How sensitive are Lagrangian coherent structures to uncertainties in data?*

Aleksandar Badza*, Trent W. Mattner, Sanjeeva Balasuriya

School of Mathematical Sciences, The University of Adelaide, Adelaide SA 5005, Australia

Abstract

Lagrangian coherent structures (LCSs) are time-varying entities which capture the most influential transport features of a flow. These can for example identify groups of particles which have greatest stretching, or which maintain a coherent jet or vortical structure. While many different LCS methods have been developed, the impact of the inevitable measurement uncertainty in realistic Eulerian velocity data has not been studied in detail. This article systematically addresses whether LCS methods are self-consistent in their conclusions under such uncertainty for nine different methods: the finite time Lyapunov exponent, hyperbolic variational LCSs, Lagrangian averaged vorticity deviation, Lagrangian descriptors, stochastic sensitivity, the transfer operator, the dynamic Laplacian operator, fuzzy c–means clustering and coherent structure colouring. The investigations are performed for two different realistic data sets: a computational fluid dynamics simulation of a Kelvin–Helmholtz instability, and oceanographic data of the Gulf Stream region. Using statistics gleaned from stochastic simulations, it is shown that the methods which detect full-dimensional coherent flow regions are significantly more robust than methods which detect lower-dimensional flow barriers. Additional insights into which aspects of each method are self-consistent, and which are not, are provided.

*Accepted to *Physica D: Nonlinear Phenomena* on 22/10/2022, with the formal publication available at: https://doi.org/10.1016/j.physd.2022.133580.

*Corresponding author

Email address: aleksandar.badza@adelaide.edu.au (Aleksandar Badza)

Keywords: Lagrangian coherent structures, jets, vortices, uncertainty quantification

1. Introduction

With the realisation several decades ago that viewing Eulerian snapshots of various fields can give misleading information about transport, a suite of methods for detecting Lagrangian coherent structures (LCSs) has come to the fore. Identifying time-varying LCSs has particular importance in turbulent flows in which coherent structures at different lengthscales are influential, but difficult to distinguish from the turbulence. Many LCS techniques have been suggested, each of which defines its own notion for "coherence" with respect to flow over a chosen time interval. Several recent review articles [1, 2, 3, 4, 5] provide descriptions and additional citations to the multitude of proposed methods. Many applications of these methods for realistic observational [6, 7, 8, 9], experimental [10, 11, 12, 13] and computational [14, 15, 16, 17, 3] velocity data have been published. Hence, the results from numerically obtaining LCSs from unsteady Eulerian velocity data are now highly influential in the analysis and prediction of fluid transport.

Interpreting exactly what an LCS is continues to be a debated issue, since the various methods for finding LCSs use different criteria. LCSs detected from one method may not be detected from another. Thus, there are ongoing attempts to distinguish and/or consolidate the many LCS methods which are being used, with a particular emphasis on interpreting exactly what is being identified [1, 2, 3]. This paper will not venture into this debate, but will introduce a crucial aspect of LCSs which has received little attention: the fact that the Eulerian velocity fields which must be used as input into any detection scheme contain *uncertainties*. These uncertainties are inevitable: they occur because of observational/measurement error (limitations of a sensor's sensitivity), subgrid error (because the velocity is often sampled on a spatio-temporal grid that is too coarse to resolve the finest scales), observational limitations (miss-

ing or limited data; e.g., further away from satellite tracks or the presence of cloud cover), modelling error (because data sometimes requires conversion; e.g., sea-surface heights are usually converted to velocities by using a geostrophic assumption), and simulation limitations (e.g., when a computational fluid dynamics simulation uses a finite number of modes). Given the inevitability of uncertainty in the Eulerian data, how confident can we be in identified structures generated from an LCS computation? How robust is this identification to uncertainties?

Our aim in this article is therefore to examine the *self-consistency* of several LCS detection methods in *realistic* data-derived scenarios with uncertainty. Under this remit we forego toy models despite their popularity in the LCS literature, but instead use two flavours of "real-world" data: (i) computational fluid dynamics (CFD) simulation of a double shear layer, featuring coherent flow structures as well as turbulence, and (ii) observational data of the Gulf Stream. These are both in the realm of flows for which the detection of LCSs is relevant but difficult, namely possessing both turbulence and coherence. We investigate nine LCS methods which are significantly different from each other: (i) finite-time Lyapunov exponents (FTLEs), which relate to stretching of fluid parcels [16]; (ii) hyperbolic variational LCSs (hyp-vLCS), which identify flow separators towards which there is extremal contraction/repulsion [18, 5]; (iii) Lagrangian averaged vorticity deviation (LAVD), which attempts to quantify vorticity in a frame-independent fashion [19]; (iv) Lagrangian descriptors, which use the arclength of Lagrangian flow trajectories [20], (v) stochastic sensitivity S^2, in which the uncertainty in eventual Lagrangian location is formalised as a field [21, 22]; (vi) coherent set detection using a probabilistic transfer operator method [23]; (vii) coherent set detection using a dynamic Laplacian operator [24, 25]; (viii) partitioning a flow domain using a fuzzy c-means (FCM) clustering algorithm [26]; and (ix) identifying kinematic flow similarity using spectral colouring techniques [27]. We initially compute the relevant LCSs for each method using the data representing each of our flow systems; these would be the 'reported LCSs' if we treat each of the methods as error-free, as is standard

practice in the literature. We might construe such a computation as being "deterministic" in the sense that for every system, the procedure for computing the LCSs treats the velocity field as if it were deterministic and accurate, and does not rely on stochastic simulations. Next, we compute the LCSs when we have introduced noise into the Eulerian velocity data, and perform many simulations. These computations are then used as a basis for a quantitative evaluation of the self-consistency of each method.

Given that the data contains uncertainties and the "truth" is unknown, the quantitative analysis we perform falls under the standard philosophy of *uncertainty quantification*: obtaining the *expected* LCSs along with an *uncertainty* determined using statistical means. We quantify these based on stochastic simulations, and in the cases when the LCSs are associated with scalar fields, use the mean and standard deviation respectively for this uncertainty quantification. When the LCSs are associated with regions and/or curves, we develop probability distributions to identify membership within sets. In all cases, we check that sufficiently many stochastic simulations are included to ensure convergence of the mean and standard deviation. Since the key in LCS methods is identifying their spatial distribution, we compute and display the spatial variation of the expected and uncertainty fields for each method.

There have been a few prior articles of the effect of noise on certain individual LCS computations: FTLEs averaged in different ways over noise realisations [28], a theoretical quantification of the error in the FTLE field [22], and trajectory-averaged scalar fields in noisy situations [29]. These studies are mostly for toy models (where the velocities are given by explicit formulæ), and are confined to *one* LCS method in each case. In contrast, BozorgMagham et al [30] use real atmospheric data is used in comparing between forecast and archive-based FTLE fields for real atmospheric data, using point-based and cross-correlation measures. Lermusiaux [31], in a review of uncertainty quantification in oceanography, provides an example of the uncertainty in the FTLE field and its ridges. In a similar spirit, Olcay et al [32] analyze the impact of Eulerian velocity data uncertainty on FTLE ridges associated with a idealized vortex pair,

as well as CFD-generated axisymmetric vortex ring. Taking things further, the present paper uses realistic velocity data, and examines self-consistency for a *range* of different LCS methods, thereby providing a comprehensive, comparative and systematic analysis on the impact of velocity uncertainty on conclusions of Lagrangian coherence.

Our results indicate a substantial variation in how the LCS methods react to uncertainties in the data. As a rule, *fields* on the entire space of initial points, as computed at either an intermediate or final stage in the LCS process, change relatively little when the uncertainties are small. Flow *regions* or *sets* which are detected as 'coherent' according to the relevant characterisation (e.g., where the LAVD exceeds a threshold, the FTLE or S^2 is less than a threshold; or where particle trajectories remain in close proximity or exhibit similar flow behaviour) are therefore relatively robust. Conversely, when seeking lower-dimensional objects (such as ridges of the FTLE field, or flow separators as identified through the hyp-vLCS approach), significant sensitivity to uncertainty is observed. In the hyp-vLCS method in particular, the inferred LCSs undergo dramatic changes with high probability, indicating their lack of self-consistency according to our definition. We provide additional insights into how the different aspects of each LCS method are impacted by uncertainty in the Eulerian data. Given the inevitability of uncertainty in velocity data in any realistic situation, we therefore provide cautions on each LCS method's usage in reliably inferring transport characteristics.

2. Lagrangian Coherent Structure Detection Methods

This section details the Lagrangian coherent structure (LCS) detection methods which will be tested against velocity uncertainties in this study. In general, we assume we have Eulerian velocity data $\boldsymbol{u}$ from an initial time t_0 for an additional time T in a spatially two-dimensional domain. This implies that Lagrangian trajectories are generated from the non-autonomous flow defined by

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x}, t) \quad \text{for} \quad t \in [t_0, t_0 + T]\,. \tag{1}$$

Let $\Omega \in \mathbb{R}^2$ be the domain at time t_0, and $\boldsymbol{x} \in \Omega$ be a general initial condition. The flow map $\mathbf{F}_{t_0}^{t_0+T}(\mathbf{x})$ then defines the location at the final time $t_0 + T$ of this initial condition subject to the flow defined in (1). The detection of LCSs is concerned with finding coherent sets or objects within Ω at time t_0, with respect to the flow over the next T time units. What defines a coherent structure varies considerably between each of these methods. We will only give brief and primarily intuitive definitions of each method in this section; for more comprehensive details, the reader is referred to a variety of review papers [1, 2, 3], or original papers that we cite. For more details regarding the computational specifics of each method relevant to this study, the reader is referred to the supplementary material.

2.1. Finite Time Lyapunov Exponent

One of the most well-known and easy-to-use methods for detecting Lagrangian coherent structures involves the use of the finite time Lyapunov exponent (FTLE), a measure of the exponential stretching of infinitesimal fluid parcels over a pre–determined finite period of time. We define $\nabla\mathbf{F}_{t_0}^{t_0+T}(\mathbf{x})$, the 2×2 gradient (or Jacobian) of the flow map at $\mathbf{x}$, and use this to define the Cauchy-Green strain tensor field

$$\mathbf{C}_{t_0}^{t_0+T}(\mathbf{x}) = \left(\nabla\mathbf{F}_{t_0}^{t_0+T}(\mathbf{x})\right)^{\top} \nabla\mathbf{F}_{t_0}^{t_0+T}(\mathbf{x}) \tag{2}$$

over Ω. The FTLE field is then given by [16]

$$\Phi_{t_0}^{t_0+T}(\mathbf{x}) = \frac{1}{|T|} \ln \sqrt{\lambda_{\max}(\mathbf{x})}, \tag{3}$$

where $\lambda_{\max}(\mathbf{x})$ is the largest eigenvalue of (2), and pertains to the maximal relative stretching of the infinitesimal fluid particle at $\mathbf{x}$, converted to an exponential-in-time rate. Some studies, consonant with the theory of stable and unstable manifolds for infinite–time flows, then seek maximal ridges of the scalar field of (3) to define LCSs which act as flow barriers along which particle stretching

capability is maximal, and thereby segregate the flow domain Ω into more coherent sub-regions. The relative simplicity of this method has made this a "go-to" method for several researchers who have used LCSs to solve problems pertaining to a wide variety of simulated and real-world flows [33, 34, 35, 36, 37, 38].

2.2. Variational LCSs

Haller [39] proposed an alternative method to the FTLE for detecting LCSs called the variational LCS (VLCS) detection method. It involves defining variational problems for the eigenbasis of (2) and solving these problems to produce geodesic curves which are tangential to relevant eigenvector fields. The basic idea is that these are material curves (or higher-dimensional analogues) to which there is maximal attraction or repulsion over the chosen flow time. An important observation that is often overlooked is that this process initially defines a collection of curves which is effectively a foliation of Ω (generically, the foliation will have a finite number of singular points), and some selection process is required for extracting a subset of these curves as being the "most influential." A variety of algorithms [17, 39, 5, 18] have been published for executing this method, with each of these using a different set of computational and selection techniques. The algorithm which will be used in this study is the "LCS Tool" algorithm [18]. This initially looks for the point in Ω which has the largest eigenvalue $\lambda_{\max}$ of the Cauchy–Green tensor (2), selects the curve from the foliation passing through this, then excludes a neighbourhood of a pre-determined size from this, looks for the point in the remaining domain with the largest $\lambda_{\max}$ value, determines the curve passing through that, and so on. An underappreciated observation is that locations where $\lambda_{\max}$ is maximal is equivalent to the FTLE being maximal (see (3)), and hence the FTLE field's variation strongly impacts the selection process. The collection of curves extracted from this process are termed "the" hyp-vLCSs. For a more comprehensive description of the steps involved in this algorithm, the reader is referred to the Supplementary Material.

2.3. Lagrangian Averaged Vorticity Deviation

Not all LCS detection methods are concerned with generating codimension–1 flow separating barriers within regions of the domain where more chaotic flow behaviour is exhibited. There are a wide variety of methods available for detecting coherent flow regions within the domain of the fluid rather than the boundaries which separate them. One such method involves the computation of Lagrangian averaged vorticity deviation (LAVD) [19], a frame-independent measure of the relative vorticity of a flow trajectory over the time interval $[t_0, t_0+T]$ against the mean vorticity over the entire domain Ω. Haller et. al [19] define the LAVD through numerical evaluation of the integral

$$\Theta_{t_0}^{t_0+T}(\mathbf{x}) = \int_{t_0}^{t_0+T} |\omega_3(\mathbf{F}_{t_0}^{s}(\mathbf{x}), s) - \overline{\omega}(s)| \, \mathrm{ds}\,, \tag{4}$$

where ω_3 is the only non-zero component of the vorticity vector $\boldsymbol{\omega} = \nabla \times \mathbf{u}$, and $\overline{\omega}(t)$ its mean over the whole domain at time t. Large values (local maxima) of the field in (4) identify coherent vortical structures.

