J. Fluid Mech. (2017), *vol.* 811, *pp.* 468–486.

doi:10.1017/jfm.2016.755

Coherent structure colouring: identification of coherent structures from sparse data using graph theory

Kristy L. Schlueter-Kuck[1] **and John O. Dabiri**[2,†]

[1]Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, USA

[2]Department of Mechanical Engineering and Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA

(Received 3 May 2016; revised 4 October 2016; accepted 6 November 2016; first published online 13 December 2016)

We present a frame-invariant method for detecting coherent structures from Lagrangian flow trajectories that can be sparse in number, as is the case in many fluid mechanics applications of practical interest. The method, based on principles used in graph colouring and spectral graph drawing algorithms, examines a measure of the kinematic dissimilarity of all pairs of fluid trajectories, measured either experimentally, e.g. using particle tracking velocimetry, or numerically, by advecting fluid particles in the Eulerian velocity field. Coherence is assigned to groups of particles whose kinematics remain similar throughout the time interval for which trajectory data are available, regardless of their physical proximity to one another. Through the use of several analytical and experimental validation cases, this algorithm is shown to robustly detect coherent structures using significantly less flow data than are required by existing spectral graph theory methods.

Key words: general fluid mechanics, low-dimensional models, nonlinear dynamical systems

1. Introduction

The concept of coherence in fluid flows has historically been used to delineate packets of fluid elements that persist while the flow evolves without significant mixing with the surrounding fluid regions. Coherence can frequently be visualized qualitatively by observing the evolution of passive tracers in a flow (e.g. Huhn *et al.* 2012; Haller 2015). However, mathematical frameworks are needed to quantify such structures objectively. Eulerian techniques for coherent structure identification include the q-criterion (Hunt, Wray & Moin 1988), λ_2-criterion (Jeong & Hussain 1995) and the Okubo–Weiss parameter (Okubo 1970; Weiss 1991). All of these methods are frame-dependent, however. Frame invariance is an important characteristic of a method for determining coherent structures. If a method identifies a structure boundary in one frame of reference, but not in another (for example, in a rotating reference frame),

† Email address for correspondence: jodabiri@stanford.edu

then the method may not be self-consistent in its characterization of fluid coherence (Haller 2005).

As an alternative, Lagrangian techniques have also been developed, many based on analysis of the deformation gradient tensor of the flow field (Haller & Yuan 2000; Shadden, Lekien & Marsden 2005). These methods use information regarding the trajectories of fluid particles, as opposed to their velocity or acceleration, which ensures frame invariance since relative particle position does not depend on the reference frame in which it is measured.

Despite the vast array of current applications, identification of coherent structures based on the deformation gradient has some limitations as well. For instance, knowledge of a discretized version of the entire flow field is typically required, of the sort obtained from computational analysis or particle image velocimetry (PIV). However, some common empirical tools for fluid flow measurement do not provide velocity data in the whole flow field. One such technique, particle tracking velocimetry (PTV) (e.g. Chang, Wilcox & Tatterson 1984; Racca & Dewey 1988), is particularly useful in applications where velocity data in three dimensions are required, or where the entire flow cannot be densely and uniformly seeded, as in studies of ocean currents. Particle tracking algorithms often result in much sparser velocity measurements than techniques such as PIV. For example, Davis (1991) reviews a wide range of studies utilizing artificial ocean drifters to study nominally two-dimensional ocean surface flows, where the number of drifters ranges from 14 to 300 per study. In three dimensions, a number of PTV studies have utilized between 800 and 5000 particle trajectories (Virant & Dracos 1997; Lüthi, Tsinober & Kinzelbach 2005; Murai *et al.* 2007; Kim, Hussain & Gharib 2013). Recently, several PIV and PTV techniques have been developed to increase the density of data collected (Elsinga *et al.* 2006; Schanz, Gesemann & Schröder 2016). Despite these advances, in many situations seeding the flow with a large number of particles can be extremely difficult, e.g. *in situ* measurements of atmospheric and oceanic flows that rely on naturally occurring particulate fields (Katija & Dabiri 2008; Ho *et al.* 2014). Hence, the need for techniques to analyse sparse data persists.

Deformation-gradient-based methods for identification of coherent structures fail for sparse trajectory data due to the assumptions inherent in analysing the deformation gradient tensor, $\partial \boldsymbol{x}/\partial \boldsymbol{X}$, where $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{X})$ maps the initial location of a fluid element, $\boldsymbol{X}$, to its location $\boldsymbol{x}$ at a later time. The principal assumption that is no longer satisfied is the initial close separation of flow trajectories, since the trajectory spacing cannot be controlled *a priori*. Moreover, the determination of the finite-time Lyapunov exponent (FTLE) field requires linearization of the flow map (Haller & Yuan 2000), which also breaks down for well-spaced flow trajectories. Therefore, an alternative approach is needed to extract coherent structures from sparse velocity data.

