time with respect to κ in terms of the *standard deviation* in both the x and y directions. This means that for small t, the effect of the submesoscale diffusivity will be larger than the effect of the strain in both the x and y directions, and this effect will be more prolonged if κ is large with respect to σ—which will be the case for examples with values found for the Latmix experiment ($\kappa \sim 0.1m^2/s$ and $\sigma \sim 10^{-6}/s$). Hence the drifters will initially spread diffusively in x and y in an isotropic fashion at a rate determined solely by κ. For larger values of t, however, the strain will eventually contribute significantly more than the diffusivity to how the drifters will spread in the x direction, as the exponentially growing terms in Equation (5.2.7) take over. Conversely in the y-direction the exponential terms

vanish in Equation (5.2.8) for large t and the drifters are simply spread around 0 with variance proportional to κ/σ.

Second Moments of $x_k(t)$ and $y_k(t)$

To determine the optimal drifter morphology, we are interested in the properties of the parameters estimated in Chapter 3, where we estimated using centre of mass coordinates. Therefore, we will calculate the positions of the drifters in centre of mass coordinates. This is found by removing the first moment from $x_k(t)$ and $y_k(t)$. The moments are defined as

$$m_{x^n y^m} = \frac{1}{K}\sum_{k=1}^{K} x_k^n y_k^m. \tag{5.2.11}$$

The first moment of $x(t)$ is found from Equation (5.2.11) by setting $n = 1$ and $m = 0$, and similarly the first moment of $y(t)$ is found by setting $n = 0$ and $m = 1$. The first moments are therefore distributed as

$$m_x(t) = m_x(0)e^{\frac{t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma}-1)}\frac{1}{K}\sum_{k=1}^{K} X_k, \tag{5.2.12}$$

$$m_y(t) = m_y(0)e^{-\frac{t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(1-e^{-t\sigma})}\frac{1}{K}\sum_{k=1}^{K} Y_k, \tag{5.2.13}$$

where $m_x(0)$ and $m_y(0)$ are the values of the first moments at time zero in x and y respectively. Defining $\tilde{x}_k(t) = x_k(t) - m_x(t)$ and $\tilde{y}_k(t) = y_k(t) - m_y(t)$, the positions with centre of mass removed are found by subtracting Equation (5.2.12) from 5.2.7 (and equivalently Equation (5.2.13) from 5.2.8), and hence are written as

$$\tilde{x}_k(t) = \tilde{x}_k(0)e^{\frac{t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma}-1)}X_k - \sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma}-1)}\frac{1}{K}\sum_{k=1}^{K} X_k, \tag{5.2.14}$$

$$\tilde{y}_k(t) = \tilde{y}_k(0)e^{\frac{t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(1-e^{-t\sigma})}Y_k - \sqrt{\frac{2\kappa}{\sigma}(1-e^{-t\sigma})}\frac{1}{K}\sum_{k=1}^{K} Y_k, \tag{5.2.15}$$

where we have defined $\tilde{x}_k(0) = x_k(0) - m_x(0)$ and $\tilde{y}_k(0) = y_k(0) - m_y(0)$.

By separating terms for drifter k from those from all the other drifters, and combining the variance for drifter k with the variance for all drifters except for k, it can be shown that Equations 5.2.14 and 5.2.15 follow Gaussian distributions

$$\tilde{x}_k(t) = \tilde{x}_k(0)e^{\frac{t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma} - 1)\left(1 - \frac{1}{K}\right)}X'_k, \tag{5.2.16}$$

$$\tilde{y}_k(t) = \tilde{y}_k(0)e^{\frac{-t\sigma}{2}} + \sqrt{\frac{2\kappa}{\sigma}(1 - e^{-t\sigma})\left(1 - \frac{1}{K}\right)}Y'_k, \tag{5.2.17}$$

where $X'_k \sim N(0,1)$ and $Y'_k \sim N(0,1)$.

The distribution of the second moments of $\tilde{x}$ and $\tilde{y}$ are of real interest to us, as we will later use them to obtain diffusivity and strain rate estimators, from which we can derive theoretical errors. We calculate the distribution of the second moment of $\tilde{x}_k(t)$ from Equation (5.2.14), this is found using Equation (5.2.11) with $n = 2$ and $m = 0$. Similarly the second moment of $\tilde{y}_k(0)$ is found from Equation (5.2.11) with $n = 0$ and $m = 2$. Finally, the cross second moment in $\tilde{x}$ and $\tilde{y}$ is found by setting $n = m = 1$ in Equation (5.2.11). The second moments are distributed as

$$m_{\tilde{x}\tilde{x}}(t) = \frac{1}{K}\frac{2\kappa}{\sigma}(e^{t\sigma} - 1)\sum_{k=1}^{K}\left[\frac{\tilde{x}_k(0)e^{\frac{t\sigma}{2}}}{\sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma} - 1)}} + X_k - \frac{1}{K}\sum_{k=1}^{K}X_k\right]^2 \tag{5.2.18}$$

$$m_{\tilde{y}\tilde{y}}(t) = \frac{1}{K}\frac{2\kappa}{\sigma}(1 - e^{-t\sigma})\sum_{k=1}^{K}\left[\frac{\tilde{y}_k(0)e^{-\frac{t\sigma}{2}}}{\sqrt{\frac{2\kappa}{\sigma}(1 - e^{-t\sigma})}} + Y_k - \frac{1}{K}\sum_{k=1}^{K}Y_k\right]^2. \tag{5.2.19}$$

$$\begin{aligned} m_{\tilde{x}\tilde{y}}(t) = \frac{1}{K}\frac{2\kappa}{\sigma}\frac{(e^{t\sigma} - 1)}{e^{t\sigma/2}}\sum_{k=1}^{K}\Bigg\{&\left[\frac{\tilde{x}_k(0)e^{\frac{t\sigma}{2}}}{\sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma} - 1)}} + X_k - \frac{1}{K}\sum_{k=1}^{K}X_k\right] \\ &\times \left[\frac{\tilde{y}_k(0)e^{-\frac{t\sigma}{2}}}{\sqrt{\frac{2\kappa}{\sigma}(1 - e^{-t\sigma})}} + Y_k - \frac{1}{K}\sum_{k=1}^{K}Y_k\right]\Bigg\}. \end{aligned} \tag{5.2.20}$$

The second moments in $\tilde{x}$ and $\tilde{y}$ can be written in terms of a noncentral chi squared

distribution. This distribution is found by pulling out the $\sqrt{2\kappa/\sigma}(e^{t\sigma}-1)$ term in Equation 5.2.18 and the $\sqrt{2\kappa/\sigma}(1-e^{-t\sigma})$ term in Equation 5.2.19 and then applying Cochran's theorem. Cochran's theorem states that $\sum_{i=1}^{K}(X_i - \frac{1}{K}\sum_{i=1}^{K}X_i)^2 \sim \chi^2_{K-1}$ when X_i are i.i.d standard normal variables. As the X_i in Equations 5.2.18 and 5.2.19 have mean $\mu_k \neq 0$, our result will be a noncentral chi square distribution, where the non-zero mean is due to the $\tilde{x}_k(0)e^{\frac{t\sigma}{2}}$ and $\tilde{y}_k(0)e^{-\frac{t\sigma}{2}}$ terms in Equations 5.2.18 and 5.2.19 respectively. The second moments can therefore be written as

$$m_{\tilde{x}\tilde{x}}(t) = \frac{1}{K}\frac{2\kappa}{\sigma}(e^{t\sigma}-1)z_{\tilde{x}}, \tag{5.2.21}$$

$$m_{\tilde{y}\tilde{y}}(t) = \frac{1}{K}\frac{2\kappa}{\sigma}(1-e^{-t\sigma})z_{\tilde{y}}, \tag{5.2.22}$$

where $z_{\tilde{x}} \sim \chi^2_{K-1}(\lambda_{\tilde{x}})$, $z_{\tilde{y}} \sim \chi^2_{K-1}(\lambda_{\tilde{y}})$ follow noncentral chi-squared distributions with noncentrality parameters found by $\lambda = \sum_{k=1}^{K}\mu_k^2$. The noncentrality parameters are therefore

$$\lambda_{\tilde{x}} = \frac{\sigma e^{t\sigma}K}{2\kappa(e^{t\sigma}-1)}m_{\tilde{x}\tilde{x}}(0), \tag{5.2.23}$$

$$\lambda_{\tilde{y}} = \frac{\sigma e^{-t\sigma}K}{2\kappa(1-e^{-t\sigma})}m_{\tilde{y}\tilde{y}}(0). \tag{5.2.24}$$

The noncentral chi-squared distribution $\chi^2_{K-1}(\lambda)$ has $K-1$ degrees of freedom, despite there being K drifters. This is due to Cochran's theorem, and can be understood because we subtracted the first moment before calculating the second moment, which removed one degree of freedom.

The noncentral chi-squared distribution has expectation $K-1+\lambda$ and variance

$2(K-1+2\lambda)$. Therefore, the second moments $m_{\tilde{x}\tilde{x}}$ and $m_{\tilde{y}\tilde{y}}$ have expectations

$$E[m_{\tilde{x}\tilde{x}}(t)] = m_{\tilde{x}\tilde{x}}(0)e^{t\sigma} + \frac{2\kappa}{\sigma K}(e^{t\sigma}-1)(K-1), \tag{5.2.25}$$

$$E[m_{\tilde{y}\tilde{y}}(t)] = m_{\tilde{y}\tilde{y}}(0)e^{-t\sigma} - \frac{2\kappa}{\sigma K}(e^{-t\sigma}-1)(K-1), \tag{5.2.26}$$

and variance

$$\mathrm{var}(m_{\tilde{x}\tilde{x}}(t)) = \frac{2\kappa}{\sigma K}(e^{t\sigma}-1)2\left[\frac{2\kappa}{\sigma}(e^{t\sigma}-1)\left(1-\frac{1}{K}\right)+e^{t\sigma}m_{\tilde{x}\tilde{x}}(0)\right], \tag{5.2.27}$$

$$\mathrm{var}(m_{\tilde{y}\tilde{y}}(t)) = \frac{2\kappa}{\sigma K}(1-e^{-t\sigma})2\left[\frac{2\kappa}{\sigma}(1-e^{-t\sigma})\left(1-\frac{1}{K}\right)+e^{-t\sigma}m_{\tilde{y}\tilde{y}}(0)\right]. \tag{5.2.28}$$

The cross second moment can be written in terms of a single Gaussian distribution. To derive this distribution we first write $m_{\tilde{x}\tilde{y}}$ as

$$\begin{aligned} m_{\tilde{x}\tilde{y}}(t) = m_{\tilde{x}\tilde{y}}(0) &+ \sqrt{\frac{1}{K}\frac{2\kappa}{\sigma}\left(1-\frac{1}{K}\right)[(e^{t\sigma}-1)m_{\tilde{x}\tilde{x}}(0)+(1-e^{-t\sigma})m_{\tilde{y}\tilde{y}}(0)]}X \\ &+ \frac{1}{K}\frac{2\kappa}{\sigma}\frac{(e^{t\sigma}-1)}{e^{t\sigma/2}}\sum_{k=1}^{K}\left[X_k-\frac{1}{K}\sum_{k=1}^{K}X_k\right]\left[Y_k-\frac{1}{K}\sum_{k=1}^{K}Y_k\right], \end{aligned} \tag{5.2.29}$$

where $X \sim N(0,1)$.

The final term of Equation (5.2.29) is the product of two normal distributions, and has $K-1$ degrees of freedom due to the summation of K normal distributions, each with the sample mean removed. The variance is therefore the product of the variance of each distribution in X and Y summed $K-1$ times, and therefore the cross second moment can be written as

$$\begin{aligned} m_{\tilde{x}\tilde{y}}(t) = m_{\tilde{x}\tilde{y}}(0) &+ \sqrt{\frac{1}{K}\frac{2\kappa}{\sigma}\left(1-\frac{1}{K}\right)[(e^{t\sigma}-1)m_{\tilde{x}\tilde{x}}(0)+(1-e^{-t\sigma})m_{\tilde{y}\tilde{y}}(0)]}X \\ &+ \left(1-\frac{1}{K}\right)\frac{2\kappa}{\sigma}\frac{(e^{t\sigma}-1)}{e^{t\sigma/2}}Y, \end{aligned} \tag{5.2.30}$$

where $Y \sim N(0,1)$. Combining normal distributions, the cross second moment can be written as

$$m_{\tilde{x}\tilde{y}}(t) = m_{\tilde{x}\tilde{y}}(0) + \sqrt{\frac{1}{K}\frac{2\kappa}{\sigma}\left(1-\frac{1}{K}\right)\left[(e^{t\sigma}-1)m_{\tilde{x}\tilde{x}}(0) + (1-e^{-t\sigma})m_{\tilde{y}\tilde{y}}(0)\right] + \left(1-\frac{1}{K}\right)^2\left(\frac{2\kappa}{\sigma}\right)^2\frac{(e^{t\sigma}-1)^2}{e^{t\sigma}}}Z, \tag{5.2.31}$$

where $Z \sim N(0,1)$.

A histogram displaying the distribution of each $m_{\tilde{x}\tilde{x}}$, $m_{\tilde{y}\tilde{y}}$ and $m_{\tilde{x}\tilde{y}}$ is displayed in Figure 5.2.1 for 10,000 simulations of nine drifters in a strain-diffusivity field with $\sigma = 7\times10^{-6}$/s and $\kappa = 0.1\text{m}^2/\text{s}$. The simulated particles trajectories were configured to have initial positions the same as in LatMix site one, were of length one day and their positions were sampled every thirty minutes. The red curves correspond to the chi-square distributions in Equations (5.2.21) and (5.2.22), and the green curve with the normal distribution in Equation (5.2.31) which confirm we have derived the exact distribution in each case. The distribution for the second moment of $x(t)$ looks to approximately follow a normal distribution, however the second moment for $y(t)$ is skewed towards lower values. These figures therefore show the stretching in x and compression in y due to the strain.

The limiting cases of Equations (5.2.21), (5.2.22), and (5.2.31) as $\kappa \to 0$ are $m_{\tilde{x}\tilde{x}}(t) \to m_{\tilde{x}\tilde{x}}(0)e^{t\sigma}$, $m_{\tilde{y}\tilde{y}}(t) \to m_{\tilde{y}\tilde{y}}(0)e^{-t\sigma}$, and $m_{\tilde{x}\tilde{y}}(t) \to m_{\tilde{x}\tilde{y}}(0)$, meaning that when diffusivity is zero then there is no random walk component and the second moments are fully deterministic. Similarly as $t \to 0$, $m_{\tilde{x}\tilde{x}}(t) \to m_{\tilde{x}\tilde{x}}(0)$, $m_{\tilde{y}\tilde{y}}(t) \to m_{\tilde{y}\tilde{y}}(0)$, and $m_{\tilde{x}\tilde{y}}(t) \to m_{\tilde{x}\tilde{y}}(0)$, that is to say that the second moment at time zero is just the initial

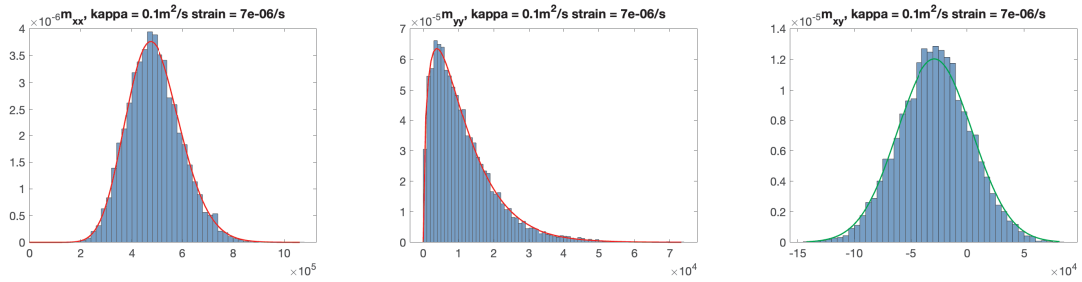

Figure 5.2.1: Histogram of second moments (in x-direction (aligned with strain) on left plot, in y-direction on centre plot, and the xy on the right plot) from 10,000 simulations of nine drifter trajectories (simulated for one day and sampled every 30 minutes) in a strain-diffusivity field, with $\sigma = 7 \times 10^{-6}/s$ and $\kappa = 0.1\text{m}^2/\text{s}$. The red curve corresponds to a chi-squared distribution with K-1 degrees of freedom and noncentrality parameters as in Equations 5.2.23 and 5.2.24. The green curve corresponds to a normal distribution, as given in Equation (5.2.31).

second moment, as would be expected.

Second moments after removing initial positions

We will show in Section 5.3 different ways to estimate diffusivity from second moments when the strain rate is zero. When the diffusivity is estimated from the second moments $m_{\tilde{x}\tilde{x}}$ and $m_{\tilde{y}\tilde{y}}$, we will show that the diffusivity estimate can sometimes be negative. This happens if the average distance to the centre of mass is large but the diffusivity is small then the second moment at the final time could be less than the initial second moment, which would result in negative diffusivity estimates.

To remedy this, we would need to increase the number of drifters or increase the length of the simulation. However this could be costly, and in some cases might

not be practical. Alternatively we could therefore remove the initial positions before calculating the second moment, which we will show in Section 5.3 will lead to guaranteed non-negative diffusivity values. To do this, define $\tilde{x}'(t) = \tilde{x}(t) - \tilde{x}(0)$ and $\tilde{y}'(t) = \tilde{y}(t) - \tilde{y}(0)$. The second moments are then distributed as

$$m_{\tilde{x}'\tilde{x}'}(t) = \frac{1}{K}\sum_{k=1}^{K}\left[\tilde{x}_k(0)(e^{\frac{t\sigma}{2}}-1) + \sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma}-1)}X_k - \sqrt{\frac{2\kappa}{\sigma}(e^{t\sigma}-1)}\frac{1}{K}\sum_{k=1}^{K}X_k\right]^2 \tag{5.2.32}$$

$$= \frac{1}{K}\frac{2\kappa}{\sigma}(e^{t\sigma}-1)z_{\tilde{x}'}, \tag{5.2.33}$$

$$m_{\tilde{y}'\tilde{y}'}(t) = \frac{1}{K}\sum_{k=1}^{K}\left[\tilde{y}_k(0)(e^{-\frac{t\sigma}{2}}-1) + \sqrt{\frac{2\kappa}{\sigma}(1-e^{-t\sigma})}Y_k - \sqrt{\frac{2\kappa}{\sigma}(1-e^{-t\sigma})}\frac{1}{K}\sum_{k=1}^{K}Y_k\right]^2 \tag{5.2.34}$$

$$= \frac{1}{K}\frac{2\kappa}{\sigma}(1-e^{-t\sigma})z_{\tilde{y}'}, \tag{5.2.35}$$

where $z_{\tilde{x}'}$ and $z_{\tilde{y}'}$ follow noncentral chi-square distributions (shown in Figure 5.2.2) with $K-1$ degrees of freedom and noncentrality parameters

$$\lambda_{\tilde{x}'} = \frac{\sigma(e^{\frac{t\sigma}{2}}-1)^2K}{2\kappa(e^{t\sigma}-1)}m_{\tilde{x}\tilde{x}}(0), \tag{5.2.36}$$

$$\lambda_{\tilde{y}'} = \frac{\sigma(e^{-\frac{t\sigma}{2}}-1)^2K}{2\kappa(1-e^{-t\sigma})}m_{\tilde{y}\tilde{y}}(0). \tag{5.2.37}$$

We note that Equations (5.2.33) and (5.2.35) are almost identical to Equations (5.2.21) and (5.2.22), the only change is to the noncentrality parameters, which are the part of the distributions which contain the initial positions (within the second moment at time zero).

In Figure 5.2.2 we display the distribution of second moments after removing the initial positions for the data in Figure 5.2.1. The second moments of $y_k(t)$ appear

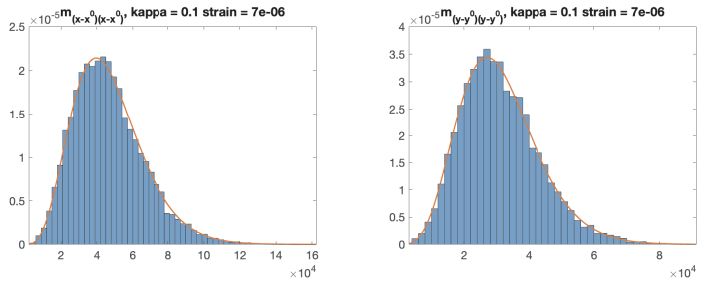

Figure 5.2.2: Histogram of re-centred second moments using the same simulation setup as Figure 5.2.1. The red curve corresponds to a chi-squared distribution with K degrees of freedom and noncentrality parameters as in Equations 5.2.36 and 5.2.37.

to be more normal than those in Figure 5.2.1, and the mode is further from zero in Figure 5.2.2. The mean of $m_{\tilde{y}'\tilde{y}'}$ is smaller than that of $m_{\tilde{y}\tilde{y}}$ due to the long tail at the upper end of the distribution in Figure 5.2.1. Due to the skewness of $m_{\tilde{y}\tilde{y}}$, if we were to sample repeatedly from the distribution, we would obtain many samples which are close to zero. We will show in Section 5.3 that this can result in negative diffusivity estimates. One way that diffusivity can be estimated from second moments is as the change in second moment with time. This requires subtracting the second moment at time zero from the second moment at time t, and so if the second moment at time t is less than that at time zero the diffusivity estimate will be negative. This is undesirable since diffusivity should be positive.

In Section 5.3 we will derive the distribution and errors for diffusivity estimates before and after removing the initial positions. We will initially do this for the simple case where $\sigma = 0$, and so we estimate the second moments of $x(t)$ and $y(t)$ where we have set the strain rate to be zero.

5.2.1 Diffusivity-Only Second Moment

We now calculate the second moment for particle trajectories with no mesoscale component. We do this for the second moments of $\tilde{x}_k(t)$ and $\tilde{y}_k(t)$ and the second moments of $\tilde{x}'_k(t)$ and $\tilde{y}'_k(t)$. These expressions are not immediate from Equations 5.2.21, (5.2.22), (5.2.33) and (5.2.35) as setting $\sigma = 0$ returns 0/0 values in numerous places and therefore we must resort to expansions to resolve the correct values.

We begin with the second moments of $\tilde{x}_k(t)$ and $\tilde{y}_k(t)$. We take a first-order Taylor series expansion of $e^{t\sigma}$ ($\approx 1+t\sigma$) and $e^{-t\sigma}$ ($\approx 1-t\sigma$) to approximate Equations (5.2.21) and (5.2.22) with strain rate set to zero as

$$m_{\tilde{x}\tilde{x},\sigma=0} = \frac{2\kappa t}{K} z_{\tilde{x},\sigma=0}, \tag{5.2.38}$$

$$m_{\tilde{y}\tilde{y},\sigma=0} = \frac{2\kappa t}{K} z_{\tilde{y},\sigma=0}, \tag{5.2.39}$$

where $z_{\tilde{x},\sigma=0}$ and $z_{\tilde{y},\sigma=0}$ follow noncentral chi-square distributions with $K-1$ degrees of freedom. We also take a Taylor expansion of $e^{t\sigma}$ in the noncentrality parameters, which we approximate as

$$\lambda_{\tilde{x},\sigma=0} = \frac{K}{2\kappa t} m_{\tilde{x}\tilde{x}}(0), \tag{5.2.40}$$

$$\lambda_{\tilde{y},\sigma=0} = \frac{K}{2\kappa t} m_{\tilde{y}\tilde{y}}(0), \tag{5.2.41}$$

when the strain rate has been set to zero.

Now, we find the second moments of $\tilde{x}'_k(t)$ and $\tilde{y}'_k(t)$. This is again done by taking a Taylor expansion of $e^{t\sigma}$, hence we obtain second moments with the strain rate set

to zero

$$m_{\tilde{x}'\tilde{x}',\sigma=0} = \frac{2\kappa t}{K} z_{\tilde{x}',\sigma=0}, \tag{5.2.42}$$

$$m_{\tilde{y}'\tilde{y}',\sigma=0} = \frac{2\kappa t}{K} z_{\tilde{y}',\sigma=0}, \tag{5.2.43}$$

where $z_{\tilde{x}',\sigma=0}$ and $z_{\tilde{y}',\sigma=0}$ are central chi-square distributions with $K-1$ degrees of freedom, mean $K-1$ and variance $2(K-1)$. Note that we have a central chi-square distribution and so $\lambda_{\tilde{x}'} = \lambda_{\tilde{y}'} = 0$. This means that the second moments of $\tilde{x}'$ and $\tilde{y}'$ do not depend on the second moments at time zero, and hence do not depend on the initial positions. The second moments only depend on the number of drifters, time length of the particle trajectories, and the value of the diffusivity.

The only difference between Equations (5.2.38) and (5.2.42) and between Equations (5.2.39) and (5.2.43) is the noncentrality parameters being nonzero when the initial positions are not removed. Therefore the second moments of $\tilde{x}$ and $\tilde{y}$ depend on the initial positions, while second moments of $\tilde{x}'$ and $\tilde{y}'$ do not.

Removing the initial positions before calculating the second moments will reduce the probability of the diffusivity estimates being negative, compared with estimating diffusivity from second moments where the initial positions have not been removed, as we shall show. Specifically, in the next section we will show that this is because the variance of the diffusivity estimates calculated from $m_{\tilde{x}'\tilde{x}'}$ and $m_{\tilde{y}'\tilde{y}'}$ is lower than that from $m_{\tilde{x}\tilde{x}}$ and $m_{\tilde{y}\tilde{y}}$, but both estimates have the same bias.

5.3 Diffusivity Estimation and Errors (no strain)

The second moments of $x(t)$ and $y(t)$ can be used to obtain estimates of the diffusivity. We calculate the distribution of this diffusivity estimate (when $\sigma = 0$) from Equations (5.2.38), (5.2.39), (5.2.42) and (5.2.43). There are three different ways in which we could estimate diffusivity in centre of mass coordinates: we could use the observed positions, $\tilde{x}$, $\tilde{y}$, we could re-centre the positions to start at the origin, $\tilde{x}'$, $\tilde{y}'$, or we could re-centre the positions and unbias the estimate. We now show the expectation, variance and mean squared error (MSE) for each method.

Throughout this section we define the mean squared distance of the drifters to the origin in the initial deployment to be

$$d_0^2 = \frac{1}{K}\sum_{k=1}^{K}\left[(\tilde{x}_k(0))^2 + (\tilde{y}_k(0))^2\right] \tag{5.3.1}$$

$$= m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0). \tag{5.3.2}$$

5.3.1 Centre of Mass Coordinates (Observed)

Calculating the diffusivity using the observed positions, after removing the centre of mass, gives a diffusivity estimate with distribution given by

$$\hat{\kappa}_{\text{obs}}(t) = \frac{m_{\tilde{x}\tilde{x},\sigma=0}(t) + m_{\tilde{y}\tilde{y},\sigma=0}(t) - m_{\tilde{x}\tilde{x},\sigma=0}(0) - m_{\tilde{y}\tilde{y},\sigma=0}(0)}{4t} = \frac{1}{2K}\kappa z - \frac{d_0^2}{4t}, \tag{5.3.3}$$

where $z \sim \chi^2_{2K-2}(\lambda_{\tilde{x},\sigma=0} + \lambda_{\tilde{y},\sigma=0})$ and has mean $2K - 2 + \lambda_{\tilde{x},\sigma=0} + \lambda_{\tilde{y},\sigma=0}$ and variance $2(2K - 2 + 2[\lambda_{\tilde{x},\sigma=0} + \lambda_{\tilde{y},\sigma=0})]$. Hence the expectation and variance of the diffusivity

estimate are

$$\mathrm{E}(\hat{\kappa}_{\mathrm{obs}}(t)) = \frac{1}{2K}\kappa\left(2K - 2 + \frac{Kd_0^2}{2\kappa t}\right) - \frac{d_0^2}{4t} \tag{5.3.4}$$

$$= \left(1 - \frac{1}{K}\right)\kappa, \tag{5.3.5}$$

$$\mathrm{var}(\hat{\kappa}_{\mathrm{obs}}(t)) = \left[\frac{1}{2K}\kappa\right]^2 2\left(2K - 2 + 2\frac{2Kd_0^2}{\kappa t}\right) \tag{5.3.6}$$

$$= \frac{\kappa^2}{K}\left(1 - \frac{1}{K}\right) + \frac{\kappa}{2Kt}d_0^2. \tag{5.3.7}$$

The bias of this diffusivity estimate is $[\mathrm{E}(\hat{\kappa}_{\mathrm{obs}}(t)) - \kappa] = -\kappa/K$ which $\to 0$ as $K \to \infty$, meaning that estimating diffusivity from the second moment is only an unbiased estimator asymptotically as the number of drifters approaches infinity.

For fixed K and t, the variance of the diffusivity estimate scales as $\mathcal{O}(d_0^2)$. The distance from the centre of mass, d_0^2, only exists in one term, and so the variance has a lower bound (for fixed K) which is found by setting d_0^2 to zero. In contrast, K is in both terms, and as the number of drifters increases, the standard deviation tends towards zero as $\mathcal{O}(1/\sqrt{K})$, consistent with the law of large numbers.

The mean squared error of the estimate is equal to the variance plus the bias squared. Hence,

$$\mathrm{MSE}(\hat{\kappa}_{\mathrm{obs}}(t)) = \frac{\kappa^2}{K}\left(1 - \frac{1}{K}\right) + \frac{\kappa}{2Kt}d_0^2 + \left(-\frac{\kappa}{K}\right)^2 \tag{5.3.8}$$

$$= \frac{\kappa^2}{K} + \frac{\kappa}{2Kt}d_0^2. \tag{5.3.9}$$

The root mean squared error (RMSE) therefore scales with d_0, and goes to a constant nonzero value as d_0 goes to zero. As the number of drifters increases, the estimate approaches the true value at $\mathcal{O}(1/\sqrt{K})$. For small t the MSE due to the initial positions could be large if the drifters are not deployed close together. Hence, we will

now show a re-centred diffusivity estimate where we will remove the initial conditions before estimating the diffusivity.

5.3.2 Centre of Mass Coordinates (Re-Centred)

If the diffusivity is instead estimated after removing initial positions in centre of mass coordinates, then its distribution follows

$$\hat{\kappa}_{\text{re-centred}}(t) = \frac{m_{\tilde{x}'\tilde{x}',\sigma=0}(t) + m_{\tilde{y}'\tilde{y}',\sigma=0}(t)}{4t} = \frac{\kappa}{2K} z', \tag{5.3.10}$$

where $z' \sim \chi^2_{2K-2}$ has mean $2K-2$ and variance $2(2K-2)$, hence the expectation and variance are

$$\mathrm{E}(\hat{\kappa}_{\text{re-centred}}(t)) = \left(1 - \frac{1}{K}\right)\kappa \tag{5.3.11}$$

$$\mathrm{var}(\hat{\kappa}_{\text{re-centred}}(t)) = \left[\frac{\kappa}{2K}\right]^2 (4K-4) = \frac{\kappa^2}{K}\left(1 - \frac{1}{K}\right). \tag{5.3.12}$$

Therefore, estimating diffusivity with or without removing initial positions have the same expectation and bias, that is the bias is again $-\frac{\kappa}{K}$. However, the variance of estimating diffusivity with or without removing initial positions is different, as there is an additional variance not present in Equation (5.3.12) attributed to the initial positions. Specifically, the variance of the re-centred estimate is equivalent to setting d_0 to zero for the previous diffusivity estimate in Equation (5.3.7). Therefore the variance of the re-centred estimates will always be less than or equal to that of the observed estimates.