2.4. Lagrangian Descriptors

The LAVD is computed by integrating the vorticity deviation along a trajectory, and helps identify rotational coherence. The process of integrating general functions over a flow trajectory defines so-called Lagrangian descriptors, initially developed for their utility in identifying stable and unstable manifolds in infinite-time flows [20]. The standard choice of the relevant function is the speed [20], and hence we use the form

$$\Psi_{t_0}^{t_0+T}(\mathbf{x}_0) = \int_{t_0}^{t_0+T} \left\|\mathbf{u}(\mathbf{F}_{t_0}^{s}(\mathbf{x}_0), s))\right\| \, \mathrm{ds}, \tag{5}$$

which effectively computes the arclength traversed by a trajectory beginning at $\mathbf{x}_0$ over the relevant time interval. We use the limits t_0 to $t_0 + T$ consistent with other comparative LCS studies [40, 2]; this is fundamental to the idea of identifying structures at time t_0 which are 'coherent' due to the flow till time $t_0 + T$, and allows comparison across all the methods that we use.

2.5. Stochastic Sensitivity

Another way of quantifying the coherence of flow trajectories is by measuring the numerical uncertainty of the eventual location of each trajectory. A model for the uncertainty in the Eulerian velocity $\boldsymbol{u}$ could be to examine the stochastic differential equation (SDE)

$$\mathrm{d}\mathbf{x}_t = \mathbf{u}(\mathbf{x}_t, t)\ \mathrm{d}t + \epsilon\boldsymbol{\sigma}(\mathbf{x}_t, t)\ \mathrm{d}\mathbf{W}_t \quad ; \quad t \in [t_0, t_0 + T] \tag{6}$$

instead of the ordinary differential equation (1). The drift function $\mathbf{u}$ is as in (1), ϵ represents the level of known or anticipated uncertainty within the velocity (assumed small in an appropriate sense), $\boldsymbol{\sigma}$ is the volatility (diffusion) matrix, which is allowed to be spatio-temporally dependent, and $\mathbf{W}_t \sim \mathrm{N}(0, 1)$ represents two-dimensional geometric Brownian motion in units of the square root of time [22]. Let $\boldsymbol{x}_t$ be a stochastic solution to (6) subject to the initial condition $\boldsymbol{x}_{t_0} = \boldsymbol{x} \in \Omega$. Balasuriya [21] defines the stochastic sensitivity (S^2) field on Ω by

$$S^2(\mathbf{x}) = \lim_{\epsilon \to 0} \max_{\phi \in [-\pi/2, \pi/2)} \mathrm{Var}\left(\begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix}^\top \frac{\mathbf{x}_{t_0+T} - \mathbf{F}_{t_0}^{t_0+T}(\mathbf{x})}{\epsilon} \right). \tag{7}$$

This variance represents the expected uncertainty in the final location, and is thus a direct measure of the *unpredictability* of a fluid particle beginning at $\boldsymbol{x}$. Balasuriya [21] develops an explicit *analytical* expression for S^2 based on only knowledge of the deterministic flow and the stochastic model (6), and does *not* require any stochastic simulation. The reader is referred to the supplementary material for more information on the steps required for numerically computing S^2 via this expression. While the original development [21] used a non-dimensional ϵ and a dimensional σ, in this paper we use the alternative scaling technique [22] which highlights the physical quantities limiting the data: we set

$$\epsilon = \sqrt{h v_r} \tag{8}$$

in which h is the spatial resolution scale, and v_r is the velocity uncertainty scale, and we set $\boldsymbol{\sigma}$ to be nondimensional. Then, the scaled version of S^2 given by

$$S_v(\mathbf{x}) = \epsilon\sqrt{S^2(\mathbf{x}_0)} = \sqrt{hv_r S^2(\mathbf{x})} \tag{9}$$

is explicitly a lengthscale associated with the uncertainty of eventual location of a trajectory beginning at $\mathbf{x}_0$. The quantities S^2 or S_v typically range over several orders of magnitude within a flow region [21], and thus their logarithms are sometimes used in visualisations. One can also define coherent flow regions within Ω at a lengthscale L using the robust sets [21]

$$R(L, v_r, h) = \{\mathbf{x} \in \Omega : S_v(\mathbf{x}) < L\} \ . \tag{10}$$

In the absence of any additional information on the uncertainty distribution, we make the default choice of $\boldsymbol{\sigma}$ being the identity matrix [21] throughout this paper.

2.6. Transfer Operator

There also exist probabilistic methods for detecting coherent sets within a flow domain. Suppose we divide the flow domain at time t_0, Ω, into $M_x \times M_y$ rectangular boxes of equal size, and divide the flow domain at time $t_0 + T$, $\widetilde{\Omega} = \mathbf{F}_{t_0}^{t_0+T}(\Omega)$ into $N_x \times N_y$ rectangular boxes of equal size. The operator $\mathbf{P}$ is defined as a $N_xN_y \times M_xM_y$ column-stochastic transition matrix where the (i, j)-th element indicates the probability of a particle beginning in box j at time t_0 ending up in box i at time $t_0 + T$ under the influence of (1). While this matrix is often defined in the literature as a row stochastic transition matrix [23, 41], we will here use the column-stochastic formulation instead. Thus $\mathbf{P}$ is a discrete approximation for the transfer or Perron–Frobenius operator [23, 41].

If $\mathbf{v}_0$ is a vector of the initial distribution of a collection of particles over the boxes within Ω at time t_0,

$$\mathbf{v}_T = \mathbf{P}\mathbf{v}_0,$$

represents the distribution of the same collection of particles throughout $\widetilde{\Omega}$ following the advection of these particles by $\mathbf{F}_{t_0}^{t_0+T}$. Since $\mathbf{P}^\top$ represents the

pullback from time $t_0 + T$ to t_0, in seeking "coherence" the idea is to look for initial distributions $\mathbf{v}_0$ such that the pullback of $\mathbf{v}_T$ approximately preserves $\mathbf{v}_0$, that is,

$$\mathbf{v}_0 \approx \mathbf{P}^\top \mathbf{P} \mathbf{v}_0.$$

In other words, we seek $\mathbf{v}_0$ corresponding to the larger λ values of the eigenvalue problem

$$\mathbf{P}^\top \mathbf{P} \mathbf{v}_0 = \lambda \mathbf{v}_0. \tag{11}$$

Since the eigenvalues of $\mathbf{P}^\top \mathbf{P}$ are non-negative and bounded by 1, and $\lambda_1 = 1$ is always the largest eigenvalue [23, 41], the eigenvector $\mathbf{v}_0$ corresponding to the second largest eigenvalue λ_2 can be considered a "highly coherent" initial distribution. The simplest implementation of the transfer operator approaches is to partition Ω based on the zero contour of this distribution [23], thereby splitting Ω into two "almost coherent" sets. Extensions of this technique (seeking an optimal partition value, taking into account next-level eigenvalues, looking for a clear spectral gap, etc) are available, and continue to be developed. We will use the most fundamental approach in this paper.

2.7. Dynamic Laplacian Operator

Another way to divide a flow domain Ω into two coherent sets involves using the Laplacian operator $\boldsymbol{\Delta}$ to define the eigenvalue problem

$$\boldsymbol{\Delta} \mathbf{v} = \lambda \mathbf{v}. \tag{12}$$

The second eigenvector of $\boldsymbol{\Delta}$ applied to an autonomous system helps divide a flow domain into two coherent sets by optimising the ratio of the boundary length of the sets to their area [24]. A similar result for non-autonomous flows requires defining the dynamic Laplacian operator [25]

$$\boldsymbol{\Delta}_D = \frac{1}{2} \left(\boldsymbol{\Delta}_0 + \mathbf{P}^\top \boldsymbol{\Delta}_T \mathbf{P} \right),$$

where $\mathbf{P}$ is the transfer operator matrix as defined in the previous section, and $\boldsymbol{\Delta}_0$ and $\boldsymbol{\Delta}_T$ are Laplacian matrix operators defined over the box configurations

of Ω at time t_0 and $\widetilde{\Omega}$ at time $t_0 + T$ respectively. We then solve the eigenvalue problem

$$\mathbf{\Delta}_D \mathbf{v} = \lambda \mathbf{v} \tag{13}$$

and use the second eigenvector $\mathbf{v}_2$ to extract two coherent sets for Ω in similar fashion to the transfer operator method [24].

2.8. Fuzzy C-Means Clustering

The aforementioned methods can all be used on flow systems where the velocity data of said system is available, either in the form of analytic equations or Eulerian snapshots of the data defined at different time intervals. However for some systems, only the data representing sparsely defined flow trajectories is available, and therefore LCSs can only be detected using trajectory-based techniques. One such method involves grouping together flow trajectories based on their spatial proximity using a fuzzy c-means (FCM) clustering algorithm [42, 43], such as the fcm algorithm available in MATLAB [1].

We have as input N trajectories of (1) each with data for a discrete set of times between t_0 and $t_0 + T$ which are equally spaced by a time step Δt and are arranged in the form

$$\mathbf{X}_j = [\mathbf{x}_j(t_0), \mathbf{x}_j(t_0 + \Delta t), \mathbf{x}_j(t_0 + 2\Delta t), \ldots, \mathbf{x}_j(t_0 + T)]\,;$$

where each $\mathbf{x}_j(t_0 + i\Delta t) = (x_j(t_0 + i\Delta t), y_j(t_0 + i\Delta t))$. Given a pre-determined or guessed number of clusters K, the idea of the fuzzy c-means clustering algorithm is to partition these N trajectories into K clusters by optimising the distance to a "centre trajectory" of each cluster. A recursive algorithm determines the K clusters and centre trajectories $\mathbf{C}_k$, along with the probability $p_{k,j}$ of trajectory j belonging to cluster k [1, 42, 43]. For further details relating to the fuzzy c-means clustering algorithm, refer to the supplementary material or to [42, 43].

2.9. Coherent Structure Colouring

The biggest drawback of the FCM clustering algorithm is the requirement of the number of coherent clusters to be specified beforehand, or just arbitrarily

guessed. In an effort to overcome this, the coherent structure colouring (CSC) method which detects LCSs using kinematic similarities of certain quantities of interest (the distance between trajectories, or the relative vorticity of trajectories) was conceptualised [27] . Given the same N sparse trajectories defined in the previous subsection, we calculate the $N \times N$ adjacency matrix $\mathbf{A}$, where the (i, j)-th element is the coefficient of variation of our quantity of interest between two trajectories $\mathbf{X}_i$ and $\mathbf{X}_j$ over all time steps $t_k = t_0 + k\Delta t$. The sums of the rows of $\mathbf{A}$ form the elements of the diagonal matrix $\mathbf{D}$, and given the graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$, one solves the generalised eigenvalue problem

$$\mathbf{L}\mathbf{v} = \lambda\mathbf{D}\mathbf{v}. \tag{14}$$

Let $\lambda_{\max}$ be the largest eigenvalue for (14), and $\mathbf{v}_{\max}$ its corresponding eigenvector. The eigenvector $\mathbf{v}_{\max}$ assigns a CSC coefficient to the final point of each trajectory, all of which are subsequently mapped back to the initial points of said trajectories. LCSs are then identified as sub-regions of Ω which correspond to similar CSC values [27].

3. Results

We will now display the results obtained from executing each of the methods described in Section 2 on two systems in which numerical data for the Eulerian velocities is available. The first of these is based on a computational fluid dynamics simulation of a flow which contains two rapidly evolving Kelvin–Helmholtz vortex layers, and the other is using direct oceanographic velocity data pertaining to the Gulf Stream in the North West Atlantic Ocean, positioned off the North East coast of the United States. Each LCS detection method will be implemented on these data sets in their regular, unperturbed form, and with the addition of artificial noise to the velocity using the stochastic differential equation (6), with noise added uniformly to the velocity at every point within a uniformly spaced initial grid of points. The noise is updated at each time-step. Our own codes were produced to execute all of the LCS detection methods,

which were tested against previously published results for analytically defined flows.

Rather than performing a few simulations for a qualitative comparison, we implement a quantitative statistical analysis on the LCS results generated from each method over sufficiently many different realisations of solutions to (6). For each of these methods (except the hyp–VLCS method and the stochastic sensitivity robust sets), we generate 100 stochastic flow maps for every initial point in Ω, perform LCS analysis on each of these realisations, and iteratively calculate the mean and standard deviation of quantities such as the FTLE, the coefficients of the transfer operator optimal vector, or the membership probability of a flow particle to a fuzzy c–means cluster. A sample size of 100 was sufficient to obtain convergence of the statistics for every method, as we demonstrate in the Supplementary Material.

Given that they are not directly associated with fields unlike the other techniques, the hyp–VLCS method and the robust sets arising from stochastic sensitivity are not amenable to analysis using means and standard deviations. For the hyp–VLCS method, we divide our flow domain into equally sized boxes or "bins", record which of these bins each hyp–VLCS traverses through for each realisation of this method and use this to produce a sample likelihood field. We also include a similar field showing the iteratively calculated variation (standard deviation) of these sample probabilities. In the case of the stochastic sensitivity robust sets, we produce a probability density function from the coordinates of the points which fall within the robust set from each stochastic simulation.