Several methods have been developed recently in an attempt to address this issue, a number of which are reviewed by Allshouse & Peacock (2015). One such method is based on braid theory (Allshouse & Thiffeault 2012), which maps two-dimensional fluid particle trajectories into three dimensions, where time is the third dimension. Plotted in this way, the entwinement of trajectories with each other can be analysed, and surfaces surrounding sets of trajectories that do not grow with time can be identified, indicating the presence of coherent structures. While the braid method is useful for two-dimensional datasets, its extension to higher spatial dimensions has not yet been achieved. Additionally, the braid-based analysis can become quite complex and computationally expensive for large numbers of trajectories.

A second method developed for use with sparse datasets is the cluster-based approach by Froyland & Padberg-Gehle (2015), and it is well suited for PTV datasets

due to its ability to handle both sparse and incomplete fluid particle trajectories. The method uses the Euclidean distance from each particle to the centre of each of a predetermined number of clusters to assign to each fluid particle a probability of belonging to each cluster while simultaneously determining the location of the cluster centres. This is accomplished using the iterative fuzzy c-means algorithm developed for use in cluster theory (Bezdek, Ehrlich & Full 1984). While the authors proved that this method can accurately detect coherent structures from a variety of benchmark flows, in addition to global ocean drifter data, it has the distinct disadvantage that the number of clusters be known *a priori*.

In order to address the need to determine the number of clusters *a priori* in the cluster-based approach of Froyland & Padberg-Gehle (2015), Hadjighasem *et al.* (2016) recently developed a method based on spectral graph theory. This method relies on the concept of a graph, or a set of nodes connected by edges, where, in this case, the nodes represent Lagrangian particles, and the edges are weighted by the inverse of the average distance between particle pairs. By examining the smallest-magnitude eigenvalues associated with the generalized eigenvalue problem of the graph Laplacian, the authors developed a heuristic for determining the number of coherent structures in the flow. The authors use this information as input to the K-means clustering approach to determine the centres of the coherent structures in the flow. The test cases used by the authors in validating this approach demonstrate its effectiveness in identifying coherent structures without knowing the number of clusters *a priori*. However, for most flows analysed, the number of trajectories used, of the order of tens of thousands, far exceeds the number of trajectories available for most PTV and ocean drifter datasets. It is also important to note that the method for determining the number of coherent structures is heuristic and therefore difficult to generalize. Other graph-theory-based methods have also been developed recently to address the issues associated with current cluster- and braid-based approaches (Banisch & Koltai 2016).

We propose an alternative graph-theory-based metric that weights the edges not by the distance between corresponding particles, but by a metric of kinematic dissimilarity, regardless of spatial proximity. This method is frame-invariant because it does not consider particle velocity, only the spatial location of each fluid particle relative to other particles in the flow. In analogy to graph colouring algorithms that partition nodes with large connecting edge weights, the present method solves an eigenvalue problem to partition fluid particle trajectories according to their kinematic dissimilarity. This approach can be considered an application of spectral graph drawing, which uses eigenvectors of matrices associated with a graph to visualize certain characteristics of the graph (Koren 2005). The present method results in a coherent structure colouring (CSC) field, where similar values of CSC indicate regions of the flow that are coherent, according to the present definition. In this way, all coherent structures in the flow can be identified simultaneously, without prior knowledge of their number. Methods for extracting individual coherent structures from the CSC field are discussed. The method was tested using three validation cases, including two canonical analytic flows (i.e. an oscillating quadruple gyre and a Bickley jet) and one experimental dataset (i.e. a two-dimensional cross-section of a high-stroke-ratio vortex ring).

The following section details the mathematical derivation of the algorithm used for coherent structure identification. Section 3 presents the results of the three test cases described above and characterizes the sensitivity of the method to certain parameters, including the number and initial location of the particles. Section 4

compares the results of the CSC method to the results of other graph-theory-based methods. Section 5 summarizes the results of the study and provides avenues for future development.

2. Methods

Coherence, defined here as the kinematic similarity of Lagrangian fluid trajectories, regardless of their spatial proximity, can be identified in flows with arbitrary time dependence using a graph-theory-based approach. The graph G is defined as the superset $G = (V, E, W)$, where V represents the set of nodes in the graph, E is the set of edges connecting the nodes, and W are the weights corresponding to the edge set. Assuming that the trajectories of a set of N Lagrangian fluid particles are known at T time steps, a graph representing the flow can be constructed, wherein each node represents a fluid particle. Unlike previous methods that weight the edges of such a graph based on the proximity of the fluid trajectories (e.g. using Euclidian distance), here we use a weight based on kinematic dissimilarity. We hypothesize, and later demonstrate, that coherent structures can be identified more robustly by quantifying the extent to which fluid trajectory kinematics are different, rather than the extent to which fluid particle trajectories remain in proximity over time. To this end, each edge, representing the connection between a pair of particles, is weighted by the standard deviation of the distance between the two fluid trajectories over their duration, normalized by the average distance between the fluid particle trajectories during the same period. The edge weights can be represented numerically using the weighted adjacency matrix $\boldsymbol{A}$, where a_{ij} contains the weight of the edge connecting particle i and particle j,

$$a_{ij} = \frac{1}{\overline{r_{ij}} T^{1/2}} \left[\sum_{k=0}^{T-1} (\overline{r_{ij}} - r_{ij}(t_k))^2 \right]^{1/2}, \tag{2.1}$$

where $r_{ij}(t_k)$ is the distance between two particles i and j at time t_k, and $\overline{r_{ij}}$ is the average distance between the two fluid particle trajectories.