Using the observed positions can result in an incorrect negative diffusivity estimate, as can be seen in Figure 5.3.1. The left panel shows the histogram of estimates

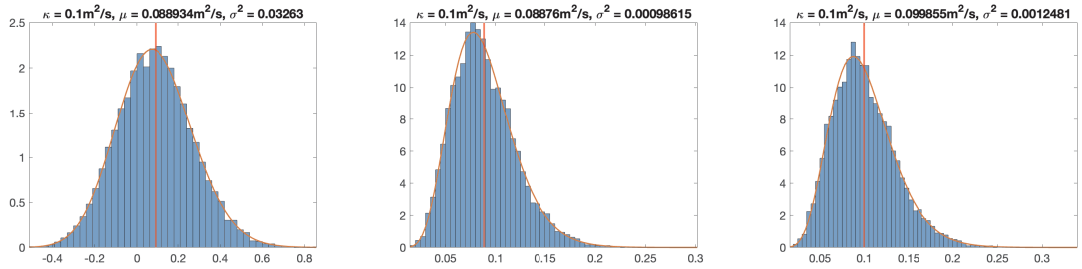

Figure 5.3.1: Histogram of diffusivity estimates calculated from the second moment, where the true diffusivity was $\kappa = 0.1\text{m}^2/\text{s}$ and $\sigma = 0$. Otherwise the same simulation setup is used as in Figure 5.2.1. The red curve corresponds to the theoretical distribution of the estimates, from Equations (5.3.3), (5.3.10) and (5.3.14). The red vertical line displays the sample mean, and is hence the mean of the blue histogram. The left panel displays estimates from observed particle trajectories, the centre panel displays estimates from the re-centred trajectories, and the right panel shows the de-biased estimates from the re-centred particle trajectories.

for the observed data (with theoretical distribution overlaid to show agreement), which has a larger variance than the centre panel which displays a histogram of the estimates from the re-centred data (again with theoretical distribution overlaid). Note that the theoretical expectation (and bias) for each of the left and centre plots is the same, as derived in Equations (5.3.5) and (5.3.11).

The rate in which the variance in Equation (5.3.12) decreases is no longer affected by the initial positions and so neither is the rate at which the diffusivity estimate approaches the true diffusivity. The variance approaches zero at the same rate as with the observed positions when we increase the number of drifters ($\mathcal{O}(1/\sqrt{K})$).

The mean squared error is simply

$$\mathrm{MSE}(\hat{\kappa}_{\text{re-centred}}(t)) = \frac{\kappa^2}{K}. \tag{5.3.13}$$

Similarly to the variance, the MSE is the same as for the observed positions, but without the part attributed to d_0^2. Hence, the same MSE could be obtained using observed positions if all drifters were deployed at the origin. The reduction in the MSE comes from a reduction in the variance, as the estimate is still biased by the same amount. We will now consider an unbiased estimate of the diffusivity to see how the error changes.

5.3.3 Centre of Mass Coordinates (Re-Centred and Unbiased)

We have seen that removing the initial positions decreases the variance of the diffusivity estimate, but has no effect on the bias. Since we know the bias of the re-centred estimator we could de-bias the estimates. The debiased diffusivity estimate would be defined and hence distributed as

$$\hat{\kappa}_{\text{unbiased}}(t) = \frac{m_{\tilde{x}'\tilde{x}',\sigma=0} + m_{\tilde{y}'\tilde{y}',\sigma=0}}{4t}\left(1-\frac{1}{K}\right)^{-1} = \frac{\kappa}{2K}\left(1-\frac{1}{K}\right)^{-1} z'. \tag{5.3.14}$$

This estimator isn't biased, and the expected value is the true diffusivity, κ which can be verified by observing that

$$\mathrm{E}(\hat{\kappa}_{\text{unbiased}}(t)) = \frac{\kappa}{2K}\left(1-\frac{1}{K}\right)^{-1}(2K-2) = \kappa, \tag{5.3.15}$$

where recall that z' is a chi-squared distribution with $2K-2$ degrees of freedom, and

hence mean equal to $2K-2$. The variance is

$$\text{var}(\hat{\kappa}_{\text{unbiased}}(t)) = \left(\frac{\kappa}{2K}\right)^2 2(2K-2) = \frac{\kappa^2}{K\left(1-\frac{1}{K}\right)} = \frac{\kappa^2}{K-1}. \tag{5.3.16}$$

The MSE is simply equal to the variance as the bias is zero, that is $\text{MSE}(\hat{\kappa}_{\text{unbiased}}(t)) = \frac{\kappa^2}{K-1}$. Hence, the bias correction actually results in a slightly larger MSE than for the re-centred estimator. This is analogous to the general choice of choosing between biased and un-biased estimators of variance and (auto-)covariance in classical statistics. Typically the biased estimator is preferred as it has a lower variance and MSE.

The de-biased estimates for diffusivity are displayed in the right panel of Figure 5.3.1. We can see that while the expectation of the de-biased estimator is almost equal to the true diffusivity, the variance is larger than that for the re-centred estimates. The observed positions produced the diffusivity estimator with both the highest bias and variance, and should therefore not be used.

5.3.4 Summary

We introduced three different possible estimators for diffusivity. As a summary, the bias, variance and MSE for each estimator is displayed in Table 5.3.1. The estimator $\hat{\kappa}_{\text{obs}}(t)$ has the highest variance and MSE. The estimator $\hat{\kappa}_{\text{unbiased}}(t)$ has the lowest bias of zero, however it has a larger variance than $\hat{\kappa}_{\text{re-centred}}(t)$ as removing the bias caused the variance to increase for $\hat{\kappa}_{\text{unbiased}}(t)$. Overall, $\hat{\kappa}_{\text{re-centred}}(t)$ has the lowest MSE, and we therefore suggest that the re-centred estimate of diffusivity is preferable for this reason.

We can use Table 5.3.1 to reinforce the results from Chapter 4 relating to the

Estimator	Expectation	Variance	Bias	MSE
$\hat{\boldsymbol{\kappa}}_{\mathbf{obs}}(\boldsymbol{t})$	$\left(1-\frac{1}{K}\right)\kappa$	$\frac{\kappa^2}{K}\left(1-\frac{1}{K}\right)+\frac{\kappa}{2Kt}d_0^2$	$-\kappa/K$	$\frac{\kappa^2}{K}+\frac{\kappa}{2Kt}d_0^2$
$\hat{\boldsymbol{\kappa}}_{\mathbf{re\text{-}centred}}(\boldsymbol{t})$	$\left(1-\frac{1}{K}\right)\kappa$	$\frac{\kappa^2}{K}\left(1-\frac{1}{K}\right)$	$-\kappa/K$	$\frac{\kappa^2}{K}$
$\hat{\boldsymbol{\kappa}}_{\mathbf{unbiased}}(\boldsymbol{t})$	κ	$\frac{\kappa^2}{K-1}$	0	$\frac{\kappa^2}{K-1}$

Table 5.3.1: A table displaying the expectation, variance, bias and MSE of each estimator from Sections 5.3.1, 5.3.2 and 5.3.3. The 'best' in each column is given in red text, where 'best' is determined by that with the expectation closest to the true value and the lowest value for each of the variance, bias and MSE.

deployment of drifters, however we note that those results were for the error of the strain rate estimates, whereas the results in Table 5.3.1 only apply to diffusivity. The MSE all three diffusivity estimators are $\mathcal{O}(\frac{1}{K})$, meaning that the RMSE is $\mathcal{O}(\frac{1}{\sqrt{K}})$, as was the case in Section 4.2 for the strain rate. Similarly, the RMSE of $\hat{\kappa}_{\text{obs}}(t)$ scales as $\frac{1}{\sqrt{t}}$, and we showed in Section 4.3 that the strain rate error was $\mathcal{O}(\frac{1}{\sqrt{t}})$. The RMSE of $\hat{\kappa}_{\text{obs}}(t)$ cannot go to zero however, as there is a term that does not depend on time, so as $t \to \infty$ the RMSE would approach a constant value, assuming κ and K are constant. We note that the MSE $\hat{\kappa}_{\text{re-centred}}(t)$ and $\hat{\kappa}_{\text{unbiased}}(t)$ do not depend on time, and hence do not display the same result as in Section 4.3, meaning that increasing the number of time points will not decrease the error of the estimators. Therefore, it is more efficient to reduce the error of diffusivity estimates by choosing an estimator with lower error (i.e. using the re-centred estimator), as opposed to needing to collect more data — at least in the case of diffusivity estimation via method-of-moments.

We have seen that the errors for diffusivity and strain rate appear to behave in the same way as we change the number of drifters or the number of observed time points. However, they do not appear behave in the same way when the distance to the centre of mass is changed. In Section 4.4 we showed that the error of the strain rate estimates decreased as drifters were placed further from the centre of mass, which is equivalent to increasing d_0^2. The RMSE of $\hat{\kappa}_{\text{obs}}(t)$ on the other hand is $\mathcal{O}(d_0)$, meaning that the error increases as the distance to the centre of mass increases. This suggests that the optimal morphology is not only dependent on which mesoscale parameters are being estimated, but also on whether we are interested in the diffusivity estimates. We note that the results for strain rate are based on simulation and the least squares estimator, whereas the diffusivity results are derived from method-of-moments estimation. Therefore, to investigate further we now provide a distribution for the method-of-moments estimator for the strain rate.

5.4 Strain Rate and Angle Estimation

In the previous section we derived the distributions and errors for method-of-moments estimators of diffusivity when the true strain rate is zero. We now derive the distribution for an estimator of the strain rate in a strain-diffusive field estimated from the observed second moments of $x(t)$ and $y(t)$. We use this distribution to quantify the error of strain rate estimates, and look at the scaling behaviour of the analytical expression of the error. The scaling behaviour is used to comment on how drifters should be deployed to reduce the error of strain rate estimates.

Throughout this section we will assume that the particle positions are given in a rotated coordinate frame where x is aligned with the stretching caused by the strain and y is perpendicular to this and aligned with where the positions are compressed by the strain. Typically drifter trajectories will (at least initially) not generally be aligned with the strain angle, and so this must be estimated first, as we shall now describe.

5.4.1 Strain Angle Estimation

We derive an expression for the strain angle using second moments. The drifter positions can then be rotated such that they are aligned with the strain angle.

The strain angle can be estimated from the second moments of the raw (unrotated) particle positions, which we denote as $m^r_{\tilde{x}\tilde{x}}(t)$, $m^r_{\tilde{y}\tilde{y}}(t)$, and $m^r_{\tilde{x}\tilde{y}}(t)$. When the drifter positions are given in a reference frame that is aligned with the strain angle the expected cross second moment doesn't change with time, that is $m_{\tilde{x}\tilde{y}}(t) = m_{\tilde{x}\tilde{y}}(0)$. We note that the cross second moment in Equation (5.2.31) had expectation $m_{\tilde{x}\tilde{y}}(0)$ as we would expect as this was derived in a rotated reference frame (with strain aligned along the x-axis), however it had a non-zero variance due to the non-deterministic nature of drifter trajectories.

We therefore wish to find the value for θ such that $m_{\tilde{x}\tilde{y}}(t) - m_{\tilde{x}\tilde{y}}(0) = 0$. We look at the rotation of $\Delta m_{\tilde{x}\tilde{y}} = m_{\tilde{x}\tilde{y}}(t) - m_{\tilde{x}\tilde{y}}(0)$ to find

$$\Delta m_{\tilde{x}\tilde{y}} = -\cos\theta\sin\theta\Delta m^r_{\tilde{x}\tilde{x}} + \cos\theta\sin\theta\Delta m^r_{\tilde{y}\tilde{y}} + \left(\cos^2\theta - \sin^2\theta\right)\Delta m^r_{\tilde{x}\tilde{y}}, \qquad (5.4.1)$$

where $\Delta m_{\tilde{x}\tilde{x}} = m_{\tilde{x}\tilde{x}}(t) - m_{\tilde{x}\tilde{x}}(0)$ and $\Delta m_{\tilde{y}\tilde{y}} = m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{y}\tilde{y}}(0)$. We set $\Delta m_{\tilde{x}\tilde{y}} = 0$,

and rewrite Equation (5.4.1) as

$$0 = -\frac{1}{2}\sin 2\theta \Delta m^r_{\tilde{x}\tilde{x}} + \frac{1}{2}\sin 2\theta \Delta m^r_{\tilde{y}\tilde{y}} + \cos 2\theta \Delta m^r_{\tilde{x}\tilde{y}}, \tag{5.4.2}$$

Therefore, the strain angle is given by

$$\tan 2\theta = \frac{2\Delta m^r_{\tilde{x}\tilde{y}}}{\Delta m^r_{\tilde{x}\tilde{x}} - \Delta m^r_{\tilde{x}\tilde{x}}}. \tag{5.4.3}$$

The second moments can be rotated by θ to align with the strain angle, and for the remainder of the chapter we will assume that this rotation has taken place. We do not comment on the error of the strain angle estimates, nor do we investigate how this error will affect further estimates through the rest of this chapter. We leave these as areas for future work.

5.4.2 Strain Rate Estimation

We now assume that the particle positions have been correctly rotated by the strain angle and derive an estimator for the strain rate in this rotated coordinate system. We will use this estimator to calculate an approximation to the errors, which will be used to comment on the optimal drifter deployment morphology.

As diffusivity is calculated from the residual flow, Equations (5.3.3), (5.3.10), and (5.3.14) can be used when the strain rate is not zero, if strain rate has first been estimated and subtracted from the drifter positions. Strain rate therefore needs to be estimated without prior knowledge of the diffusivity, and so diffusivity needs to be eliminated from the equations to be able to estimate strain rate. To eliminate diffusivity, we divide the expected second moment of $\tilde{x}_k(t)$ by the expected second

moment of $\tilde{y}_k(t)$ from Equations (5.2.25) and (5.2.26). This gives us a statistic which is distributed as

$$\frac{E[m_{\tilde{x}\tilde{x}}(t)] - m_{\tilde{x}\tilde{x}}(0)e^{t\sigma}}{E[m_{\tilde{y}\tilde{y}}(t)] - m_{\tilde{y}\tilde{y}}(0)e^{-t\sigma}} = -\frac{(e^{t\sigma}-1)}{(e^{-t\sigma}-1)}. \tag{5.4.4}$$

We multiply out to obtain

$$(E[m_{\tilde{x}\tilde{x}}(t)] - m_{\tilde{x}\tilde{x}}(0)e^{t\sigma})(e^{-t\sigma}-1) = -(E[m_{\tilde{y}\tilde{y}}(t)] - m_{\tilde{y}\tilde{y}}(0)e^{-t\sigma})(e^{t\sigma}-1). \tag{5.4.5}$$

Multiplying out the brackets, and collecting terms,

$$E[m_{\tilde{x}\tilde{x}}(t)](e^{-t\sigma}-1) - m_{\tilde{x}\tilde{x}}(0)(1-e^{t\sigma}) = -E[m_{\tilde{y}\tilde{y}}(t)](e^{t\sigma}-1) + m_{\tilde{y}\tilde{y}}(0)(1-e^{-t\sigma}), \tag{5.4.6}$$

$$\{E[m_{\tilde{x}\tilde{x}}(t)] + m_{\tilde{y}\tilde{y}}(0)\}(e^{-t\sigma}-1) = -\{E[m_{\tilde{y}\tilde{y}}(t)] + m_{\tilde{x}\tilde{x}}(0)\}(e^{t\sigma}-1). \tag{5.4.7}$$

Hence,

$$\frac{E[m_{\tilde{x}\tilde{x}}(t)] + m_{\tilde{y}\tilde{y}}(0)}{E[m_{\tilde{y}\tilde{y}}(t)] + m_{\tilde{x}\tilde{x}}(0)} = -\frac{(e^{t\sigma}-1)}{(e^{-t\sigma}-1)} = e^{t\sigma.} \tag{5.4.8}$$

Therefore, the strain rate can be estimated from observed second moments as

$$\hat{\sigma} = \frac{1}{t}\log(m_{\tilde{x}\tilde{x}}(t) + m_{\tilde{y}\tilde{y}}(0)) - \frac{1}{t}\log(m_{\tilde{y}\tilde{y}}(t) + m_{\tilde{x}\tilde{x}}(0)). \tag{5.4.9}$$

In Figure 5.4.1 we show in a histogram a collection of strain rate estimates over 100,000 repeated simulations using the method of Chapter 3. The red curve, on the other hand, corresponds to a kernel density estimate of the estimates obtained using Equation (5.4.9) over the same 100,000 repeat simulations, where the true strain rate was set to be 7×10^{-6} and is highlighted with the yellow vertical line. We see that Equation (5.4.9) appears to be a reasonable estimate for the strain rate (and have

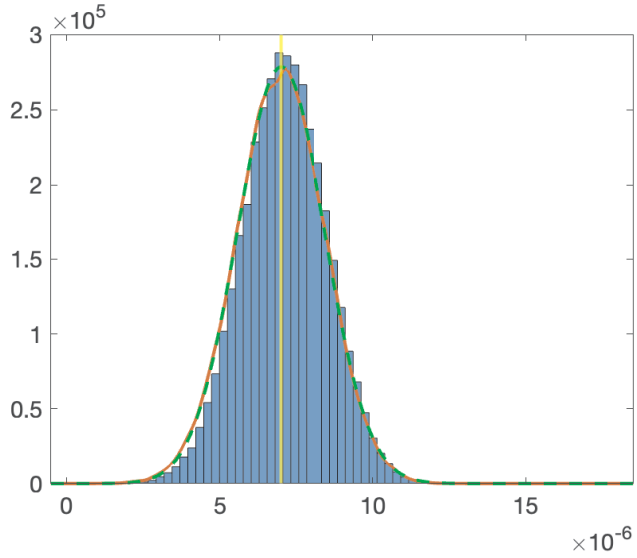

Figure 5.4.1: Histogram of strain estimates using least squares regression from Chapter 3 over 100,000 repeated simulations, using the same simulation setup as in Figure 5.2.1. The red curve corresponds to a kernel density estimate from the same simulations from Equation (5.4.9), the green curve corresponds to a normal approximation with parameters given in Equations (5.4.13) and (5.4.14), and the yellow line shows the true strain rate of 7×10^{-6}/s.

a similar distribution to the least squares method of Chapter 3), and because it has been derived under expectation it should produce unbiased estimates.

Recall that in Equations (5.2.21) and (5.2.22) we showed that the second moments $m_{\tilde{x}\tilde{x}}(t)$ and $m_{\tilde{y}\tilde{y}}(t)$ both followed exact chi-squared distributions under the assumptions of our strain-diffusive model. The distribution of $\hat{\sigma}$ in Equation (5.4.9) therefore follows an addition of two independent non-central log chi-squared distributions. This probability density has no known analytical form and must be computed numerically (see Pav (2015) for more details on the log chi-squared distribution). However a numerical form will not reveal scaling behaviours with respect to different parameters. There instead we proceed by approximating this density via a normal distribution as we shall now show.

5.4.3 Normal Approximation of Strain Rate Estimator

We can analyse the distribution of the method-of-moments estimator for the strain rate, given in Equation (5.4.9), to better understand its scaling behaviour with respect to different parameters of the drifter deployment. However, we cannot directly obtain the expectation and variance of the estimator from Equation (5.4.9) as this is not analytically solvable and instead would have to be computed numerically for given parameter values. We will therefore instead approximate the distribution of the estimator in Equation (5.4.9) with a normal distribution to obtain a theoretical form for the error as a function of different parameter values.

The log of a chi-squared distribution approaches a normal distribution faster with K than a chi-squared distribution. Initially it therefore appears to be better to write

Equation (5.4.9) as a sum of two log chi-squared distributions before approximating as a normal distribution. At first glance, we know that $m_{\tilde{x}\tilde{x}}(t)$ and $m_{\tilde{y}\tilde{y}}(t)$ follow noncentral chi-square distributions, and that $m_{\tilde{x}\tilde{x}}(0)$ and $m_{\tilde{y}\tilde{y}}(0)$ are constants. However, we cannot combine $m_{\tilde{x}\tilde{x}}(t)$ with $m_{\tilde{y}\tilde{y}}(0)$, or $m_{\tilde{y}\tilde{y}}(t)$ with $m_{\tilde{x}\tilde{x}}(0)$, to form a single chi-square distribution. This is because the constant term would need to be added to the mean, but the mean of the noncentral chi-square distribution is determined by the noncentrality parameter, which is the sum of the degrees of freedom and the squared means of the squared random variables which were summed to form the chi-square distribution. Since $m_{\tilde{y}\tilde{y}}(0)$ and $m_{\tilde{x}\tilde{x}}(0)$ are not the mean of any of the squared random variables, then it cannot be added to the mean of the chi-squared distribution, and we cannot combine the terms to form a chi-squared distribution.

We therefore will approximate the second moments inside Equation (5.4.9) as normal distributions before taking the log and summing. To do this, we use the expectation and variance of $m_{\tilde{x}\tilde{x}}(t)$ and $m_{\tilde{y}\tilde{y}}(t)$ from Equations (5.2.25), (5.2.26), (5.2.27), and (5.2.28). Taking a normal approximation, we approximate Equation (5.4.9) as

$$\begin{aligned}\hat{\sigma}t \approx &\log\left[\sqrt{\text{var}(m_{\tilde{x}\tilde{x}}(t))}X + E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)\right] \\ &- \log\left[\sqrt{\text{var}(m_{\tilde{y}\tilde{y}}(t))}Y + E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)\right],\end{aligned} \tag{5.4.10}$$

where $X \sim N(0,1)$ and $Y \sim N(0,1)$. To take the log and combine both terms into a single distribution, we will take a Taylor expansion. To do this we must first write

Equation (5.4.10) as

$$\hat{\sigma}t \approx \log\left[\frac{\sqrt{\text{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)}X + 1\right] + \log[E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)]$$
$$- \log\left[\frac{\sqrt{\text{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)}Y + 1\right] - \log[E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)]. \quad (5.4.11)$$

Taking a Taylor expansion, we get

$$\hat{\sigma}t \approx \frac{\sqrt{\text{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)}X + \log[E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)]$$
$$- \frac{\sqrt{\text{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)}Y - \log[E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)]. \quad (5.4.12)$$

Hence we find that

$$E(\hat{\sigma}) \approx \frac{1}{t}\log\left[\frac{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)}\right], \quad (5.4.13)$$

$$\text{var}(\hat{\sigma}) \approx \frac{1}{t^2}\left[\frac{\text{var}(m_{\tilde{x}\tilde{x}}(t))}{\{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)\}^2} + \frac{\text{var}(m_{\tilde{y}\tilde{y}}(t))}{\{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)\}^2}\right]. \quad (5.4.14)$$

The expectation of $\hat{\sigma}$ is approximated to be simply σ, suggesting an unbiased estimator. This is because the expectation in Equation (5.4.13) is equivalent to that found in Equation (5.4.8) which was used to form an estimator for σ.

Finally, we take a normal approximation of the distribution of the $\hat{\sigma}$ in Equation (5.4.9) to assume that the strain rate estimates are distributed as

$$\hat{\sigma}_{Norm} \sim N\left(\sigma, \text{var}(\hat{\sigma})\right), \quad (5.4.15)$$

where the $\hat{\sigma}_{Norm}$ denotes the normal approximation of the strain rate.

The normal approximation for the strain rate distribution is shown in the green curve of Figure 5.4.1, and appears to be a good fit to the histogram, for the chosen parameter values which were motivated by observations and settings from the LatMix

experiment. However, strain and diffusivity vary across the ocean, if they didn't then estimating them would be a much simpler problem! We therefore require the strain rate estimates from Equation (5.4.9) as well as the normal approximation for their distribution in Equation (5.4.15) to be a good estimate in terms of bias and variance for different possible strain rate and diffusivity values. To show that Equations (5.4.9) and (5.4.15) are approximately valid for different parameter values, we will now look at the way that these equations were derived and any possible limitations that may cause.

5.4.4 Limitations of Normal Approximation to Method-of-Moments Strain Rate Estimator

We made approximations when we derived the distributional form of Equation (5.4.15), and so for any parameter values where these approximations might not hold, the normal approximation could be unreliable. In this section we discuss the assumptions we made and how they affect the normal approximation from Equation (5.4.15).

The normal approximation assumes that the strain rate estimates themselves are normally distributed, and this is the case when the degrees of freedom in the chi-squared distribution (of the second moments) is large. Therefore, when we have a small number of drifters, the normal approximation in Equation (5.4.15) may no longer be suitable. In Figure 5.4.2 we repeat the simulations from Figure 5.4.1, but reduce the number of drifters from nine to three. We see that the normal approximation gives approximately the same distribution of estimates as the kernel density estimate

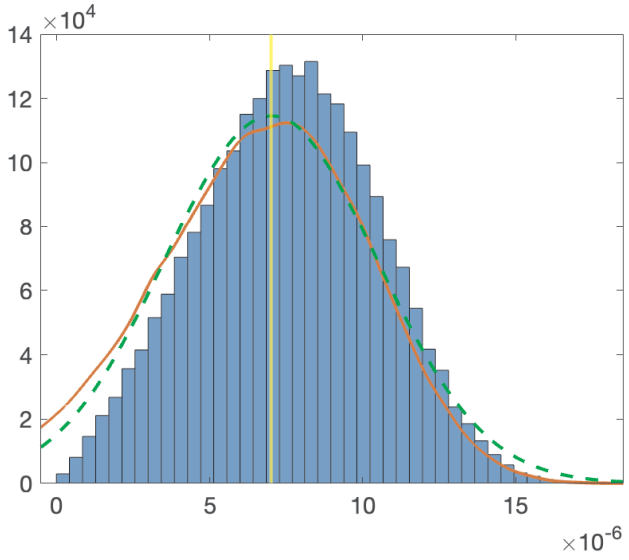

Figure 5.4.2: As in Figure 5.4.1, but reducing the number of drifters to $K = 3$.

of Equation (5.4.9), and so is a good approximation. The expected strain rate from the normal approximation is equal to the true strain rate, as would be expected from Equation (5.4.9). We note that the estimates from the least squares estimate from Chapter 3, appear to have a slightly different distribution to those from Equation (5.4.9) and the normal approximation. This is likely to be due to a difference between the definition of the estimates—the least squares estimates of the strain rate are defined to be strictly positive, whereas estimates from Equation (5.4.9) can be either positive or negative. A negative strain rate is equivalent to rotating the angle of a positive strain rate by 90 degrees, and then the negative estimates in Figure 5.4.2 could be mirrored at $\sigma = 0$. Further simulations (not shown) indicate that negative strain rates are seen for $K \leq 3$ in Figure 5.4.2, for $K > 3$ however the least squares and method-of-moment estimators appear to be increasingly equivalent. These results may vary as model parameters vary from those in our simulations.

Approximating the distribution of Equation (5.4.9) as a normal distribution in Equation (5.4.15) required us to take a Taylor approximation in Equation (5.4.12). The expectation and variance of the normal approximation to the strain rate estimates are therefore only likely to hold close to the expansion points which were $\tilde{x} = 0$ and $\tilde{y} = 0$. In Equation (5.4.12) we took a Taylor expansion of

$$\log\left[\frac{\sqrt{\mathrm{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)}X + 1\right] \quad \text{and} \quad \log\left[\frac{\sqrt{\mathrm{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)}Y + 1\right] \tag{5.4.16}$$

to get

$$\frac{\sqrt{\mathrm{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)}X \quad \text{and} \quad \frac{\sqrt{\mathrm{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)}Y. \tag{5.4.17}$$

The Taylor expansion breaks down when these terms become large, and as X and Y are normal distributions with $\mathcal{O}(1)$ values, the Taylor expansion will break down when

$$\frac{\sqrt{\mathrm{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)} = \frac{\frac{2\kappa}{\sigma K}(e^{t\sigma} - 1)\sqrt{2\left(K - 1 + 2\frac{\sigma K}{2\kappa(1-e^{-t\sigma})}m_{\tilde{x}\tilde{x}}(0)\right)}}{\frac{2\kappa}{\sigma K}(e^{t\sigma} - 1)\left(K - 1 + \frac{\sigma K}{2\kappa(1-e^{-t\sigma})}m_{\tilde{x}\tilde{x}}(0)\right) + m_{\tilde{y}\tilde{y}}(0)} \tag{5.4.18}$$

and

$$\frac{\sqrt{\mathrm{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)} = \frac{\frac{2\kappa}{\sigma K}(1 - e^{-t\sigma})\sqrt{2\left(K - 1 + 2\frac{\sigma K}{2\kappa(e^{t\sigma}-1)}m_{\tilde{y}\tilde{y}}(0)\right)}}{\frac{2\kappa}{\sigma K}(1 - e^{-t\sigma})\left(K - 1 + \frac{\sigma K}{2\kappa(e^{t\sigma}-1)}m_{\tilde{y}\tilde{y}}(0)\right) + m_{\tilde{x}\tilde{x}}(0)} \tag{5.4.19}$$

are large.

Equations (5.4.18) and (5.4.19) will never become larger than 1 for any choice of parameters where $K > 3$, as we will now show. We will rewrite Equations (5.4.18) and (5.4.19) in terms of dimensionless parameters for the time scale $\bar{t}$ and the length

scales $\bar{x}$ and $\bar{y}$, which we define as follows

$$\bar{t} = \sigma t, \qquad \bar{x} = \frac{\tilde{x}}{\sqrt{(2\kappa)/\sigma}}, \qquad \bar{y} = \frac{\tilde{y}}{\sqrt{(2\kappa)/\sigma}}. \tag{5.4.20}$$

We note that changing the value of t is synonymous with changing σ in the time scale as $e^{(2t)\sigma} \equiv e^{t(2\sigma)}$. The length scale is chosen such that if $m_{\bar{y}\bar{y}}(0) = 1$, then the initial positions are in equilibrium such that the strain and diffusivity are acting to keep $m_{\bar{y}\bar{y}}(t)$ constant at $(2\kappa)/\sigma$.

Substituting the dimensionless variables, Equation (5.4.18) is written as

$$\frac{\sqrt{\text{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)} = \frac{\frac{1}{K}(e^{\bar{t}} - 1)\sqrt{2\left(K - 1 + \frac{2K}{(1-e^{-\bar{t}})} m_{\bar{x}\bar{x}}(0)\right)}}{\frac{1}{K}(e^{\bar{t}} - 1)\left(K - 1 + \frac{K}{(1-e^{-\bar{t}})} m_{\bar{x}\bar{x}}(0)\right) + m_{\bar{y}\bar{y}}(0)} \tag{5.4.21}$$

and Equation (5.4.19) as

$$\frac{\sqrt{\text{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)} = \frac{\frac{1}{K}(1 - e^{-\bar{t}})\sqrt{2\left(K - 1 + \frac{2K}{(e^{\bar{t}}-1)} m_{\bar{y}\bar{y}}(0)\right)}}{\frac{1}{K}(1 - e^{-\bar{t}})\left(K - 1 + \frac{K}{(e^{\bar{t}}-1)} m_{\bar{y}\bar{y}}(0)\right) + m_{\bar{x}\bar{x}}(0)}. \tag{5.4.22}$$

To show that Equations (5.4.21) and (5.4.22) never become large enough for the Taylor expansion to break down, we start by looking at the limit as $K \to \infty$. We write Equations (5.4.21) and (5.4.22) as

$$\frac{\sqrt{\text{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)} = \frac{(e^{\bar{t}} - 1)\frac{1}{\sqrt{K}}\sqrt{2\left(1 - \frac{1}{K} + \frac{2}{(1-e^{-\bar{t}})} m_{\bar{x}\bar{x}}(0)\right)}}{(e^{\bar{t}} - 1)\left(1 - \frac{1}{K} + \frac{1}{(1-e^{-\bar{t}})} m_{\bar{x}\bar{x}}(0)\right) + m_{\bar{y}\bar{y}}(0)} \tag{5.4.23}$$

and

$$\frac{\sqrt{\text{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)} = \frac{(1 - e^{-\bar{t}})\frac{1}{\sqrt{K}}\sqrt{2\left(1 - \frac{1}{K} + \frac{2}{(e^{\bar{t}}-1)} m_{\bar{y}\bar{y}}(0)\right)}}{(1 - e^{-\bar{t}})\left(1 - \frac{1}{K} + \frac{1}{(e^{\bar{t}}-1)} m_{\bar{y}\bar{y}}(0)\right) + m_{\bar{x}\bar{x}}(0)}. \tag{5.4.24}$$

Hence both Equations (5.4.18) and (5.4.19) are $\mathcal{O}(1/\sqrt{K})$.

Looking at the scaling as we change the initial positions of the drifters, Equation (5.4.18) scales as $\mathcal{O}(1/m_{\bar{y}\bar{y}}(0))$ and $\mathcal{O}(1/\sqrt{m_{\bar{x}\bar{x}}(0)})$. Equation (5.4.19) scales as $\mathcal{O}(1/\sqrt{m_{\bar{y}\bar{y}}(0)})$ and $\mathcal{O}(1/m_{\bar{x}\bar{x}}(0))$.