3.1. Simulated Kelvin–Helmholtz Flow

First, we will look at numerical velocity data simulated from the Navier–Stokes equations with an initial condition set up to generate Kelvin–Helmholtz type vortex layers on the doubly periodic flow domain $\Omega = [0, 2\pi) \times [0, 2\pi)$; inspired by several computational fluid dynamics experiments documented in the literature [44, 45, 46, 47]. To generate a typical flow in which both coherence

and transience is present, we choose the initial condition

$$\boldsymbol{u}_0(x,y) = \begin{bmatrix} \frac{1}{2}\left(\tanh\left[\frac{1}{\delta}(y-\frac{\pi}{2})\right] - \tanh\left[\frac{1}{\delta}(y-\frac{3\pi}{2})\right]\right) \\ \frac{1}{100}\left(\sin\left[\alpha x\right] + \frac{1}{2}\sin\left[(\alpha-1)x\right] + \frac{1}{2}\sin\left[(\alpha+1)x\right]\right) \end{bmatrix}, \tag{15}$$

where $\delta = 0.05$ and $\alpha = 9$, the closest integer to the most unstable wavenumber mode $0.22/\theta_0 \approx 0.44/\delta$ [48]. For concreteness in our subsequent uncertainty analysis, we treat the lengthscale, time and velocities as being scaled respectively by 1 m, 1 s and $1\,\mathrm{m\,s^{-1}}$. Our own validated MATLAB code was used to produce CFD simulations from the initial condition (15) using a fast Fourier transform algorithm. We generate Eulerian velocity data on a uniform 1024×1024 spatial grid within the double-periodic domain Ω using a third-order Runge–Kutta integration method from time 0 to 20, with time-step $\Delta t = 0.001$ s and velocity data captured after time intervals of length 0.1 s. We also apply modest viscosity to these simulations using a Reynolds number $\mathrm{Re} = 10^4$, computed using the lengthscale $L = 1\,\mathrm{m}$, the maximal velocity $1\,\mathrm{m/s}$ and kinematic viscosity $\nu = 10^{-4}\,\mathrm{m^2/s}$.

A basic description of the generated flow is in order. It consists of a channel of horizontal particle flow centred along $y = \pi\,\mathrm{m}$ with velocity magnitude $\sim 1\,\mathrm{m/s}$, very small velocities in the outer layers of the flow domain where $y < \pi/2\,\mathrm{m}$ and $y > 3\pi/2\,\mathrm{m}$, and two Kelvin-Helmholtz vortex layers centred along $y = \pi/2\,\mathrm{m}$ and $y = 3\pi/2\,\mathrm{m}$ at initial time $t = 0$ s. Beginning from $t \approx 5$ s, these vortices begin to combine and grow in size, increasing the width of the vortex layers. This continues until $t \approx 15$ s, by which time four large vortices have been created and the width of the coherent flow channel centred at $y = \pi\,\mathrm{m}$ has approximately halved.

We perform LCS analysis on this system using all of the methods listed in Section 2 on the flow period ranging from $t_0 = 12$ s to $t_0 + T = 15$ s. The level of velocity uncertainty v_r which will be applied to all stochastic simulations has been set to $1.63 \times 10^{-2}\ \mathrm{ms^{-1}}$ which represents approximately a 1.6% velocity uncertainty. Given that the data resolution is $h = 2\pi/1024 \approx 6.1 \times 10^{-3}\,\mathrm{m}$, this gives a value of $\epsilon = \sqrt{hv_r} = 10^{-2}\ \mathrm{ms^{-1/2}}$ for the SDE (6).

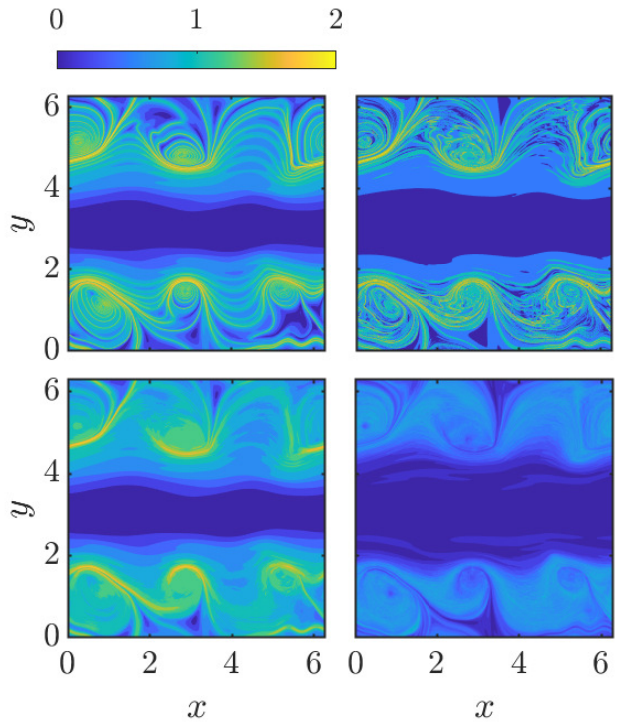

Figure 1: The deterministic FTLE of the Kelvin–Helmholtz flow from $t_0 = 12$ s to $t_0+T = 15$ s (top-left), one realisation of the FTLE with noise applied to the flow (top-right) and the mean (bottom-left) and twice the standard deviation (bottom-right) of 100 realisations of the stochastic FTLE.

The first LCS detection method to be examined is the FTLE, which appears to be modestly robust to velocity uncertainty as shown in Figure 1. The panels represent the FTLE scalar field generated from the deterministic Kelvin-Helmholtz flow (top left), one realisation of the FTLE with noise applied to the flow (top right), the mean (bottom left) and twice the standard deviation (bottom right) of the FTLE obtained from 100 stochastic realisations. We plot twice the standard deviation to enhance the visibility of key structures within this scalar field. We will follow this structuring of panels for (almost) all the figures in this section. Under noise perturbation, the FTLE field wobbles, and some of the maximal ridges become jagged and less smooth. From the mean

FTLE field, the central flow channel, within which the FTLE is low and particle movement more robust, more or less maintains its shape. However, due to a large variation in FTLE values within the upper and lower vortex layers, many of the maximal ridges from the deterministic FTLE field have been ironed out. The standard deviation field indicates that errors are more likely near the higher FTLE values, i.e., near the ridges. That being said, most of the maximal ridges corresponding to the globally largest values of the FTLE have more or less maintained their shape.

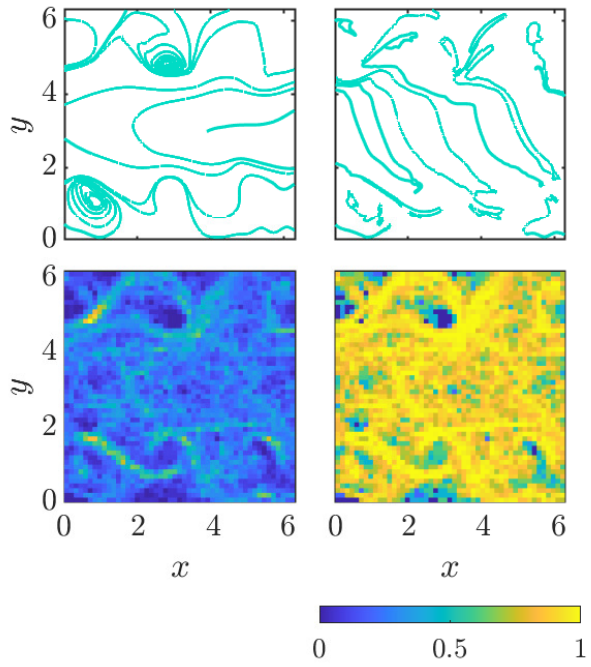

Figure 2: Deterministic hyp–VLCSs of the Kelvin–Helmholtz flow from $t_0 = 12$ s to $t_0 + T = 15$ s (top–left), one realisation of these hyp–VLCSs with noise applied to the flow (top–right) and the likelihood (bottom–left) and twice the variation in this likelihood (bottom–right) of a hyp–VLCS passing through one of $50 \times 50 = 2500$ bins computed from 100 realisations of stochastic hyp–VLCSs.

While the FTLE method is reasonably robust to stochastic noise, the variational hyperbolic LCS method is considerably affected, as demonstrated by

Figure 2. In the top row, we show the deterministic hyp–VLCSs along with one stochastic realisation. The LCS curves in the noisy simulation are significantly different from the deterministic one. A quantitative analysis across many simulations is shown in the lower row. The left panel shows the likelihood of a hyp–VLCS passing through $50 \times 50 = 2,500$ equally sized bins (we only choose this many boxes so that the width of each of these boxes is approximately the same as the separation radius parameter ρ for the LCS Tool algorithm - the reader is referred to the Supplementary Material for further information), along with a similar field showing twice the variation of these sample probabilities. The likelihood field has few discernible features, save for some hook–shaped curves which surround the coherent vortices. The abundance of bins with a low probability of being traversed by a hyp–VLCS (indicated by blue coloured cells) indicates a general lack of certainty in identifying features. This is further verified by the variation field which is high throughout most of the domain. Thus, the shape and locations of hyp–VLCSs is *highly* susceptible to any uncertainty in the data.

Next we consider the LAVD, which is structurally robust to velocity uncertainty as displayed in Figure 3. The vortices which correspond to the largest values of the LAVD do not change position under noise, however their size and LAVD amplitude do. The mean and standard deviation fields show that the LAVD is robust in the sense that it is able to locate the centres and approximate shapes of coherent vortices well, despite the variation in the size of the LAVD field. The relative error—the ratio of the standard deviation to the mean—is fairly uniformly distributed across the domain.

We next examine the Lagrangian descriptor method, which as Figure 4 demonstrates, is similarly robust. The noise causes minor distortions of the structures within the flow layers, while preserving the shape of the central flow channel. The mean, as in the FTLE, is a diffused version of the deterministic field. The standard deviation is relatively low, and appears to be concentrated not only between the coherent vortices, but also around the centres of these vortices.

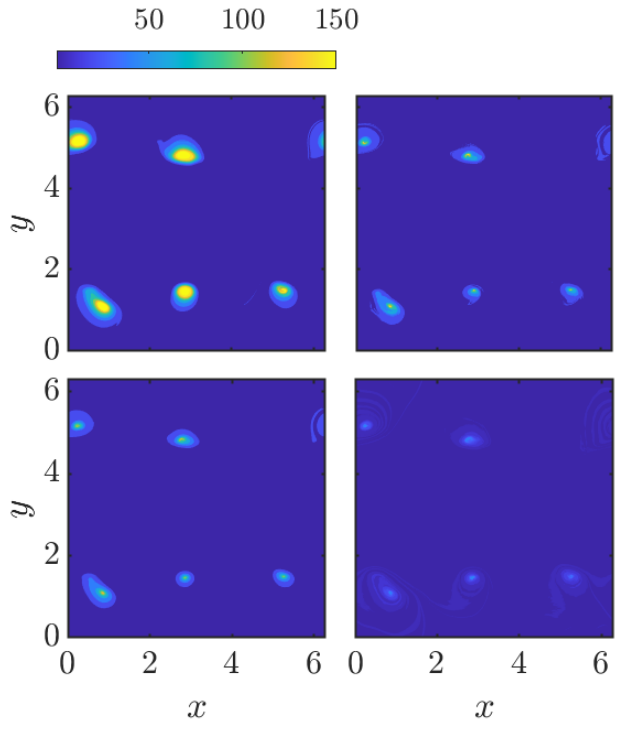

Figure 3: LAVD diagnostics for the Kelvin–Helmholtz flow arranged in the same fashion as seen in Figure 1.

Figure 5 shows that the generated values of the stochastic sensitivity are mildly compromised under the influence of velocity uncertainty. (Due to the known large variation in the values of S^2 [21], we plot the scaled quantity $\ln(S_v) = \ln\left(\sqrt{hv_r S^2}\right)$ rather than the raw S^2 for improved visibility.) Stochastic sensitivity, as it is a measure of unpreditability, is largest within the vortex layers above and below the flow channel, and its maximal ridges are similar to that of the FTLE field shown in Figure 1. Similar to the FTLE, the ridges deform under the influence of stochastic noise. The mean ridges from the 100 stochastic realisations are consistently located with respect to the deterministic ones, though some deterministic ridges have disappeared indicating their uncertainty. The standard deviation is noticeably high, particularly along the

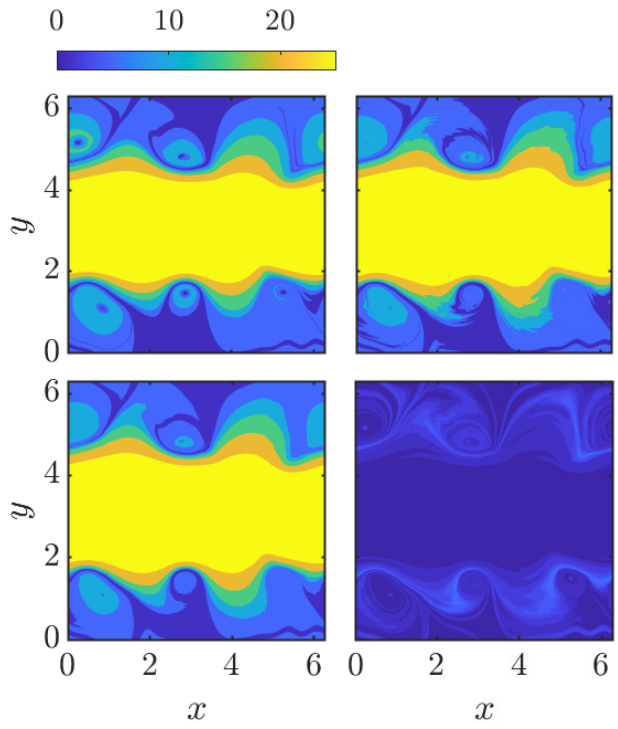

Figure 4: Lagrangian descriptor diagnostics for the Kelvin–Helmholtz flow arranged in the same fashion as seen in Figure 1.

ridges where $\ln(S_v)$ is largest. This can be attributed to large values and the high spatial gradients of S^2 near these ridges [21], which will amplify velocity uncertainties.