Graph colouring is a labelling of nodes in a graph such that node pairs with large edge weights are assigned dissimilar values (Muñoz *et al.* 2005). This makes graph colouring a natural tool for coherent structure identification based on the kinematic dissimilarity metric in (2.1). We pose this as the one-dimensional problem of maximizing

$$z = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)^2 a_{ij}, \tag{2.2}$$

where N is the number of rows and columns in the weighted adjacency matrix $\boldsymbol{A}$ (i.e. the number of particles) and $\boldsymbol{X}$ is a row vector containing the value of CSC associated with each particle. By maximizing z we are in effect determining CSC values such that fluid particle trajectories that are kinematically dissimilar (i.e. where the weight of the edge between them a_{ij} is large) are assigned CSC values that are as different as possible. Following Hall (1970), we define the degree matrix $\boldsymbol{D}$, which contains the row sums of the adjacency matrix along the diagonal, as

$$d_{ij} = \begin{cases} 0, & i \neq j, \\ \displaystyle\sum_{k=1}^{N} a_{ik}, & i = j. \end{cases} \tag{2.3}$$

We also define the graph Laplacian, $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$. The maximization problem can then be manipulated as follows:

$$z = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - x_j)^2 a_{ij} \tag{2.4}$$

$$= \frac{1}{2}\left(\sum_{i=1}^{N} x_i^2 \sum_{k=1}^{N} a_{ik} - 2\sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j a_{ij} + \sum_{j=1}^{N} x_j^2 \sum_{m=1}^{N} a_{mj}\right) \tag{2.5}$$

$$= \sum_{i=1}^{N} x_i^2 \sum_{k=1}^{N} a_{ik} - \sum_{j=1}^{N}\sum_{i=1}^{N} x_i x_j a_{ij} \tag{2.6}$$

$$= \boldsymbol{X}'\boldsymbol{L}\boldsymbol{X}. \tag{2.7}$$

In order to avoid the trivial case where $x_1 = x_2 = \cdots = x_N = 0$, and to bound $\boldsymbol{X}$ to finite values, the constraint $\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} = 1$ is imposed (another finite constraint can be imposed without loss of generality). Given this constraint, the maximization problem can be written in Lagrangian form as $z = \boldsymbol{X}'\boldsymbol{L}\boldsymbol{X} - \lambda(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} - 1)$. As a necessary condition for z to be a local maximum, $\mathrm{d}z/\mathrm{d}\boldsymbol{X} = 0$, yielding $2\boldsymbol{L}\boldsymbol{X} - 2\lambda\boldsymbol{D}\boldsymbol{X} = 0$, which can be written as

$$\boldsymbol{L}\boldsymbol{X} = \lambda\boldsymbol{D}\boldsymbol{X}, \tag{2.8}$$

which is the generalized eigenvalue problem for the graph Laplacian. The generalized eigenvector $\boldsymbol{X}$ that maximizes z is the eigenvector corresponding to the maximum eigenvalue of (2.8) (Hall 1970). Each element of $\boldsymbol{X}$ assigns that value of CSC to the corresponding fluid particle at the final time of the interval over which particle trajectories were compared. The CSC vector can be mapped to the space of the original flow with arbitrary dimensionality based on the spatial locations of the particles. Interpolation between the particles facilitates construction of the corresponding CSC field. Thus, regions in the flow with a similar value of CSC indicate coherence.

3. Results

The effectiveness of the CSC algorithm is demonstrated using three example flows. The first, a quadruple gyre, is an extension of the double gyre, which is frequently used in vortex detection algorithm verification (Allshouse & Peacock 2015; Froyland & Padberg-Gehle 2015). Both the steady and the unsteady cases are examined. The second verification case is the Bickley jet, which introduces complexities due to the presence of multiple coherent vortices as well as a meandering jet. Finally, we apply the CSC method to sparse trajectories derived from a PIV dataset of a long-stroke-ratio vortex ring, where secondary and tertiary rings in addition to a trailing jet form behind the primary vortex ring. This shows the robustness of the algorithm to errors associated with experimental data. For the two analytical validation cases, a fifth-order Runge–Kutta method was used to determine fluid trajectories. For the PIV dataset, particle velocities were determined using linear interpolations between velocity vectors and particles were advected using the Euler method.

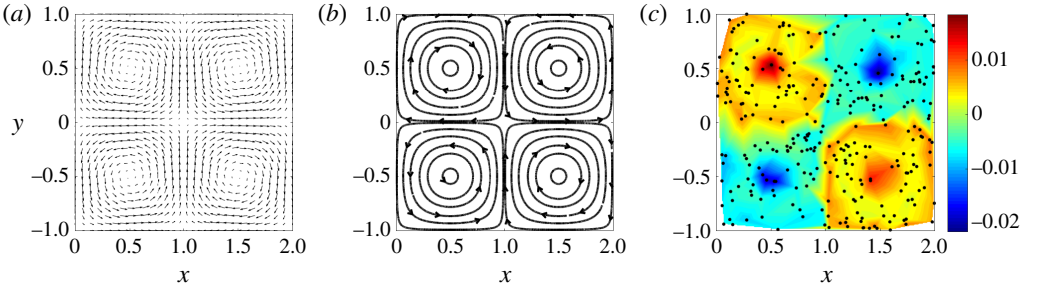

FIGURE 1. Steady quadruple gyre flow: (*a*) velocity vector field, (*b*) streamlines and (*c*) CSC using 300 particles, particle locations indicated by black dots.