Finally we look at how Equations (5.4.18) and (5.4.19) scale as we change the timescale, $\bar{t}$. As $\bar{t} \to \infty$, Equation (5.4.21) becomes

$$\frac{\sqrt{\text{var}(m_{\tilde{x}\tilde{x}}(t))}}{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)} \to \frac{-\sqrt{2\left(K-1+2Km_{\bar{x}\bar{x}}(0)\right)}}{-\left(K-1+Km_{\bar{x}\bar{x}}(0)\right)} \tag{5.4.25}$$

and Equation (5.4.22) approaches

$$\frac{\sqrt{\text{var}(m_{\tilde{y}\tilde{y}}(t))}}{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)} \to \frac{\sqrt{2\left(K-1\right)}}{K-1+Km_{\bar{x}\bar{x}}(0)}. \tag{5.4.26}$$

Neither Equation (5.4.21) nor Equation (5.4.22) contain $m_{\bar{y}\bar{y}}(0)$ as $\bar{t} \to \infty$, and so in the long time limit, the initial positions in y have no effect on the Taylor series expansion. Both Equations (5.4.25) and (5.4.26) remain small (< 1) for all $K \geq 3$, and we require at least 3 drifters to obtain statistically significant estimates, as we showed in simulation in Section 4.2.

We have shown that Equations (5.4.18) and (5.4.19) never become large, and the Taylor series expansion will not break down for $K \geq 3$. Equations (5.4.13) and (5.4.14) can therefore be used to obtain good estimates for the expectation and variance of estimates of the strain rate from Equation (5.4.9). The expectation and variance can then be used to obtain an approximation of the error and the scaling behaviour of the estimates as we shall now explore.

5.4.5 Comparison with scaling behaviour of least-squares estimator

In Chapter 4 we studied the error of strain rate estimates using the least-squares estimator from Chapter 3, and the effect on these errors that different parameters had under simulation. We now repeat some of this investigation but using the theoretical normal approximation to the method-of-moments estimator in Equation (5.4.15), rather than simulated estimates from least-squares.

The method-of-moments strain rate estimator is unbiased, therefore the MSE is equal to the variance. We can write the MSE for the method-of-moment estimator in Equation (5.4.15) as

$$\text{MSE}(\hat{\sigma}_{Norm}) = \frac{1}{t^2}\left\{\frac{\text{var}(m_{\tilde{x}\tilde{x}}(t))}{\{E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0)\}^2} + \frac{\text{var}(m_{\tilde{y}\tilde{y}}(t))}{\{E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)\}^2}\right\}. \quad (5.4.27)$$

This is simply $1/t^2$ times the sum of the square of Equations (5.4.18) and (5.4.19), which we studied the scaling behaviour of in Section 5.4.4. To study the scaling behaviour of Equation (5.4.27), we write the error as a single fraction rearranging Equation (5.4.8) to get

$$E(m_{\tilde{x}\tilde{x}}(t)) + m_{\tilde{y}\tilde{y}}(0) = e^{t\sigma}\left[E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)\right]. \quad (5.4.28)$$

The variance (and equivalently the MSE) of $\hat{\sigma}$ can therefore be written as

$$t^2\text{var}(\hat{\sigma}) \approx \frac{e^{-2t\sigma}\text{var}(m_{\tilde{x}\tilde{x}}(t)) + \text{var}(m_{\tilde{y}\tilde{y}}(t))}{\left[E(m_{\tilde{y}\tilde{y}}(t)) + m_{\tilde{x}\tilde{x}}(0)\right]^2}. \quad (5.4.29)$$

Substituting $\text{var}(m_{\tilde{x}\tilde{x}}(t))$, $\text{var}(m_{\tilde{y}\tilde{y}}(t))$ and $E(m_{\tilde{y}\tilde{y}}(t))$ from Equations (5.2.27),

(5.2.28) and (5.2.26), the variance becomes

$$\text{var}(\hat{\sigma}) \approx \frac{4}{t^2}\frac{1}{K}\frac{\frac{2\kappa}{\sigma}(1-e^{-t\sigma})}{\left(1-\frac{1}{K}\right)\frac{2\kappa}{\sigma}(1-e^{-t\sigma}) + m_{\tilde{x}\tilde{x}}(0) + e^{-\sigma t}m_{\tilde{y}\tilde{y}}(0)}. \tag{5.4.30}$$

The variance can be written in terms of the dimensionless variables, $\bar{t}$, $\bar{x}$, and $\bar{y}$ as defined in Equation (5.4.20). Dividing by σ^2, the variance becomes

$$\frac{\text{var}(\hat{\sigma})}{\sigma^2} \approx \frac{4}{\bar{t}^2}\frac{1}{K}\frac{1}{\left(1-\frac{1}{K}\right) + \frac{m_{\bar{x}\bar{x}}(0)}{1-e^{-\bar{t}}} + \frac{m_{\bar{y}\bar{y}}(0)}{e^{\bar{t}}-1}}. \tag{5.4.31}$$

The dimensionless variance can be used to obtain scaling behaviour which can be compared with the scaling behaviour for the least squares estimator from Chapter 4.

We start by considering the scaling behaviour as the number of drifters changes. Equation (5.4.31) scales as $\mathcal{O}(1/K)$, and therefore the RMSE scales as $\mathcal{O}(1/\sqrt{K})$. This is the same result as observed under simulation in Section 4.2.

Now we consider the scaling behaviour as we change the initial positions of the drifters. Equation (5.4.31) is $\mathcal{O}(1/m_{\bar{x}\bar{x}}(0))$ and $\mathcal{O}(1/m_{\bar{y}\bar{y}}(0))$. In Section 4.4 we didn't demonstrate the precise scaling behaviour with the initial positions, however we showed that increasing the distance from the centre of mass decreased the error, which is consistent with the result for Equation (5.4.31).

Finally, we look at the scaling behaviour as we change the number of time points by increasing the length of the experiment. Equation (5.4.31) is $\mathcal{O}(1/\bar{t}^2)$ for large t. In the limit $t \to \infty$, the variance approaches

$$\frac{\text{var}(\hat{\sigma})}{\sigma^2} \approx \frac{4}{\bar{t}^2}\frac{1}{K}\frac{1}{\left(1-\frac{1}{K}\right) + m_{\bar{x}\bar{x}}(0)}. \tag{5.4.32}$$

Therefore, for large time scales, the error depends only upon the starting positions in x (but not y), and is minimised by increasing the value of $m_{\bar{x}\bar{x}}(0)$ to be as large as possible.

The behaviour of a cluster of drifters can change over time, and the optimal deployment configuration depends of the phase of the experiment in which we wish to estimate the parameters. Sundermeyer and Price (1998) model mixing over three different phases based on the size of a tracer patch. The small-time phase is used when the tracer is smaller than the mesoscale, the intermediate-time phase when the tracer is the size of the mesoscale, and the long-time limit where the trace is bigger than the mesoscale. We will look at the scaling at each of these phases, and have already shown the scaling behaviour at the long time limit. To consider the small time limit, we will take a Taylor expansion at $\bar{t} = 0$, to obtain

$$\frac{\text{var}(\hat{\sigma})}{\sigma^2} \approx \frac{4}{\bar{t}} \frac{1}{K} \frac{1}{\bar{t}\left(1 - \frac{1}{K}\right) + m_{\bar{x}\bar{x}}(0) + m_{\bar{y}\bar{y}}(0)}. \tag{5.4.33}$$

In the small-time limit, the variance scales as $\mathcal{O}(1/\bar{t})$. This is because $\bar{t} << 1$, and so the contribution of $\bar{t}(1 - 1/K)$ to the denominator of Equation (5.4.33) is much less than that of $m_{\bar{x}\bar{x}}(0)$ and $m_{\bar{y}\bar{y}}(0)$, and hence $\bar{t}\left(1 - \frac{1}{K}\right) + m_{\bar{x}\bar{x}}(0) + m_{\bar{y}\bar{y}}(0)$ scales as $\mathcal{O}(1)$. Therefore the scaling behaviour with the number of time points changes depending on the value of $\bar{t}$. In the small-time limit, changing $m_{\bar{x}\bar{x}}(0)$ has the same effect on the error as changing $m_{\bar{y}\bar{y}}(0)$. To minimise the error we want both $m_{\bar{x}\bar{x}}(0)$ and $m_{\bar{y}\bar{y}}(0)$ to be as big as possible. Hence the optimal morphology for the small-time limit is different than that of the long-time limit where only the value of $m_{\bar{x}\bar{x}}(0)$ had an effect on the error.

To look at the intermediate time limit, we require $\bar{t} \approx 1$. We set $\bar{t} = 1$ in Equation (5.4.31), to get

$$\frac{\text{var}(\hat{\sigma})}{\sigma^2} \approx \frac{4}{\bar{t}^2} \frac{1}{K} \frac{1}{\left(1 - \frac{1}{K}\right) + \frac{m_{\bar{x}\bar{x}}(0)e}{e-1} + \frac{m_{\bar{y}\bar{y}}(0)}{e-1}}. \tag{5.4.34}$$

In the medium-time limit, there is a shift towards increasing $m_{\bar{x}\bar{x}}(0)$ having a greater effect on the error than increasing $m_{\bar{y}\bar{y}}(0)$. As $\bar{t}$ increases, then $m_{\bar{x}\bar{x}}(0)$ will have increasingly more effect in decreasing the error than $m_{\bar{y}\bar{y}}(0)$. Therefore, it is optimal to align the drifters with the strain angle (and hence along $m_{\bar{x}\bar{x}}(0)$) in order to minimise the error of the strain estimates, consistent with the findings of Chapter 4.

We note that the above results about the initial positions assume that we have prior knowledge of the strain angle. We require $m_{\bar{x}\bar{x}}(0)$ (and $m_{\bar{y}\bar{y}}(0)$ depending on the phase of the experiment) to be as large as possible to minimise the error, however the drifters need to be deployed close enough together such that the strain rate is approximately constant. Therefore, there will be a practical upper limit on the size of $m_{\bar{x}\bar{x}}(0)$ and $m_{\bar{y}\bar{y}}(0)$, and this limit will likely vary for different parts of the ocean and at different times. The optimal initial configuration for a strain-only experiment will be to deploy drifters aligned with the strain angle where $m_{\bar{x}\bar{x}}(0)$ is as large as possible, but still ensuring the strain rate experienced across the drifters is approximately constant. This will provide strain rate estimates with the lowest possible error. However the above analysis only accounts for the error of the strain rate estimates, and as we showed in Section 4.4, if we wanted to estimate other mesoscale parameters, then the optimal morphology and experimental setup might be different. This analysis did not consider the error of estimating diffusivity, and so in the next section we will look at the error of estimating diffusivity when the strain rate is non-zero.

5.5 Diffusivity Estimation (with strain)

We provide a diffusivity estimator using method-of-moments where the strain rate is non-zero. This differs from the estimators in Section 5.3 which required the strain rate (and other mesoscale parameters) to be zero. We use this estimator to determine the scaling behaviour, which we use to discuss how to deploy drifters to reduce the error of the diffusivity estimates.

The diffusivity can be estimated by adding the expected second moments in $\tilde{x}$ and $\tilde{y}$ from Equations (5.2.25) and (5.2.26) to obtain

$$\frac{4\kappa}{\sigma K}(K-1) = \frac{E[m_{\tilde{x}\tilde{x}}(t)] - m_{\tilde{x}\tilde{x}}(0)e^{t\sigma}}{e^{t\sigma}-1} - \frac{E[m_{\tilde{y}\tilde{y}}(t)] - m_{\tilde{y}\tilde{y}}(0)e^{-t\sigma}}{e^{-t\sigma}-1}. \tag{5.5.1}$$

Substituting $e^{t\sigma}$ from Equation (5.4.8) and changing expectations for observed second moments, the method of moments diffusivity estimator is

$$\hat{\kappa} = \frac{\sigma K}{2(K-1)}\left[\frac{m_{\tilde{x}\tilde{x}}(t)m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)}{m_{\tilde{x}\tilde{x}}(t) - m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)}\right]. \tag{5.5.2}$$

The expected value of $\hat{\kappa}$ is κ. This is because the diffusivity estimator was derived under expectation. Therefore, estimating diffusivity using method-of-moments in the presence of strain provides better estimates in terms of bias than when the strain rate is zero. In Section 5.3 the diffusivity estimator which had the lowest MSE was biased and had expectation $(1-1/K)\kappa$.

Because the estimator is unbiased, the MSE of $\hat{\kappa}$ is equal to the variance, which we now compute. Equation (5.5.2) is a product of distributions, and it is not straightforward to calculate the variance of Equation (5.5.2) in its current form. We will instead work with the log of the diffusivity to obtain an approximation.

5.5.1 Lognormal Approximation of Diffusivity Estimator

To obtain an estimate of the variance of Equation (5.5.2), we want to estimate the diffusivity as a single distribution. It is easier to find the distribution of $\log(\hat{\kappa})$, which we write as

$$\log(\hat{\kappa}) = \log\left(\frac{\sigma K}{2(K-1)}\right) + \log\left[m_{\tilde{x}\tilde{x}}(t)m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)\right] - \log\left[m_{\tilde{x}\tilde{x}}(t) - m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)\right]. \tag{5.5.3}$$

We wish to approximate Equation (5.5.2) with a single distribution. We therefore approximate the distributions of each of the second moments with a normal distribution,

$$m_{\tilde{x}\tilde{x}}(t) = \mu_{\tilde{x}} + \sigma_{\tilde{x}}X, \tag{5.5.4}$$

$$m_{\tilde{y}\tilde{y}}(t) = \mu_{\tilde{y}} + \sigma_{\tilde{y}}Y, \tag{5.5.5}$$

where we have defined $\mu_{\tilde{x}} = E(m_{\tilde{x}\tilde{x}}(t))$, $\mu_{\tilde{y}} = E(m_{\tilde{y}\tilde{y}}(t))$, $\sigma_{\tilde{x}}^2 = \mathrm{var}(m_{\tilde{x}\tilde{x}}(t))$, and $\sigma_{\tilde{y}}^2 = \mathrm{var}(m_{\tilde{y}\tilde{y}}(t))$ to be the expectation and variances of the second moments, as defined in Equations (5.2.25), (5.2.26), (5.2.27), and (5.2.28).

To calculate the diffusivity as a single distribution, we look at each term of Equation (5.5.3) separately, starting with the second term which we can write as

$$\begin{aligned}
&\log\left[m_{\tilde{x}\tilde{x}}(t)m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)\right] \\
&= \log[\sigma_{\tilde{x}}\sigma_{\tilde{y}}XY + \sigma_{\tilde{x}}\mu_{\tilde{y}}X + \sigma_{\tilde{y}}\mu_{\tilde{x}}Y + \mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)] \qquad (5.5.6) \\
&= \log[\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)] + \log\left[\frac{\sigma_{\tilde{x}}\sigma_{\tilde{y}}XY + \sigma_{\tilde{x}}\mu_{\tilde{y}}X + \sigma_{\tilde{y}}\mu_{\tilde{x}}Y}{\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)} + 1\right]. \qquad (5.5.7)
\end{aligned}$$

Taking a Taylor series expansion, we can approximate this term as

$$\log\left[m_{\tilde{x}\tilde{x}}(t)m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)\right] \approx \log[\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)] + \frac{\sigma_{\tilde{x}}\sigma_{\tilde{y}}XY + \sigma_{\tilde{x}}\mu_{\tilde{y}}X + \sigma_{\tilde{y}}\mu_{\tilde{x}}Y}{\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)}. \quad (5.5.8)$$

The third term of Equation (5.5.3) can be written as

$$\begin{aligned}
&\log\left[m_{\tilde{x}\tilde{x}}(t) - m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)\right] \\
&= \log[\sigma_{\tilde{x}}X - \sigma_{\tilde{y}}Y + \mu_{\tilde{x}} - \mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)] \quad (5.5.9) \\
&= \log[\mu_{\tilde{x}} - \mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)] + \log\left[\frac{\sigma_{\tilde{x}}X - \sigma_{\tilde{y}}Y}{\mu_{\tilde{x}} - \mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)} + 1\right]. \quad (5.5.10)
\end{aligned}$$

We approximate this term using a Taylor series expansion to be

$$\log\left[m_{\tilde{x}\tilde{x}}(t) - m_{\tilde{y}\tilde{y}}(t) - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)\right] \approx \log[\mu_{\tilde{x}} - \mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)] + \frac{\sigma_{\tilde{x}}X - \sigma_{\tilde{y}}Y}{\mu_{\tilde{x}} - \mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)}. \quad (5.5.11)$$

We can now combine all the terms in Equation (5.5.3) as a single distribution. We can approximate the distribution of XY to be unit normal, as each of X and Y are unit normal, and so the product will have zero mean, and variance one. While this product doesn't exactly follow a normal distribution, it's a close approximation. Therefore, we can estimate

$$\log(\hat{\kappa}) \approx \log\left[\frac{\sigma K}{2(K-1)}\frac{\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)}{\mu_{\tilde{x}} - \mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)}\right] + \sqrt{\frac{\sigma_{\tilde{x}}^2\sigma_{\tilde{y}}^2 + \sigma_{\tilde{x}}^2\mu_{\tilde{y}}^2 + \sigma_{\tilde{y}}^2\mu_{\tilde{x}}^2}{[\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)]^2} + \frac{\sigma_{\tilde{x}}^2 + \sigma_{\tilde{y}}^2}{[\mu_{\tilde{x}} - \mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)]^2}}X', \quad (5.5.12)$$

where $X' \sim N(0,1)$.

Using the fact that the expectation of $\hat{\kappa}$ in Equation (5.5.2) is κ, the log term on the right of Equation (5.5.12) is equal to $\log(\kappa)$. We can take the expectation of Equation (5.5.2) to write

$$(\sigma K)(\bar{\mu}_{\tilde{x}}\bar{\mu}_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)) = 2\kappa(K-1)\left[\bar{\mu}_{\tilde{x}} - \bar{\mu}_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0) + m_{\tilde{y}\tilde{y}}(0)\right], \quad (5.5.13)$$

and hence the terms inside the square root of Equation (5.5.12) can be combined into a single fraction, giving

$$\log(\hat{\kappa}) \approx \log(\kappa) + \sqrt{\frac{\sigma_{\tilde{x}}^2\sigma_{\tilde{y}}^2 + \sigma_{\tilde{x}}^2\mu_{\tilde{y}}^2 + \sigma_{\tilde{y}}^2\mu_{\tilde{x}}^2 + \left(\frac{2\kappa(K-1)}{\sigma K}\right)^2(\sigma_{\tilde{x}}^2 + \sigma_{\tilde{y}}^2)}{[\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)]^2}}X'. \quad (5.5.14)$$

Hence $\hat{\kappa} \sim \text{lognormal}(\log(\kappa), \text{var}[\log(\hat{\kappa})])$, where

$$\text{var}[\log(\hat{\kappa})] = \frac{\sigma_{\tilde{x}}^2\sigma_{\tilde{y}}^2 + \sigma_{\tilde{x}}^2\mu_{\tilde{y}}^2 + \sigma_{\tilde{y}}^2\mu_{\tilde{x}}^2 + \left(\frac{2\kappa(K-1)}{\sigma K}\right)^2(\sigma_{\tilde{x}}^2 + \sigma_{\tilde{y}}^2)}{[\mu_{\tilde{x}}\mu_{\tilde{y}} - m_{\tilde{x}\tilde{x}}(0)m_{\tilde{y}\tilde{y}}(0)]^2}. \quad (5.5.15)$$

A lognormal(a,b^2) distribution has expectation $e^{a+b^2/2}$ and variance $[e^{b^2}-1](e^{2a+b})$. The approximation in Equation (5.5.14) has expectation $E(\hat{\kappa}) = \kappa e^{\text{var}[\log(\hat{\kappa})]/2}$. Therefore, the lognormal distribution produces biased estimates, except when $\log(\hat{\kappa}) = 0$, and in this case we would just get the true diffusivity every time. Both the expectation and the variance of $\hat{\kappa}$ are minimised by minimising $\text{var}[\log(\hat{\kappa})]$.

We can use $\text{var}[\log(\hat{\kappa})]$ to place error bars around estimates of κ. We do this by calculating the quantiles from the lognormal distribution. Writing Equation (5.5.14) as $\log(\hat{\kappa}) = \log(\kappa) + \text{var}\left[\log(\hat{\kappa})\right]X'$, the quantiles of the distribution can be calculated as $q = \exp[\log(\hat{\kappa}) + \sqrt{2\text{var}\left[\log(\hat{\kappa})\right]}\,\text{erf}^{-1}(2p-1)]$. The value of $\sqrt{2}\,\text{erf}^{-1}(2p-1)$ is equivalent to the quantile of a normal distribution with zero mean. Therefore, a 90% confidence interval is found by $[\exp\{-1.64\text{var}\left[\log(\hat{\kappa})\right]\}\kappa, \exp\{1.64\text{var}\left[\log(\hat{\kappa})\right]\}\kappa]$, and a

95% confidence interval is found by $[\exp\{-1.96\text{var}\,[\log(\hat{\kappa})]\}\kappa, \exp\{1.96\text{var}\,[\log(\hat{\kappa})]\}\kappa]$. The exact confidence interval will depend on the values of $\text{var}\,[\log(\hat{\kappa})]$ and κ. This means that we can be 90% confident that an estimate of κ from Equation (5.5.2) is approximately (with the approximation coming from the assumptions made to fit the lognormal distribution) no bigger than $\exp\{1.64\text{var}\,[\log(\hat{\kappa})]\}$ times κ, and we can be 95% confident that the estimate isn't bigger than $\exp\{1.96\text{var}\,[\log(\hat{\kappa})]\}$ times the size of κ.

5.5.2 Scaling Behaviour of Lognormal Approximation

We now wish to use the distribution of diffusivity estimates in Equation (5.5.14) to determine how in an idealised setting drifters should be deployed to reduce the error of the estimates. To know the best way to deploy drifters, we want to look at how the number of drifters, the length of the experiment, and the initial positions affect the error. We look at the scaling behaviour of $\text{var}[\log(\hat{\kappa})]$, as this also tells us about how $\text{var}(\hat{\kappa})$ scales.

We write $\text{var}[\log(\hat{\kappa})]$ in Equation (5.5.15) in terms of the dimensionless variables as given in Equation (5.4.20). To do this we define dimensionless expectation and

variances in $\tilde{x}$ and $\tilde{y}$ to be

$$\bar{\mu}_{\tilde{x}} = \frac{\mu_{\tilde{x}}}{2\kappa/\sigma} = m_{\bar{x}\bar{x}}(0)e^{\bar{t}} + (e^{\bar{t}} - 1)\left(1 - \frac{1}{K}\right), \tag{5.5.16}$$

$$\bar{\mu}_{\tilde{y}} = \frac{\mu_{\tilde{y}}}{2\kappa/\sigma} = m_{\bar{y}\bar{y}}(0)e^{-\bar{t}} - (e^{-\bar{t}} - 1)\left(1 - \frac{1}{K}\right), \tag{5.5.17}$$

$$\bar{\sigma}^2_{\tilde{x}} = \frac{\sigma^2_{\tilde{x}}}{(2\kappa/\sigma)^2} = \frac{2}{K}(e^{\bar{t}} - 1)\left[(e^{\bar{t}} - 1)\left(1 - \frac{1}{K}\right) + m_{\bar{x}\bar{x}}(0)e^{\bar{t}}\right] = \frac{2}{K}(e^{\bar{t}} - 1)\bar{\mu}_{\tilde{x}}, \tag{5.5.18}$$

$$\bar{\sigma}^2_{\tilde{y}} = \frac{\sigma^2_{\tilde{y}}}{(2\kappa/\sigma)^2} = \frac{2}{K}(1 - e^{-\bar{t}})\left[(1 - e^{-\bar{t}})\left(1 - \frac{1}{K}\right) + m_{\bar{y}\bar{y}}(0)e^{-\bar{t}}\right] = \frac{2}{K}(1 - e^{-\bar{t}})\bar{\mu}_{\tilde{y}}. \tag{5.5.19}$$

The variance of $\log(\hat{\kappa})$ in dimensionless variables can therefore be written as

$$\mathrm{var}[\log(\hat{\kappa})] = \frac{\bar{\sigma}^2_{\tilde{x}}\bar{\sigma}^2_{\tilde{y}} + \bar{\sigma}^2_{\tilde{x}}\bar{\mu}^2_{\tilde{y}} + \bar{\sigma}^2_{\tilde{y}}\bar{\mu}^2_{\tilde{x}} + \left(\frac{K-1}{K}\right)^2(\bar{\sigma}^2_{\tilde{x}} + \bar{\sigma}^2_{\tilde{y}})}{[\bar{\mu}_{\tilde{x}}\bar{\mu}_{\tilde{y}} - m_{\bar{x}\bar{x}}(0)m_{\bar{y}\bar{y}}(0)]^2}. \tag{5.5.20}$$

We start by looking at the scaling behaviour of Equation (5.5.20) with the number of drifters, K. We see from Equations (5.5.16)–(5.5.19) that $\bar{\mu}_x$ and $\bar{\mu}_y$ are $\mathcal{O}(1)$ and that $\bar{\sigma}^2_x$ and $\bar{\sigma}^2_y$ are $\mathcal{O}(1/K)$. Therefore Equation (5.5.20) is also $\mathcal{O}(1/K)$, and the RMSE scales as $\mathcal{O}(1/\sqrt{K})$. This is the same scaling behaviour that we observed for the strain rate, and so both strain rate and diffusivity can be considered together when choosing the number of drifters in a deployment.

Now we look at the scaling of $\mathrm{var}[\log(\hat{\kappa})]$ as we change $\bar{t}$. As $\bar{t} \to \infty$, $\bar{\mu}_y \to (1-1/K)$ and $\bar{\sigma}^2_y \to (2/K)(1-1/K)$, meaning that there is no dependence on the initial positions in y in the long-time limit. This same scaling does not happen for $\bar{\mu}_x$ or $\bar{\sigma}_x$, as these approach ∞ at rate $m_{\bar{x}\bar{x}}(0)$. This means that for large time scales, $\mathrm{var}[\log(\hat{\kappa})]$ depends only on the initial positions in $\bar{x}$ and not the initial positions in $\bar{y}$, and that it will decrease to a non-zero value at rate 1 with $m_{\bar{x}\bar{x}}$.

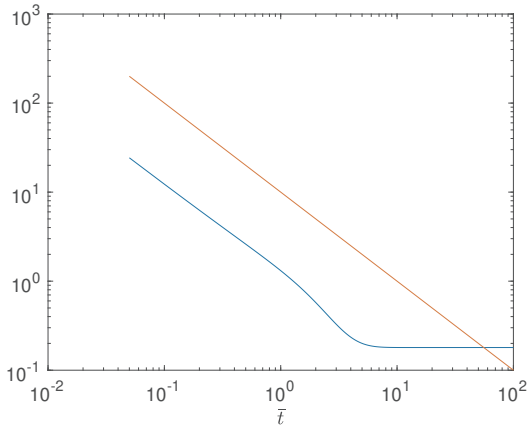

Figure 5.5.1: The scaling of var[log($\hat{\kappa}$)] in Equation (5.5.20) with $\bar{t}$ (blue line) when $\sigma = 7 \times 10^{-6}$, $\kappa = 0.1/s$ and the time between observations is 30 minutes. The red line is a reference line of $1/\bar{t}$, which is the scaling in the small-time limit.

To look at the scaling behaviour in the small-time limit, we take a Taylor expansion of $\bar{\mu}_x$, $\bar{\mu}_y$, $\bar{\sigma}_x^2$, and $\bar{\sigma}_y^2$ around $\bar{t} = 0$ to obtain

$$\bar{\mu}_{\tilde{x}} \approx m_{\bar{x}\bar{x}}(0)(1+\bar{t}) + \bar{t}\left(1 - \frac{1}{K}\right), \tag{5.5.21}$$

$$\bar{\mu}_{\tilde{y}} \approx m_{\bar{y}\bar{y}}(0)(1-\bar{t}) + \bar{t}\left(1 - \frac{1}{K}\right), \tag{5.5.22}$$

$$\bar{\sigma}_{\tilde{x}}^2 \approx \frac{2\bar{t}}{K}\left[\bar{t}\left(1 - \frac{1}{K}\right) + m_{\bar{x}\bar{x}}(0)(1+\bar{t})\right], \tag{5.5.23}$$

$$\bar{\sigma}_{\tilde{y}}^2 \approx \frac{2\bar{t}}{K}\left[\bar{t}\left(1 - \frac{1}{K}\right) + m_{\bar{y}\bar{y}}(0)(1-\bar{t})\right]. \tag{5.5.24}$$

Substituting only the highest order term in $\bar{t}$ for each of Equations (5.5.21)–(5.5.24), we get that var[log($\hat{\kappa}$)] is $\mathcal{O}(1/\bar{t})$, and hence the RMSE is $\mathcal{O}(1/\sqrt{\bar{t}})$. This is again the same result as for the strain rate, and so changing the time the experiment runs for has the same effect on estimating strain rate and diffusivity.

Finally, we look at the scaling of var[log($\hat{\kappa}$)] as we change the initial positions. Figure 5.5.2 shows var[log($\hat{\kappa}$)] for different values of $m_{\bar{x}\bar{x}}(0)$ and $m_{\bar{y}\bar{y}}(0)$, at $\bar{t} = 0.6, 2$

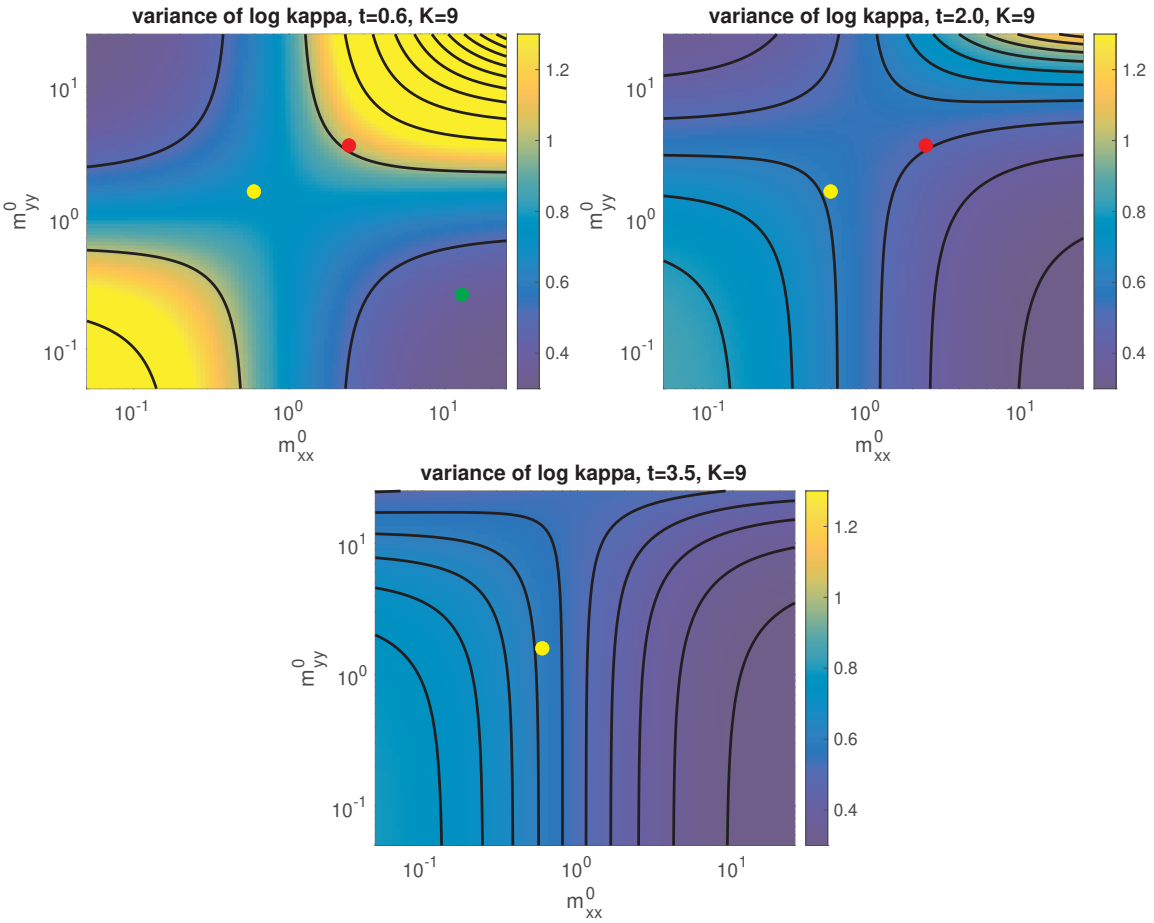

Figure 5.5.2: The value of var[log($\hat{\kappa}$)] for different values of $m_{\bar{x}\bar{x}}(0)$ and $m_{\bar{y}\bar{y}}(0)$, at $\bar{t} = 0.6, 2$, and 3. The coloured dots refer to the initial positions from the LatMix experiment, transformed into dimensionless variables using the values for σ and κ from Chapter 3, rotated by σ. The color of the dots correspond as follows: red - Site 1 initial positions, yellow - Site 2 initial positions scaled by the strain estimate for the first three days of the experiment, green - the positions at the start of day 3 at Site 2, scaled by the strain estimate for the final three days. Each dot is only displayed for times less than $\bar{t}$ for that site (see Table 5.5.1).

and 3.5. The colour corresponds to the value of var[log($\hat{\kappa}$)], which changes for different initial positions, as well as changing with time. For lower time points, either a large value of $m_{\bar{x}\bar{x}}(0)$ and a small value of $m_{\bar{y}\bar{y}}(0)$, or a small value of $m_{\bar{x}\bar{x}}(0)$ and a large value of $m_{\bar{y}\bar{y}}(0)$ minimises var[log($\hat{\kappa}$)]. As time increases, this relationship gradually evolves until changing the value of $m_{\bar{y}\bar{y}}(0)$ has minimal impact on var[log($\hat{\kappa}$)], but increasing $m_{\bar{x}\bar{x}}(0)$ will reduce var[log($\hat{\kappa}$)]. Therefore, to reduce the variance of the estimates, the bottom right corner of these plots is the most optimal across the three timescales we plotted. Specifically, that corresponds to a large value of $m_{\bar{x}\bar{x}}(0)$, and a small value of $m_{\bar{y}\bar{y}}(0)$. Therefore, to minimise the variance of estimating diffusivity, it is the most optimal to deploy drifters spread apart along the axis aligned with the strain angle, and to minimise their spread in the axis orthogonal to the strain angle.