The robust sets, as defined in (10), are much more self-consistent, as Figure 6 indicates. In the left panel we display the deterministic robust set R (with velocity uncertainty $v_r = 1.63\times10^{-2}$ and grid-spacing $h = 6.1\times10^{-3}$) of regions of flow where the eventual lengthscale of uncertainty is less that $L = 0.25$m. The right panel is a likelihood, representing the sample probability density of being selected in a robust set from the 100 realisations. The central flow channel is included with high probability notwithstanding the velocity uncertainty. The jet being significantly more robust than the vortex centres is understandable in

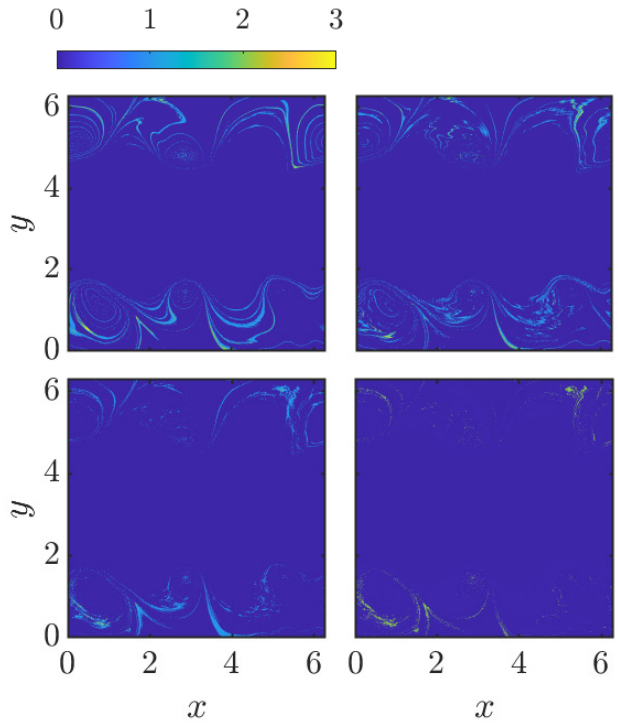

Figure 5: Diagnostics for the scaled stochastic sensitivity S_v for the Kelvin–Helmholtz flow arranged in the same fashion as seen in Figure 1. The natural logarithm of S_v has been used to enhance the visibility of key structures within each contour plot.

this Kelvin–Helmholtz flow in which the vortex layers are unstable.

The scalar field associated with the optimal vector derived from the transfer operator method is mostly robust to velocity uncertainty, according to the results in Figure 7 (see the Supplementary Material for more details regarding how the optimal vector is produced). The lower panel shows the mean and twice the standard deviation generated from 100 stochastic realisations. If partitioning at the zero contour (as in the original idea of partitioning into two coherent sets [23], the central flow channel along with most of the flow domain lying between the vortex layers appears to form one coherent set, while the two vortex layers form the other coherent set. Applying stochastic noise produces the same coherent sets, except with some fuzziness. The mean optimal vector field appears

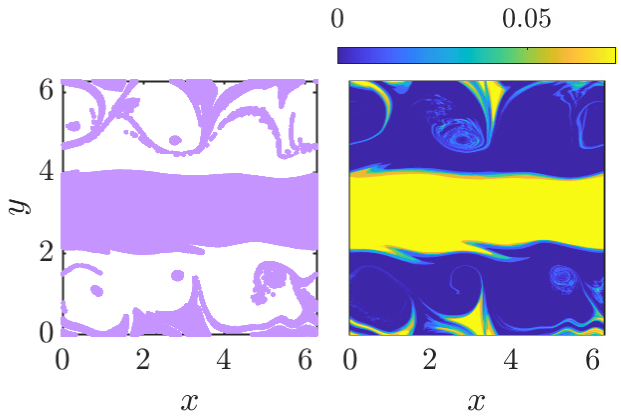

Figure 6: The deterministic robust set R at lengthscale 0.25 m of the Kelvin–Helmholtz flow from $t_0 = 12$ s to $t_0 + T = 15$ s (left) and the sample probability density function of the points which form this set from 100 stochastic realisations of S^2 (right).

a diffused version of the deterministic field, while the standard deviations of the vector coefficients appear to be low with the largest values scattered throughout the central flow channel instead of within the vortex layers. The coherent sets detected from this one eigenvector are unable to detect the smaller structures within the vortex layer unlike in other methods; hence the lack of uncertainty in the larger vortex layer region. Distinguishing these structures relies on considering eigenvectors corresponding to eigenvalues much lower in the ordering. We remark, however, that in this complicated flow the first 10 eigenvalues are all within one thousandth (10^{-3}) of each other with consistently small gaps between consecutive eigenvalues, making it difficult to use the transfer operator method to detect these individual structures. The fact that the flow is "more turbulent" in this region is likely relevant, with the presence of structures at finer and finer scales. Thus, as the flow gets more turbulent, the transfer operator approach is likely an impractical approach for distinguishing smaller-scale coherent structures.

The second eigenvector of the dynamic Laplacian operator is similarly robust, as the results in Figure 8 show. As was the case with the transfer operator

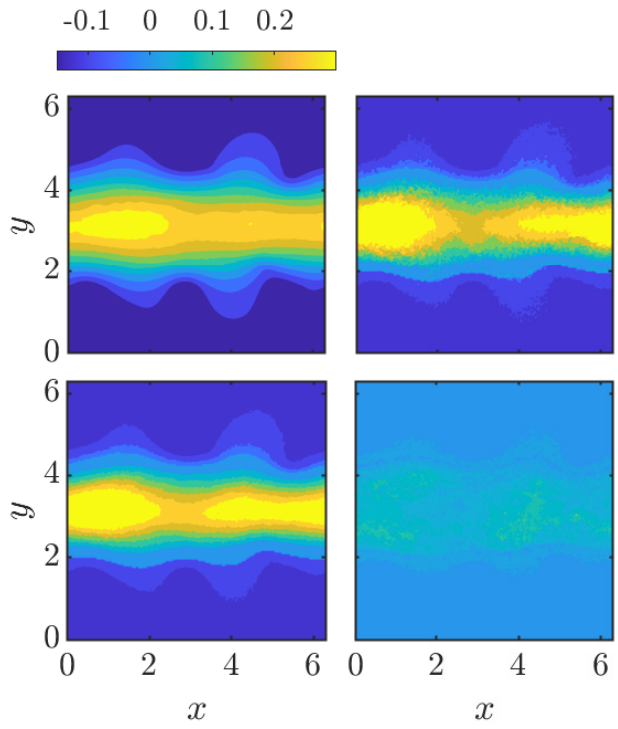

Figure 7: Transfer operator optimal vector diagnostics for the Kelvin–Helmholtz flow arranged in the same fashion as seen in Figure 1.

optimal vector field (see Figure 7), the central flow channel is identified as the first coherent set, and the two vortex layers as the other. Unlike the transfer operator optimal vector, the dynamic Laplacian operator second eigenvector is noticeably less fuzzy under the influence of stochastic noise. This is most likely due to smoothening generated via adding the Laplacian operator. Indeed, this method conceptualises the explicit addition of diffusive noise in the original calculations, and hence the 'deterministic' and mean fields (two left panels in the figure) do end up being remarkably similar. The standard deviation field indicates that the error is reasonably low and, as was the case with the transfer operator method, is at its largest within the flow channel rather than along the boundaries between the flow channel and the vortex layers.

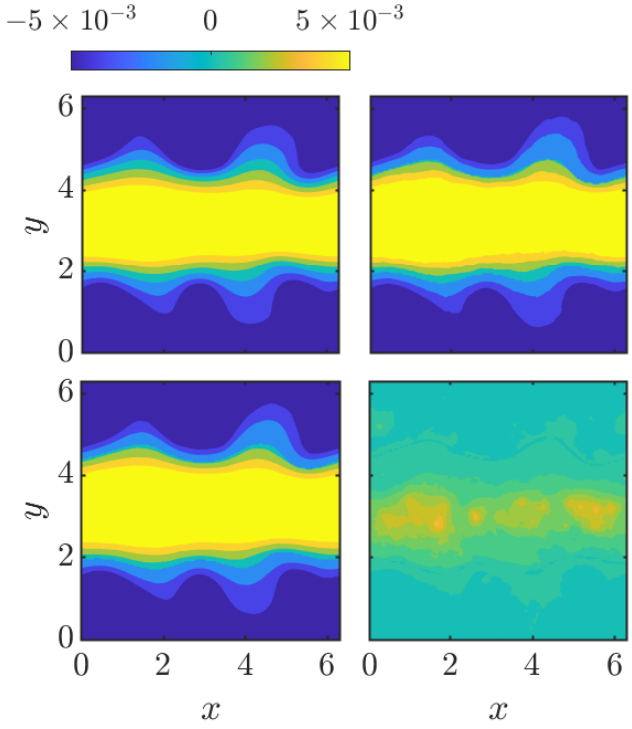

Figure 8: Dynamic Laplacian operator second eigenvector diagnostics for the Kelvin–Helmholtz flow arranged in the same fashion as seen in Figure 1.

The FCM Clustering algorithm also shows self-consistency under the influence of noise, as observable from Figure 9. Here, we program the FCM Clustering algorithm to partition Ω into four coherent clusters. These clusters have been identified as the central flow channel, the upper and lower vortex layers; and two disjoint wave–like sets positioned between the flow channel and each vortex layer. Applying noise only appears to have a minor impact on the membership probabilities and shapes of each cluster, and the mean membership probabilities are more or less consistent with the deterministic results. From the standard deviation fields, the largest variation in membership probability is observable primarily along the boundaries of each cluster as expected, with only some very minor exceptions identifiable.

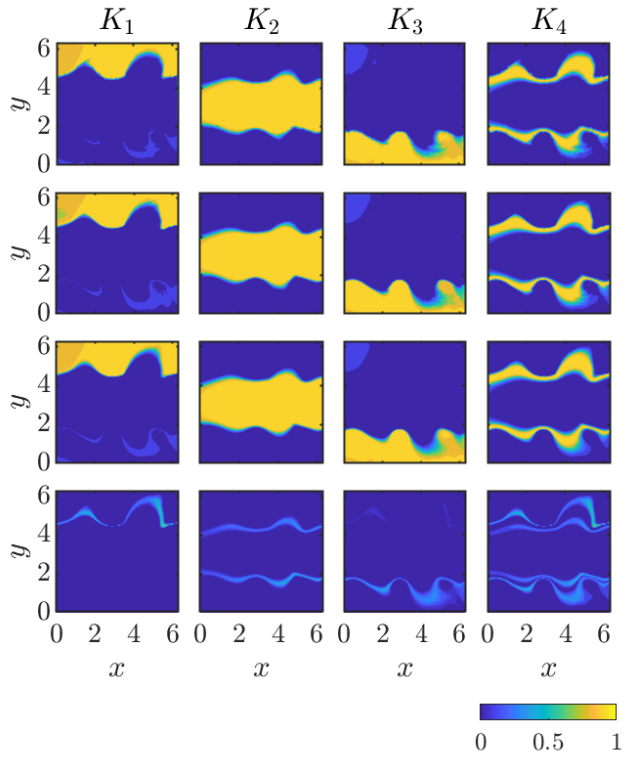

Figure 9: The deterministic membership probability fields for four FCM clusters K_1, K_2, K_3 and K_4 relevant to the Kelvin–Helmholtz flow from $t_0 = 12$ s to $t_0 + T = 15$ s (first row); one realisation of each of these fields with noise applied to the flow (second row) and the mean (third row) and twice the standard deviation (fourth row) of these probabilities generated from 100 realisations of the FCM clustering method.