3.1. *Quadruple gyre*

First, we examine the characteristics of the CSC algorithm using the analytical quadruple gyre flow. This flow is defined by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\pi A \sin(\pi f)\cos(\pi y), \tag{3.1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\pi A \cos(\pi f)\cos(\pi y)(2ax + b), \tag{3.2}$$

where x and y are the spatial coordinates, t is time and

$$a = \epsilon \sin(\omega t), \quad b = 1 - 2\epsilon \sin(\omega t), \quad f = ax^2 + bx. \tag{3.3a–c}$$

Here we examine two parameter cases: the steady case where $A = 0.1$ and $\epsilon = 0$; and the unsteady case where $A = 0.1$, $\epsilon = 0.1$ and $\omega = 2\pi/10$. Figure 1 shows the velocity vector field, streamlines and CSC for the steady quadruple gyre, tracking only 300 particles over 40 time units. CSC is able to clearly delineate the four quadrants of the flow, and assigns a high value to the gyre centres. Gyres in opposite corners have approximately the same value of colouring, due to their identical rotational orientation (clockwise in the upper left and lower right, and anticlockwise in the upper right and lower left quadrants). This is a result of having a measure of coherence that does not conflate kinematic similarity and physical proximity. The latter does not necessarily imply the former, and *vice versa*. Large weights in the present adjacency matrix correspond to fluid particles that are kinematically dissimilar, and given that the weights correspond to the standard deviation of the distance between two particles divided by their average distance, the mean distance between particles and the standard deviation in their distances both contribute to coherence as defined by this algorithm.

The flow becomes significantly more complex when periodic oscillatory unsteadiness is introduced. When comparing the coherence colouring to the FTLE of the same flow computed using 65 000 particles, seen in figure 2(*c*), it is clear that the largest positive value of coherence colouring correctly identifies all four vortex cores. The negative values of CSC correspond to the regions in which particles have switched quadrants due to the oscillatory nature of the flow, as indicated by the red dots in figure 2(*a*). In this case, the largest kinematic dissimilarity in the flow is between

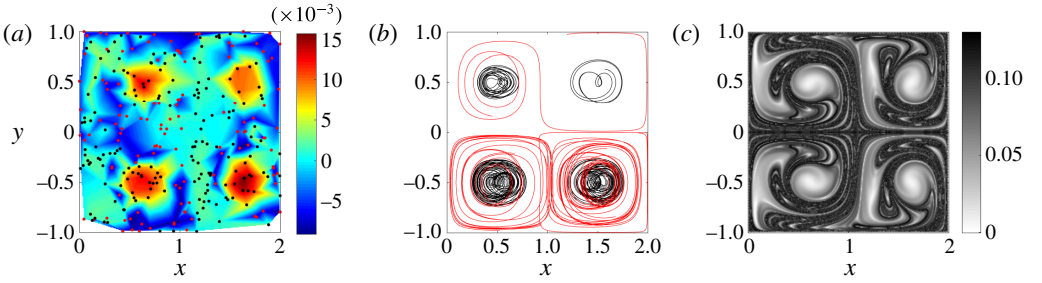

FIGURE 2. Unsteady quadruple gyre, $\epsilon = 0.1$, $A = 0.1$ and $T = [2.5, 42.5]$. (*a*) CSC using 300 particles; black dots show final locations of particles that remained in their initial quadrant, and red dots show final locations of particles that switched quadrants during the coherent structure calculation time interval. (*b*) Fluid trajectories for particles with highest (black) and lowest (red) CSC values. (*c*) FTLE field, calculated over the time interval $T = [2.5, 42.5]$, using 65 000 particles.

those particles that remain near the centre of the quadrant in which they started, and those particles that switch quadrants. This is highlighted by the particle traces shown in figure 2(*b*), which shows the trajectories of the particles with the largest positive value of colouring (in black) and those with the largest negative value of colouring (in red). This result is in contrast to the steady case, in which the sign of vortex rotation was the predominant distinguishing feature. Notably, the CSC algorithm can be applied recursively to the subset of particles with similarly high values of CSC in figure 2(*a*), in order to recover the vortex orientation information in figure 1(*c*). This is demonstrated in figure 3. Here we see that, for an initial CSC field calculated with 3000 particles, all four gyre centres are identified with a high CSC value. In order to detect more subtle differences between the four gyre cores, a threshold value of CSC > 0.0009 is set, indicated by the solid black lines surrounding the gyre cores in figure 3(*a*). When the CSC algorithm is applied using only the particles exceeding this threshold value, the information distinguishing the gyres that rotate clockwise from those that rotate anticlockwise (as in the steady quadruple gyre case) is recovered.