The coloured dots in Figure 5.5.2 correspond to the initial positions of the LatMix experiment. Because we are working in dimensionless variables, the points have been scaled by our estimates for σ and κ from Chapter 3. The strain rate wasn't constant in time at Site 2, and therefore we have two points corresponding to splitting the time series in half. Therefore, the red dot corresponds to the initial positions at Site 1, scaled by $2\kappa/\sigma$. The yellow dot corresponds to the initial positions at Site 2, scaled by the strain rate for days 0–3, and the green dot corresponds to the positions of the drifters at the start of day 3, scaled by the strain rate estimate for days 3–6. The initial positions have been rotated by the strain angle estimate for each site from Chapter 3. The estimates used for σ, θ, and κ are those displayed in Table 3.5.1, where each parameter was estimated over the entire 6 day experiment, with the exception of the strain rate at Site 2, which is taken from Figure 3.5.2. The values of σ, θ, and κ that

	κ (m^2/s)	σ (/s)	T (days)	L (m)	T_E/T	$m_{\bar{x}\bar{x}}(0)$	$m_{\bar{y}\bar{y}}(0)$	var[log($\hat{\kappa}$)]
Site 1 (days 0–6)	0.19	4.5×10^{-6}	2.6	290	2.33	2.4	3.6	0.48
Site 2 (days 0–3)	1.9	1.5×10^{-5}	0.77	490	3.9	0.59	1.6	0.59
Site 2 (days 3–6)	1.90	2.3×10^{-6}	5.03	1280	0.60	12.9	0.26	0.38

Table 5.5.1: A table of the value of var[log($\hat{\kappa}$)] for each site of the Latmix experiment (splitting Site 2 in half corresponding to different strain rates) for the values of κ and σ that are given in the table, and their corresponding dimensionless initial positions, $m_{\bar{x}\bar{x}}$ and $m_{\bar{y}\bar{y}}$. Also given are the time scales T, defined as $1/\sigma$, the length scale L, defined as $\sqrt{2\kappa/\sigma}$, and the experiment length in non-dimensional time units T_E/T.

were used to calculate the initial positions in Figure 5.5.2 are shown in Table 5.5.1. The table also displays the time scale T, which is equivalent to $1/\sigma$, the length scale L, defined as $\sqrt{2\kappa/\sigma}$, the experiment length in dimensionless units T_E/T, equal to the length of the experiment divided by T, the non dimensional initial positions $m_{\bar{x}\bar{x}}(0)$ and $m_{\bar{y}\bar{y}}(0)$, and the value of var[log($\hat{\kappa}$)] at the final time point of each experiment (T_E/T). Note that T_E/T is equivalent to the value of $\bar{t}$ at the end of the experiment (either 3 or 6 days). Each point is only displayed on the figure for values of $\bar{t}$ which are relevant, if the experiment time T_E/T is less than the value of $\bar{t}$ then that point is not shown in the figure.

Of the three points corresponding to the LatMix experiment, the one with the lowest variance is the green dot – which represents the second half of the experiment at Site 2. The values used for $m_{\tilde{x}\tilde{x}}(0)$ and $m_{\tilde{y}\tilde{y}}(0)$ are the position of the drifter at the start of day 3, and therefore these positions were the result of the drifters movement through days 0-3 and are not the initial deployment positions. The initial deployment positions (scaled by the value of strain for the first three days) are shown in the yellow dot. Therefore, the drifter has moved itself into a morphology where the variance of estimating κ has reduced when compared with the initial drifter morphology. This is due to the effects of the strain on the drifters, stretching them along the x axis, and compressing them along the y axis. Our parameter estimates for the LatMix campaign in Chapter 3 could potentially have been improved if the drifters had been deployed along the axis aligned with the strain angle, and more compressed in the orthogonal axis. In the case of Site 2, this could have resulted with diffusivity estimates where var[log($\hat{\kappa}$)] was reduced by a factor of two, and in Site 1 var[log($\hat{\kappa}$)] could be reduced

by up to a factor of three. These factors correspond to choosing a drifter deployment with $m_{\bar{x}\bar{x}}(0) = 10$ and $m_{\bar{y}\bar{y}}(0) = 0.1$, which is in the area of the figure where var$[\log(\hat{\kappa})]$ is minimised all three values of $\bar{t}$.

5.6 Discussion and Conclusions

In this chapter we have showed that the scaling behaviour for the distribution of strain rate estimates, as we change the number of drifters and length of simulation, are the same analytically using method-of-moments as found in simulations using least squares from Chapter 4. The results in Chapter 4 about the optimal deployment configuration were confirmed theoretically in Section 5.4 and reinforced in Section 5.5 as they minimise the error for both strain rate and diffusivity estimates.

In Section 5.3 we gave method-of-moment estimators for diffusivity. We considered three different diffusivity estimators in the absence of strain and found that removing the initial positions before calculating the second moments gave the diffusivity estimates with the lowest error. These estimates were biased, however performing a bias correction increased the resulting RMSE. The error of the re-centred diffusivity estimator scales as $\mathcal{O}(1/\sqrt{K})$, and isn't changed by increasing the length of the simulation or by changing the initial deployment positions of the drifters. Therefore, to decrease the error of the diffusivity estimates in the absence of strain, more drifters must be deployed.

In Section 5.4 we introduced a method-of-moments estimator of the strain rate in a strain-diffusive flow. The strain rate estimator from Sundermeyer et al. (2020)

in Equation (5.1.3) is equivalent to setting $m_{\tilde{y}\tilde{y}}(t)$ and $m_{\tilde{y}\tilde{y}}(0)$ to zero in Equation (5.4.9). Hence Sundermeyer et al. (2020) estimate strain similarly to our proposed method, however they only use drifter position from the major axis, whereas we use both the major and minor axes.

The scaling behaviour of Equation (5.1.3) from Sundermeyer et al. (2020) is the same as our strain rate estimator in Equation (5.4.9) (setting $m_{\tilde{y}\tilde{y}}(0) = 0$ in d_0^2), however the mean and variance are not the same. This can be seen in Figure 5.6.1 which repeats Figure 5.4.1, where we have added a kernel density estimate of the distribution of estimates from Equation (5.1.3) from the same 100,000 Monte Carlo simulations. The curve corresponding to Sundermeyer et al. (2020) (Equation (5.1.3)) is more variable than that corresponding to Equation (5.4.9), and the mean appears to be more than the true strain rate, meaning that Equation (5.1.3) appears to produce more biased estimates than Equation (5.4.9). Hence, the inclusion of the second moment in y, as well as the second moment in x in Equation (5.4.9) will reduce the RMSE of the estimates. Therefore, a cheaper way to reduce the error of the estimates is not increasing the number of drifters or running the experiment for longer, it is rather to include more of the available data in the estimation procedure! For the case of strain rate estimation, if drifter trajectories are available in two dimensions, then both dimensions should be used in the estimator if possible, and so we propose the strain rate should be estimated using our estimator of Equation (5.4.9).

In Section 5.5 we derived an estimator for diffusivity when strain is present, but has first been estimated. We gave an analytical equation for the error of these diffusivity estimates, however we did not consider how the error of the strain rate estimates from

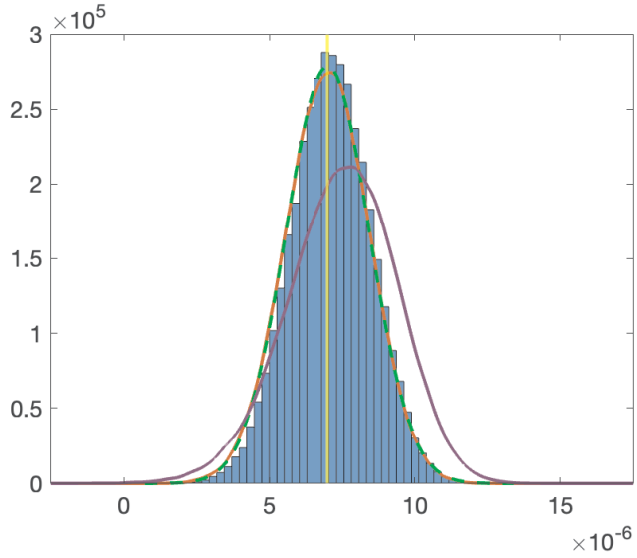

Figure 5.6.1: As in Figure 5.4.1, with an additional curve (purple) corresponding to the distribution of estimates from the method by Sundermeyer et al. (2020) in Equation (5.1.3).

Section 5.4 would affect the error of diffusivity estimates (joint estimation and studying their correlated errors is reserved for future work). We showed that changing the initial drifter deployment positions has different impacts upon the scaling behaviour of the error of our estimates depending on the phase of the experiment, but the overall best deployment morphology was to align the drifters with the strain angle, consistent with the simulation findings of Chapter 4. For small time scales, increasing the value of $\bar{t}$ decreases the error, but this is only up to a point, as the error will asymptote to a constant value with $\bar{t}$. We note that we cannot change the value of σ in practice, and so the only way to increase $\bar{t}$ is to let the experiment run for longer. Therefore, to reduce the error of the diffusivity estimates either more drifters are needed, the drifters should be deployed aligned with the strain angle, or the experiment needs to

run for longer.

Overall, increasing the number of drifters is the most consistent way to decrease the error, as this is the only parameter that decreases the error across all three estimators (strain rate and diffusivity with/without strain). Increasing $\bar{t}$ reduces the error for the diffusivity estimator with strain, but only for small values of $\bar{t}$, for longer deployments then the length of the experiment doesn't impact the error. However, increasing the experiment length and the number of drifters can be costly, and therefore it is important for oceanographers to deploy these drifters in an optimal configuration. We have showed both analytically in this chapter, as well as in simulation studies in Chapter 4, that deploying drifters further apart will decrease the error of strain rate estimates (assuming the strain rate is constant over this distance). For large timescales, the initial positions in $\tilde{y}$ (the axis aligned orthogonal to the strain) don't affect the error, and hence the most optimal way to deploy drifters in a strain-diffusivity field is to align them with the axis aligned with the strain angle, spread across this axis.

The results throughout this chapter are based on an idealised linear strain-diffusivity model with Gaussian white-noise forcing. Specifically, we assume that drifter positions follow an Ornstein-Uhlenbeck process, and are normally distributed. We don't include other terms in our model, such as vorticity or divergence, and so the impact of other features must be negligible, or have already been separated, for these results to be applicable. We assume a scale separation between different parts of our model, specifically we assume that diffusivity is a submesoscale process, and that the strain is mesoscale. If we were to observe, for example, mesoscale diffusivity it is possible that this diffusivity could alias into the mesoscale components of the model. The work in

this chapter is applicable to small scale drifter deployments that are used to measure mesoscale strain and submesoscale diffusivity.

An area for future work would be to extend these results to a strain-divergence-vorticity model, as well as to more complex models with nonlinear mesoscale effects and/or non-white noise forcing terms. This chapter and Chapter 4 use only simulation results and analytical data, and so another area for further work would be to analyse these results using other real-world data in addition to Latmix such as, for example, GLAD drifters.

Chapter 6

Estimating diffusivity from non-clustered Drifters

6.1 Introduction

Throughout this thesis we have considered position data from clustered drifters in a small region of the ocean. Drifter deployments can also happen on a larger scale with drifters spread through the ocean. In the Global Drifter Program data we often have single drifters in a region at a given time, rather than clusters of drifters. We require drifters to be clustered in order to realistically estimate mesoscale flow, however we can still estimate diffusivity from single drifters, as we shall show in this chapter where we investigate how to reduce the error of such estimates.

In Chapters 3–5 diffusivity was estimated from the residual velocity after removing the mesoscale flow. As a result, diffusivity summarised the submesoscale part of the flow. Diffusivity from global drifters can be estimated either directly from the

displacement of particle trajectories, or the mean flow can be removed before diffusivity is estimated. How we estimate diffusivity changes how we interpret what it is telling us. Submesoscale diffusivity as estimated in Chapter 3 will be much smaller ($\sim 0.1m^2/s$) than large-scale diffusivity where we have not removed the mesoscale part of the flow (as large as $\sim 10^4m^2/s$). Diffusivity can therefore be thought of as a parameterisation of the unknown part of the flow that hasn't been explained by mean flow, submesoscale strain and vorticity, etc.

This chapter focuses on estimating diffusivity using particle trajectories from the Global Drifter Program. The Global Drifter Program provides information on the oceans across the globe, with observations over 40 years. It can therefore be used to provide an insight into how diffusivity changes across both space and time. We cannot exactly estimate diffusivity for every point in the ocean continuously. Instead oceanographers usually split the ocean into a number of boxes, assuming diffusivity to be approximately constant in each box. The size of the spatial bin determines both the resolution and the error of the estimates—smaller bins provide a higher resolution but the variance is increased as less data is available inside each bin, whereas larger bins are more biased as they smooth over larger areas, but the variance will typically be lower. This leads to an idea of finding an optimal balance between the bias and the variance, to minimise overall error, which will be discussed in detail throughout this chapter.

In this chapter we first introduce a core method for estimating diffusivity based on Taylor (1922) which uses the periodogram at frequency zero. We show that this method to estimate diffusivity has large errors—larger than the actual size of the

diffusivity that we are estimating—and that these errors are due to large variances. We therefore present a new method of estimating diffusivity which involves splitting a longer time series into many new smaller time series and calculating diffusivity as the smoothed average over these chunks—motivated by the notion of Bartlett or Welch-type smoothed periodogram estimates from the field of time series analysis. This reduces the variance but increases the bias of the estimates, hence introducing a trade-off between bias and variance which requires us to optimally choose the number of chunks that we split the time series into.

In Section 6.2 we introduce the definition of diffusivity, and compare a number of different single-drifter estimators. We introduce two simple time series processes that we could use to model the particle velocities, a white noise process, or an AR(1) process. We then derive a distribution for the periodogram-based estimator for drifters following either a white noise or AR(1) process, and calculate the bias and variance for each of these distributions. In Section 6.3 we propose a new way to estimate diffusivity which involves splitting a time series into a number of smaller chunks and estimating diffusivity as the average across all chunks. In Section 6.4 we demonstrate how to choose the number of chunks to minimise the error. We do this by fitting a parametric model to drifter trajectories and deriving theoretical forms for the bias and variance. In Section 6.5 we apply our new estimator to particle trajectories in a simulation environment where we compare it to the periodogram-based estimator against the known diffusivity. We show that our estimator is able to reduce the error of the estimates. We then apply our estimator to global drifter program drifters in Section 6.6. Finally in Section 6.7 we draw some conclusions and comment on some

areas for future work.

6.2 Benchmark methods for single particle diffusivity estimation

In this section we will introduce three different single particle diffusivity estimators and use these to provide motivation for developing a new estimator with a lower error. We will show that increasing the length of the time series used doesn't reduce the errors of these estimators to zero, and instead we need a large number of drifters to reduce the error which are not always available in practice.

6.2.1 Definition of Single Particle Diffusivity

In this section we will give a formal definition of diffusivity, which will be used in Section 6.2.2 to introduce different diffusivity estimators. We will show that diffusivity can be defined using the autocovariance function and the spectral density, and that both definitions are equivalent.

In Section 5.3 we estimated diffusivity by taking the time derivative of the particle second moments. The second moment is equivalent to the absolute dispersion when calculated using a single particle, and diffusivity is defined as the time derivative of the absolute dispersion. For single particles, diffusivity can be thought of as being an absolute diffusivity, as it tells us about the position of a particle relative to its original position. On the other hand, when diffusivity is calculated from multiple particles, we are usually calculating a relative diffusivity, which tells us about the position of

particles relative to each other.

In this section we will build our theory and methodology for one-dimensional velocities and diffusivity (in the x/u direction for positions and velocities respectively). The same principles apply by extension in the y/v directions, and later when we apply to real data we shall average the diffusivity estimates in each direction to obtain one isotropic diffusivity estimate (though they could of course be estimated and reported separately). LaCasce (2008) defines the diffusivity in one dimension at time t as

$$\kappa(t) = \int_0^t E[u(t)u(\tau)]d\tau, \tag{6.2.1}$$

where $u(\cdot)$ is the velocity of a particle and $E[u(\cdot)]$ denotes the expectation or average velocity over multiple particles. When the flow is stationary and zero mean, this equation is analogous to the integral of the velocity autocovariance. Specifically, for a stationary and zero mean flow, the expected velocity is the same at all time points, and we can write $E[u(t)] = E[u(t-\tau)] = 0$. Hence the autocovariance can be written as $\gamma(\tau) = E[u(t)u(t-\tau)] - E[u(t)]E[u(t-\tau)] = E[u(t)u(t-\tau)]$. Therefore, after a change of variables, Equation 6.2.1 may be written as

$$\kappa(t) = \int_0^t \gamma(\tau)d\tau. \tag{6.2.2}$$

After long times, the diffusivity will approach the long-term diffusivity, which we define to be

$$\kappa_\infty = \int_0^\infty \gamma(\tau)d\tau. \tag{6.2.3}$$

The diffusivity can be written in terms of the power spectrum, as we showed in Equation (3.2.8). For a stationary time series, the power spectral density given in

Equation (3.2.7) is equivalent to

$$S(f) = \int_{-\infty}^{\infty} \gamma(\tau) e^{-i2\pi f\tau} d\tau. \tag{6.2.4}$$

Hence, the diffusivity can be written in terms of the power spectrum by setting the frequency to be zero in Equation (6.2.4). As the autocovariance is symmetric ($\gamma(\tau) = \gamma(-\tau)$), the diffusivity can therefore be written as

$$\kappa_\infty = \frac{1}{2} S(f = 0). \tag{6.2.5}$$

We have shown that Equation (6.2.5) is equivalent to Equation (6.2.3) and the diffusivity can be defined using the autocovariance or the spectral density at frequency zero. In order to estimate diffusivity we therefore require estimators for the autocovariance and/or the spectral density, which we will introduce in the next section.

6.2.2 Estimating Single Particle Diffusivity

The equations defining diffusivity in Section 6.2.1 are defined for continuous time, whereas drifters are observed at discrete time points. Suppose we wish to estimate diffusivity from a sampled drifter trajectory with N velocity observations $u_1, \ldots, u_N$, recorded at positions $x_1, \ldots, x_N$, where Δ is the regular time interval between successive observations (which will be 1 hour or 6 hours for the global drifter program depending on which data product we use). To be able to estimate diffusivity, we must first estimate either the autocovariance or the power spectrum that were defined in Section 6.2.1.

Autocovariance Estimation

The autocovariance can be estimated in two ways from observations, producing biased or unbiased estimates. The unbiased estimator is

$$\hat{\gamma}_\tau^{(u)} = \frac{1}{N-|\tau|} \sum_{t=1}^{N-|\tau|} (u_t - E[u_t])(u_{t+|\tau|} - E[u_{t+|\tau|}]), \tag{6.2.6}$$

for $\tau = 0, \ldots, N-1$, where $E[u_{(\cdot)}]$ is the expected velocity at time $(\cdot)$, and the superscript (u) refers to this being the unbiased estimator. The biased estimator for the autocovariance is typically preferred

$$\hat{\gamma}_\tau^{(b)} = \frac{1}{N} \sum_{t=1}^{N-|\tau|} (u_t - E[u_t])(u_{t+|\tau|} - E[u_{t+|\tau|}]), \tag{6.2.7}$$

as this often reduces the root mean square error (RMSE) of the estimator (Percival et al., 1993). Here the superscript (b) refers to this being the biased estimator for autocovariance. In Section 4.4 the biased diffusivity estimator was preferred over the unbiased diffusivity estimator as it resulted in a lower error overall. The same applies to estimating the autocovariance from drifter observations as we shall show later.

Spectral Density Estimation

The spectral density is defined for continuous data. When we have velocity observations sampled at discrete time points, the spectral density needs to be estimated. One way in which the spectral density can be estimated is using the periodogram, which is defined as (Percival et al., 1993)

$$\hat{S}^{(P)}(f) = \frac{\Delta}{N} \left| \sum_{i=1}^{N} u_i e^{-i2\pi f \Delta} \right|^2, \tag{6.2.8}$$

where the notation $\hat{S}^{(P)}(f)$ refers to this being the periodogram estimator for spectral density.

Diffusivity Estimation

The diffusivity at time $t = N\Delta$ can be approximated by taking the value of $\hat{S}^{(P)}(f)$ at frequency zero using all N observations:

$$\hat{\kappa}^{(P)}(t) = \frac{\Delta}{2N}\left|\sum_{i=1}^{N} u_i\right|^2 . \tag{6.2.9}$$

The superscript (P) in the left hand side of Equation (6.2.9) denotes that this is the periodogram-based estimate of diffusivity, and we use this notation to distinguish this estimate of diffusivity from others that will be introduced later in this section. This is a discrete time approximation to the true time-dependent diffusivity given in Equation (6.2.1), and if t is large, can be used to approximate the long-term diffusivity, κ_∞, given in Equation (6.2.5).

This estimator is equivalent to the re-centred diffusivity estimator in Equation (5.3.10), but for a single particle, as we will now show. The observed velocity u_i at time index $i \in \{1, ..., N\}$, is equivalent to $(x_i - x_{i-1})/\Delta$. After summing all velocities $u_1, u_2, \cdots, u_N$ in Equation (6.2.9), we are therefore just left with $(x_N - x_0)/\Delta$, where x_0 denotes the initial position at time $t = 0$ (which can be found from the data by taking $x_0 = x_1 - u_1\Delta$). The diffusivity estimator can therefore be written as

$$\hat{\kappa}^{(P)}(t) = \frac{1}{2N\Delta}|x_N - x_0|^2 = \frac{1}{2t}|x_N - x_0|^2. \tag{6.2.10}$$

Hence the diffusivity depends only on the initial and final positions x_0 and x_N, and not the path taken to get there. We are only working with diffusivity in a single

dimension, hence we have a 1/2 scaling for diffusivity, whereas in Chapter 5 diffusivity was estimated in two dimensions and so the scaling was 1/4. The different scaling can be understood by taking the two-dimensional diffusivity to be the average diffusivity over each dimension. Therefore, the only difference between Equations (6.2.10) and (5.3.10) is that (6.2.10) is defined for a single drifter and so the re-centred second moment is just $|x_N - x_0|^2$. Hence the estimator can be interpreted as being the squared displacement of a particle from its starting position divided by time, which is why the units are m^2/s. This means that diffusivity provides us with an area in which we could expect to find a particle at each time point after its deployment. It is clear to see from Equation (6.2.9) that $\hat{\kappa}^{(P)}(t)$ is in the correct units of m^2/s as u_t is assumed to be given in SI units of metres (m) per second (s), as is Δ which is given in seconds (s).

We can also estimate diffusivity directly from Equation (6.2.1) using a Riemann approximation to the integral. We will use the trapezium rule which approximates the integral of a function $f(x)$ between x_0 and x_n, such that $x_0 < x_1 < \cdots < x_{n-1} < x_n$, to be

$$\int_{x_0}^{x_n} f(x)dx = \frac{1}{2}h[f(x_0) + f(x_n) + 2\{f(x_1) + f(x_2) + ... + f(x_{n-1})\}], \qquad (6.2.11)$$

where h is the step size. Hence, the Riemann approximation to Equation (6.2.1) for a single drifter time series is

$$\hat{\kappa}^{(R)}(t) = \Delta u_N \left[\sum_{\tau=1}^{N} u_\tau - \frac{1}{2}(u_1 + u_N)\right]. \qquad (6.2.12)$$

Here the superscript (R) refers to this estimator being a Riemann approximation. This estimator can be written in terms of x_t, as we did with the periodogram estimator in

Equation (6.2.10). The Riemann diffusivity estimator is equivalent to

$$\hat{\kappa}^{(R)}(t) = \frac{(x_N - x_{N-1})}{2\Delta}(x_N + x_{N-1} - x_1 - x_0), \tag{6.2.13}$$

which is quite different from the periodogram estimator in Equation (6.2.10).

The diffusivity can also be approximated from Equation (6.2.2), using either the biased or unbiased estimate of the covariance. Taking a Riemann approximation of Equation (6.2.2) gives us the diffusivity estimator

$$\hat{\kappa}^{(\gamma)}(t) = \Delta \left[\sum_{\tau=0}^{N-1} \hat{\gamma}_\tau - \frac{1}{2}(\hat{\gamma}_0 + \hat{\gamma}_t) \right], \tag{6.2.14}$$

where the superscript (γ) refers to this being an estimator which uses the sample autocovariance. It can be shown that when the biased estimator is used for the autocovariance, then this estimator is equivalent to the periodogram-based estimator given in Equation (6.2.9) when the mean flow is zero. This equivalence is shown analytically using the relation

$$\left(\sum_{i=1}^{n} \alpha_i \right)^2 = 2 \sum_{k=0}^{n} \sum_{j=1}^{n-k} \alpha_j \alpha_{j+k} - \sum_{i=1}^{n} \alpha_i^2. \tag{6.2.15}$$

From the above relation and the diffusivity estimators in Equations (6.2.9) and (6.2.14), setting $E(u_t) = 0$, we get that

$$\frac{\Delta}{2}\hat{S}^{(P)}(f=0) = \frac{\Delta}{2N}\left(\sum_{t=1}^{N} u_t \right)^2 = \frac{\Delta}{N} \sum_{\tau=0}^{t} \sum_{t=1}^{N-\tau} u_t u_{i-\tau} - \frac{\Delta}{2N} \sum_{t=1}^{N-\tau} u_t^2, \tag{6.2.16}$$

where $\hat{S}^{(P)}(f)$ is the periodogram, as defined in Equation (6.2.8). This is equivalent to

$$\frac{\Delta}{2}\hat{S}^{(P)}(f=0) = \Delta \sum_{\tau=0}^{t} \hat{\gamma}_\tau^{(P)} - \frac{\Delta}{2}\hat{\gamma}_0^{(P)}. \tag{6.2.17}$$

Hence, this gives us the relation

$$\hat{\kappa}^{(P)}(t) = \hat{\kappa}^{(\gamma)}(t) + \frac{\Delta}{2}\hat{\gamma}_t^{(P)}, \tag{6.2.18}$$

and the only difference between the two estimators is $\frac{\Delta}{2}\gamma_t^{(P)}$ which converges to zero for large T. Therefore we can assume these estimators to be equal when mean flow has been removed. This only holds for the biased estimator, not the unbiassed estimator and hence leaves us with three different ways in which diffusivity could be estimated (not including those from second moments in Section 4.4). Specifically, the estimators are $\hat{\kappa}^{(P)}(t)$, $\hat{\kappa}^{(R)}(t)$ and $\hat{\kappa}^{(\gamma)}(t)$ when $\hat{\kappa}^{(\gamma)}(t)$ uses the unbiased autocovariance estimator.

6.2.3 Diffusivity of White Noise and AR(1) Velocity Processes

Throughout the remainder of this chapter we will use two discrete-time statistical processes to model particle trajectory velocities. These will be used to analyse the error of the diffusivity estimates. The first of these processes is a Gaussian white noise process which is perhaps the simplest way to model particle velocities. The second is an AR(1) process.

In this section we will introduce a white noise and an AR(1) velocity process and derive their theoretical diffusivities. We then compare diffusivity estimates via simulation for the three estimators that were introduced in the previous section against the theoretical diffusivity.

Diffusivity of a White Noise Velocity Process

A Gaussian discrete-time white noise velocity process is defined as

$$u_t \overset{iid}{\sim} N(0, \sigma_I^2), \tag{6.2.19}$$

where u_t is the drifter velocity at index time t, and σ_I^2 is the variance of the process.

A white noise process has $\text{Cov}\,(u_t, u_{t+\tau}) = 0$, and is therefore a stationary process with autocovariance

$$\gamma_\tau = \begin{cases} \sigma_I^2, & \text{if } \tau = 0, \\ 0, & \text{if } \tau \neq 0. \end{cases} \tag{6.2.20}$$

The spectrum of a discrete-time white noise process is calculated from the discrete-time equivalent of Equation (6.2.4) to be

$$S(f) = \Delta \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{-i2\pi f\tau} = \Delta\gamma_0 = \Delta\sigma_I^2. \tag{6.2.21}$$

Hence the spectrum of a discrete-time white noise process is flat and is proportional to its variance at all lags.

The diffusivity of a white noise velocity process is calculated using its autocovariance, as defined in Equation (6.2.20). Using the same concept as Equation (6.2.14) for discrete-time processes, the diffusivity of a white noise process is therefore

$$\kappa(t) = \sum_{\tau=0}^{t} \gamma_\tau - \frac{1}{2}(\gamma_0 + \gamma_t) = \frac{\Delta\sigma_I^2}{2}. \tag{6.2.22}$$

For white noise velocities, the diffusivity immediately converges to its long-time limit, i.e. $\kappa(t) = \kappa_\infty$, which is consistent with applying the definition of Equation (6.2.5) to Equation (6.2.21) to obtain half the spectral density at frequency zero as the long-term diffusivity.

Diffusivity of an AR(1) Velocity Process

The second model we use for drifter velocities is an AR(1) process, where each velocity is expressed as a linear combination of its previous sampled velocity plus a white noise term. An AR(1) process is a discrete-time analogue of the Ornstein-Uhlenbeck process, the latter of which has been proposed by oceanographers to model Lagrangian particle flow (Griffa et al., 1995; Lilly et al., 2017; Berloff and McWilliams, 2002; Veneziani et al., 2004). The AR(1) process can be defined as

$$u_t = \phi u_{t-1} + \epsilon_t, \tag{6.2.23}$$

where $\epsilon_t \overset{iid}{\sim} N(0, \sigma_I^2)$ is a white noise signal, and ϕ is the autoregressive parameter which acts as a damping parameter if $|\phi| < 1$.