The CSC scalar field also shows robustness to noise, as demonstrated by Figure 10. We can once again identify the coherent flow channel and the vortex layers, along with approximate outlines of the actual vortices themselves. In a flow with clearly defined vortices which do not merge and remain consistent in size, such as the Bickley jet [27], one would be able to clearly identify these vortices structures using the CSC method. In the current Kelvin–Helmholtz system, the vortices at time $t_0 = 12$ s are the product of smaller vortices having merged as the flow progressed over the preceding 12 seconds, and are not well

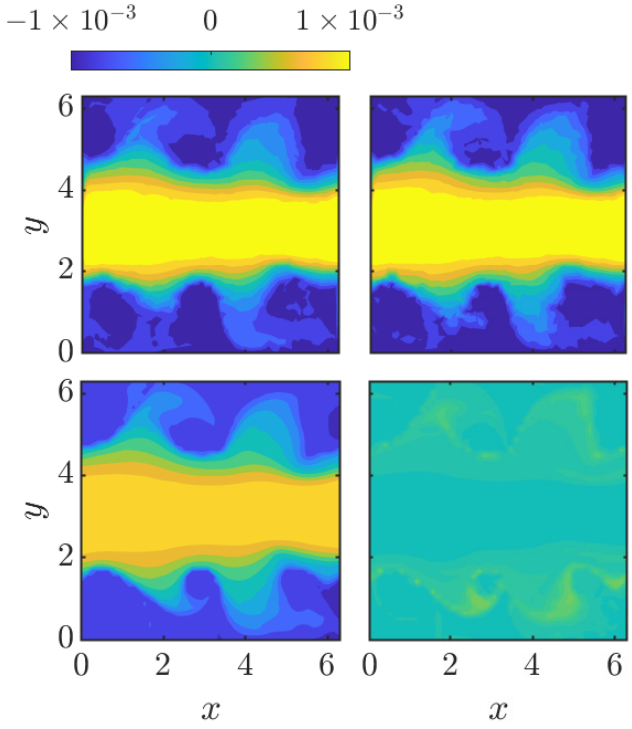

Figure 10: CSC diagnostics for the Kelvin–Helmholtz flow arranged in the same fashion as seen in Figure 1.

detected by the CSC method. (The LAVD, in contrast, is primed for explicit vorticity detection, and performs well in this endeavour as is seen in Figure 3.) The statistics from the 100 stochastic realisations (lower panels) confirms that the CSC is self-consistent to the influence of noise.

3.2. Gulf Stream Data

Next, we consider oceanographic velocity data representing the Gulf Stream, which is highly influential to climate patterns in the Northern and Western hemispheres [49]. We use absolute geostrophic velocity data compiled by SSALTO/DUACS and supplied by the AVISO Satellite Altimetry Data service (https://www.aviso.altimetry.fr) to perform LCS analysis on the Gulf Stream. We use data on $\Omega = [65°\text{W}, 35°\text{W}] \times [32°\text{N}, 48°\text{N}]$, at a spatial resolution of $h = 1/8$

degrees, and time ranging from t_0 = midnight, January 15th 2015 to $t_0 + T$ = midnight, April 15th 2015 (UTC) with a temporal resolution of 1 day, to be consistent with other studies [50]. We use the velocity uncertainty $v_r = 1.63 \times 10^{-2}\text{ms}^{-1}$ or roughly, 1.27×10^{-2} degrees latitude/longitude per day, corresponding to a value of $\epsilon = \sqrt{hv_r} \approx 4 \times 10^{-2}$ degrees per square root day for the SDE (6).

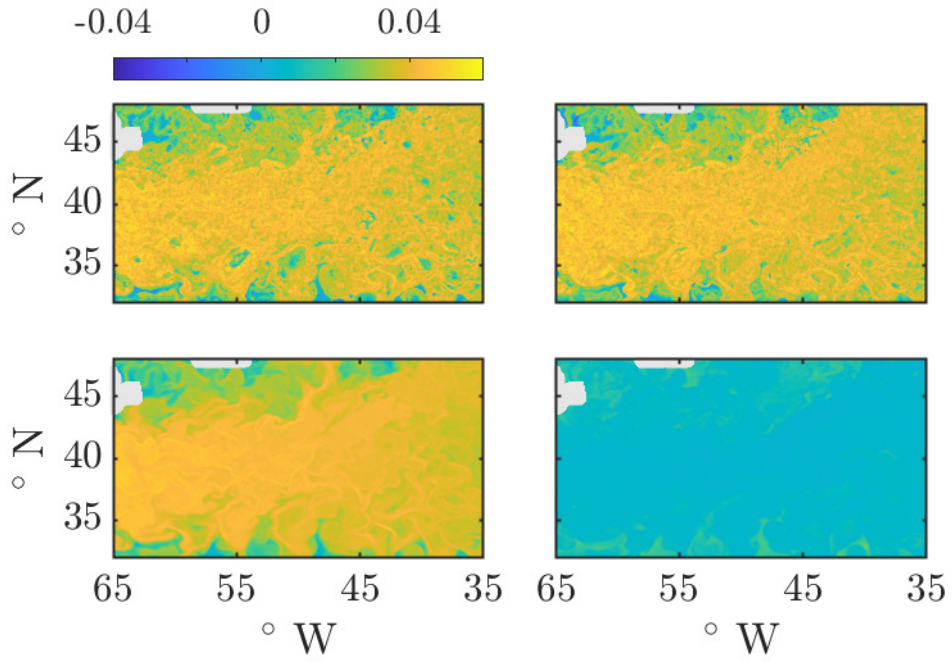

Figure 11: The deterministic FTLE of the Gulf Stream flow from t_0 = midnight, January 15, 2015 to $t_0 + T$ = midnight, April 15, 2015 (UTC) (top-left), one realisation of the FTLE with noise applied to the flow (top-right) and the mean (bottom-left) and standard deviation (bottom-right) of 100 realisations of the stochastic FTLE.

We begin in Figure 11 with the FTLE method, which once again shows self–consistency against uncertainty. We have included the deterministic FTLE (top left), one realisation of the FTLE generated following the application of stochastic noise to the data (top right), and the mean (bottom left) and standard deviation (bottom right) of 100 stochastic realisations of the FTLE (there is no

need to enhance the standard deviation field for improved visibility this time). The other Figures in this Section (almost always) conform to this ordering of panels. We demonstrate in the Supplementary Material that 100 realisations is sufficient to obtain convergence of these statistics across this, and all, LCS methods. At this scale, maximal ridges appear to cover most of the chosen domain. We observe smaller (and sometimes negative) values of the FTLE within coherent eddies produced by the Gulf Stream, along the Northern and Southern boundaries of the domain and within the Gulf of St Lawrence in the North West corner of the domain. Like with the Kelvin–Helmholtz system, the presence of stochastic noise causes the FTLE field to wobble slightly, changing the shapes of some of the maximal ridges. The flow patterns remain generally consistent in the stochastic simulation shown. The mean FTLE field appears to be a smoothened version of the deterministic field with most of the maximal ridges still readily identifiable and all of the coherent eddies observed from low FTLE values having been obscured. The standard deviation is reasonably low throughout the domain, with larger values observable around the Northern and Southern edges of the domain and closer to land in the North Western portion of the domain. This is likely due to the large exodus of particles from the domain over the 90 day flow period and a large number of particles "washing up" on land, which will change with every new stochastic flow map realisation. These larger standard deviation values outweigh those observable closer to the maximal ridges of the FTLE field, likely owing to the fact that most of the domain is dominated by large FTLE values and the error is therefore consistent.

Figure 12 indicates that the variational LCSs have once again been significantly compromised by the presence of noise. As for the Kelvin–Helmholtz system, hyp–VLCSs produced for the Gulf Stream completely deform under the influence of even a relatively small amount of stochastic noise. We produce a statistical analysis of the 100 simulations by dividing the Gulf Stream domain into $241 \times 159 = 38319$ equally sized bins (chosen after some experimentation with the number of bins defined). From the sample probability field, some consistent structures are identifiable by green coloured streaks within the field. Most

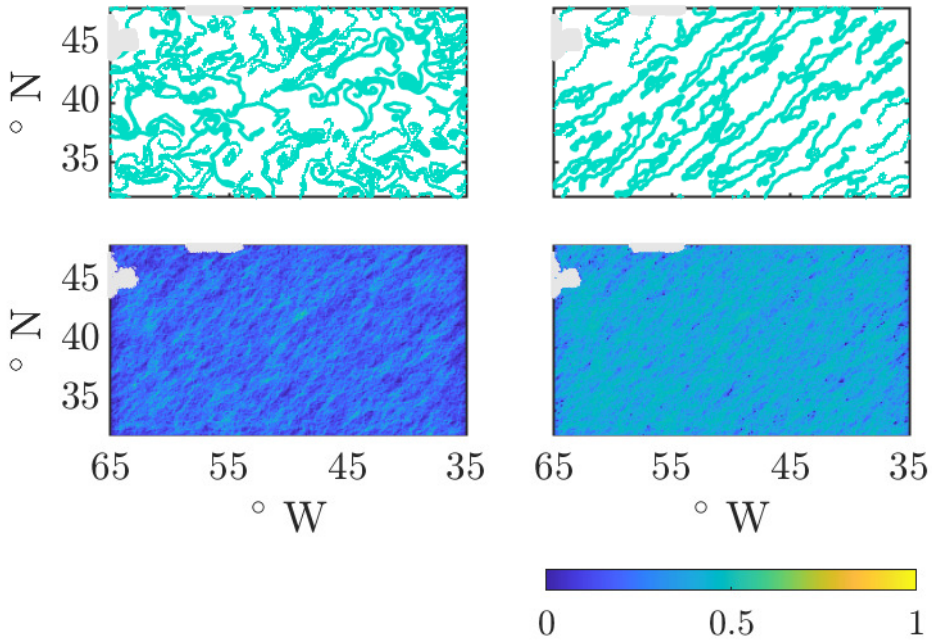

Figure 12: Deterministic hyp–VLCSs of the Gulf Stream flow from t_0 = midnight, January 15, 2015 to t_0+T = midnight, April 15, 2015 (UTC) (top–left), one realisation of these hyp–VLCSs with noise applied to the flow (top–right) and the likelihood (bottom–left) and the variation in this likelihood (bottom–right) of a hyp–VLCS passing through one of $241 \times 159 = 38319$ bins computed from 100 realisations of stochastic hyp–VLCSs.

of these streaks appear to look like diagonal lines, which is likely attributable again to the exodus of particles over the large period of flow considered. The fact that most of the bins in the likelihood Figure are coloured dark or light blue indicates that the spread of the hyp–VLCSs generated covers approximately the whole flow domain and therefore verifies the volatility of these structures under velocity uncertainty. This is further highlighted by the variation in these sample probabilities, which is consistently high over the entire domain. It is therefore difficult to think of the hyp-VLCSs computed from the deterministic flow as being reliable.

Figure 13 demonstrates the structural robustness of the LAVD method, con-

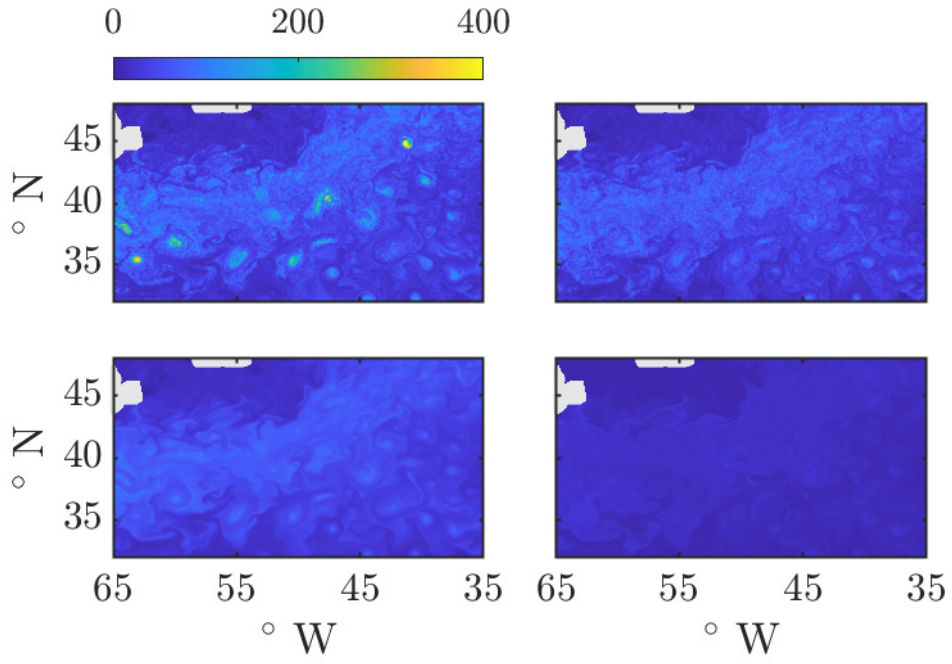

Figure 13: LAVD diagnostics for the Gulf Stream arranged in the same fashion as seen in Figure 11.

sistent with the Kelvin–Helmhotz results (see Figure 3). The vortices (Gulf Stream eddies) detected by the LAVD maintain their position under noise, but change shape slightly while being associated with changed LAVD values. The mean LAVD field appears to be a smoothened version of the deterministic LAVD field with extremal LAVD values reduced, while the standard deviation of the LAVD appears to be larger near the centres of key coherent eddies and within the Gulf Stream itself where large LAVD values are also observable.

Once again, the Lagrangian descriptor method shows structural robustness as shown in Figure 14. The mean is a diffused version of the deterministic Lagrangian descriptor, and the standard deviation is low.