For this dataset, a cluster-based method using fuzzy *c*-means clustering would correctly identify the four quadrants of the steady quadruple gyre as four separate coherent structures, if it is assumed *a priori* that there are four clusters present. For the unsteady case, assuming the presence of four structures, the four gyre cores would again be correctly identified using this method. However, if more than four coherent structures are assumed, the clustering method will detect the four cores and a number of additional structures, some of which may correspond to fluid parcels that did not switch quadrants but are not adjacent to the gyre cores (Allshouse & Peacock 2015). A braid-based analysis for this flow would probably identify eight structures, again assuming an extension of the results of the double gyre system in Allshouse & Peacock (2015). In summary, the cluster-based method requires *a priori* knowledge of the number of structures present, and the braid-based analysis cannot be easily extended to three dimensions and is computationally expensive. The CSC method addresses all of these issues.

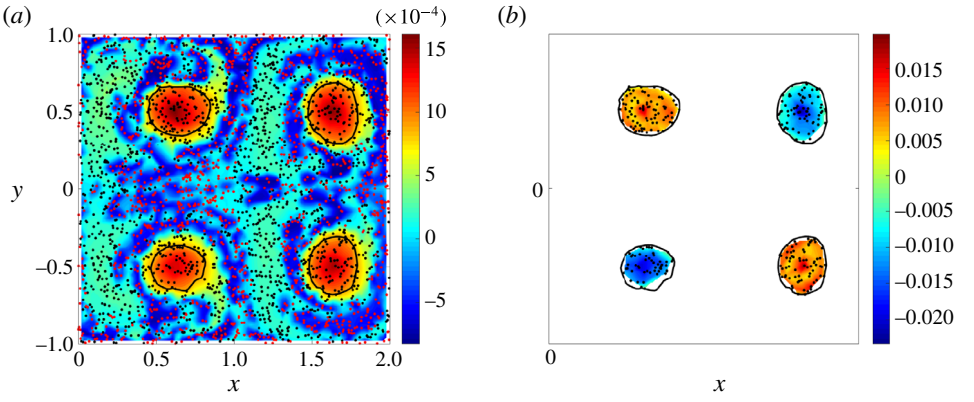

FIGURE 3. Unsteady quadruple gyre, $\epsilon = 0.1$, $A = 0.1$, $T = [2.5, 42.5]$ and 3000 total particles. (*a*) CSC; black dots show final locations of particles that remained in their initial quadrant, and red dots show final locations of particles that switched quadrants during the coherent structure calculation time interval. Black lines trace contours of CSC = 0.0009. (*b*) CSC, algorithm applied recursively only to particles whose initial CSC value was greater than 0.0009.

3.2. *Bickley jet*

The Bickley jet, another analytical example, is frequently used as a model of zonal jets in the Earth's atmosphere (Rypina, Brown & Beron-Vera 2007). It is a periodic flow comprising a spatially undulatory jet with counter-rotating vortices above and below. The flow is described by the streamfunction $\psi = \psi_0 + \psi_1$, where

$$\psi_0 = c_3 y - UL \tanh(y/L), \tag{3.4}$$

$$\psi_1 = UL \operatorname{sech}^2(y/L) \sum_{n=1}^{3} \epsilon_n \cos(k_n(x - \sigma_n t)). \tag{3.5}$$

We use similar values of the parameters as in Hadjighasem *et al.* (2016): $U = 62.66$ m s^{-1}, $L = 1770$ km, $k_n = 2n/r_0$, $c = [0.1446U, 0.205U, 0.461U]$, $\sigma = c - c(3)$ and $\epsilon = [0.0075, 0.15, 0.3]$; and the flow is computed on the interval $x = [0, 20 \times 10^6]$ m, $y = [-3 \times 10^6, 3 \times 10^6]$ m, over the time interval $t = [0, 40]$ days, divided into 607 discrete time steps. The flow was considered periodic in x. For calculation of the CSC, particles were initialized randomly in the domain and advected with the flow. The particles were followed over the entire time interval, even if they left the domain, analogous to how ocean drifters are tracked. The velocity magnitude of the flow overlaid with streamlines is shown in figure 4, along with the FTLE field calculated using 48 000 particles.

Figure 5 shows the results of the CSC algorithm using only 480 particles: (*a*) shows the CSC field overlaid with black dots indicating the final particle positions, and (*b*) shows particle tracks for the highest positive colouring values (in black) and the highest negative colouring values (in red). Without specifying the number of coherent structures *a priori*, the algorithm is able to accurately detect the centres of the three vortices above the jet and two full and two half vortices below. However, if the seeding density was too low, such that one of the vortices contained no particles, or only a few, the structure could not be identified. The jet itself is aligned with the most

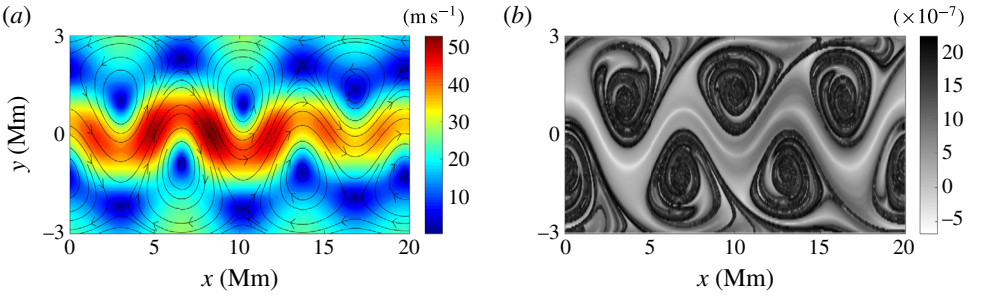

FIGURE 4. Bickley jet: (*a*) velocity magnitude with sample streamlines; and (*b*) FTLE field, calculated over the time interval $t = [0, 3456 \times 10^3]$ s using 48 000 particles.