The process is zero mean so the autocovariance is calculated as $\gamma = E(u_t u_{t-\tau})$. Calculating the variance first, we get that

$$\gamma_0 = E(u_t^2) = E([\phi u_{t-1} + \epsilon_t]^2) = \phi^2 \gamma_0 + \sigma_I^2. \tag{6.2.24}$$

And so the variance of an AR(1) process is

$$\gamma_0 = \frac{\sigma_I^2}{1 - \phi^2}. \tag{6.2.25}$$

We can calculate the autocovariance of an AR(1) process at lag τ as

$$\begin{aligned} \gamma_\tau &= E(X_t X_{t-\tau}) = E(\phi X_{t-1} X_{t-\tau} + \epsilon_t X_{t-\tau}) \\ &= \phi \gamma_{\tau-1} = \phi^\tau \gamma_0. \end{aligned} \tag{6.2.26}$$

The autocovariance of an AR(1) process is therefore

$$\gamma_\tau = \phi^\tau \frac{\sigma_I^2}{1 - \phi^2}, \tag{6.2.27}$$

when $|\phi| < 1$ such that the process is stationary.

The spectrum of an AR(1) process is calculated as

$$S(f) = \Delta \sum_{\tau=-\infty}^{\infty} \phi^{\tau} \frac{\sigma_I^2}{1-\phi^2} e^{-i2\pi f\tau} = \frac{\Delta\sigma_I^2}{1 - 2\phi\cos(2\pi f) + \phi^2}. \tag{6.2.28}$$

We now calculate the diffusivity of an AR(1) process from its autocovariance as defined in Equation (6.2.27). Using Equation (6.2.14), the diffusivity of an AR(1) process is therefore calculated as

$$\begin{aligned}
\kappa_t &= \sum_{\tau=0}^{t} \gamma_\tau - \frac{1}{2}(\gamma_0 + \gamma_t) \\
&= \Delta \sum_{\tau=0}^{t} \phi^{\tau} \frac{\sigma_I^2}{1-\phi^2} - \frac{\Delta}{2}\left[\frac{\sigma_I^2}{1-\phi^2} + \phi^t \frac{\sigma_I^2}{1-\phi^2}\right] \\
&= \frac{\Delta\sigma_I^2}{1-\phi^2}\left[\frac{\phi^{t+1}-1}{\phi-1} - \frac{1}{2}\left(1+\phi^t\right)\right].
\end{aligned} \tag{6.2.29}$$

For an AR(1) process the diffusivity converges to a fixed value as $t \to \infty$,

$$\kappa_\infty = \frac{\Delta\sigma_I^2}{1-\phi^2}\left[\frac{1}{1-\phi} - \frac{1}{2}\right] = \frac{\Delta\sigma_I^2}{2(1-\phi)^2}, \tag{6.2.30}$$

which again we can see is equivalent to half the value of the spectral density at frequency zero in Equation (6.2.28). We note that this only holds when $|\phi| < 1$, as this is when the process is stationary.

For an AR(1) process the diffusivity changes over time, and is not immediately the long-term diffusivity. This would generally be the case for other time series models other than white noise, as well as what is expected to be seen in the ocean.

Comparison of Diffusivity Estimators

The different diffusivity estimators introduced in Section 6.2.2 do not all give the same estimate of the diffusivity even though they are all based on the same original

definition from Equation (6.2.1). We will compare the error of the three estimators for particle trajectories generated from a white noise process and an AR(1) process.

To estimate the error of each diffusivity estimator from Section 6.2.2 we simulated 5000 time series each from a Gaussian white noise process and an AR(1) process with damping parameter $\phi = 0.7$ where $N = 100, \sigma_I = 1, \Delta = 1$s. The true value of the diffusivity in the white noise and AR(1) cases are therefore $0.5\text{m}^2/\text{s}$ and $5.55\text{m}^2/\text{s}$ respectively. We then estimated the diffusivity from each estimator from the time series, and used these to calculate the bias, standard deviation, and RMSE which were then normalised by the true diffusivity. In Table 6.2.1 we can see that $\hat{\kappa}^{(P)}$ has the lowest RMSE for both the white noise and AR(1) processes. The absolute value of the bias isn't lowest for $\hat{\kappa}^{(P)}$ for the white noise process, however the standard deviation of this estimator is consistently the lowest across both processes. The superior performance of $\hat{\kappa}^{(P)}$ is expected as $\hat{\kappa}^{(\gamma)}$ uses an unbiased (and hence highly variable) autocovariance estimate, and $\hat{\kappa}^{(R)}$ uses a Riemann approximation that weights heavily on the final velocity (u_N) rather than spread information across all velocities (see Equation (6.2.13)). Based on this, we shall assume $\hat{\kappa}^{(P)}$ to be generally better than $\hat{\kappa}^{(\gamma)}$ and $\hat{\kappa}^{(R)}$ in terms of error, but recognise that for different time series models this might not necessarily be true. Overall, we note already that standard deviations are greater than 1 meaning estimated diffusivities have larger standard deviations than their true value, motivating the need for a new and less variable estimator.

For the remainder of this chapter, we will now focus on the periodogram-based estimator for diffusivity and understanding its properties.

	White Noise			AR(1)		
Estimate	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\kappa}^{(\gamma)}$	-0.0045	3.7307	3.7438	0.1087	3.1377	3.1414
$\hat{\kappa}^{(R)}$	0.0124	13.3736	13.3976	0.0348	5.7915	5.5959
$\hat{\kappa}^{(P)}$	0.0246	1.4607	1.4613	-0.0232	1.4360	1.4388

Table 6.2.1: Average bias, standard deviation and root mean square error (calculated from 5000 repeated simulations) for three different diffusivity estimators, for particles with velocities following a Gaussian white noise process and AR(1) process. All results normalised by the true diffusivity. The choice of parameters was $\sigma_I = 1$, $\phi = 0.7$, $\Delta = 1$, and $N = 100$.

6.2.4 Distribution of Diffusivity Estimates

We wish to derive a analytical form for the error of estimates from $\hat{\kappa}^{(P)}$. We shall do this for particle trajectories with velocities that follow either a white noise or AR(1) processes. This will enable us to determine the scaling behaviour of the estimator, similar to the ideas developed in Chapter 5. Note that both these processes are an over-simplification of velocities from drifters, and the results are therefore not directly indicative of the error from real ocean drifter paths—instead we shall use our results to guide construction of a new estimation technique.

White Noise Velocity Process

The diffusivity estimator, $\hat{\kappa}^{(P)}$, of a Gaussian white noise velocity process given in Equation (6.2.19), can be written in terms of a chi-squared distribution, as we now show. The sum of N i.i.d. normal variables is distributed as $\sum_{t=1}^{N} u_t \sim N(0, N\sigma_I^2)$. Hence the diffusivity estimator in Equation (6.2.9) can be written as

$$\hat{\kappa}^{(P)}(t) = \frac{\Delta}{2N}\left|\sum_{t=1}^{N} u_t\right|^2 = \frac{\Delta\sigma_I^2}{2}Z^2, \tag{6.2.31}$$

where $Z \sim N(0,1)$. Since $Z^2 \sim \chi_1^2$, we get that

$$\frac{2}{\Delta\sigma_I^2}\hat{\kappa}^{(P)}(t) \sim \chi_1^2. \tag{6.2.32}$$

Therefore the diffusivity estimator of a single white noise process follows a re-scaled χ_1^2 distribution, which is shown in Figure 6.2.1. The estimate has expectation and variance given by

$$\mathrm{E}\left[\hat{\kappa}^{(P)}(t)\right] = \frac{\Delta\sigma_I^2}{2} = \kappa(t) \tag{6.2.33}$$

$$\mathrm{var}\left[\hat{\kappa}^{(P)}(t)\right] = \frac{\Delta^2\sigma_I^4}{2} = 2\kappa(t)^2. \tag{6.2.34}$$

The periodogram diffusivity estimator applied to a white noise velocity process produces unbiased estimates. However, the variance does not depend on the length of the time series, N, and so adding additional data points does not improve the estimate. Furthermore, the RMSE of the estimate is

$$\mathrm{RMSE}\left[\hat{\kappa}^{(P)}(t)\right] = \frac{\Delta\sigma_I^2}{\sqrt{2}} = \sqrt{2}\kappa(t). \tag{6.2.35}$$

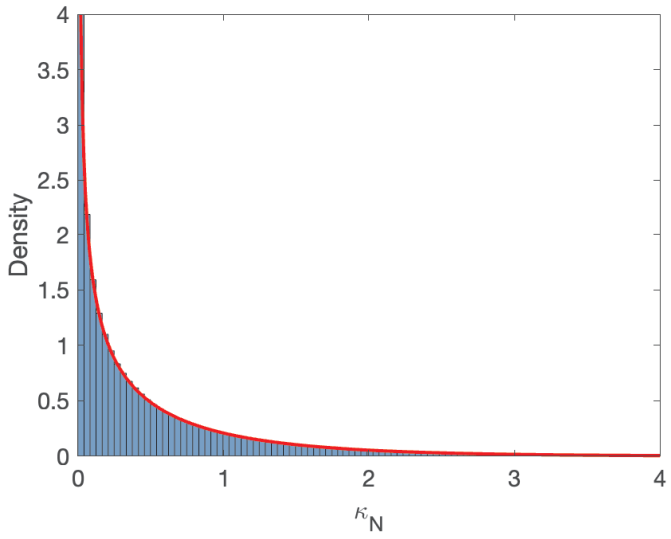

Figure 6.2.1: Distribution of $\hat{\kappa}^{(P)}(t)$ for a white noise velocity process. The histogram shows the density found by simulating 5000 white noise processes, with $\sigma_I = 1$, $\Delta = 1$, of length $N = 1000$, and finding the diffusivity of each individual time series. The red curve shows the PDF of the re-scaled chi-squared distribution given in Equation (6.2.32).

The RMSE is larger than the value that we are estimating, this means that we would get a smaller RMSE by simply estimating the diffusivity to be zero and getting an RMSE of $\kappa(t)$. The periodogram-based estimator is therefore clearly a poor estimator of diffusivity for white noise velocities, despite being the best of the three proposed in Section 6.2.2. The high error is solely due to the variance of the estimator, and in Section 6.3 we will propose an estimator which reduces this variance.

AR(1) Velocity Process

We can show that the diffusivity estimates of an AR(1) velocity process also follow a chi-squared distribution. We start by writing the AR(1) process as

$$u_t = \phi u_{t-1} + \epsilon_t = \phi^t u_0 + \sum_{n=1}^{t} \phi^{t-n} \epsilon_n. \tag{6.2.36}$$

The expectation and variance of an AR(1) process are

$$E(u_t) = 0, \qquad \text{var}(u_t) = \frac{\sigma_I^2}{1-\phi^2}. \tag{6.2.37}$$

To derive the distribution of the diffusivity estimator, we write the sum of velocities from Equation (6.2.36) up to N observations as

$$\sum_{t=1}^{N} u_t = \frac{\phi(\phi^N - 1)}{\phi - 1} u_0 + \sum_{k=1}^{N} \frac{\phi^{N+1-k} - 1}{\phi - 1} \epsilon_k, \tag{6.2.38}$$

and calculate the variance of the sum of velocities to be

$$\text{var}\left(\sum_{t=1}^{N} u_t\right) = \frac{\phi^2(\phi^N - 1)^2}{(\phi-1)^2} \frac{\sigma_I^2}{1-\phi^2} + \sum_{k=1}^{N} \left(\frac{\phi^{N+1-k} - 1}{\phi - 1}\right)^2 \sigma_I^2 \tag{6.2.39}$$

$$= \frac{2\sigma_I^2 \phi(\phi^N - 1) + N(1-\phi^2)}{(\phi-1)^2(1-\phi^2)}. \tag{6.2.40}$$

Defining $Z = \frac{\sum_{t=1}^{N} u_t}{\sqrt{\text{var}(\sum_{t=1}^{N} u_t)}} \sim N(0,1)$, the diffusivity can be written as as

$$\hat{\kappa}^{(P)}(t) = \frac{\Delta}{2N}\left|\sum_{t=1}^{N} u_t\right|^2 = \frac{\Delta}{2N}\text{var}\left(\sum_{t=1}^{N} u_t\right) Z^2. \tag{6.2.41}$$

As was the case for a white noise process, the diffusivity follows a chi-squared distribution, and is distributed as

$$\frac{2N(\phi-1)^2(1-\phi^2)}{\Delta\sigma_I^2[2\phi(\phi^N-1)+N(1-\phi^2)]}\hat{\kappa}^{(P)}(t) \sim \chi_1^2. \tag{6.2.42}$$

A histogram of the diffusivity of an AR(1) velocity process is shown in Figure 6.2.2 for different choices of the damping parameter, ϕ. We see that the damping parameter vastly changes the distribution, and therefore will change the error of estimates. The expectation and variance of the estimator are

$$\text{E}\left[\hat{\kappa}^{(P)}(t)\right] = \frac{\Delta\sigma_I^2[2\phi(\phi^N-1)+N(1-\phi^2)]}{2N(\phi-1)^2(1-\phi^2)} \tag{6.2.43}$$

$$\text{var}\left[\hat{\kappa}^{(P)}(t)\right] = \frac{\Delta^2\sigma_I^4[2\phi(\phi^N-1)+N(1-\phi^2)]^2}{2N^2(\phi-1)^4(1-\phi^2)^2}. \tag{6.2.44}$$

For an AR(1) process, the diffusivity estimator is biased, unlike for a white noise process. Specifically, the bias of estimating κ_∞ by $\hat{\kappa}(t)$ is

$$\text{Bias}\left[\hat{\kappa}^{(P)}(t)\right] = \frac{\Delta\sigma_I^2\phi(\phi^N-1)}{N(\phi-1)^2(1-\phi^2)}. \tag{6.2.45}$$

The bias is $\mathcal{O}(\frac{1}{N})$. However, a large number of observations could be required to make the bias of the estimator sufficiently small. This is not necessarily realistic for modelling the ocean because as a drifter moves though the ocean it will leave the area for which we are estimating the diffusivity, meaning that the number of observations within a region are finite. The RMSE follows from combining the variance and bias

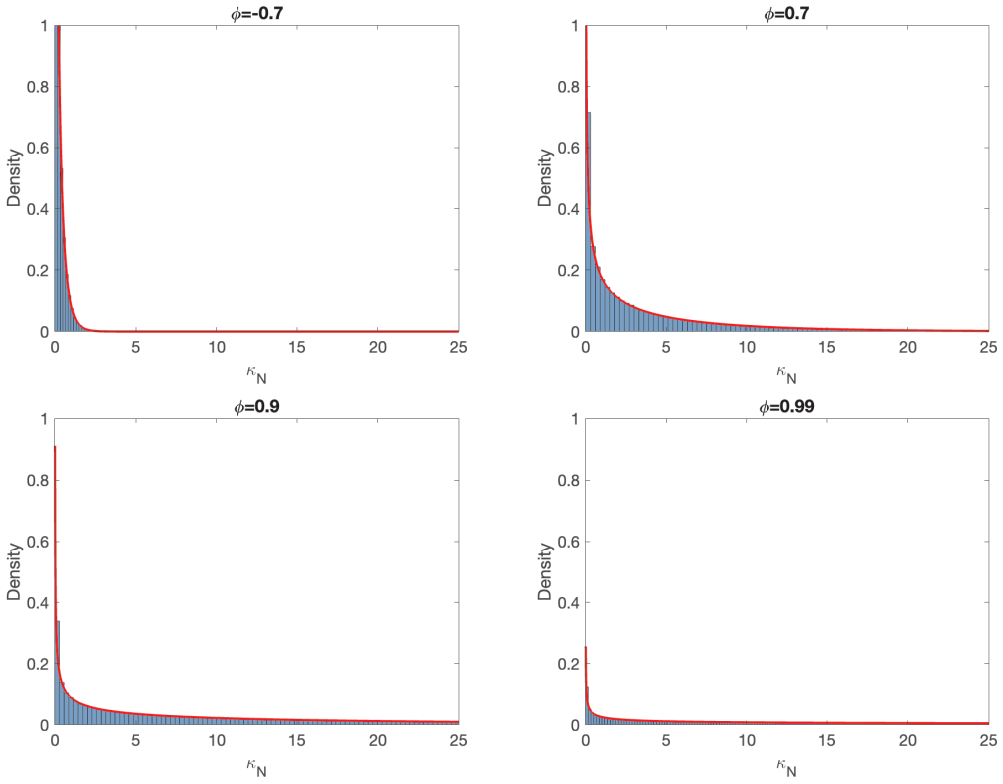

Figure 6.2.2: Distribution of the periodogram-based estimator of the diffusivity of an AR(1) velocity process, as given in Equation (6.2.42). The histogram shows the density found by simulating 200,000 AR(1) processes of length $N = 10$ where $\sigma_I = 1$, and finding the diffusivity of each individual time series. The four plots correspond to different values of ϕ, where ϕ is chosen to be one of $-0.7, 0.7, 0.9$ and 0.99. The red curve shows the PDF of the scaled chi-squared distribution.

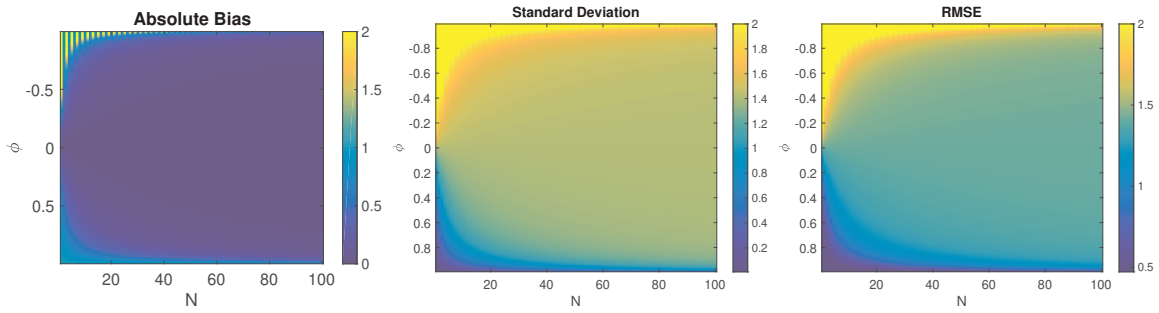

Figure 6.2.3: Normalised bias (left), standard deviation (centre) and RMSE (right) of the periodogram-based diffusivity estimator with AR(1) velocities for $\phi = [-1, 1]$. For each plot we have set $\sigma_I = 1$, $\Delta = 1$ and considered all values of ϕ for which the process is stationary, and for $N = 1, 2, \ldots, 100$. The plots have been normalised by the long-term diffusivity of an AR(1) process.

from Equations (6.2.44) and (6.2.45)

$$\text{RMSE}\left[\hat{\kappa}^{(P)}(t)\right] = \frac{\Delta\sigma_I^2\sqrt{\phi^2(\phi^N-1)^2 + \left[\phi(\phi^N-1) + \frac{N}{2}(1-\phi^2)\right]^2}}{N(\phi-1)^2(1-\phi^2)}. \tag{6.2.46}$$

In the limit as $N \to \infty$ the RMSE goes to $\Delta\sigma_I^2/\sqrt{2}(\phi-1)^2 = \sqrt{2}\kappa_\infty$. This means that the error doesn't go to zero with increased observations. As was the case for the white noise process, the RMSE would be lower if we were to estimate the diffusivity to simply be zero when we only have data from a single time series.

Figure 6.2.3 shows the bias, standard deviation and RMSE of the diffusivity estimator for an AR(1) velocity process, normalised by κ_∞, for different damping parameters and different numbers of observations. For all values of ϕ and N the RMSE will approach $\sqrt{2}\kappa_\infty$, as expected from Equation (6.2.46). This corresponds to the normalised RMSE being $\sqrt{2}$, which can be seen in the Figure for each different damping parameter whenever N is sufficiently large. Hence for an AR(1) process, in order

to improve the estimator we seek to reduce both bias and variance, but primarily variance since for large N the bias vanishes, but the variance remains.

6.2.5 Combining Multiple Drifter Time Series

The goal of this chapter is to reduce the RMSE in order to improve diffusivity estimates. However, as we saw in the last section, for both Gaussian white noise and AR(1) processes the RMSE of the diffusivity is greater than the diffusivity itself. The variance in the estimates is not improved by increasing N for either Gaussian white noise or AR(1) processes. Thus we need to consider other approaches to reducing variance.

One way in which we can reduce variance, and hence reduce the RMSE, is to calculate diffusivity using trajectories from multiple drifters. The Global Drifter Program has a large number of drifters, and so there is typically information from more than one drifter in each part of the ocean, although not necessarily at the same time. In this section we look at the theoretical properties of averaging diffusivity estimates over multiple drifters.

We write the diffusivity estimate calculated by averaging over multiple drifters as

$$\hat{\kappa}_M^{(P)}(t) = \frac{1}{M}\sum_{j=1}^{M} \hat{\kappa}_j^{(P)}(t), \tag{6.2.47}$$

where $\hat{\kappa}_j^{(P)}(t)$ are diffusivity estimates from individual drifter time series as in Equation (6.2.9), M is the total number of time series and j is an index representing individual time series.

We show how averaging over multiple drifters, as in Equation (6.2.47), improves

the diffusivity estimates of the white noise and AR(1) processes. We will repeat the derivations in Section 6.2.4 to calculate the bias, variance and RMSE for the diffusivity estimates for each process.

White Noise Velocity Process

First we consider the white noise velocity process. We showed in Equation (6.2.32) that the diffusivity estimates follow a chi-squared distribution with one degree of freedom. If we take M independent time series of length N, and average the diffusivity estimates of each time series, we show that the periodogram-based estimate at time $t = N\Delta$ follows a χ^2_M distribution. We have from Equation (6.2.32) that the diffusivity of a single particle following a white noise process is distributed as $\frac{2}{\Delta\sigma_I^2}\hat{\kappa}_j^{(P)}(t) \sim \chi^2_1$. The average diffusivity over M particles can be found by writing $z_m \sim \chi^2_1$ and applying Equation (6.2.47),

$$\hat{\kappa}_M^{(P)}(t) = \frac{\Delta}{M}\sum_{m=1}^{M}\frac{\sigma_I^2}{2}z_m = \frac{\Delta}{M}\frac{\sigma_I^2}{2}z_{nm}, \tag{6.2.48}$$

where $z_{nm} \sim \chi^2_M$. Therefore the diffusivity is distributed as

$$\frac{2M}{\Delta\sigma_I^2}\hat{\kappa}_M^{(P)}(t) \sim \chi^2_M. \tag{6.2.49}$$

Hence the estimate follows a chi-squared distribution with M degrees of freedom, as shown in Figure 6.2.4. The figure shows the distribution of the estimates for different values of M. As $M \to \infty$ the distribution will converge towards a normal distribution centred around the true diffusivity value by the central limit theorem.

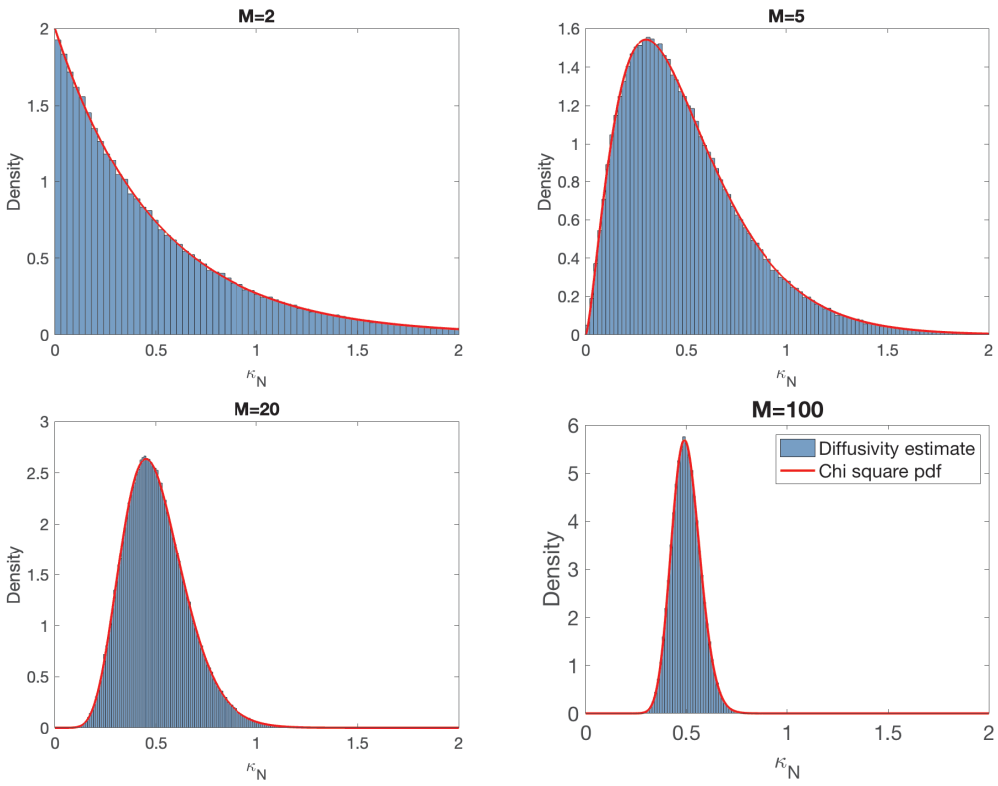

Figure 6.2.4: Histogram of diffusivity estimates using periodogram estimator, for a white noise velocity process with M = 2, 5, 20 and 100. Here we simulated 200,000 white noise processes of length $N = 2$, where $\sigma_I = 1$, $\Delta = 1$, such that $\kappa_\infty = 0.5$. The the red line corresponds to a chi-squared distribution from Equation (6.2.49).

The expectation and variance are,

$$\mathrm{E}\left[\hat{\kappa}_M^{(P)}(t)\right] = \frac{\Delta\sigma_I^2}{2} = \kappa(t) \tag{6.2.50}$$

$$\mathrm{var}\left[\hat{\kappa}_M^{(P)}(t)\right] = \frac{\Delta^2\sigma_I^4}{2M} = \frac{2\kappa(t)^2}{M}. \tag{6.2.51}$$

Hence, the estimates are unbiased, and for $M \to \infty$ the variance goes to zero. The RMSE of the estimate is

$$\mathrm{RMSE}\left[\hat{\kappa}_M^{(P)}(t)\right] = \frac{\Delta\sigma_I^2}{\sqrt{2M}} = \frac{\sqrt{2}\kappa(t)}{\sqrt{M}}. \tag{6.2.52}$$

When $M \geq 3$, then $\mathrm{RMSE}[\hat{\kappa}_M^{(P)}(t)] < \kappa_\infty$, meaning that at least three time series are needed before the periodogram-based estimator yields meaningful estimates of the long-term diffusivity.

AR(1) Velocity Process

We show that the diffusivity estimate of an AR(1) velocity process also follows a chi-squared distribution with M degrees of freedom. We calculate the diffusivity from Equation (6.2.47) to be

$$\hat{\kappa}_M^{(P)}(t) = \frac{1}{M}\sum_{m=1}^{M} \frac{\Delta 2\phi(\phi^N - 1) + N(1-\phi^2)}{2N(\phi-1)^2(1-\phi^2)} z_m^2, \tag{6.2.53}$$

$$= \frac{\Delta}{M} \frac{2\phi(\phi^N - 1) + N(1-\phi^2)}{2N(\phi-1)^2(1-\phi^2)} z_{nm}, \tag{6.2.54}$$

where $z_m^2 \sim \chi_1^2$ and $z_{nm} \sim \chi_M^2$. Therefore the diffusivity of an AR(1) process is distributed as

$$\frac{2NM(\phi-1)^2(1-\phi^2)}{\Delta[2\phi(\phi^N-1) + N(1-\phi^2)]} \hat{\kappa}_M^{(P)}(t) \sim \chi_M^2. \tag{6.2.55}$$

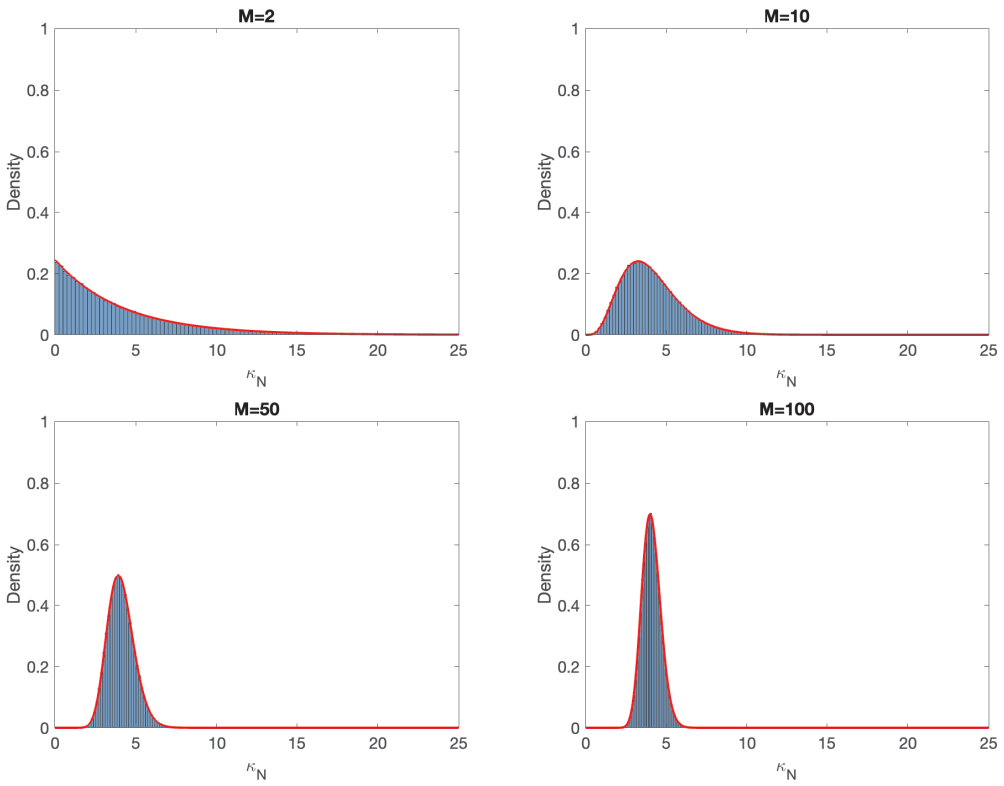

Figure 6.2.5: Histogram of diffusivity estimates using periodogram estimator, for AR(1) velocity process with $\phi = 0.7$ and $M = 2, 10, 50, 100$. Here we sampled 200,000 AR(1) processes of length $N = 10$ with $\sigma_I = 1$, $\Delta = 1$, such that $\kappa_\infty = 5.55$. The red curve corresponds to the distribution of Equation (6.2.55).

Diffusivity estimates of an AR(1) process averaged over M drifters are shown in Figure 6.2.5 for different choices of M. As was the case for the white noise process, as $M \to \infty$ the distribution of the diffusivity estimates approaches a normal distribution centred around the true value.

The expectation and variance of the diffusivity estimates are

$$\mathrm{E}\left[\hat{\kappa}_M^{(P)}(t)\right] = \frac{\Delta\sigma_I^2[2\phi(\phi^N - 1) + N(1-\phi^2)]}{2N(\phi-1)^2(1-\phi^2)} \tag{6.2.56}$$

$$\mathrm{Var}\left[\hat{\kappa}_M^{(P)}(t)\right] = \frac{\Delta^2\sigma_I^4[2\phi(\phi^N - 1) + N(1-\phi^2)]^2}{2MN^2(\phi-1)^4(1-\phi^2)^2}. \tag{6.2.57}$$

The bias is the same as when the diffusivity was calculated from a single time series, and so averaging over multiple time series will not reduce the bias. However the variance depends on the number of time series, M, and so as we average the diffusivity estimates from multiple drifters, the variance decreases as expected. The RMSE is

$$\mathrm{RMSE}\left[\hat{\kappa}_M^{(P)}(t)\right] = \Delta\sqrt{\frac{2\sigma_I^4 M\phi^2(\phi^N-1)^2 + \sigma_I^4[2\phi(\phi^N-1)+N(1-\phi^2)]^2}{2MN^2(\phi-1)^4(1-\phi^2)^2}}. \tag{6.2.58}$$

Figure 6.2.6 shows the bias, standard deviation and RMSE of the estimator, normalised by the long-term diffusivity, for different values of N and M. We see that as we increase N the bias will tend towards 0, similarly as we increase M the variance will approach 0. Hence, in order for the RMSE decrease to 0, we require large values for both N and M. In practice this is not realistic, we cannot obtain long time series from a single drifter, as the drifter moves freely around the ocean into different regions, and so the data must be split into different parts of the ocean. Similarly, we cannot have always have a large number of drifters in a given region.