The stochastic sensitivity method is mildly susceptible to noise, as shown in Figure 15. Once again, we plot $\ln(S_v)$ due to large variations in the raw

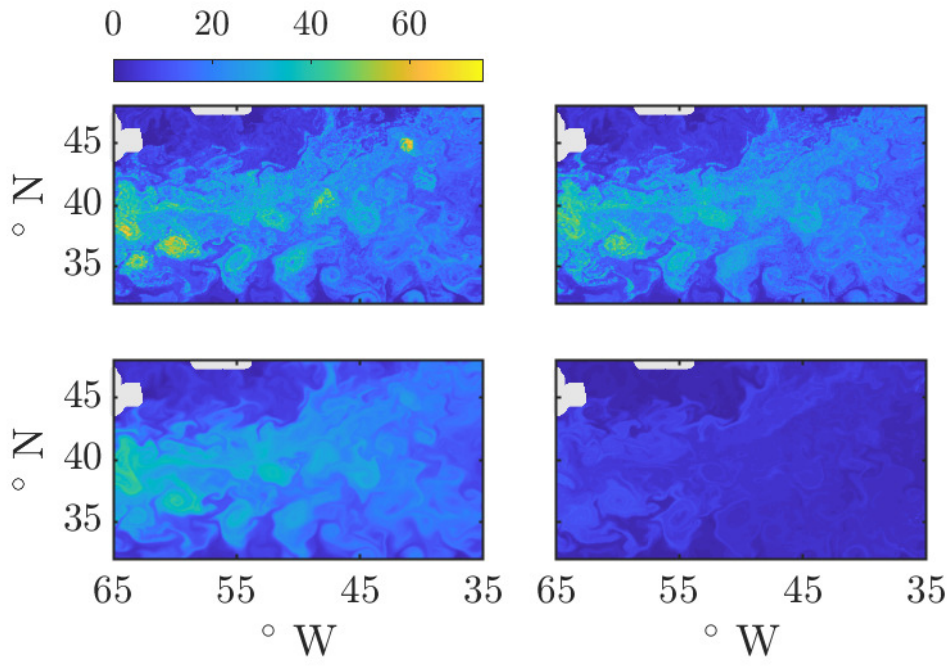

Figure 14: Lagrangian descriptor diagnostics for the Gulf Stream arranged in the same fashion as seen in Figure 11.

quantity S^2. The mean field from 100 realisations appears to be a smoothened out version of the deterministic field with a large portion of extremal values observable throughout the Gulf Stream. The standard deviation is once again large close to large values of S^2. We observe larger values of the mean and standard deviation in the western part of the region, possibly caused by the inevitable oversampling of initial conditions in this region (initial conditions further east are more likely to pass out of the domain because of the direction of flow, particularly when noise is added, and thus are discounted from the calculations). Hence, while this is chosen to produce comparable figures to published work [50], this particular choice of domain is not optimal for LCS computations for a 90 day flow.

On the other hand, the stochastic sensitivity robust sets are self–consistent

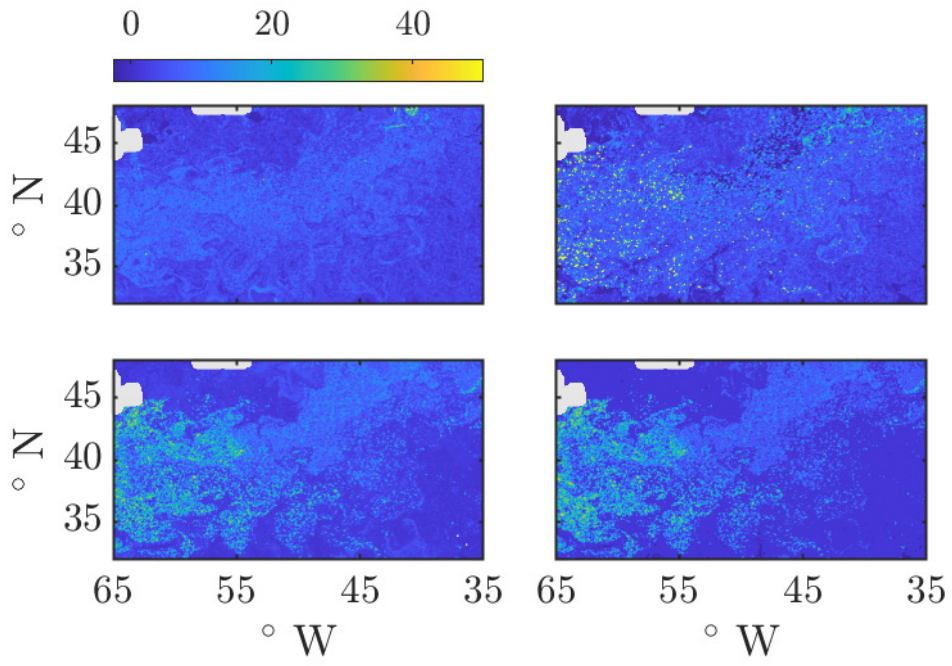

Figure 15: Diagnostics for the scaled stochastic sensitivity S_v for the Gulf Stream arranged in the same fashion as seen in Figure 11. The natural logarithm of S_v has been used to enhance the visibility of key structures within each contour plot.

under the influence of noise as indicated in Figure 16. We display the deterministic robust set R at the level of lengthscale uncertainty of 6 degrees (with $v_r = 1.27 \times 10^{-2}$ degrees per day and $h = 1/8$ degrees) in the left panel. We remark that the white regions in the left panel identify the Northeastern bending of the Gulf Stream much clearer than any of the other methods in this section (save perhaps the CSC). This is because the large stretching rates in the jet, coupled with the intense vorticity in the adjacent regions, renders the eventual fluid parcel locations much less predictable. The likelihood function from the 100 stochastic realisations (right panel) indicates that the robust set is more likely to contain portions of the flame shaped objects located along the Southern edge of the domain along with the inner portion of the Gulf of St Lawrence (as indicated by yellow and green shaded portions of the domain). Fluid parcels

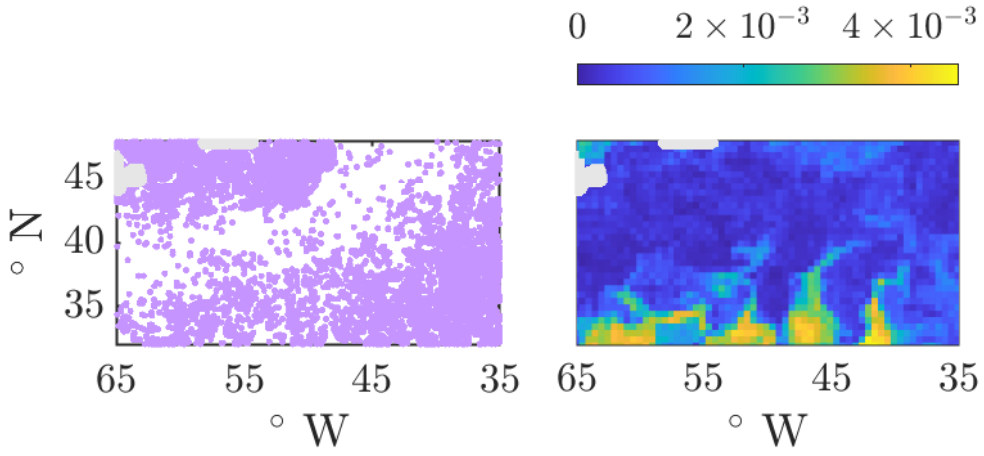

Figure 16: The robust set R at lengthscale 6 degrees for the Gulf Stream from t_0 = midnight, January 15, 2015 to t_0+T = midnight, April 15, 2015 (UTC) (left) and the sample probability density function of being in a robust set from 100 stochastic realisations (right).

beginning in these regions have the most predictable final locations, whereas regions in and adjacent to the jet are identified as *not* being in the robust set. The flame-shaped regions are consistent with low stretching rates, while the less robust regions are consistent with larger stretching and vorticity (see Figures 11) 13).

Figure 17 indicates that the transfer operator optimal vector is also reliable. Consistent with Froyland et al [50], we consider the third singular vector of the transfer operator and the third eigenvector of the dynamic Laplacian operator to find two coherent sets for this system. The claim is that the main jet region of the Gulf Stream is identifiable along the zero contour of the transfer operator optimal vector field, which is used to divide the flow domain Ω into two coherent sets [50]. (However, this produces an essentially East–West curve, in contrast to the Northeastern bending of the jet visible in Figure 16.) Consistent with our CFD application, applying noise to the Gulf Stream velocity causes fuzziness in the optimal vector field (see Figure 7), though in a more unique observation,

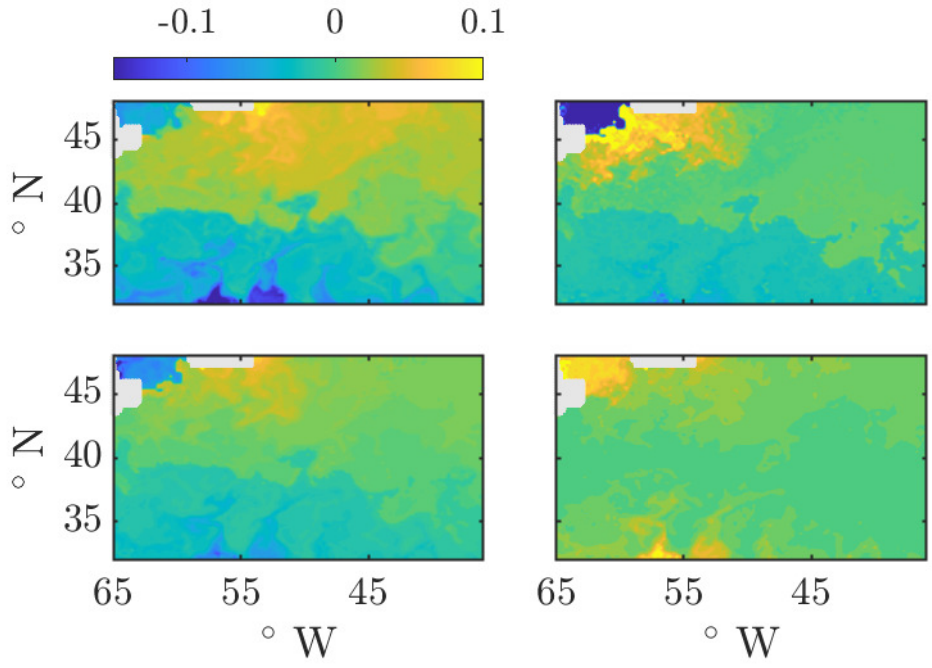

Figure 17: Transfer operator optimal vector diagnostics for the Gulf Stream arranged in the same fashion as seen in Figure 11.

entries of the optimal vector closer to land and the Northern and Southern edges of the domain, while maintaining their sign, have changed considerably in value. The standard deviation is reasonably low overall with larger quantities observable within the Gulf of St Lawrence and along the Northern and Southern edges of the domain. As with most of the previously discussed methods, this can be attributed to the exodus of particles from the domain over our 90 day flow period. The standard deviation around the zero contour lines of the optimal vector field remain relatively low in comparison, providing a sense of confidence and reliability in the results produced from this method.

The third eigenvector of the dynamic Laplacian also shows resistance to velocity uncertainty as seen in Figure 18. However, the size of the eigenvector is less extreme within the Gulf of St Lawrence and along the Northern and

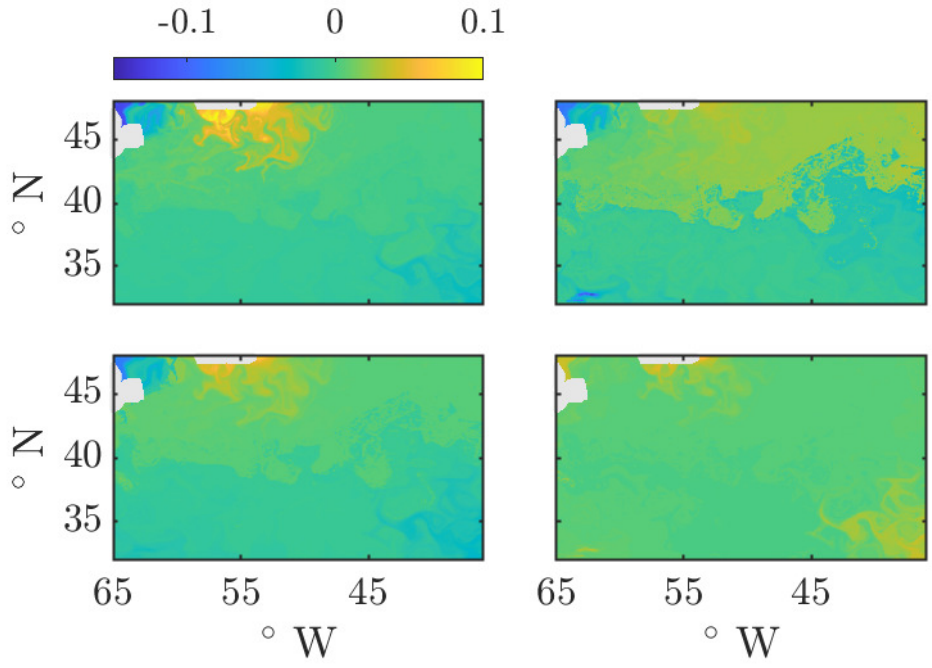

Figure 18: Dynamic Laplace operator third eigenvector diagnostics for the Gulf Stream arranged in the same fashion as seen in Figure 11.

Southern edges of the domain in comparison to the transfer operator. Applying noise results in only a minor change to the formation of the two sets, with noticeable fuzziness and a change in shape of the zero contour as we move Eastwards. Further West, no such changes in the zero contour can be observed. Most of the other observations we have made regarding the transfer operator method also hold in this case.