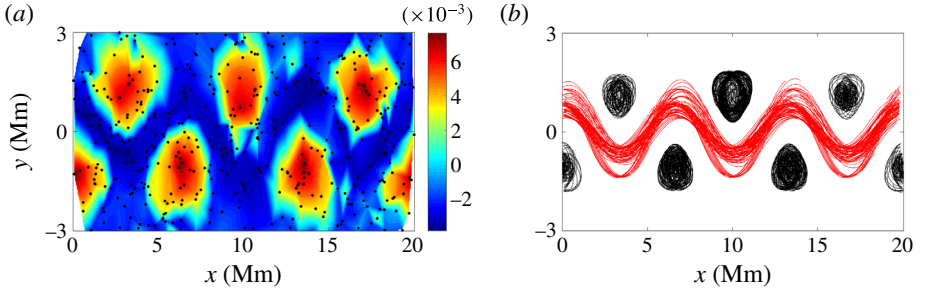

FIGURE 5. Bickley jet: (*a*) CSC contours overlaid with dots indicating final particle positions, 480 particles, $t = [0, 3456 \times 10^3]$ s; and (*b*) particle tracks for particles with highest (black) and lowest (red) CSC values.

negative colouring contours, indicating that the largest kinematic dissimilarity detected is between the jet and the vortices flanking it. It is noteworthy that the eigenvector associated with the largest eigenvalue of the generalized eigenvalue problem (i.e. the CSC) provides information about all of the coherent structures simultaneously, as opposed to N-cut approaches that recover one structure per eigenvalue among those selected heuristically. This is in part because the present algorithm avoids unnecessarily restricting the definition of coherence to particles that are physically close; even the relative kinematics of particles that are far apart provides useful information regarding the coherent structures in the flow.

The Bickley jet flow can also be used to characterize the robustness of the CSC method to broken or incomplete fluid particle trajectory information. Two types of incomplete datasets are examined, each corresponding to a specific experimental circumstance in which data might be lost. First, we examine the case in which a fluid particle trajectory is lost and later recovered, but is still identified as the same particle as before the data loss. This could be the case for ocean drifters, e.g. when a drifter temporarily goes offline with associated data loss. Additionally, certain PTV algorithms are capable of linking broken trajectories if datasets are sparse enough and particle trajectory crossings do not occur when breaks occur (Li *et al.* 2008). To characterize the response of the CSC method to this type of data loss, a dataset containing full trajectories of 480 particles was manipulated to randomly remove 10 %, 50 % and 90 % of the trajectory data. When considering the weight of a pair of particles, the CSC algorithm only considers the overlapping time interval in

which both particles are present. If a particle is not present during the time step for which the CSC field is computed, the trajectory information for that particle is not considered in the calculation of the adjacency matrix. Hence, the size of the adjacency matrix is $n_p \times n_p$, where n_p is the number of particles present in the domain during the time step in which the CSC field is calculated. For this analysis, it was ensured that all 480 particles were present in the final time step so that their trajectory information could be used to calculate the CSC field. This was done so that the analysis would show the effect of intermittent data loss as opposed to the effect of total number of particles. The results are shown in figure 6. From this figure it is evident that, in the case where pieces of particle trajectories can be linked and identified as broken pieces of the same trajectory, intermittent data loss does not adversely affect the robustness of the CSC algorithm, as long as there are a sufficient number of particles present in the time step for which the CSC field is calculated.

In order to characterize the CSC method in cases where broken trajectories cannot be reconstructed and the particle tracks must be treated as independent fluid particles, a baseline group of 480 particles was again examined. A portion of the position data was then randomly deleted, and every time a break in the position data occurred, the remainder of the track was recharacterized as a separate particle trajectory. The results from this analysis are shown in figure 7. For the case with 480 unbroken particle trajectories, shown in figure 5(*a*), all particles have a trajectory length that spans the entire time domain. For the shorter particle trajectories shown in figure 7, it is evident that, as the average particle trajectory length is shortened, the flanking vortices appear to blend together into two large coherent structures, one below the jet and one above. This result can be understood by considering the length of the fluid particle trajectories relative to the eddy turnover time. Based on the flow velocity around closed streamlines for the Bickley jet flow at $t = 0$, the eddy turnover time is estimated to be approximately 279×10^3 s, and the time interval $t = [0, 3456 \times 10^3]$ s is equivalent to approximately 12.4 eddy turnover times. Thus, while it is clear that the CSC algorithm is capable of handling broken trajectories of this type, an average trajectory length of approximately 2.6 eddy turnover times, as in figure 7(*a*), is necessary to distinguish individual coherent structures. Otherwise, there is not enough information available to effectively characterize the flow, even if the total observation time is a larger multiple of the eddy turnover time.