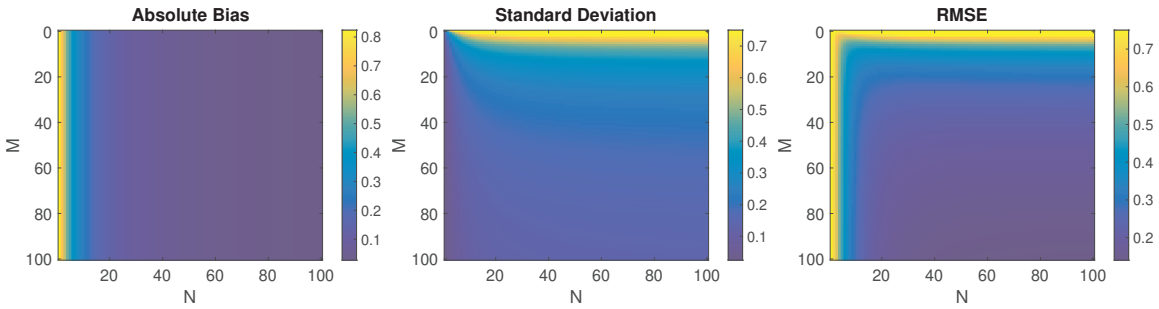

Figure 6.2.6: Normalised bias, variance and RMSE of periodogram-based estimator for $\phi = 0.7$. For each plot we have set $\sigma_I = 1$ and $\Delta = 1$, such that $\kappa_\infty = 5.55$, and then considered a range of values for N and M. The colour on each plot shows the value relating to its title, for example on the left hand plot the colour is the bias. The plots have been normalised by the long-term diffusivity, and capped at 0.75 for visualisation purposes.

This means that while theoretically averaging over M time series can give a good estimate, in practice we require techniques that work with a given small and finite set of drifter data. This motivates finding a better way to estimate diffusivity in the ocean with a fixed sample size.

6.3 Reducing Variance from a Single Time Series

In Section 6.2 we showed that the error of diffusivity estimates for a white noise velocity process cannot be reduced when estimated from a single drifter by changing the length of the observed time series N. For an AR(1) velocity process, the error of the diffusivity estimates asymptote to a constant value as N is increased, and therefore cannot be reduced to zero. When the diffusivity is found as an average

from multiple time series, at least three time series are required before the estimates become statistically significant, which isn't possible in all regions of the ocean.

In this section we introduce a method of estimating diffusivity which minimises the RMSE when estimating from a single time series. By minimising the RMSE for just one drifter we can find a method that works when we only have one time series, but can be expanded to multiple drifters if applicable to further improve the estimate.

In Section 6.2.4 we showed that the diffusivity estimate from a single time series has a high RMSE due to the variance. For the white noise velocity process, the error is completely due to variance as the estimator is unbiased, and the RMSE doesn't depend on the length of the time series, as it is a constant value irrespective of N. When averaging over M time series, the RMSE reduces as M increases.

Similarly for the AR(1) velocity process, for large N diffusivity estimates become unbiased, and so the RMSE of the long-term diffusivity is mostly due to the large variance, which approaches a constant non-zero value as $N \to \infty$. However, as the estimator is biased for smaller N, both bias and variance should be considered when constructing a new method for estimating the diffusivity.

We propose a new way to estimate the diffusivity which involves splitting a single time series into smaller chunks, and treating each chunk as a new time series. We then estimate the diffusivity separately for each segment and then average these estimates to obtain a new estimator. This method can be thought of as calculating the rate of squared displacement gradually over sections over the time series and averaging them to get the rate of dispersion of the overall time series. Hence we are finding local diffusivity along segments of the time series, and then averaging these to find the

overall diffusivity over a larger spatial distance. This method smooths out the effects of areas which have unusually high or low diffusivity compared with the areas of the ocean around them, and so physical effects such as tides should have less impact on the estimated diffusivity.

This idea is related to the concept of smoothing spectral density estimates, using techniques such as Welch's overlapped segment averaging (WOSA) (Percival et al., 1993). This method reduces variance, in the same way as we showed in Section 6.2.5 that averaging diffusivity estimates over a greater number of time series will reduce the variance. However, we note that diffusivity estimates calculated as an average over chunks from a single time series will have a higher bias than if we were just to calculate diffusivity from the full time series. This is because we are now working with time series of a shorter length, and we showed in Section 6.2 shorter time series yield a higher bias.

We will derive the distribution of diffusivity estimates from a single time series chunked into smaller sections for a white noise and AR(1) velocity process respectively, and use these to derive an expression for the error. We will show that the error can be reduced from a single time series by applying this method.

6.3.1 Illustration: Variance Reduction of Diffusivity Estimates from a White Noise Velocity Process

We begin by looking at the simple case of a white noise velocity process, and split the time series into multiple segments. To illustrate the core idea, we start with the

extreme and simplified case of splitting the time series into N chunks each of size 1.

For a white noise velocity process, estimating diffusivity as an average of chunks of length 1 is equivalent to estimating the diffusivity to be half the sample variance, which we write as

$$\hat{\kappa}_N^{(CP)}(t) = \frac{\Delta}{2N} \sum_{i=1}^{N} |u_i|^2, \tag{6.3.1}$$

where $\hat{\kappa}_m^{(CP)}(t)$ denotes the periodogram-based estimate of a time series of length N which has been chunked into m new time series, where in the case of Equation (6.3.1), the number of chunks m has been set to equal to the number of observations in the original time series N.

This can be written in terms of the chi-squared distribution. We let $U_i = u_i/s \sim N(0,1)$. This gives

$$\hat{\kappa}_N^{(CP)}(t) = \frac{\Delta}{2N} \sum_{i=1}^{N} |sU_i|^2 = \frac{\Delta\sigma_I^2}{2N} Z^2, \tag{6.3.2}$$

where $Z^2 = \sum_{i=1}^{N} U_i^2 \sim \chi_N^2$. Hence the estimate is distributed as

$$\frac{2N}{\Delta\sigma_I^2} \hat{\kappa}_N^{(CP)}(t) \sim \chi_N^2. \tag{6.3.3}$$

This distribution can be seen in Figure 6.3.1. We see that this distribution is less variable than the distribution shown in Figure 6.2.1, where the diffusivity was estimated across the full length of the time series, with the estimates distributed evenly around the true diffusivity of $0.5\text{m}^2/\text{s}$.

The expectation and variance of the diffusivity estimate where we have split the

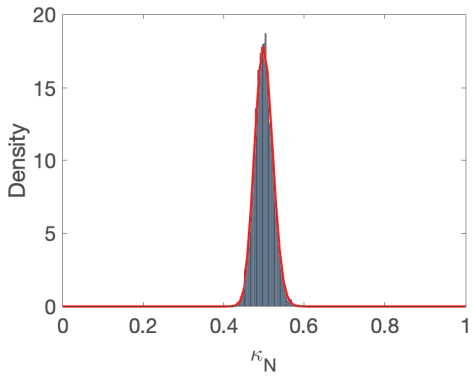

Figure 6.3.1: Diffusivity estimate of a white noise velocity process found by splitting a time series into N chunks of length 1. The histogram has been found by simulating 10,000 time series of length $N = 1,000$ with $\sigma_I = 1$, and $\Delta = 1$s, such that $\kappa_\infty = 0.5$, and finding the average diffusivity estimate of each time series of length 1. The red curve represents the pdf of the chi-squared distribution given in Equation (6.3.3).

time series into N chunks of length 1 are

$$\mathrm{E}\left[\hat{\kappa}_N^{(CP)}(t)\right] = \frac{\Delta\sigma_I^2}{2} = \kappa(t) \tag{6.3.4}$$

$$\mathrm{var}\left[\hat{\kappa}_N^{(CP)}(t)\right] = \frac{\Delta^2\sigma_I^4}{2N}. \tag{6.3.5}$$

As with the periodogram-based diffusivity estimator for a white noise velocity process, this estimator is unbiased. However, the distribution of the estimate is now dependent on the length of the time series, and so the variance of the estimate will decrease as N increases.

From the expectation and variance we can calculate the RMSE to be

$$\mathrm{RMSE}\left[\hat{\kappa}_N^{(CP)}(t)\right] = \frac{\Delta\sigma_I^2}{\sqrt{2N}} = \sqrt{\frac{2}{N}}\kappa_\infty. \tag{6.3.6}$$

After splitting the time series into chunks, the error now depends on the length of the

time series, and hence the RMSE can be reduced from a single time series, unlike the error for diffusivity estimates from the periodogram estimator in Section 6.2.4.

6.3.2 Average Chunked Periodogram Estimator

The idea of partitioning a single time series into multiple segments can be generalised into different chunk sizes, as well as different time series processes, as we shall now describe. Specifically, we partition a time series into m non-overlapping chunks of length $n = \lfloor N/m \rfloor$. The diffusivity for each new smaller time series is then estimated, and averaged of the m segments. This estimator will in general increase bias compared with the periodogram-based estimator because the length of the time series is decreased, however it will also decrease the variance as we now have multiple time series to average over.

We initially define the Average Chunked Periodogram (ACP) diffusivity estimator to be

$$\hat{\kappa}_m^{(CP)}(t) = \frac{1}{m} \sum_{j=1}^{m} \hat{\kappa}_j^{(P)}(t), \tag{6.3.7}$$

where $\hat{\kappa}_j^{(P)}(t), j = 1, \ldots, m_i$ are diffusivity estimates from each time series chunk as defined in Equation (6.2.9).

We note that depending on the values of N and m it might not be possible to split a time series into equal sized chunks using all the data. This could mean that some data are not used, which seems wasteful. If this is the case we can introduce two different chunk lengths—either length $\lfloor N/m \rfloor$ or $\lfloor N/m \rfloor + 1$. Specifically, we take $m_1 = N - nm$ chunks of size $n_1 = n + 1$, and $m_2 = m - N + nm$ of size $n_2 = n$. We

have that $m_1 + m_2 = m$ and $m_1 n_1 + m_2 n_2 = N$, and no data are wasted.

Then the ACP diffusivity estimator is fully defined as

$$\hat{\kappa}_m^{(CP)}(t) = \frac{1}{m}\left[\sum_{j=1}^{m_1} \hat{\kappa}_{n_1,j}^{(P)}(t) + \sum_{j=1}^{m_2} \hat{\kappa}_{n_2,j}^{(P)}(t)\right], \tag{6.3.8}$$

where $\hat{\kappa}_{n_i,j}^{(P)}(t), i = 1, 2, j = 1, \ldots, m_i$ are diffusivity estimates from each time series chunk where the first m_1 chunks are length n_1 and the second m_2 chunks are length n_2. Note that setting $m = N$ recovers the diffusivity estimator of Equation (6.3.1) proposed for white noise velocities, and that if all chunks are of the same size then $m_1 = m_2 = m$ and Equation (6.3.7) is recovered.

6.3.3 ACP Estimates from an AR(1) Velocity Process

We now apply the ACP estimator to an AR(1) velocity process and show how splitting an AR(1) time series into chunks affects the RMSE of the diffusivity estimates. For this process it is not as straightforward as the white noise process to deduce the optimal number of chunks into which to split the time series to minimise the RMSE, and how this is done will be shown in Section 6.4.

From Equation (6.2.42) diffusivity estimates for each chunk from an AR(1) velocity process follow

$$\hat{\kappa}_{n_i,j}^{(P)}(t) = \frac{\Delta\sigma_I^2[2\phi(\phi^{n_i} - 1) + n_i(1 - \phi^2)]}{2n_i(\phi - 1)^2(1 - \phi^2)} z^2, \tag{6.3.9}$$

for $j = 1, \ldots, m$ and $z^2 \sim \chi_1^2$. Combining Equations (6.3.8) and (6.3.9), the AR(1)

ACP diffusivity estimator approximately follows

$$\hat{\kappa}_m^{(CP)}(t) \approx \frac{\Delta\sigma_I^2}{m}\left[\frac{2\phi(\phi^{n_1}-1)+n_1(1-\phi^2)}{2n_1(\phi-1)^2(1-\phi^2)}z_{m_1}+\frac{2\phi(\phi^{n_2}-1)+n_2(1-\phi^2)}{2n_2(\phi-1)^2(1-\phi^2)}z_{m_2}\right], \tag{6.3.10}$$

where $z_{m_i} \sim \chi^2_{m_i}$. Equation (6.3.10) is an approximation of the diffusivity distribution as we have assumed that the $\hat{\kappa}^{(P)}_{n_i,j}(t)$ terms in Equation (6.3.9) are independent. This is only approximately true, and the accuracy of this approximation is discussed at the end of this section.

We calculate the expected value of the ACP estimator to be

$$E\left[\hat{\kappa}_m^{(CP)}(t)\right] \approx \frac{\Delta\sigma_I^2}{m_1+m_2}\left[\frac{m_1n_2[2\phi(\phi^{n_1}-1)+n_1(1-\phi^2)]+m_2n_1[2\phi(\phi^{n_2}-1)+n_2(1-\phi^2)]}{2n_1n_2(\phi-1)^2(1-\phi^2)}\right]. \tag{6.3.11}$$

The bias can be calculated from the expectation to be

$$\text{bias}\left[\hat{\kappa}_m^{(CP)}(t)\right] \approx \frac{\Delta\sigma_I^2[m_1n_2\phi(\phi^{n_1}-1)+m_2n_1\phi(\phi^{n_2}-1)]}{(m_1+m_2)n_1n_2(\phi-1)^2(1-\phi^2)}. \tag{6.3.12}$$

The bias is dependent on both the number of chunks and the length of these chunks. For the case when $m_1 = m_2$ then all chunks are the same length and the bias reduces to the form given earlier in Equation (6.2.45) when we had just a single time series. Hence the bias is largely unaffected by the number of chunks, and is instead reduced by increasing the length of each chunk. This means that to reduce bias we would need to minimise the number of chunks.

The variance is

$$\text{var}\left[\hat{\kappa}_m^{(CP)}(t)\right] \approx \frac{\Delta^2\sigma_I^4}{(m_1+m_2)^2}\left\{\frac{m_1n_2^2\left[2\phi(\phi^{n_1}-1)+n_1(1-\phi^2)\right]^2+m_2n_1^2\left[2\phi(\phi^{n_2}-1)+n_2(1-\phi^2)\right]^2}{2n_1^2n_2^2(\phi-1)^4(1-\phi^2)^2}\right\}. \tag{6.3.13}$$

The variance also depends on both the number and length of chunks. However the variance is minimised by increasing the number of chunks. Hence we observe a trade-off between the bias and variance to minimise the RMSE.

The RMSE can be calculated as

$$\text{RMSE}\left[\hat{\kappa}_m^{(CP)}(t)\right] \approx \sqrt{\text{bias}\left[\hat{\kappa}_m^{(CP)}(t)\right]^2+\text{var}\left[\hat{\kappa}_m^{(CP)}(t)\right]}. \tag{6.3.14}$$

The RMSE is minimised by increasing both the number of chunks n_1, n_2, and the length of these chunks m_1, m_2. The number and length of the chunks are not independent of one another, and cannot both be simultaneously be increased. We therefore need to find the optimal length of a chunk which balances the bias and variance to find the overall minimised RMSE.

This error is only an approximation to the true error as we assumed in Equation (6.3.10) that the chunks are independent, which is not the case since they were formed from a single time series following an AR(1) process where the velocity at time t depends on the velocity at time $t-1$. We therefore need to verify that this assumption is appropriate and show that the effect that this dependence has on the error is small.

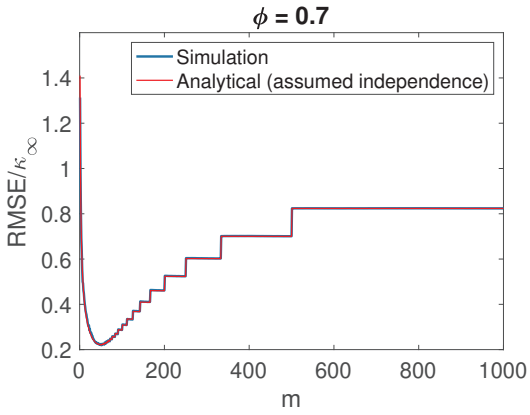

Figure 6.3.2: The analytical RMSE where independence is assumed, and simulated RMSE where chunks are dependent, both normalised by κ, for an AR(1) process with $\phi = 0.7$, $\sigma_I = 1$, $\Delta = 1$, and $N = 1000$. Values are shown for a range of chunk numbers m ranging from 1 to 1000.

Dependence Between Chunks

We have assumed up to now that chunks are independent of one another, and so we note that the RMSE given in Equation (6.3.14) is only approximate. To consider whether this assumption is appropriate, we simulate 1000 time series of length 1000 from an AR(1) velocity process with damping parameter $\phi = 0.7$, and compare the RMSE of computing Equation (6.3.8) with the analytical expression of Equation (6.3.14), as shown in Figure 6.3.2 for a range of values for m (the number of chunks). We see that the simulated and analytical errors are almost identical for all values of m, meaning that the assumption, at first glance, is reasonable.

To further test this assumption, we found the simulated and analytical RMSE for $\phi \in [-0.9, 0.9]$ in steps of 0.1, and produced a histogram of the difference, as shown in Figure 6.3.3. We see that the histogram shows the difference to be extremely small.

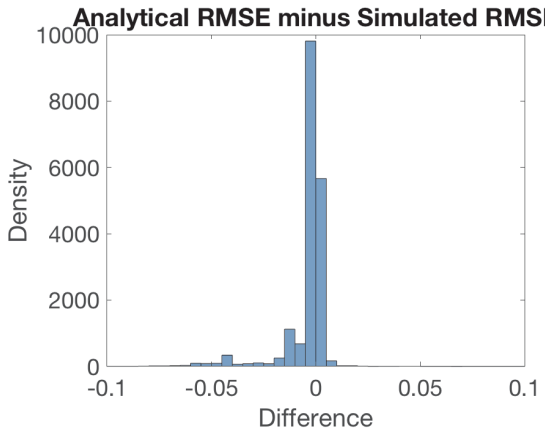

Figure 6.3.3: A histogram displaying the difference in the RMSE of the analytical RMSE where independence is assumed, and simulated RMSE where chunks are dependent. The error is calculated from 1000 simulated AR(1) processes with $N = 1000$ for ϕ between -0.9 and 0.9 in intervals of 0.1 for all possible values od m, with $\sigma_I = 1$ and $\Delta = 1$.

The average difference between the RMSE from simulations and the analytic approximations is only 0.0284% of the value of the true diffusivity. The apparent small difference between the simulations and analytical expressions is promising. Analytically bounding the error induced by this approximation is left for future work.

6.4 Window length selection for ACP Estimator

The window length (or equivalently the number of chunks) of the ACP estimator will need to be chosen to minimise the RMSE of the resulting estimate. We discussed in Section 6.3.3 that minimising the RMSE will result in a trade-off between the bias and variance as they cannot be simultaneously minimised—reducing the bias increases

the variance and vice versa. In fact this trade-off can be observed in the experiment performed in Figure 6.3.2 for an AR(1) velocity process where for $N = 1000$ it appears a value of m around 50 is optimal (such that corresponding window lengths are around 20). The choice of the window length is process dependent, and in this section we discuss how the chunk length can be chosen.

To emphasise an important distinction when we refer to "chunking"—a drifter will move through different regions of the ocean during its lifetime, and therefore will need to be partitioned into segments where the currents and diffusivity can be assumed to be approximately stationary. We assume this pre-processing step has occurred, and through this chapter we assume that all time series have been split into stationary segments. Choosing the window length in this section instead refers to splitting these stationary segments into chunks that the diffusivity estimates are averaged in the ACP diffusivity estimator to reduce error.

To determine this optimal window length, a closed form expression for the error is required. This is because the bias in Equation (6.3.12) cannot be estimated directly from the data since we do not know the true diffusivity. We therefore turn to parametric stochastic models to calculate the RMSE. One such model that could be used is an AR(1) velocity process, and this is the model that we will use. We note that other models could also be used, and an analytical form for the error would need to be found.

6.4.1 Using an AR(1) process as a rough approximation to low frequency motion

An AR(1) process can be used to approximate ocean drifter velocities. Matérn and Ornstein-Uhlenbeck processes have also been used to model velocities by, for example, Lilly et al. (2017), and are a good fit for low frequencies below the frequencies of tides and inertial oscillations. We choose to use the AR(1) process as it is the discrete-time analogue of an Ornstein-Uhlenbeck processes, and no information is lost in the discretisation as drifter velocities are regularly sampled in any case, so fitting to a discrete-time stochastic process has intuitive appeal and benefits.

An Ornsterin-Uhlenbeck process (or equivalently an AR(1) process) provides a good fit to ocean velocity observations at low frequencies, as has been shown in Sykulski et al. (2017a). Higher frequencies do not fit to an Ornstein-Uhlenbeck or AR(1) processes due to physical effects such as tides, wind and inertial oscillations, however the ACP diffusivity estimator only uses frequency zero, and so we do not require a good fit for higher frequencies since they are not directly used. Therefore, an AR(1) process is a reasonable approximation to ocean velocities at low frequency, and can be used for tuning the window length of our diffusivity estimator. We emphasise that the final estimator of diffusivity is ultimately non-parametric, it is only the window length choice where we use a parametric model to select its value.

6.4.2 Optimal window length for an AR(1) Velocity Process

We use an AR(1) velocity process to approximate the optimal window length for the ACP estimator. The optimal window lengths n_1, n_2 are chosen to minimise the RMSE from Equation (6.3.14), which takes the form

$$\text{RMSE}\left[\hat{\kappa}_m^{(CP)}(t)\right] \approx \frac{\Delta\sigma_I^2}{(m_1+m_2)n_1^2n_2^2(\phi-1)^4(1-\phi^2)^2}\Bigg[\left\{[m_1n_2\phi(\phi^{n_1}-1)+m_2n_1\phi(\phi^{n_2}-1)]\right\}^2 + 0.5\left\{m_1n_2^2\left[2\phi(\phi^{n_1}-1)+n_1(1-\phi^2)\right]^2+m_2n_1^2\left[2\phi(\phi^{n_2}-1)+n_2(1-\phi^2)\right]^2\right\}\Bigg]^{1/2}. \tag{6.4.1}$$

Figure 6.4.1 shows the RMSE of the ACP estimator of an AR(1) process of length 1000 for $\phi = -0.7, 0.7, 0.9$ and 0.99 and $\sigma_I = 1$, for different values of m (the number of chunks). Note that here we plot over m and not N, to see the performance of our estimator for different chunk lengths for a fixed value of N. The blue lines correspond to splitting the time series into chunks which all have the same length, and discarding any remaining observations. The red curves correspond to using all drifter velocity observations, and splitting the time series into chunks of lengths n_1 and n_2. The jumping in the blue lines is due to splitting the drifter velocities into equal chunks as some observations are lost if m doesn't divide N exactly. The red curves are much more smooth than the blue lines, and generally have lower RMSE, and hence the RMSE is better minimised if all observations are used in the estimation process as expected.

For each value of ϕ there is a clear global minimum on the plot, and so we can

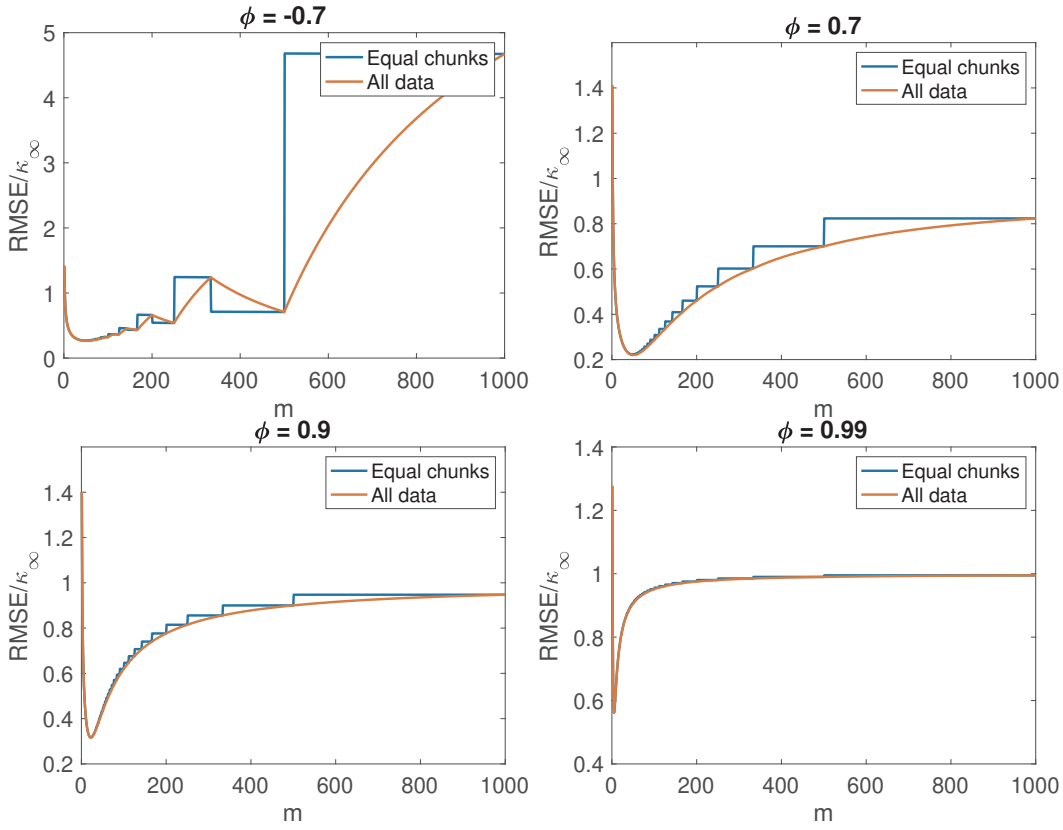

Figure 6.4.1: RMSE for $\phi = -0.7, 0.7, 0.9$ and 0.99 plotted against the number of chunks, m, from a time series of length $N = 1000$, with $\Delta = 1$, using either equal chunk sizes, or chunks of different sizes where every drifter velocity observation is used.

choose a value of m for which the error is minimised. The RMSE when $m = 1$ is equivalent to the periodogram-based estimate using just one time series that we considered in the previous section in Equation (6.2.9). For positive values of ϕ it is the least optimal to set $m = 1$, and so even if we did not chose the value of m which minimises the RMSE it is still optimal to chunk the drifter velocities into multiple time series as opposed to obtaining the periodogram-based estimate directly from the time series.

The RMSE is a tradeoff between the bias and the variance, as shown in Figure 6.4.2, which shows the bias, variance and RMSE of an AR(1) velocity process of length 1000 with $\phi = 0.7$ split into m equal chunks. As the time series is divided into a greater number of chunks, the variance reduces as we are averaging over more time series, however the bias will increase since each time series will not be as long. The value of m which minimises the RMSE should be chosen to optimally balance the bias and variance.

We can approximate the optimal number of chunks for any AR(1) process with $\sigma_I = 1$ for $|\phi| < 1$ and for all N using Equations (6.3.12)-(6.3.14). Figure 6.4.3 shows the optimal value of m which we denote m_{opt} normalised by the length of the time series. If we were to take the mean value of m_{opt} over all $N \leq 1000$ and all ϕ in intervals of 0.01 from -0.99 to 0.99, we get that it is optimal to have $m \approx N/5$. However this takes into account more extreme values, such as where N is small, and ϕ is close to zero. Omitting these values, it appears from inspection that the optimal choice of m_{opt} is around $N/20$.

For $\phi = 0$, which is equivalent to a white noise process, we get that $m_{\text{opt}} = N$

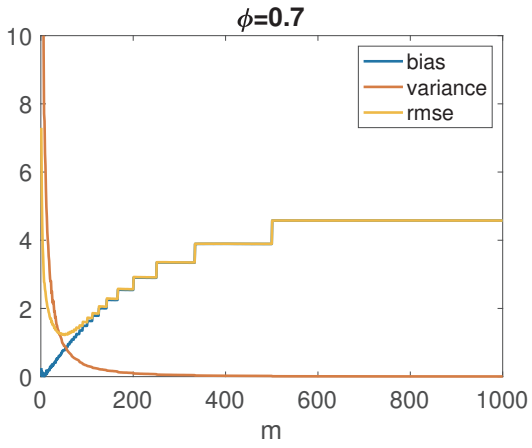

Figure 6.4.2: Bias variance trade-off demonstrated by plotting bias, variance and RMSE for $\phi = -0.7, 0.7, 0.9$ and 0.99, $\Delta = 1$, and $\sigma_I = 1$, plotted against the number of chunks, m, from a time series of length $N = 1000$, comparing using equal chunks, or chunks of different sizes where every drifter velocity observation is used.

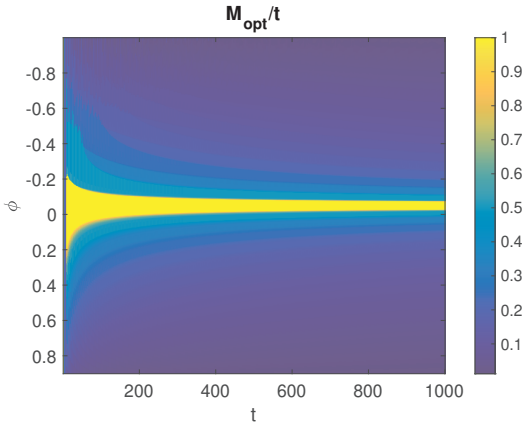

Figure 6.4.3: The normalised optimal number of chunks m_{opt}/t for $\phi \in (-1, 1)$ for a time series of length $t \in (1, 1000)$ and $\sigma_I = 1$.

which is equivalent to chunking the time series into N new time series of length 1, as discussed in Section 6.3.1.

We note that the optimal value of m in Figure (6.4.3) assumes that the white noise signal in the AR(1) process has variance $\sigma_I = 1$, but this only affects the scaling behaviour of the RMSE (see Equation 6.4.1) and hence does not affect the optimal choice of m, therefore the key variables one has to know are ϕ (the autocorrelation at lag-1 between subsequent velocity observations spaced Δ apart) and the time series length N.

Relationship to decorrelation time

The optimal chunk length, $n_{\text{opt}} = N/m_{\text{opt}}$, is related to the concept of decorrelation time, which we denote t_{decorr}, and we define as the time taken for the autocorrelation of the velocity process to become less than $1/e$. The autocorrelation is defined as

$$R(\tau) = \frac{\gamma_\tau}{\gamma_0}. \tag{6.4.2}$$

From Equations (6.2.25) and (6.2.27) which define the variance and autocovariance of an AR(1) process respectively, the autocorrelation of an AR(1) process is simply

$$R(\tau) = \phi^\tau. \tag{6.4.3}$$

The decorrelation time for an AR(1) process is therefore the time when

$$\phi^\tau < 1/e. \tag{6.4.4}$$

For a stationary AR(1) process with positive damping parameter ($0 < \phi < 1$), the decorrelation time is therefore

$$\tau > -\frac{1}{\log(\phi)}, \tag{6.4.5}$$

where the log refers to the natural logarithm. The decorrelation time doesn't hold when the damping parameter is zero or negative due to the nature of the logarithm, but in general we expect drifter velocities to exhibit positive autocorrelation.

Repeating Figure 6.4.3 but plotting the optimal chunk length multiplied by Δ, and scaled by the decorrelation time, $n_{\text{opt}}\Delta/t_{\text{decorr}}$, produces a plot which is approximately flat colour for values of t greater than 300s. Therefore, for time series with $t > 300$, the optimal chunk length is approximately proportional to the decorrelation time. For the particle trajectories simulated in Figure 6.4.3, the median optimal chunk length is $13t_{\text{decorr}}/\Delta$.