Figure 19 demonstrates that the FCM Clustering method is also self–consistent. This time we program this method to detect five clusters, and we note that each cluster does not have an obvious interpretation in terms of the jet or eddies. The method has essentially divided the domain Ω into partitions of approximately equal size (the same is true when we experimented with different numbers of clusters). That said, the results are self–consistent with noise having a relatively

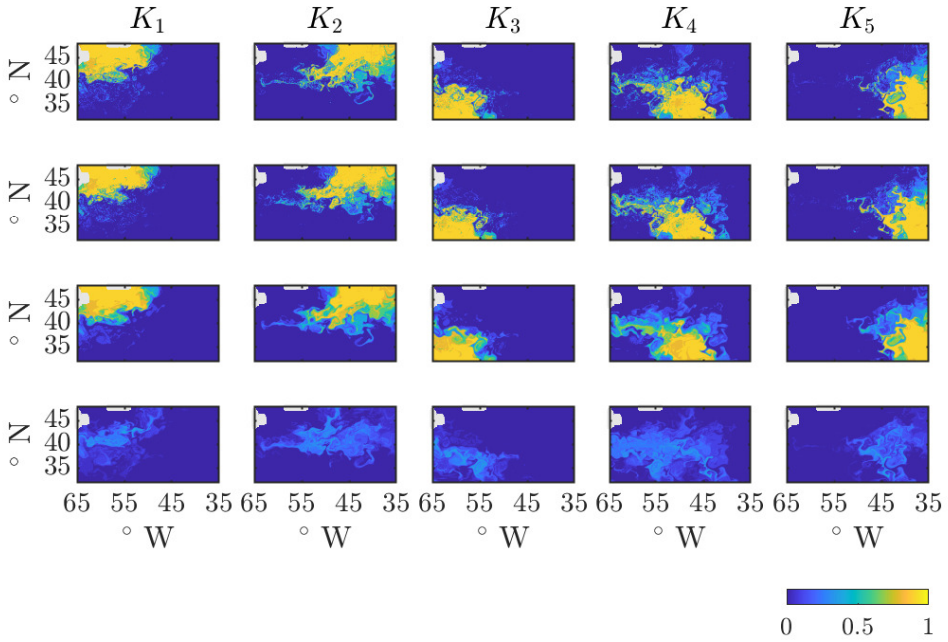

Figure 19: The deterministic membership probability fields for five FCM clusters K_1, K_2, K_3, K_4 and K_5 relevant to the Gulf Stream from t_0 = midnight, January 15, 2015 to $t_0 + T$ = midnight, April 15, 2015 (UTC) (first row); one realisation of each of these fields with noise applied to the flow (second row) and the mean (third row) and standard deviation (fourth row) of these probabilities generated from 100 realisations of the FCM clustering method.

minimal impact on the membership probability fields for each cluster, and the mean membership probabilities looking approximately like a smoothened out version of the deterministic results. The standard deviations of the realisations for each cluster are also relatively low, and are concentrated around the edges of each cluster which lie closer to the centre of the domain. The spatial spread of the standard deviation is considerably large (particularly for the second and fourth FCM clusters), however this again can be attributed to the added uncertainty stemming from the large length of the flow interval.

The functionality of the CSC method also appears reasonably unaffected by noise, as indicated in Figure 20. Structures corresponding to strongly positive

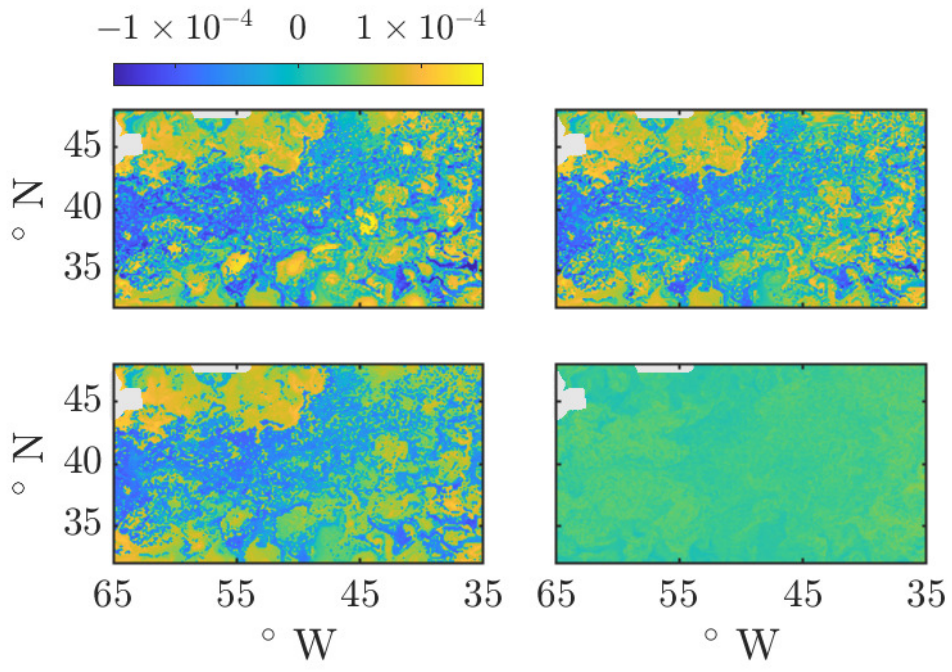

Figure 20: CSC diagnostics for the Gulf Stream arranged in the same fashion as seen in Figure 11.

CSC coefficients represent eddies and coherent parcels of flow closer to land or within the Gulf of St Lawrence, while strongly negative CSC coefficients help identify the Northeastern bending of the Gulf Stream jet, along with streaks of thin flow which separate coherent eddies. Adding noise does not have much visible impact, with the mean being comparable to the deterministic field, and standard deviation being small and uniform.

4. Concluding Remarks

LCS detection is important particularly in flows which possess both turbulent and coherent features, since detecting the influential coherent structures amidst complicated flow is difficult. We have examined in this paper Eulerian velocity data from two such flows, as opposed to from simpler flows (idealised

and/or nonturbulent) where the concept of LCS detection can be argued to be redundant. Ultimately, the functionality of any LCS detection method will be compromised to some degree by the presence of numerical uncertainty within the Eulerian velocity data. We have in this paper performed a systematic analysis of the impact of this uncertainty for a range of LCS methods.

A common observation for the LCS methods which produce fields on Ω is that the mean of the stochastic fields usually looks like a diffused version of the deterministic field. This implies that these fields exhibit self-consistency with respect to random perturbations in the Eulerian data. This is true across both our systems for the FTLE, S^2, LAVD, Lagrangian descriptor, CSC, transfer-operator, and dynamic Laplacian fields. The latter field is interesting in that the mean stochastic field appears to be much closer to the deterministic one than for the other methods, as is clear in Figure 8 in particular. The fact that the dynamic Laplacian includes diffusion within its definition likely contributes to this fact. While the fields appear diffused, the standard deviation plots indicate that the errors tend to not be spatially uniform. For the FTLE and S^2, there are large deviations in the fields in regions where the deterministic field is large. These observations are intuitively reasonable. Since a large FTLE implies a large exponential stretching rate for the deterministic flow, noise added to such a system will invariably be amplified far more at large FTLE values. Similarly, if there was a large expectation of uncertainty in the eventual location of a trajectory (i.e., if S^2 is large), inclusion of noise throughout the time of flow should exacerbate this uncertainty. Consequently, the self-consistency of the FTLE and S^2 fields with respect to noise is diminished at very large values. Put another way, if reported FTLE or S^2 values are very large from a deterministic calculation, then the actual values reported cannot be considered to be too reliable. Thus, codimension–1 entities such as maximal ridges of the FTLE field cannot be considered robust under Eulerian velocity uncertainty. However, it is not completely clear whether such ridges can be identified in a well-defined way, or whether it makes any sense to try to use ridges as proxies of stable/unstable manifolds in infinite-time flows [39, 51], because realistic data-driven flows are

of finite-time and possess different time-scales of motion. On the other hand, the FTLE and S^2 *fields* have meaning (respectively quantifying stretching [16] and unpredictability in final position [21]), and these remain robust in the sense that their stochastic mean is effectively a diffused version of the deterministic field.

It is worth emphasising that the methods which detect coherent flow *regions* rather than the *boundaries* which separate them are more robust under these conditions as co-dimension–0 entities would tend to simply wobble as opposed to co-dimension–1 entities which can disappear entirely. Thus, methods such as set-selection using the transfer or dynamic Laplacian operator, robust sets from stochastic sensitivity, local maxima locations of the LAVD to identify coherent vortices, sets via fuzzy c-means clustering, or sets with FTLE values less than a threshold to identify flow regions which have lower stretching, will in general be fairly robust to uncertainty in the Eulerian data.

Out of all of the LCS detection methods examined in this paper, the hyp–VLCS method has the most concerning performance as Figures 2 and 12 indicate. Some stochastic simulations (pictured in the top right panels of these figures) are *radically* different from the deterministic (top left) results. Our quantitative analysis on the statistics indicates that these simulations are not atypical, with the stochastic hyp–VLCS spreading out to cover almost the whole flow domain. There are few (if any) discernible curves when the mean of the stochastic realisations is plotted (bottom left panel), and the likelihood (bottom right) of finding hyp-vLCS curves is effectively constant everywhere. Indeed, the simulated curves tend to fill the entire domain with no gaps as the separation radius parameter is made smaller (see the Supplementary Material for more of a discussion related to this computational parameter). This method's high susceptibility to noise can be inferred to be the result of the many computational steps required: numerical differentiation of flow maps, integration of eigenvector fields and interpolation of various quantities required for regulation of the resultant solution curves. However, also (and possibly more) important is the selection method of curves from the foliation [17, 39, 18]. We recall that in the

LCS Tool algorithm [18] this is performed by first locating a point where the larger eigenvalue $\lambda_{\max}$ if the Cauchy–Green tensor (2) is maximum, finding the relevant curve through this point, and then excluding regions within a specified separation radius and continuing this process for the remaining part of Ω. An underappreciated fact is that the maximal $\lambda_{\max}$ location is *exactly* the maximal FTLE location, as is clear from (3). Thus, the FTLE comes into play in the variational LCS selection process, and as discussed previously the FTLE field at large values is particularly susceptible to uncertainty in the Eulerian velocity field. Consequently, the LCS Tool algorithm in seeking hyp-vLCSs inherits this uncertainty in addition to uncertainties in the remaining calculations. The ordered selection process of curves in a stochastic simulation can jump completely to a different region from the deterministic one, because the FTLE field's value (at these large values) has changed radically. Exclusion of a region (associated with the separation radius parameter) will further impact the next selection, and consequently the computed hyp-vLCSs (see Figures 2 and 12) can be markedly different. The inevitable uncertainty in Eulerian velocity data implies that the hyp-vLCS method using the LCS Tool algorithm is therefore highly non-robust.

The methods which have shown the greatest self–consistency (with respect to the *value* of the relevant field, in addition to shapes of objects) are the dynamic Laplacian operator, CSC, FCM Clustering and stochastic sensitivity robust sets. The LAVD's identification of vortical structures is self–consistent (though the values of the LAVD field change quite drastically). Indeed, essentially all the vortices visible in the deterministic implementation are also visible in the mean stochastic ones (see Figures 3 and 13). The relative errors (thought of as the ratio between the standard deviation and the mean) appears to be fairly uniform across the domain, indicating that factors such as excessive stretching in certain regions do not necessarily inordinately impact noise amplification. Likewise, the Lagrangian descriptor method shows good self–consistency, with only the identification of smaller vortices being susceptible to noise. While the FCM clustering method does show a great deal of self–consistency under the influence of additional velocity uncertainty, it must be noted that this method comes with

the significant drawback of the requirement to pre–determine the number of clusters anticipated to exist within a flow domain. The transfer operator optimal vector becomes noticeably more fuzzy when computed for noisy data though still manages to capture important structures. The dynamic Laplacian operator is able to divide a flow domain into two coherent sets without the boundaries between these sets becoming too fuzzy, as any uncertainties within the operator are ironed out by the second derivative finite differencing information embedded within the matrices used to produce this operator (we do observe some fuzziness in the Gulf Stream flow, see Figure 18, but this is owed more to the large interval of flow time rather than velocity uncertainty). The stochastic sensitivity robust sets are able to pinpoint the most consistent flow regions of a domain while taking into account a specific level of velocity uncertainty along with a lengthscale selected especially for a particular application. The CSC method identifies a wide variety of coherent objects present within a flow system from one relatively simple calculation, and these objects are seen to retain their shape remarkably well under the influence of stochastic velocity uncertainty.

Our analysis indicates that when applying LCS methods to realistic data situations, the results need to be interpreted with much caution because of the presence of uncertainty in the Eulerian velocity data. Irrespective of the argument as to whether a given LCS method identifies "coherent structures" or not, if the method is not robust to uncertainties in the Eulerian data, it cannot be considered reliable. We expect our results in this paper to help practitioners to choose an LCS method which is appropriate for their application, while being self-consistent.

Acknowledgements

This research is supported by an Australian Government Research Training Program Scholarship, and with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide. The authors acknowledge that the Gulf Stream data products were processed by SSALTO/DUACS

and distributed by AVISO+ (https://www.aviso.altimetry.fr) with support from CNES. The third author acknowledges with thanks partial support from the Australian Research Council via grant DP200101764. The authors are grateful to Oliver Junge for advice related to the numerical implementation of the dynamic Laplacian method and Gary Froyland for recommending an appropriate spatial domain and flow time interval for the Gulf Stream.