It is also useful to analyse and understand the response of the method to a large number of particles, approaching the quantity used for non-sparse methods such as FTLE analysis. As such, a CSC analysis of a Bickley jet seeded with 12 000 particles was examined, and the resulting CSC field is shown in figure 8. The features of this CSC field are similar to what would be seen by clustering-based methods, if thresholding of the CSC were used to separate the vortex cores into distinct structures. There are also similarities with what would be seen if vortices were extracted from the flow using the forward and reverse FTLE fields, including the five isolated vortex cores and two half cores. Although not demonstrated here, the subsequent extraction of the coherent structures from the CSC field can be performed in a manner similar to the FTLE field analysis. For example, one option is to use thresholding of the CSC field to determine boundaries of the coherent structures. Additionally, the spatial gradients of the CSC field can be used to separate coherent structures from the background flow.

In assessing the feasibility of high-resolution CSC analysis, computational time is an important factor to consider. Table 1 provides a summary of computational run times on a single processor for the Bickley jet with three seeding densities.

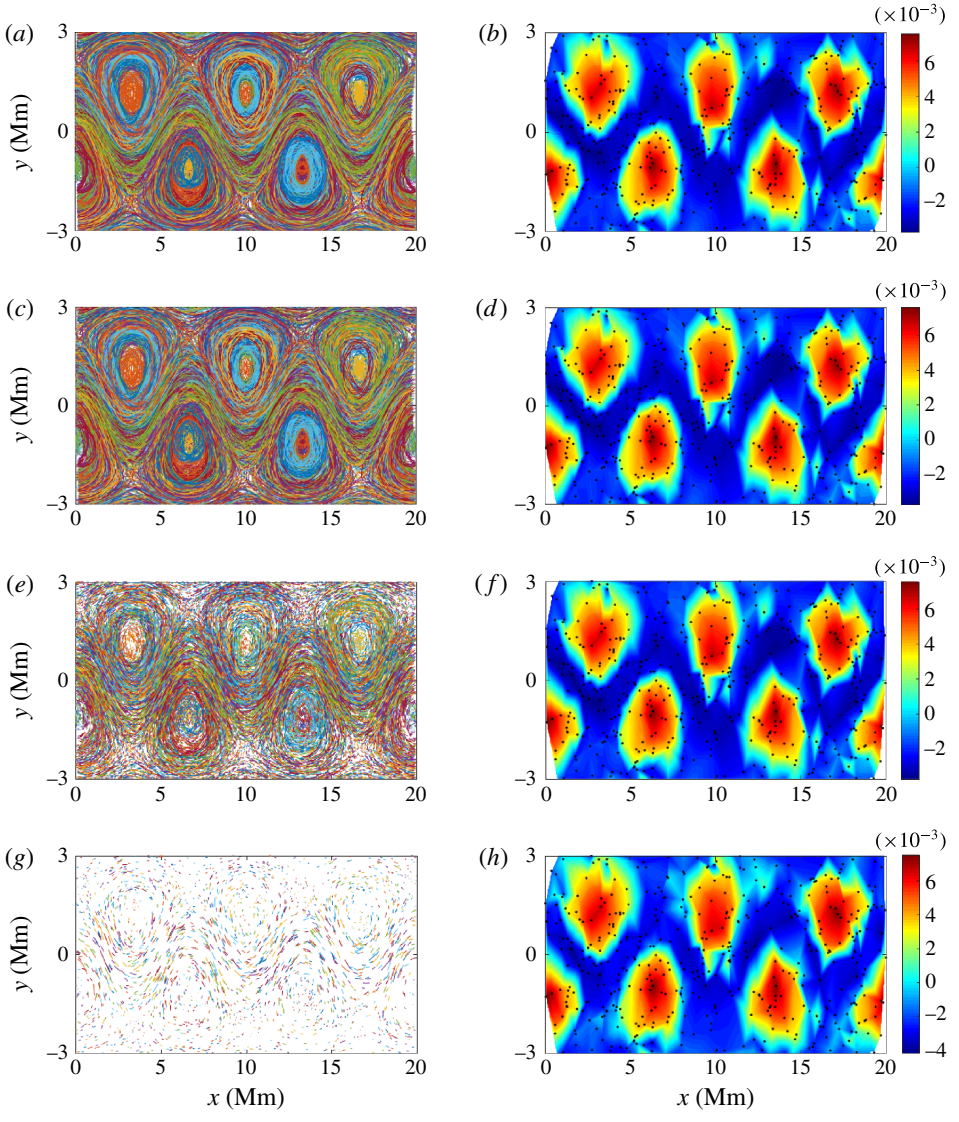

FIGURE 6. Bickley jet, 480 particles: (*a*) unbroken particle trajectories; (*b*) CSC field for unbroken particle trajectories; (*c*) particle trajectories, 10 % of position data deleted; (*d*) CSC field for case where 10 % of particle position data is deleted; (*e*) particle trajectories, 50 % of position data deleted; (*f*) CSC field for case where 50 % of particle position data is deleted; (*g*) particle trajectories, 90 % of position data deleted; (*h*) CSC field for case where 90 % of particle position data is deleted. Black dots indicate final particle position.