6.5 Application of ACP Diffusivity Estimator to drifters

The ACP estimator can be applied to large scale particle trajectories such as the Global Drifter Program. In order to apply the estimator, a number of choices must be made. These choices include: the value of the AR(1) damping parameter ϕ to tune the number of chunks, the initial length of the time series N, the width of spatial bins, whether or not to allow drifter velocities from neighbouring bins, and if a trajectory passes through multiple bins then which bin should the estimate be placed in, or

should the estimate belong to all bins?

We leave the choice of damping parameter, the initial length of the time series and the width of spatial bins to the practitioner, and as an area of future research to establish their optimal tuning. As noted in Chapter 2, Zhurbas and Oh (2004) and Koszalka et al. (2011) use varying bin sizes. In Section 6.6 we will explore different initial time series lengths and bin sizes, but make no attempt to choose these optimally. We choose to place our diffusivity estimates within the spatial bin corresponding to the position of the drifter halfway through the time series. We place the diffusivity estimate just in that single bin but allow the velocity observations from outside the spatial bin to contribute towards the estimate.

An alternative way we could have assigned trajectories to bins would be to split the drifter trajectories into time series that are contained solely within each spatial bin, regardless of the number of observations. Faster moving drifters, for example those close to the equator or on the boundary currents, move through bins more quickly resulting in shorter time series. We chose to instead estimate with fixed length time series so that the estimates are more comparable in terms of the variance from a single time series, but acknowledge that this may cause biases. Some parts of the ocean are more densely populated with drifters, and so in these areas the diffusivity estimates are expected to have lower variances as estimates are averaged over multiple drifters. The chunk length could therefore take the number of drifters in the spatial grid into consideration, and use this number as the number of time series, rather than just the number of time series from an individual drifter. This type of 'optimal tuning' is left for future work.

6.5.1 Mean Flow Estimation

One factor which affects diffusivity estimates is the differing mean flow across the ocean. Because we are not working with clustered drifters, as we did in Chapters 3–5, only the mean flow can be removed, and not mesoscale features such as strain and vorticity, although some mesoscale features will be captured as part of the mean flow.

To estimate diffusivity without mean flow, we can't simply calculate the mean across each time series and remove that to estimate diffusivity, as this will trivially always return a diffusivity of zero. This is because of the way we estimate diffusivity in Equation (6.2.9), as we now show. If we remove the mean velocity across the time series, then the diffusivity estimator becomes

$$\hat{\kappa}^{(P)}(t) = \frac{\Delta}{2N} \left| \sum_{i=1}^{N} \left(u_i - \frac{1}{N} \sum_{j=1}^{N} u_j \right) \right|^2 . \tag{6.5.1}$$

Expanding out gives

$$\hat{\kappa}^{(P)}(t) = \frac{\Delta}{2N} \left| \left(\sum_{i=1}^{N} u_i - \sum_{j=1}^{N} u_j \right) \right|^2 = 0. \tag{6.5.2}$$

This happens due to the way diffusivity is calculated, as we're calculating the square of the sample mean. Therefore, we require an alternative method to estimate mean flow.

To estimate the mean flow, we take an average of all velocity observations within a spatial grid. This mean is subtracted from the velocity of each individual drifter, where the mean removed corresponds to the spatial position of the drifter at each time point and can therefore change throughout the time series. The diffusivity can then be estimated in the same way as in Section 6.3.2.

This method for estimating mean flow uses all observations in time, and does not account for temporal variations to the background velocity at a given spatial location. An example where the mean flow varies over time is the Gulf Stream whose position and amplitude oscillates in time (Zhai et al., 2008). Therefore our mean flow estimates mix spatial and temporal scales. Possible alternatives would be to provide mean flow estimates that vary seasonally or by decade. The choice of the time period over which to estimate the mean flow would introduce another trade-off between the bias and variance of mean flow estimates and is left to further work.

6.5.2 Comparison of ACP Diffusivity Estimator with Periodogram Estimator via Box Simulations

We have previously shown theoretically that the ACP estimator will reduce the error of diffusivity estimates by reducing the variance. We will now demonstrate this error reduction via an idealised simulation environment, where the mean flow and diffusivity are known quantities. We will consider the simple case of white noise velocities with constant mean flow and diffusivity. We then look at a more complicated flow made up from trigonometric functions, but with a homogeneous diffusivity.

Throughout the simulations, we will simulate 10,000 particles across a box with dimensions defined to be 100km×100km. The drifters are equally spaced at the start of the simulation, arranged in a grid shape, with a particle positioned every 1km. We will take hourly observations, equivalent to setting $\Delta = 3600$s. The box is doubly periodic, and so when a particle reaches the boundary of the box, it will wrap around

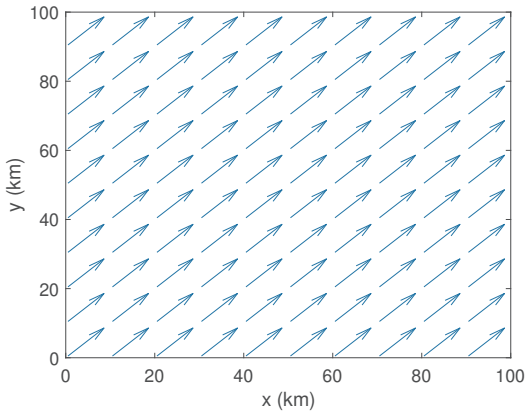

Figure 6.5.1: Mean flow velocity field with $u_t = v_t = 1\text{m/s}$.

as if it were on a torus. Away from the poles the box can be thought of as a simplified representation of the globe.

White Noise Positions with Constant Mean Flow

We begin with the simple case of particles where the velocities follow a two-dimensional white noise process with additional mean flow, and hence the positions follow a two-dimensional random walk with drift. The simulation lasts for $t = 500$ hours with the observation interval set to $\Delta = 1$ hour such that $N = 50$. The mean flow velocity is set to $u_t = v_t = 1\text{m/s}$ (shown in Figure 6.5.1), and the diffusivity is set to $1000\text{m}^2/\text{s}$ in each x and y direction (with zero correlation or cross-diffusivity).

We estimate the mean flow in each direction, using the method described in Section 6.5.1, on a spatial grid of resolution 1km×1km. Figure 6.5.2 shows box plots mean flow estimates aggregated across the 100km×100km box, with each box and whisker covering 20km in either x or y, and all 100km in the other direction. The reason for separating into 20km regions will become apparent in experiments that follow where

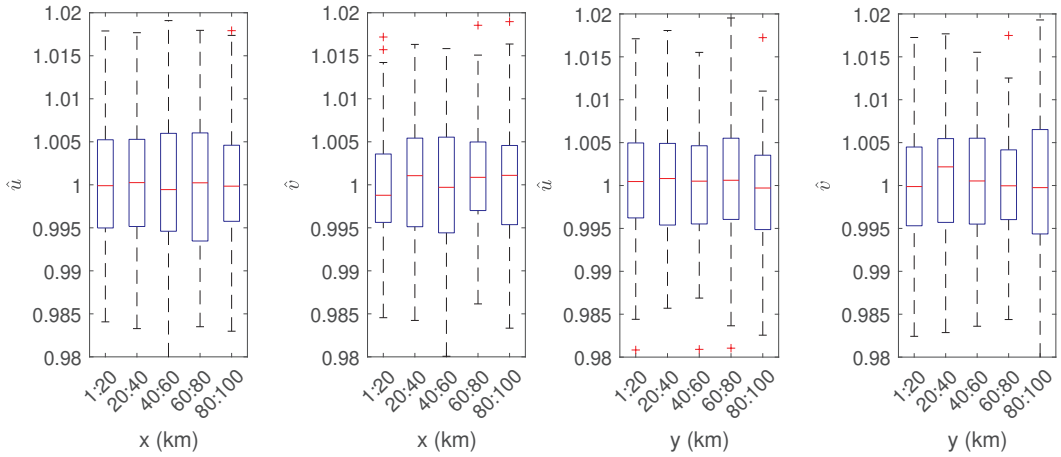

Figure 6.5.2: Box plot of mean flow estimates from particles simulated with mean flow as in Figure 6.5.1, where diffusivity is set as $1000\text{m}^2/\text{s}$. The velocity observations were organised into 1km×1km grids based on their positions, and the mean calculated for each grid. Each box plot corresponds to aggregating across 20km grids in either x (left panels) or y (right panels), with the red horizontal line corresponding to the mean.

the mean flow is inhomogeneous. The simulations revealed that if we were to estimate mean flow in 1km×1km grids following Section 6.5.1, the average bias would be zero, and standard deviation 0.033m/s. Therefore for the simple case of trajectories with constant mean flow and white noise velocities, this method for estimating the mean flow performs well.

We now subtract the estimated mean flow from each grid from the velocity, and estimate diffusivity using the periodogram estimator from Equation (6.2.9) and our proposed ACP estimator from Equation (6.3.7), where we have chosen a chunk length of 50 hours (corresponding to 10 chunks containing 50 observations each). We report the two dimensional diffusivity, taken as the average of the diffusivity in the x and y

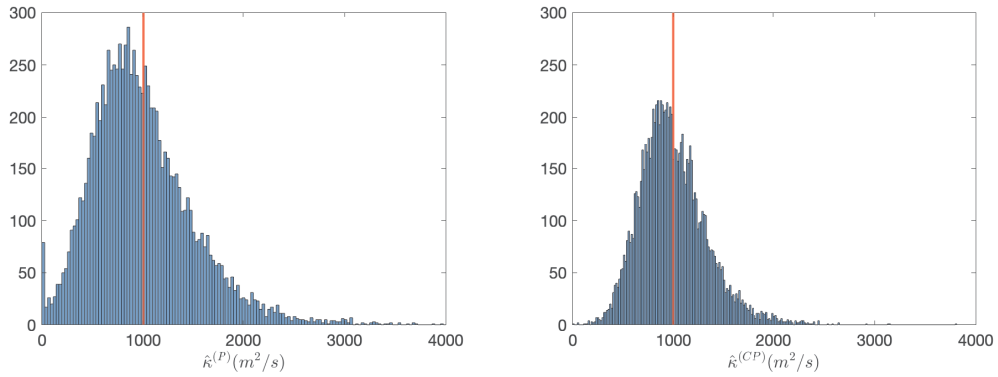

Figure 6.5.3: Histogram of diffusivity estimate for each spatial grid, for particles simulated as in Figure 6.5.2 using the periodogram diffusivity estimator (left) and ACP estimator (right). The red vertical line corresponds to the mean diffusivity estimate.

directions (Lilly et al., 2017). Figure 6.5.3 shows histograms of the periodogram and ACP diffusivity estimates, where each diffusivity estimate corresponds to a particular 1km×1km grid-point. The periodogram and ACP estimates have mean 1003m^2/s and 996m^2/s respectively, and standard deviations 518m^2/s and 334m^2/s. Therefore, the bias of each estimate is almost negligible whereas the variance is completely dominating the RMSE. Specifically the periodogram estimates have RMSE 518m^2/s, and the ACP estimates have RMSE 334m^2/s.

For the simple case of particles with velocities following a white noise process, the ACP estimator was able to reduce the error of diffusivity estimates compared with estimating with the periodogram estimator. While the bias was greater with the ACP estimator, this increase is very small and so the significant decrease in variance was able to reduce the overall RMSE. We note that drifters in the ocean follow a much

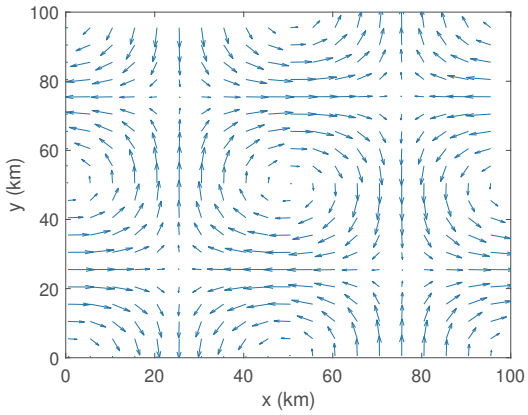

Figure 6.5.4: Mean flow velocity field from Equation (6.5.3).

more complicated path than that which can be described by a white noise velocity process with constant mean flow, and so we shall now apply both estimators to more complicated simulated trajectories.

Stream function Mean Flow

We now use our simulation environment to test the ACP estimator on more complicated particle trajectories. We will use a stream function (Lamb, 1924) to define the velocity field, choosing a periodic function so that the boundaries are divergence free. We define the stream function

$$\psi(x, y) = U \cos\left(\frac{2\pi x}{D_X}\right) \cos\left(\frac{2\pi y}{D_Y}\right), \tag{6.5.3}$$

where $D_X = D_Y = 100$km define the size of the box in x and y, and we set $U = 10,000m/\sigma_I^2$, meaning that the velocity in each direction will be ~ 0.1m/s. The stream function is shown in Figure 6.5.4.

We can calculate the velocities in each direction as

$$u = -\psi_y = U\frac{2\pi}{D_Y}\cos\left(\frac{2\pi x}{D_X}\right)\sin\left(\frac{2\pi y}{D_Y}\right) \tag{6.5.4}$$

$$v = \psi_x = -U\frac{2\pi}{D_X}\sin\left(\frac{2\pi x}{D_X}\right)\cos\left(\frac{2\pi y}{D_Y}\right). \tag{6.5.5}$$

The stochastic velocities again follow a two-dimensional white noise process with the diffusivity is set as $200\text{m}^2/\text{s}$ in each x and y direction (with no correlation), and $N = 1000$hours. For this simulation, while the observed time between observations is still $\Delta = 1$ hour, due to the complexity of the stream function, we simulate positions every 30 seconds to update the positions based on intermediate positions rather than just the previous recorded position. If we were not to do this, then drifters will not follow the stream lines, and this results in regions of convergence/divergence, which should be otherwise infeasible due to the definition of a stream function which is always divergence free.

Figure 6.5.5 displays box plots of mean flow estimates (compare with Figure 6.5.2 for a constant mean flow), calculated by taking the mean of all velocity estimates in a 1km×1km grid. We can see in Figure 6.5.4 that the mean flow changes across space, and these changes can be seen in the estimates in Figure 6.5.5. While the mean of each box is approximately the same, the interquartile range varies between boxes reflecting the changing magnitude of velocities in different regions.

As in the previous section, we remove the mean flow from the particle velocities and calculate diffusivity using the periodogram and ACP estimators, obtaining an estimate for each spatial grid and a chunk length of 50 hours. Again we report the two dimensional diffusivity, taken as the average of the diffusivity in the x and y direc-

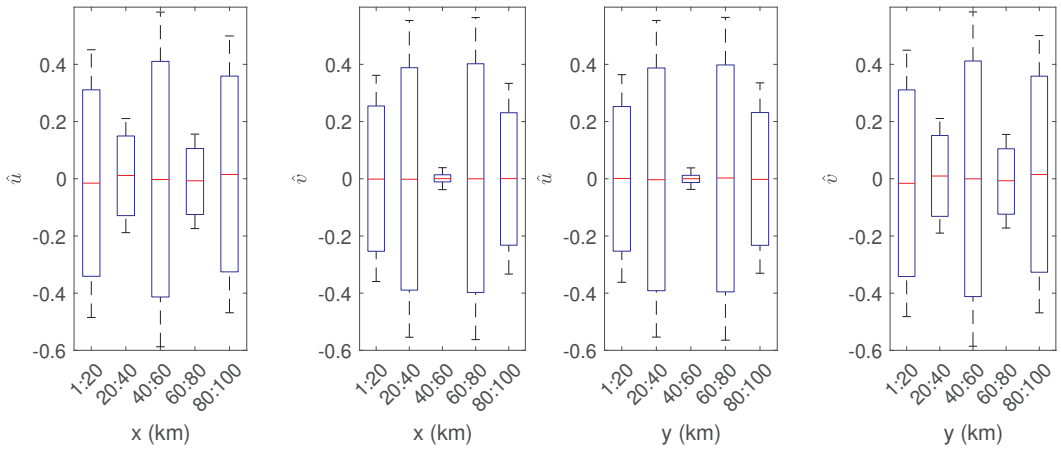

Figure 6.5.5: Box plot of mean flow estimates from particles simulated with mean flow as in Figure 6.5.4, where diffusivity is set as $200\text{m}^2/\text{s}$. The mean flow was estimated in 1km×1km grids. Each box plot corresponds to 20 grids in either x (left panels) or y (right panels), with the red horizontal line corresponding to the mean.

tions. Figure 6.5.6 shows the a histogram of diffusivity estimates. Despite the more complicated background flow which changed across space, the diffusivity estimates are almost unbiased, with each histogram having mean $201\text{m}^2/\text{s}$ and therefore bias $1\text{m}^2/\text{s}$. The periodogram estimates have standard deviation, and RMSE of $68\text{m}^2/\text{s}$, and the ACP estimator has standard deviation and RMSE of $47\text{m}^2/\text{s}$. If we were to scale the RMSE by the true value of the diffusivity then for the ACP estimator we get 0.235, and for the white noise case the scaled RMSE was 0.334. Similarly for the periodogram estimator the scaled RMSE is 0.34 for the stream function flow and 0.58 for the white noise. Therefore the scaled RMSE is significantly lower with the ACP estimator.

We have tested the ACP estimator on two different simulations with varying levels of complexity, and in both cases it performed better than the periodogram estimator

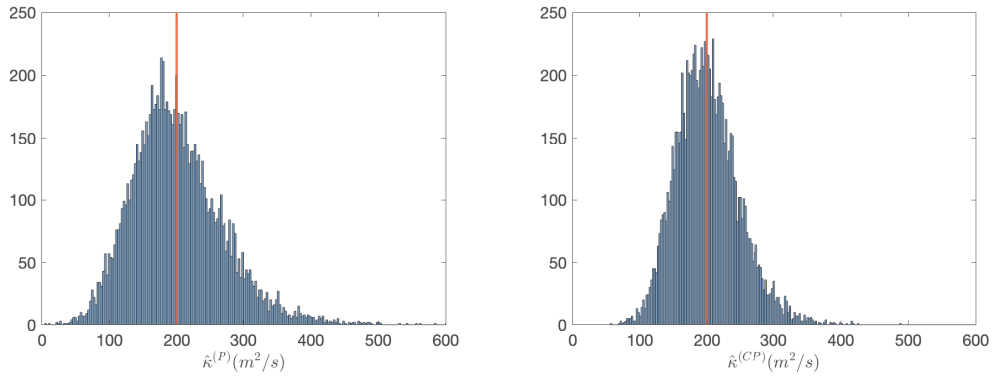

Figure 6.5.6: Histogram of diffusivity estimate for each spatial grid, for particles simulated as in Figure 6.5.4 using the periodogram diffusivity estimator (left) and ACP estimator (right). The red vertical line corresponds to the mean diffusivity estimate.

in terms of RMSE, as we expected. Further, the bias of the ACP estimator was very similar to that of the periodogram estimator, and the decrease in variance was significant. Changing the chunk length is likely to change the bias and variances to those we have found, however we have shown that the ACP estimator is able to produce reasonable diffusivity estimates in a simulation environment.

Further work would be to repeat the work in this section to include time variation since the ocean circulation changes with time. Including time variation in Equation (6.5.3) would mimic Rossby waves. If we were to estimate a time varying simulation as if it were constant in time, then we would expect the moving wave to be weakly seen in all locations that it has been during the simulation. The structures in the simulation would be smoothed across the different spatial locations, due to the temporal evolution. Therefore, we would need to include temporal evolution into our

estimation procedures.

6.6 Diffusivity Estimates using Global Drifter Program Drifters

We now apply the ACP estimator to drifter trajectories from the Global Drifter Program, which we introduced in more detail in Section 2.2.3. The work presented in this section is a preliminary analysis as we do not fine-tune the window length of the ACP estimator, or attempt to incorporate temporal variability into our mean flow or diffusivity estimates.

The window length of the drifter trajectories should be chosen as outlined in Section 6.4.2. We suggest that segments are chosen to be at least as long as the decorrelation timescale. In an AR(1) process the decorrelation time is the time that there is no significant memory between two time points. As we are using an AR(1) process to approximate ocean velocities when choosing the chunk lengths, a window length at least as long as the decorrelation time would contain at least two near-independent pieces of information. We note however, that Figure 6.4.1 showed the ACP estimator to have the largest RMSE when we had only a single chunk (equivalent to the periodogram based estimator), and hence applying the ACP estimator without optimising the number of chunks should still produce diffusivity estimates with a lower error than the periodogram diffusivity estimator.

We work with the 6-hourly product ($\Delta = 6$ hours) and split the drifter trajectories into time series of either 50 or 100 days ($N = 200$ or 400), and apply the ACP

estimator with chunks that are 12.5 days, or 50 observations, long (such that the number of chunks is m=16 or 32). We note that these window lengths have not been chosen to fully optimise the error, however both are longer than the decorrelation timescales found in Sundermeyer and Price (1998) (5-10 days). We split the globe into either $0.5° \times 0.5°$ bins or $1° \times 1°$ bins, and assign the diffusivity estimates to the grid that the drifter was in at the middle of the time window. Once estimates are found using all of the drifter data, we average the estimates across the grids to obtain a single diffusivity estimate for each spatial grid. These estimates assume that diffusivity is constant over time, and does not consider any seasonal patterns or general trend, but this could be considered in future work.

6.6.1 Mean Velocity Estimates

We begin by estimating the mean velocity, following the method described in Section 6.5.1, using $0.5° \times 0.5°$ and $1° \times 1°$ bins. The estimates are shown in Figure 6.6.1. The spatial resolution doesn't appear to have significantly improved when estimating using smaller bins, however the difference in the standard deviation is also small. If we were to instead use much larger bins we would lose some spatial resolution, such as the stripes of faster drifters near the equator and boundary currents.

6.6.2 Diffusivity Estimates

We now apply our diffusivity estimator to GDP drifters, as well as the periodogram estimator for comparison. Figure 6.6.2 shows diffusivity estimates from the peri-

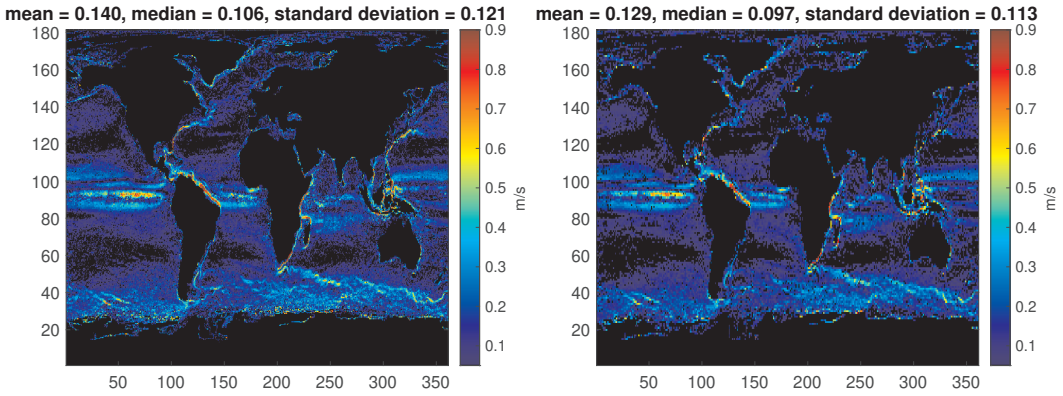

Figure 6.6.1: Mean flow estimates across the globe using GDP drifters, estimated following the method from Section 6.5.1 with bins of size $0.5° \times 0.5°$ (left) and $1° \times 1°$ (right).

odogram and ACP estimators using $0.5° \times 0.5°$ bins. We report the two dimensional diffusivity, taken as the average of the diffusivity in the x and y directions. We produce estimates separately from the ACP estimator for each direction, before averaging as this allows different chunk lengths to be used in each direction if required. The ACP estimator split the time series into chunks of length 12.5 days, hence the ACP diffusivity estimates are being estimated from 4 times as many time series as the periodogram estimator. This results in the standard deviation for the ACP estimates being approximately half that of the periodogram estimates. It also results in smaller diffusivity estimates, with the mean estimate across the globe being half that of those from the periodogram estimates. Figure 6.6.2 shows a notable resulting decrease in the noise when using the ACP estimator, when compared with the periodogram estimator, allowing finer scale structures to be resolved.

In Figure 6.6.3 we show the periodogram and ACP estimates when estimating using $1° \times 1°$ bins, and an initial time series of 50 days. The means are similar to

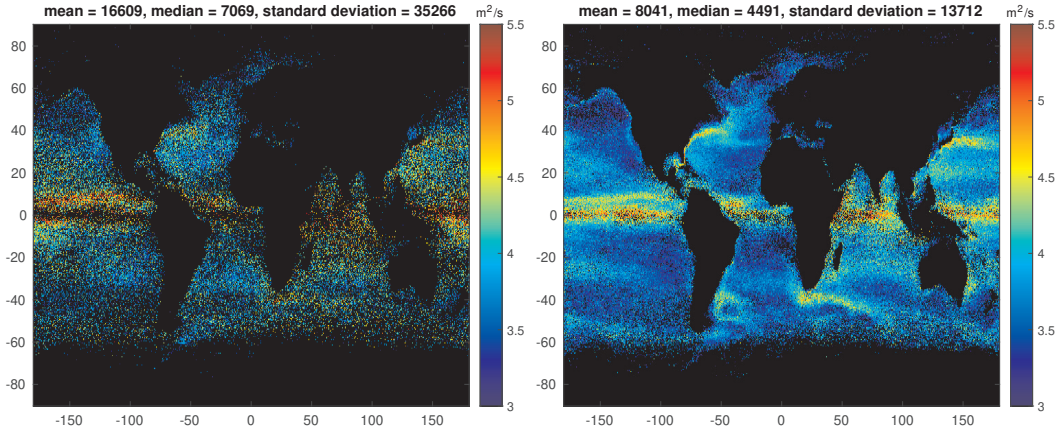

Figure 6.6.2: Diffusivity estimates using the periodogram (left) and ACP (right) estimators plotted on $\log_{10}$ scale, with $0.5° \times 0.5°$ bins, and an initial time series of length 100 days. The ACP estimator split the time series into chunks of length 12.5 days. The colour scale has been capped between 10^3 and $10^{5.5}$ for visualisation purposes.

the corresponding plots of Figure 6.6.2, however the standard deviations are lower. The most noticeable difference is the decreased noisiness of both plots compared with Figure 6.6.2. The larger bins combined with shorter initial time series combines more time series per spatial bin allowing the standard deviation to significantly drop, although some finer structures may be lost from the larger resolution.

The periodogram and ACP diffusivity estimators produce estimates which are distinctly different. As we do not know what the true diffusivity is in the ocean we are unable to give a definite answer for which estimates are closer to the truth, however we can see the standard deviation is much lower for the ACP estimates. For both estimators the average standard deviation is larger than the mean diffusivity estimate, and so the decrease in the standard deviation is expected to almost always outweigh the increased bias on average. Therefore, we expect the ACP estimates to

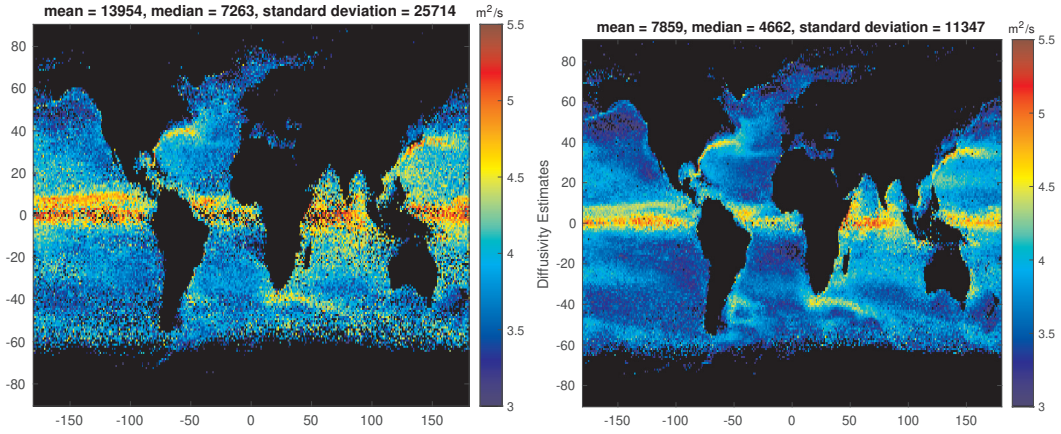

Figure 6.6.3: Diffusivity estimates using the periodogram (left) and ACP (right) estimators plotted on $\log_{10}$ scale, with $1° \times 1°$ bins, and an initial time series of 50 days. The ACP estimator split the time series into chunks of length 12.5 days. The scale has been fixed between 10^3 and $10^{5.5}$.

have a lower error than those from the periodogram estimator.

Our estimates are larger than those found in the literature, however the estimates from the ACP estimator are closer to what we would expect. Koszalka et al. (2011) obtained diffusivity estimates up to approximately $3 \times 10^3 \mathrm{m}^2/\mathrm{s}$, which are lower than the estimates we show, however we fixed the scale of our plots with a minimum of $10^3 \mathrm{m}^2/\mathrm{s}$ which is the maximum diffusivity that Koszalka et al. (2011) are reporting. They estimate diffusivity in the Nordic Seas, which are away from the major currents where we report our highest estimates.

Zhurbas and Oh (2004) estimate diffusivity to be closer to our estimates, with their estimates increasing up to $2.8 \times 10^4 \mathrm{m}^2/\mathrm{s}$, and the majority of their estimates being approximately $5 \times 10^3 \mathrm{m}^2/\mathrm{s}$. Our estimates range from $10^3 - 10^4 \mathrm{m}^2/\mathrm{s}$, except at the major currents where they are larger, and therefore are approximately in agreement

with Zhurbas and Oh (2004). We would expect to have larger diffusivity estimates than Zhurbas and Oh (2004) as they estimate diffusivity using the minor principal component, whereas we estimate diffusivity in both axes aligned with the longitude and latitude. Our method could be combined with theirs to measure the ACP estimator along the minor principal component, and this avenue of investigation is reserved for future work.

Another area for further work is to smooth out the mean flow and diffusivity estimates across the different spatial bins. Our estimation procedures assume that the mean flow/diffusivity estimate in each bin is independent of that in any other bin, whereas we would expect adjacent bins to have correlated mean flow/diffusivities. Therefore, the decorrelation scales could be calculated spatially for each bin and used to optimally smooth the estimates.

6.6.3 Optimal Chunk Length from Global Drifter Program Drifters

In the previous section we didn't optimise the number of chunks before estimating diffusivity. If we had optimised this number then the estimates would have had an even lower error. The optimal chunk length depends on the length of the initial time series that we are using to estimate diffusivity, as well as the autocorrelation structure, as we have shown before.