Supplementary Material

Please refer to the Supplementary Material for further details regarding the computational specifics of each LCS detection method and analysis of the convergence of stochastic LCS realisations to sufficient mean quantities.

Declaration of Interests

The authors report no conflict of interest.

References

[1] M. R. Allshouse, T. Peacock, Lagrangian based methods for coherent structure detection, Chaos: An Interdisciplinary Journal of Nonlinear Science 25 (9) (2015) 097617. doi:10.1063/1.4922968.

[2] A. Hadjighasem, M. Farazmand, D. Blazevski, G. Froyland, G. Haller, A critical comparison of Lagrangian methods for coherent structure detection, Chaos: An Interdisciplinary Journal of Nonlinear Science 27 (5) (2017) 053104. doi:10.1063/1.4982720.

[3] S. Balasuriya, N. T. Ouellette, I. I. Rypina, Generalized Lagrangian coherent structures, Physica D: Nonlinear Phenomena 372 (2018) 31–51. doi:10.1016/j.physd.2018.01.011.

[4] S. Shadden, Lagrangian coherent structures, in: R. Grigoriev (Ed.), Transport and Mixing in Laminar Flows: From Microfluidics to Oceanic Currents, Wiley, 2011.

[5] G. Haller, Lagrangian Coherent Structures, Annual Review of Fluid Mechanics 47 (2015) 137–162. doi:10.1146/annurev-fluid-010313-141322.

[6] P. J. Nolan, H. Foroutan, S. D. Ross, Pollution Transport Patterns Obtained Through Generalized Lagrangian Coherent Structures, Multidisciplinary Digital Publishing Institute: Atmosphere 11 (2) (2020) 168. doi:10.3390/atmos11020168.

[7] D. Schmale, S. Ross, High–Flying Microbes, Scientific American 316 (2) (2017) 40–45. doi:10.1038/scientificamerican0217-40.

[8] J. Bettencourt, V. Rossi, E. Hernández-García, M. Marta-Almeida, C. López, Characterization of the structure and cross-shore transport properties of a coastal upwelling filament using three-dimensional finite-size Lyapunov exponents, Journal of Geophysical Research: Oceans 122 (9) (2017) 7433–7448. doi:10.1002/2017JC012700.

[9] G. Froyland, R. M. Stuart, E. van Sebille, How well–connected is the surface of the global ocean?, Chaos: An Interdisciplinary Journal of Nonlinear Science 24 (3) (2014) 033126. doi:10.1063/1.4892530.

[10] D. Kelley, M. Allshouse, N. Ouellette, Lagrangian coherent structures separate dynamically distinct regions in fluid flows, Physical Review E 88 (2013) 013017. doi:10.1103/PhysRevE.88.013017.

[11] A. von Kameke, S. Kastens, S. Rüttinger, S. Herres-Pawlis, M. Schlüter, How coherent structures dominate the residence time in a bubble wake: An experimental example, Chemical Engineering Science 207 (2) (2019) 317–326. doi:10.1016/j.ces.2019.06.033.

[12] S. Gowen, T. Solomon, Experimental studies of coherent structures in an advection-reactiondiffusion system, Chaos: An Interdisciplinary Journal of Nonlinear Science 25 (8) (2015) 087403. doi:10.1063/1.4918594.

[13] S. Raben, S. Ross, P. Vlachos, Experimental determination of three-dimensional finite-time Lyapunov exponents in multi-component flows, Experiments in Fluids 55 (2014) 1824. doi:10.1007/s00348-014-1824-3.

[14] A. Farghadan, F. Coletti, A. Arzani, Topological analysis of particle transport in lung airways: Predicting particle source and destination, Computers in Biology and Medicine 115 (C) (2019) 103497. doi:10.1016/j.compbiomed.2019.103497.

[15] H.-Y. Cheng, X.-P. Long, B. Ji, Y. Zhu, J.-J. Zhou, Numerical investigation of unsteady cavitating turbulent flows around twisted hydrofoil from the Lagrangian viewpoint, Journal of Hydrodynamics 28 (4) (2016) 709–712. doi:10.1016/S1001-6058(16)60674-1.

[16] S. C. Shadden, F. Lekien, J. E. Marsden, Definition and properties of Lagrangian coherent structures from finite–time Lyapunov exponents in two–dimensional aperiodic flows, Physica D: Nonlinear Phenomena 212 (3–4) (2005) 271–304. doi:10.1016/j.physd.2005.10.007.

[17] M. Farazmand, G. Haller, Computing Lagrangian coherent structures from their variational theory, Chaos: An Interdisciplinary Journal of Nonlinear Science 22 (1) (2012) 013128. doi:10.1063/1.3690153.

[18] K. Onu, F. Huhn, G. Haller, LCS Tool: A computational platform for Lagrangian coherent structures, Journal of Computational Science 7 (2015) 26–36. doi:10.1016/j.jocs.2014.12.002.

[19] G. Haller, A. Hadjighasem, M. Farazmand, F. Huhn, Defining coherent vortices objectively from the vorticity, Journal of Fluid Mechanics 795 (2016) 136–173. doi:10.1017/jfm.2016.151.

[20] A. M. Mancho, S. Wiggins, J. Curbelo, C. Mendoza, Lagrangian descriptors: A method for revealing phase space structures of general time dependent dynamical systems, Communications in Nonlinear Science and Numer-

ical Simulation 18 (12) (2013) 3530–3557. doi:10.1016/j.cnsns.2013.05.002.

[21] S. Balasuriya, Stochastic Sensitivity: A Computable Lagrangian Uncertainty Measure for Unsteady Flows, SIAM Review 62 (4) (2020) 781–816. doi:10.1137/18M1222922.

[22] S. Balasuriya, Uncertainty in finite–time Lyapunov exponent computations, Journal of Computational Dynamics 7 (2) (2020) 313–337. doi:10.3934/jcd.2020013.

[23] G. Froyland, N. Santitissadeekorn, A. Monahan, Transport in time–dependent dynamical systems: Finite–time coherent sets, Chaos: An Interdisciplinary Journal of Nonlinear Science 20 (4) (2010) 043116. doi:10.1063/1.3502450.

[24] G. Froyland, Dynamic isoperimetry and the geometry of Lagrangian coherent structures, Nonlinearity 28 (10) (2015) 3587–3622. doi:10.1088/0951-7715/28/10/3587.

[25] G. Froyland, E. Kwok, A dynamic Laplacian for identifying Lagrangian coherent structures on weighted Riemannian manifolds, Journal of Nonlinear Science 30 (1) (2017) 1889–1971. doi:10.1007/s00332-017-9397-y.

[26] G. Froyland, K. Padberg-Gehle, A rough–and–ready cluster–based approach for extracting finite–time coherent sets from sparse and incomplete trajectory data, Chaos: An Interdisciplinary Journal of Nonlinear Science 25 (8) (2015) 087406. doi:10.1063/1.4926372.

[27] K. L. Schlueter-Kuck, J. O. Dabiri, Coherent structure colouring: identification of coherent structures from sparse data using graph theory, Journal of Fluid Mechanics 811 (2017) 468–486. doi:10.1017/jfm.2016.755.

[28] H. Guo, W. He, T. Peterka, H.-W. Shen, S. Collis, J. Helmus, Finite–Time Lyapunov Exponents and Lagrangian Coherent Structures in Uncer-

tain Unsteady Flows, IEEE Transactions on Visualization and Computer Graphics 22 (6) (2016) 1672–1682. doi:10.1109/TVCG.2016.2534560.

[29] F. Balibrea-Iniesta, C. Lopesino, S. Wiggins, A. M. Mancho, Lagrangian Descriptors for Stochastic Differential Equations: A Tool for Revealing the Phase Portrait of Stochastic Dynamical Systems, International Journal of Bifurcation and Chaos 26 (13) (2016) 1630036. doi:10.1142/S0218127416300366.

[30] A. BorzorgMagham, S. Ross, D. S. III, Real-time prediction of atmospheric Lagrangian coherent structures based on forecast data: An application and error analysis, Physica D: Nonlinear Phenomena 258 (2013) 47–60.

[31] P. Lermusiaux, Uncertainty estimation and prediction for interdisciplinary ocean dynamics, Journal of Computational Physics 217 (2006) 176–199.

[32] A. Olcay, T. Pottebaum, P. Krueger, Sensitivity of Lagrangian coherent structure identification to flow field resolution and random errors, Chaos: An Interdisciplinary Journal of Nonlinear Science 20 (2010) 017506.

[33] D. Garaboa-Paz, J. Eiras-Barca, V. Pérez–Muñuzuri, Climatology of Lyapunov exponents: the link between atmospheric rivers and large–scale mixing variability, Earth System Dynamics 8 (2017) 865–873. doi:10.5194/esd-8-865-2017.

[34] M. Rockwood, Y. Huang, M. Green, Tracking coherent structures in massively–separated and turbulent flows, Physical Review Fluids 3 (1) (2018) 014702. doi:10.1103/PhysRevFluids.3.014702.

[35] M. Leclair, R. Lowe, Z. Zhang, G. Ivey, T. Peacock, Uncovering Fine–Scale Wave–Driven Transport Features in a Fringing Coral Reef System via Lagrangian Coherent Structures, Fluids 5 (4) (2020) 190. doi:10.3390/fluids5040190.

[36] H. Lin, Y. Xiang, S. Qin, H. Xu, H. Liu, Lagrangian analysis of the fluid transport induced by the interaction of two co–axial co–rotating vortex

rings, Journal of Hydrodynamics 32 (6) (2020) 1080–1090. doi:10.1007/s42241-020-0074-5.

[37] K. Suara, M. Khanarmuei, A. Ghosh, Y. Yu, H. Zhang, T. Soomere, R. J. Brown, Material and debris transport patterns in Moreton Bay, Australia: The influence of Lagrangian coherent structures, Science of the Total Environment 721 (3) (2020) 137715. doi:10.1016/j.scitotenv.2020.137715.

[38] C. Teeraratkul, Z. Irwin, S. C. Shadden, D. Mukherjee, Computational investigation of blood flow and flow–mediated transport in arterial thrombus neighborhood, Biomechanics and Modeling in Mechanobiology 20 (2) (2021) 701–715. doi:10.1007/s10237-020-01411-7.

[39] G. Haller, A variational theory of hyperbolic Lagrangian coherent structures, Physica D: Nonlinear Phenomena 240 (7) (2011) 574–598. doi:10.1016/j.physd.2010.11.010.

[40] G. García-Sánchez, A. M. Mancho, S. Wiggins, A bridge between invariant dynamical structures and uncertainty quantification, Communications in Nonlinear Science and Numerical Simulation 104 (2022) 106016. doi:10.1016/j.cnsns.2021.106016.

[41] G. Froyland, An analytic framework for identifying finite–time coherent sets in time–dependent dynamical systems, Physica D: Nonlinear Phenomena 250 (2013) 1–19. doi:10.1016/j.physd.2013.01.013.

[42] J. C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c–means clustering algorithm, Computers and Geosciences 10 (2–3) (1984) 191–203. doi:10.1016/0098-3004(84)90020-7.

[43] J. C. Bezdek, R. J. Hathaway, M. J. Sabin, W. T. Tucker, Convergence theory for fuzzy c–means: Counterexamples and repairs, IEEE Transactions on Systems, Man, and Cybernetics 17 (5) (1987) 873–877. doi:10.1109/TSMC.1987.6499296.

[44] M. Lesieur, C. Staquet, P. Le Roy, P. Comte, The mixing layer and its coherence examined from the point of view of two–dimensional turbulence, Journal of Fluid Mechanics 192 (1988) 511–534. doi:10.1017/S002211208800196X.

[45] R. W. Metcalfe, S. A. Orszag, M. E. Brachet, S. Menon, J. J. Riley, Secondary instability of a temporally growing mixing layer, Journal of Fluid Mechanics 184 (1987) 207–243. doi:10.1017/S0022112087002866.

[46] H. G. Lee, J. Kim, Two–dimensional Kelvin–Helmholtz instabilities of multi–component fluids, European Journal of Mechanics B/Fluids 49 (A) (2015) 77–88. doi:10.1016/j.euromechflu.2014.08.001.

[47] P. W. Schroeder, V. John, P. L. Lederer, C. Lehrenfeld, G. Lube, J. Schöberl, On reference solutions and the sensitivity of the 2D Kelvin–Helmholtz instability problem, Computers and Mathematics with Applications 77 (4) (2019) 1010–1028. doi:10.1016/j.camwa.2018.10.030.

[48] T. W. Mattner, Large–eddy simulations of turbulent mixing layers using the stretched-vortex model, Journal of Fluid Mechanics 671 (2011) 507–534. doi:10.1017/S002211201000580X.

[49] Y. Liu, C. Wilson, M. A. Green, C. W. Hughes, Gulf stream transport and mixing processes via coherent structure dynamics, Journal of Geophysical Research: Oceans 123 (4) (2018) 3014–3037. doi:10.1002/2017JC013390.

[50] G. Froyland, C. P. Rock, K. Sakellariou, Sparse eigenbasis approximation: Multiple feature extraction across spatiotemporal scales with application to coherent set identification, Communications in Nonlinear Science and Numerical Simulation 77 (8) (2019) 81–107. doi:10.1016/j.cnsns.2019.04.012.

[51] S. Balasuriya, R. Kalampattel, N. Ouellette, Hyperbolic neighbourhoods as organizers of finite-time exponential stretching, Journal of Fluid Mechanics 807 (2016) 509–545.