Figure 6.6.4 shows an approximation to the optimal number of chunks in each direction for time series of lengths 50 days and 100 days. We approximate the optimal

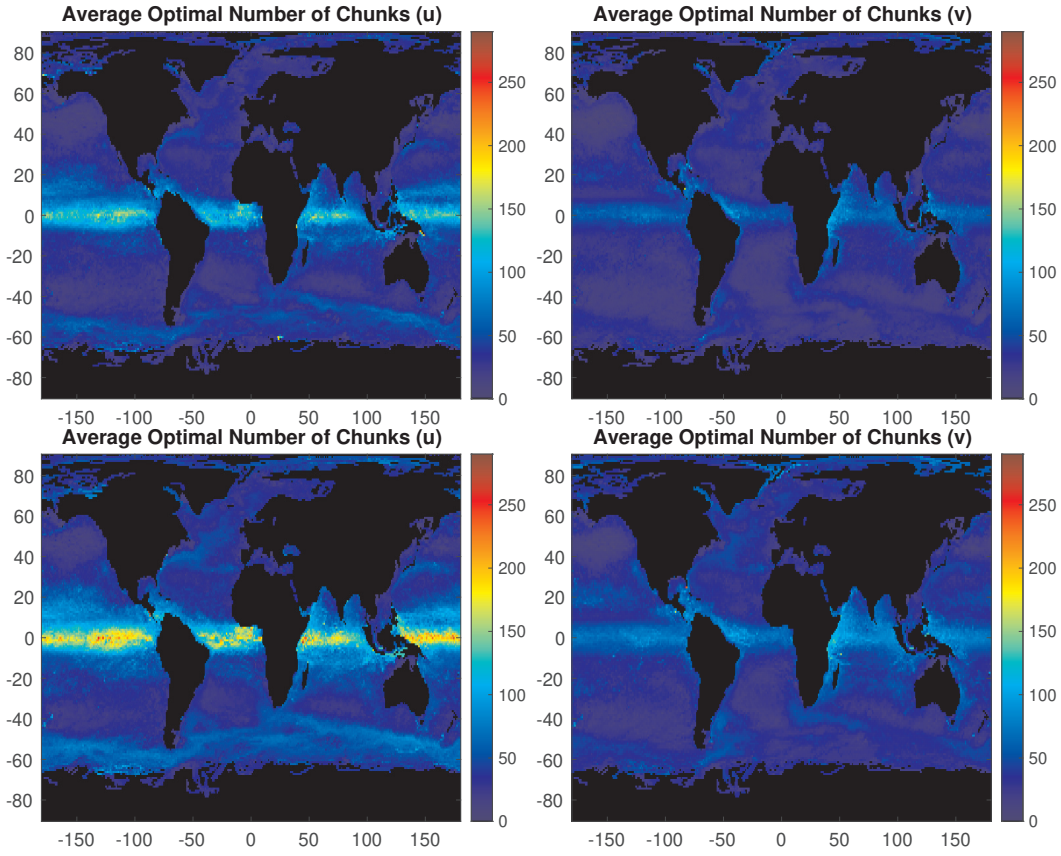

Figure 6.6.4: The optimal number of chunks found from GDP drifters in the u (left column) and v (right column) directions for initial time series of length 50 days (top row) and 100 days (bottom row), found by fitting an AR(1) process to the drifter trajectories and minimising the RMSE.

number of chunks by fitting an AR(1) process to the drifter velocities and then minimising the RMSE of the ACP estimator using Equation (6.4.1). The chunk length for each bin is found by calculating the optimal chunk length for each individual time series and then taking the average. We choose to use different chunk lengths in the x and y directions as the fast moving drifters near the equator mean we can estimate diffusivity with a larger number of shorter chunks in the x direction, but the drifters are not as quickly moving in the y direction and so fewer longer chunks are needed. We would expect faster moving drifters to have less memory (once the mean is removed), and therefore a lower value for ϕ. From Figure 6.4.3 we know that smaller damping parameters mean that the optimal number of chunks is higher than for larger values of ϕ. The choice to have different numbers of chunks in each direction means that the variance from the number of drifters will vary in each direction, but that we are reducing the error overall.

6.7 Conclusion

Diffusivity can be estimated many different ways, and some estimation methods were introduced earlier in Section 2.4. Many estimation techniques derive from Taylor (1922), and these methods are in general equivalent to estimating diffusivity as the integral of the velocity autocorrelation. We have shown that some implementations, while equivalent to estimating using the periodogram as $t \to \infty$, perform worse than estimating from the periodogram in terms of RMSE. Further to this, we showed that it would be better to estimate the diffusivity to be zero than to produce an estimate

from a single drifter using the periodogram estimator.

The large errors that we showed for the periodogram estimator are due to the variance being large, and therefore we introduced the ACP estimator which was proposed as a method to estimate diffusivity with reduced variance. This method requires splitting a time series into many smaller time series and taking the average diffusivity across the chunks. It introduces a trade-off between the bias and variance of an estimate, and requires an optimal window length to be chosen to minimise the RMSE.

We use theoretical forms of bias and variance from an AR(1) velocity process which is fitted to drifter trajectories in order to determine the window length. Unlike Griffa et al. (1995) we do not use the parametric model to directly estimate diffusivity, instead we use the model to refine our non-parametric estimator to each unique drifter trajectory, before then applying the estimator directly to the velocity observations to estimate diffusivity.

We applied the ACP estimator to simulated trajectories as well as global drifter program drifters and showed that it performed better in terms of RMSE than the periodogram estimator. Our estimates of diffusivity are in general higher than those by Koszalka et al. (2011) and Zhurbas and Oh (2004), but the ACP estimates are closer to their estimates than those from the periodogram estimator. The meaning of these estimates can vary depending on how they are calculated, Zhurbas and Oh (2004) estimate diffusivity using the minor principal component whereas we estimate a two-dimensional averaged estimate. Therefore, we can compare with them to see that our estimates are approximately of the right order, however we do not expect to obtain estimates as low as theirs due to the assumptions made in different estimators.

There are a number of choices that we needed to make to estimate diffusivity from the GDP. The first of these decisions was what size spatial bins to split the oceans into. We used relatively small bins of sizes $0.5° \times 0.5°$ and $1° \times 1°$. We chose small bins (versus other studies in the literature) because the ACP estimator decreases the variance, and therefore increases the statistical significance of the estimates, allowing finer resolutions to be realistically resolved. We split each time series into at least 4 chunks, and so based on the result from Section 6.2.4 that we require at least three time series to ensure statistical significance, then all the estimates from the ACP estimator should be significant. The same cannot be said for the periodogram based estimates, as these do not take any measure to ensure the significance of the estimates other than averaging across multiple drifters (which may not be available). Therefore for more sparse regions of the ocean the number of chunks in the ACP estimator can be chosen to ensure that the resulting estimates are significant.

We know that diffusivity is not constant across the ocean, therefore it would be unrealistic to use the whole historic data from each drifter to choose a window length. However, how do we choose an initial partition to obtain the optimal window length for each part of the ocean? Do we wish to only use the part of a trajectory within a certain bin to estimate diffusivity within that bin, or can observations in neighbouring bins contribute to the estimate? These are decisions that need to be made before estimating with the ACP estimator. We chose initial time series of either 100 or 50 days, as this allowed us chunks of length 12.5 days which still allowed the diffusivity to converge towards the long-term diffusivity. However, optimally choosing the initial time series length, as well as other tunable decisions, is an area that we

leave for further work.

Another important piece of further work is fitting the AR(1) velocity process to drifter trajectories to obtain estimates for the damping parameter. The periodogram and ACP estimators only use frequency zero, and therefore the process only needs to fit at the lowest frequencies, rather than to all frequencies. This reduces the impact of processes such as tides and waves on the choice of the AR(1) parameter. We also note that an alternative parametric model could be used to determine the chunk length, so long as the RMSE can be determined. An AR(1) process is more simple than an oscillatory Ornstein-Uhlenbeck process or a Matérn process, and so these might capture more features of drifter trajectories and result in a more optimal choice for the chunk length. The ACP estimator should yield diffusivity estimates with lower RMSE than the periodogram estimator regardless of the number of chunks chosen (where just one chunk equal to the entire time series is equivalent to the periodogram estimator). Therefore while optimising the number of chunks chosen is expected to further improve the error, the ACP estimator should still be preferred over the periodogram estimator without optimal tuning.

In Section 6.6 we estimated mean flow and diffusivity using all available trajectories in the GDP, and therefore we didn't account for any temporal variation. By neglecting temporal changes we were able to use all drifters in each spatial grid, as opposed to just those within a given time window, and therefore we reduced the variance of our estimates. However, if the temporal changes were significant then we introduced bias by estimating mean flow and diffusivity as averages across all time points. Choosing temporal windows to estimate over poses a similar trade-off between bias and variance

to our ACP estimator, and an optimal choice is likely to further reduce the error of diffusivity estimates.

Chapter 7

Conclusions and Further Work

7.1 Conclusions

Estimating the path a drifter will take through the ocean requires us to estimate various different components of the flow. The different components we can feasibly estimate can be determined by what data is available, i.e. we require clustered drifters to be able to realistically measure mesoscale flow. What we wish to learn can also aid the decision of which parameters to estimate, and whether the estimates are expected to be statistically significant. Throughout this thesis we have introduced methods to estimate background, mesoscale, and submesoscale flow, with an emphasis on reducing the error of the estimates.

In Chapter 3 we introduced a way to estimate mesoscale flow components which involves taking a Taylor series of drifter velocity observations and using least-squares regression to estimate strain, vorticity and divergence. We used a hierarchy of models to estimate combinations of parameters depending on which features we believe to be

present in the flow, as well as which parameters we are able to produce statistically significant estimates for. Our method to estimate the mesoscale features is similar to that by Okubo and Ebbesmeyer (1976), however their framework only permits estimation of parameters at every observational time point. We instead implemented rolling windows to estimate each parameter within a time window in which we assume that parameter to be constant. The use of windows allowed us to improve the significance of our estimates by reducing the variance, as well as smoothing out physical features such as internal waves.

We used our estimator to provide estimates for strain rate and submesoscale diffusivity using drifter observations from the LatMix experiment. As was the case in Shcherbina et al. (2015) and Sundermeyer et al. (2020), we obtained different estimates for the parameters dependent on assumptions made. When we assumed that there is no vorticity our estimates of the strain rate were different to when we estimated both strain and vorticity together. It is therefore important to choose the model that best fits the data in order to get the most accurate estimates for our parameters. We proposed two different methods in Chapter 3 to select the hierarchy that best fits the drifter trajectories and minimises the error.

The error of parameter estimates is determined by many different factors, such as the sampling rate and the overall length of the observational time window. Not all of these inputs can be changed after an experiment has taken place and the data has been collected, as the length of the experiment, the sampling frequency, and the initial positions of the drifters all affect the error, as well as the number of drifters. These are all choices that must be made as part of the experimental set up, before

the deployment takes place.

In Chapter 4 we explored via simulation how each of these choices affects the errors, with an emphasis on the initial drifter positions. We found it was always more optimal to run the deployment for longer and use more drifters, but this can be costly and so it is important to also optimise how these drifters are deployed. We found that the optimal drifter "morphology" is dependent on the number of parameters being estimated, as well as which parameters we estimate. If we wish to estimate only a single mesoscale parameter then it was optimal to deploy drifters in a straight line (preferably aligned with strain angle), with the only exception to this being the strain-only model if the drifters are deployed orthogonal to the strain angle, where instead a more isotropic deployment morphology was preferred. Conversely, when all mesoscale parameters were being estimated then an isotropic arrangement had the lowest error, and when any two parameters (from strain/divergence/vorticity) were estimated then it depended on which combination in terms of which morphology had the lowest error. It is specifically the parameters being *estimated*, rather than the features present in the data that determines the optimal initial drifter arrangement. Therefore, without any prior knowledge of the mesoscale part of the flow, an isotropic morphology should be used to minimise expected error of parameter estimates.

In Chapter 5 we extended the analysis of Chapter 4 but used a parametric model to calculate theoretical errors without having to resort to simulations. This allowed us to verify the optimal initial drifter morphology for different strain rates and diffusivities to those used in Chapter 4. This analysis assumed a strain only model and so was less extensive than Chapter 4, but was useful nevertheless as, for example, many authors

estimate only strain from the LatMix experiment, and so the results of Chapter 5 would have been applicable to determining a more optimal drifter deployment for the LatMix experiment, as well as placing error bars on the parameter estimates.

To derive a theoretical form for the error we assumed that drifters followed an Ornstein-Uhlenbeck process and used method-of-moments to fit a noncentral chi-squared distribution to the second moment of drifter positions. These second moments were then used to derive estimators for strain rate and diffusivity. We fitted a distribution to each estimator which was used to calculate the expected error. The error of both the strain rate and diffusivity estimates were minimised when the second moment in the direction aligned with the strain angle was minimised, while changing the second moment in the direction orthogonal to the strain angle had no effect on the error. Therefore, it is optimal to deploy drifters spread out along the axis aligned with the strain, which is consistent with the simulation results from Chapter 4.

Finally, in Chapter 6 we explored the error of estimating large-scale diffusivity on a global scale. We fitted white noise and AR(1) processes to drifter velocities and derived the distributions of the diffusivity estimates. We used the distributions to calculate the expected error of diffusivity estimates from each process and found that when diffusivity is estimated from a single time series, it would be more optimal to estimate the diffusivity to be zero than to use the estimate. Specifically, we would require to average diffusivity estimates from at least three time series for the estimate to becomes statistically significant. This is because the estimates have a larger standard deviation than the true value of the diffusivity.

We proposed a new method to estimate diffusivity where each time series was

split into a number of chunks and the diffusivity was calculated as an average of the diffusivity of each chunk. This decreases the variance of the estimate but increases the bias, hence introducing a trade off between bias and variance. The number of chunks that the time series is split into can be chosen by fitting an AR(1) process to drifter velocities and choosing the number of chunks which minimises the RMSE. The RMSE is maximised when one single chunk containing the original time series is chosen, which is equivalent to estimate diffusivity from the periodogram estimator. Therefore this estimator should always reduce the error, even if the most optimal number of chunks is not used.

In this thesis we have proposed methods to estimate mesoscale and submesoscale flow from clustered drifters, as well as a method to estimate large-scale diffusivity from global drifters. These estimators have all been proposed as a way to estimate flow parameters while reducing the error of such estimates. We also investigated how these estimates could be further reduced by deploying drifters in a more optimal morphology which is determined by which flow parameters we wish to estimate. Reducing the error of parameter estimates enables us to build a clearer picture of how particles move through the ocean without having to deploy more drifters, but we note that our models are more simple than true ocean dynamics, and describing ocean flow remains an important area of active research.

7.2 Further Work

In addition to elements of future work identified throughout the thesis, we now focus on three particular areas that would warrant further investigation building on the results of this thesis.

One area for further work would be to expand our methods for estimating mesoscale and submesoscale parameters to allow for their spatial variation. In this thesis we only allowed for parameters to evolve temporally (see e.g. Section 3.4.3 using splines), but we always assumed parameters to be fixed spatially in the area covered by a drifter deployment. Our methods for estimating parameters using either least squares regression (Chapters 3 and 4) or method-of-moments (Chapter 5) therefore made potentially unrealistic assumptions of spatially fixed mesoscale and submesoscale parameters. This might result in a bias of the parameter estimates, the size of which is determined by how much the parameters truly change across the spatial region sampled by the drifters. Implementing our smoothed time-windowed estimator in Chapter 3 is likely to increase this bias versus estimating each parameter at every sampled time point, as the drifters will spread through the time window which could result in the true parameter values changing through space. One possible way to incorporate spatial variation into the least squares estimator would be to model Equation (3.2.4) using a second-order stream function. A stream function is incompressible, and so the model defines the fluid to be divergence-free (which is not the case in Equation 3.2.4). The use of a stream function allows the addition of third-order terms which can then be used to model spatial variation. Preliminary investigation using a simplified version of

Equation (3.2.4) suggests that the use of a higher order stream function could capture the variation in strain rate estimates, as well as possibly reducing the size of the error bars when compared with a lower order stream function.

Another area for further work is to apply the least squares estimator from Chapter 3 and the method-of-moments estimator in Chapter 5 to other regional drifter programs. Comparing with estimates in the literature for both the least squares and method-of-moments estimators will provide a better idea of how well the estimators perform for real data. In Chapter 3 we applied the estimator to drifter trajectories from the LatMix experiment, but it could be applied to other drifter programs such as GLAD, LASER, or CALYPSO, amongst others. Similarly, the second-moment estimator in Chapter 5 was only tested on simulated particle trajectories, and so further work would involve testing it on real drifter deployments, such as those listed above.

The final area for future work that we wish to highlight is to develop the large-scale diffusivity estimator from Chapter 6 to account for other flow features than just the mean background flow. Our approach has been to estimate diffusivity from the residual velocities after other flow features have been estimated and subtracted, and diffusivity hence quantifies the diffusive properties of the unknown part of the drifter velocities. Therefore estimating and removing more components of the velocities should reduce the value of diffusivity estimates. Features that could be estimated and removed involve tides (e.g. Foreman et al., 1995), inertial oscillations (e.g. Sykulski et al., 2016), and eddies (e.g. Lilly and Pérez-Brunius, 2021; Sallée et al., 2008). Furthermore, we could estimate different features (including mean flow) seasonally to reduce any bias to the estimates due to seasonal variations.

Bibliography

Pavel S Berloff and James C McWilliams. Material Transport in Oceanic Gyres. Part II: Hierarchy of Stochastic Models. *Journal of Physical Oceanography*, 32(3): 797–830, 2002.

Francisco J Beron-Vera and J H LaCasce. Statistics of Simulated and Observed Pair Separations in the Gulf of Mexico. *Journal of Physical Oceanography*, 46(7):2183–2199, 2016. ISSN 0022-3670. doi: 10.1175/jpo-d-15-0127.1.

Z. I. Botev, J. F. Grotowski, and D. P. Kroese. Kernel Density Estimation via Diffusion. *The Annals of Statistics*, 38(5):2916–2957, 2010. ISSN 0090-5364. doi: 10.1214/10-aos799.

Oliver Buehler and Miranda Holmes-Cerfon. Particle Dispersion by Random Waves in Rotating Shallow Water. *Journal of fluid mechanics*, 638:5–26, 2009.

Meghan F Cronin, Noel A Pelland, Steven R Emerson, and William R Crawford. Estimating Diffusivity from the Mixed Layer Heat and Salt Balances in the North Pacific. *Journal of Geophysical Research: Oceans*, 120(11):7346–7362, 2015.

Russ E Davis. Observing the General Circulation with Floats. *Deep Sea Research Part A. Oceanographic Research Papers*, 38:S531–S571, 1991.

Jeffrey J. Early and Adam M Sykulski. Smoothing and Interpolating Noisy GPS Data with Smoothing Splines. *Journal of Atmospheric and Oceanic Technology*, 2020. doi: 10.1175/jtech-d-19-0087.1.

Jeffrey J Early, M Pascale Lelong, and Miles A Sundermeyer. A Wave-Vortex Decomposition for Rotating Boussinesq Flows in Bounded Domains. *Journal of Fluid Mechanics*, 2021.

Shane Elipot, Rick Lumpkin, Renellys C Perez, Jonathan M Lilly, Jeffrey J Early, and Adam M Sykulski. A Global Surface Drifter Data Set at Hourly Resolution. *Journal of Geophysical Research: Oceans*, 121(5):2937–2966, 2016.

Shane Elipot, Adam Sykulski, Rick Lumpkin, Luca Centurioni, and Mayra Pazos. A Dataset of Hourly Sea Surface Temperature From Drifting Buoys. *arXiv preprint arXiv:2201.08289*, 2022.

Michael GG Foreman, William R Crawford, and Richard F Marsden. De-Tiding: Theory and Practice. *Coastal and estuarine studies*, pages 203–203, 1995.

TW Fulton. The Surface Currents of the North Sea. *The Scottish Geographical Magazine*, 13(12):636–645, 1897.

Christopher Garrett. On the Initial Streakness of a Dispersing Tracer in Two- and Three-Dimensional Turbulence. *Dynamics of Atmospheres and Oceans*, 7(4):265–277, 1983. doi: 10.1016/0377-0265(83)90008-8.

Rafael C Gonçalves, Mohamed Iskandarani, Tamay Özgökmen, and W Carlisle Thacker. Reconstruction of Submesoscale Velocity Field from Surface Drifters. *Journal of Physical Oceanography*, 49(4):941–958, 2019. ISSN 0022-3670. doi: 10.1175/jpo-d-18-0025.1.

Annalisa Griffa, Kenneth Owens, Leonid Piterbarg, and Boris Rozovskii. Estimates of Turbulence Parameters from Lagrangian Data Using a Stochastic Particle Model. *Journal of Marine Research*, 53(3):371–401, 1995.

Britta D Hardesty, Joseph Harari, Atsuhiko Isobe, Laurent Lebreton, Nikolai Maximenko, Jim Potemra, Erik Van Sebille, A Dick Vethaak, and Chris Wilcox. Using Numerical Model Simulations to Improve the Understanding of Micro-Plastic Distribution and Pathways in the Marine Environment. *Frontiers in marine science*, 4:30, 2017.

Peter H Haynes. Vertical Shear Plus Horizontal Stretching as a Route to Mixing. In *From Stirring to Mixing in a Stratified Ocean, Proceedings, 'Aha Huliko'a Hawaiian Winter Workshop*, pages 73—80, Honolulu, HI, University of Hawaii at Manoa, 2001.

P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations. Vol. 23.* Springer Science & Business Media, 2013.

I Koszalka, JH LaCasce, M Andersson, KA Orvik, and C Mauritzen. Surface Circulation in the Nordic Seas from Clustered Drifters. *Deep Sea Research Part I: Oceanographic Research Papers*, 58(4):468–485, 2011.

Inga Monika Koszalka and Joseph H LaCasce. Lagrangian Analysis by clustering. *Ocean dynamics*, 60(4):957–972, 2010.

K Kanhai La Daana, Rick Officer, Olga Lyashevska, Richard C Thompson, and Ian O'Connor. Microplastic Abundance, Distribution and Composition along a Latitudinal Gradient in the Atlantic Ocean. *Marine pollution bulletin*, 115(1-2):307–314, 2017.

JH LaCasce and A Bower. Relative Dispersion in the Subsurface North Atlantic. *Journal of marine research*, 58(6):863–894, 2000.

JHt LaCasce. Statistics from Lagrangian Observations. *Progress in Oceanography*, 77 (1):1–29, 2008.

Horace Lamb. *Hydrodynamics*. University Press, 1924.

M. Pascale Lelong, Yannis Cuypers, and Pascale Bouruet-Aubertot. Near-Inertial Energy Propagation inside a Mediterranean Anticyclonic Eddy. *Journal of Physical Oceanography*, 2020. doi: 10.1175/jpo-d-19-0211.1.

Jonathan M. Lilly. Kinematics of a Fluid Ellipse in a Linear Flow. *Fluids*, 3(1):16, 2018. doi: 10.3390/fluids3010016.

Jonathan M Lilly and Paula Pérez-Brunius. Extracting Statistically Significant Eddy Signals from Large Lagrangian Datasets Using Wavelet Ridge Analysis, With Application to the Gulf of Mexico. *Nonlinear Processes in Geophysics*, 28(2):181–212, 2021.

Jonathan M Lilly, Adam M Sykulski, Jeffrey J Early, and Sofia C Olhede. Fractional Brownian Motion, the Matérn Process, and Stochastic Modeling of Turbulent Dispersion. *Nonlinear Processes in Geophysics*, 24(3):481–514, 2017.

John Lodise, Tamay Özgökmen, Rafael C Gonçalves, Mohamed Iskandarani, Björn Lund, Jochen Horstmann, Pierre-Marie Poulain, Jody Klymak, Edward H Ryan, and Cedric Guigand. Investigating the Formation of Submesoscale Structures Along Mesoscale Fronts and Estimating Kinematic Quantities Using Lagrangian Drifters. *Fluids*, 5(3):159, 2020.

Rick Lumpkin and Mayra Pazos. Measuring Surface Currents with Surface Velocity Program Drifters: The Instrument, its Data, and Some Recent Results. *Lagrangian analysis and prediction of coastal and ocean dynamics*, 39:67, 2007.

Rick Lumpkin, Tamay Özgökmen, and Luca Centurioni. Advances in the Application of Surface Drifters. *Annual Review of Marine Science*, 9:59–81, 2017.

Amala Mahadevan, Ananda Pascual, Daniel L. Rudnick, Simón Ruiz, Joaquín Tintoré, and Eric D'Asaro. Coherent Pathways for Vertical Transport from the Surface Ocean to Interior. *Bulletin of the American Meteorological Society*, preprint(2020): 1–21, 2020. ISSN 0003-0007. doi: 10.1175/bams-d-19-0305.1.

Pearn P Niiler, Andrew S Sybrandy, Kenong Bi, Pierre M Poulain, and David Bitterman. Measurements of the Water-Following Capability of Holey-Sock and TRISTAR Drifters. *Deep Sea Research Part I: Oceanographic Research Papers*, 42(11-12): 1951–1964, 1995.

PP Niller, RE Davis, and HJ White. Water-Following Characteristics of a Mixed Layer Drifter. *Deep Sea Research Part A. Oceanographic Research Papers*, 34(11): 1867–1881, 1987.

Im Sang Oh, Victor Zhurbas, and WonSun Park. 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, 2000.

J C Ohlmann, M J Molemaker, B Baschek, B Holt, G Marmorino, and G Smith. Drifter Observations of Submesoscale Flow Kinematics in the Coastal Ocean. *Geophysical Research Letters*, 44(1):330–337, 2017. doi: 10.1002/2016gl071537.

JC Ohlmann, PF White, AL Sybrandy, and PP Niiler. GPS–Cellular Drifter Technology for Coastal Ocean Observing Systems. *Journal of Atmospheric and Oceanic Technology*, 22(9):1381–1388, 2005.

Akira Okubo and Curtis C Ebbesmeyer. Determination of Vorticity, Divergence, and Deformation Rates from Analysis of Drogue Observations. *Deep Sea Research and Oceanographic Abstracts*, 23(4):349–352, 1976.

Steven E Pav. Moments of the Log Non-Central Chi-Square Distribution. *arXiv preprint arXiv:1503.06266*, 2015.

Jenna Pearson, Baylor Fox-Kemper, Roy Barkan, Jun Choi, Annalisa Bracco, and James C. McWilliams. Impacts of Convergence on Structure Functions from Surface Drifters in the Gulf of Mexico Impacts of Convergence on Structure Functions from

Surface Drifters in the Gulf of Mexico. *Journal of Physical Oceanography*, 49(3): 675–690, 2019. ISSN 0022-3670. doi: 10.1175/jpo-d-18-0029.1.

Donald B Percival, Andrew T Walden, et al. *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.

Andrew C Poje, Tamay M Özgökmen, Bruce L Lipphardt, Brian K Haus, Edward H Ryan, Angelique C Haza, Gregg A Jacobs, A J H M Reniers, Maria Josefina Olascoaga, Guillaume Novelli, Annalisa Griffa, Francisco J Beron-Vera, Shuyi S Chen, Emanuel Coelho, Patrick J Hogan, Albert D Kirwan, Helga S Huntley, and Arthur J Mariano. Submesoscale Dispersion in the Vicinity of the Deepwater Horizon Spill. *Proceedings of the National Academy of Sciences of the United States of America*, 111(35):12693–12698, 2014. doi: 10.1073/pnas.1402452111.

Peter B Rhines and Sirpa Häkkinen. Is the Oceanic Heat Transport in the North Atlantic Irrelevant to the Climate in Europe. *ASOF Newsletter*, 1:13–17, 2003.

T Rossby. Evolution of Lagrangian Methods in Oceanography. *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, pages 1–38, 2007.

JB Sallée, Kevin Speer, R Morrow, and Rick Lumpkin. An Estimate of Lagrangian Eddy Statistics and Diffusion in the Mixed Layer of the Southern Ocean. *Journal of Marine Research*, 66(4):441–463, 2008.

R Seager. The Source of Europe's Mild Climate. *American Scientist*, 94(4):334, 2006.

Andrey Y Shcherbina, Eric A D'Asaro, Craig M Lee, Jody M Klymak, M Jeroen Molemaker, and James C McWilliams. Statistics of Vertical Vorticity, Divergence,

and Strain in a Developed Submesoscale Turbulence Field. *Geophysical Research Letters*, 40(17):4706–4711, 2013.

Andrey Y Shcherbina, Miles A Sundermeyer, Eric Kunze, Eric D'Asaro, Gualtiero Badin, Daniel Birch, Anne-Marie E G Brunner-Suzuki, Jörn Callies, Brandy T Kuebel Cervantes, Mariona Claret, Brian Concannon, Jeffrey J. Early, Raffaele Ferrari, Louis Goodman, Ramsey R Harcourt, Jody M Klymak, Craig M Lee, M Pascale Lelong, Murray D Levine, R C Lien, Amala Mahadevan, James C McWilliams, M Jeroen Molemaker, Sonaljit Mukherjee, Jonathan D Nash, Tamay Özgökmen, Stephen D Pierce, Sanjiv Ramachandran, Roger M Samelson, Thomas B Sanford, R Kipp Shearman, Eric D Skyllingstad, K Shafer Smith, Amit Tandon, John R Taylor, Eugene A Terray, Leif N Thomas, and James R Ledwell. The LatMix Summer Campaign: Submesoscale Stirring in the Upper Ocean. *Bulletin of the American Meteorological Society*, 96(8):1257 – 1279, 2015. doi: 10.1175/bams-d-14-00015.1.

Miles A Sundermeyer. LatMix 2011 and 2012 Dispersion Analysis. Technical report, University of Massachusetts Dartmouth New Bedford United States, 2017.

Miles A Sundermeyer and James R Ledwell. Lateral Dispersion over the Continental Shelf: Analysis of Dye Release Experiments. *Journal of Geophysical Research*, 106 (C5):9603–9621, 2001.

Miles A Sundermeyer and James F. Price. Lateral Mixing and the North Atlantic Tracer Release Experiment: Observations and Numerical Simulations of Lagrangian

Particles and a Passive Tracer. *Journal of Geophysical Research*, 103(C10):21481–21497, 1998.

Miles A Sundermeyer, Daniel A Birch, James R Ledwell, Murray D Levine, Stephen D Pierce, and Brandy T Kuebel Cervantes. Dispersion in the Open Ocean Seasonal Pycnocline at Scales of 1-10 km and 1-6 days. *Journal of Physical Oceanography*, 50(2):415–437, 2020. ISSN 0022-3670. doi: 10.1175/jpo-d-19-0019.1.

HU Sverdrup, MW Johnson, RH Fleming, et al. *The Oceans: Their Physics, Chemistry, and General Biology*, volume 7. Prentice-Hall New York, 1942.

Mark S Swenson and Pearn P Niiler. Statistical Analysis of the Surface Circulation of the California Current. *Journal of Geophysical Research: Oceans*, 101(C10): 22631–22645, 1996.

AL Sybrandy and PP Niiler. *The WOCE/TOGA SVP Lagrangian Drifter Construction Manual*. University of California, San Diego, Scripps Institution of Oceanography, 1991.

Adam Sykulski, Jonathan Lilly, Sofia Olhede, and Jeffrey Early. Analysis of the Global Surface Drifter Dataset using Stochastic Processes. In *EGU General Assembly Conference Abstracts*, page 18280, 2017a.

Adam M Sykulski, Sofia C Olhede, Jonathan M Lilly, and Eric Danioux. Lagrangian Time Series Models for Ocean Surface Drifter Trajectories. *Journal of the Royal Statistical Society Series C*, 65(1):29–50, 2016.

Adam M Sykulski, Sofia Charlotta Olhede, Jonathan M Lilly, and Jeffrey J Early. Frequency-Domain Stochastic Modeling of Stationary Bivariate or Complex-Valued Signals. *IEEE Transactions on Signal Processing*, 65(12):3136–3151, 2017b.

JB Tait. *The Surface Water Drift in the Northern and Middle Areas of the North Sea and in the Faroe-Shetland Channel*. Number 1. HM Stationery off., 1930.

Geoffrey I Taylor. Diffusion by Continuous Movements. *Proceedings of the london mathematical society*, 2(1):196–212, 1922.

Joaquin A Trinanes, M Josefina Olascoaga, Gustavo J Goni, Nikolai A Maximenko, David A Griffin, and Jan Hafner. Analysis of Flight MH370 Potential Debris Trajectories Using Ocean Observations and Numerical Model Results. *Journal of Operational Oceanography*, 9(2):126–138, 2016.

Erik Van Sebille, Stephen M Griffies, Ryan Abernathey, Thomas P Adams, Pavel Berloff, Arne Biastoch, Bruno Blanke, Eric P Chassignet, Yu Cheng, Colin J Cotter, et al. Lagrangian Ocean Analysis: Fundamentals and Practices. *Ocean Modelling*, 121:49–75, 2018.

Milena Veneziani, Annalisa Griffa, Andy M Reynolds, and Arthur J Mariano. Oceanic Turbulence and Stochastic Models from Subsurface Lagrangian Data for the Northwest Atlantic Ocean. *Journal of physical oceanography*, 34(8):1884–1906, 2004.

Guilherme S Vieira, Irina I Rypina, and Michael R Allshouse. Uncertainty Quantification of Trajectory Clustering Applied to Ocean Ensemble Forecasts. *Fluids*, 5(4): 184, 2020.

Xiaoming Zhai, Richard J Greatbatch, and Jan-Dirk Kohlmann. On the seasonal variability of eddy kinetic energy in the gulf stream region. *Geophysical Research Letters*, 35(24), 2008.

Victor Zhurbas and Im Sang Oh. Lateral Diffusivity and Lagrangian Scales in the Pacific Ocean as Derived from Drifter Data. *Journal of Geophysical Research: Oceans*, 108(C5), 2003.

Victor Zhurbas and Im Sang Oh. 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), 2004.