3.3. *Vortex ring*

Next, we examine a PIV dataset of a forming vortex ring with a high maximum stroke ratio. The vortex ring is created in a water tank using a piston forced at speed U a distance $L = Ut$ through a hollow cylinder of diameter D, which in turn ejects fluid from an axisymmetric nozzle with a sharp edge. The shear layer formed

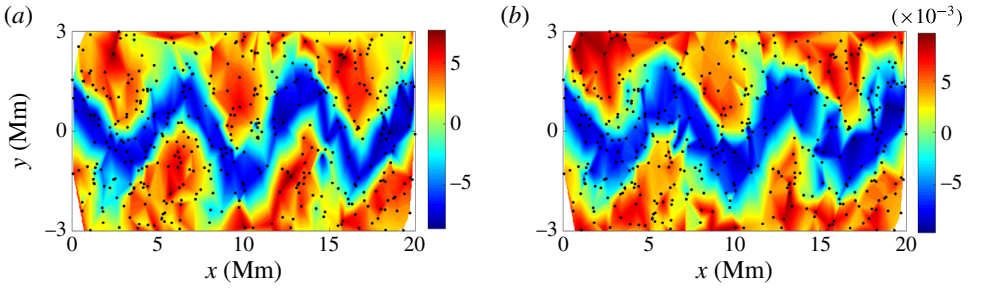

FIGURE 7. Bickley jet CSC fields, 480 particles: (*a*) average trajectory length of 2.6 eddy turnover times; and (*b*) average trajectory length of 1.7 eddy turnover times.

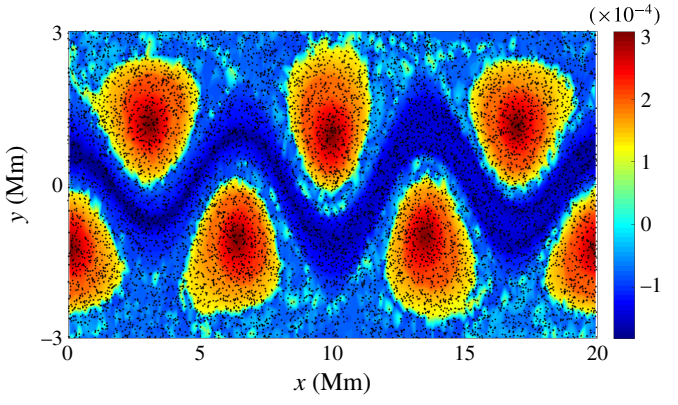

FIGURE 8. Bickley jet CSC field, 12 000 particles; black dots indicate final particle positions.

	480 particles	2400 particles	12 000 particles
Adjacency matrix calculation (s)	2.8	79.1	2345.2
Eigen-decomposition (s)	0.3	1.9	213.2

TABLE 1. Run times for Bickley jet flow on a single processor.

inside the nozzle due to the motion of fluid through it rolls up at the nozzle exit, forming a vortex ring. If the maximum stroke ratio, $t^*_{max} = Ut_{max}/D$, where t_{max} is the total time over which the piston is moving, is greater than 4, then a trailing jet and potentially secondary and tertiary vortex rings are formed behind the primary vortex ring (Gharib, Rambod & Shariff 1998). The dataset examined here is of a vortex ring formed using a piston travelling with a constant velocity until a maximum stroke ratio of 8. The Reynolds number based on the diameter is approximately 1800. Details of the experimental set-up and acquisition and processing of the PIV images for this dataset can be found in Schlueter-Kuck & Dabiri (2016).

Figure 9(*a*) shows the vorticity field at $t^* = 10.2$, after the piston has stopped moving, where $t^* = Ut/D$ is the non-dimensional time, equivalent to the number of piston diameters that the piston has travelled. At this point, the leading vortex ring, as well as secondary and tertiary vortex rings and a trailing jet, are clearly visible.

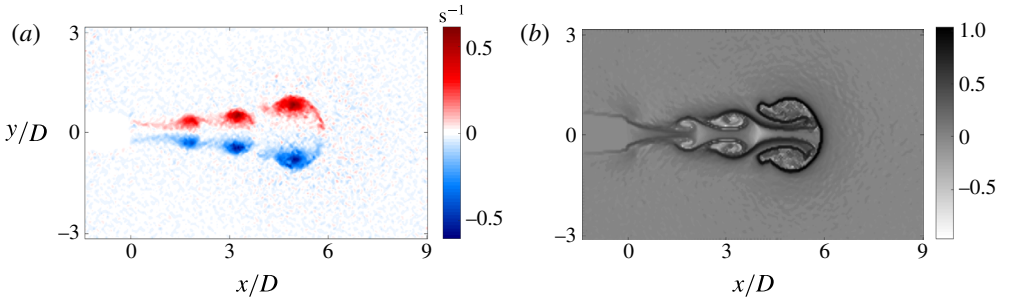

FIGURE 9. Vortex ring, $t^*_{max} = 8$: (*a*) vorticity field, $t^* = 10.2$; and (*b*) FTLE field, calculated over the time interval $t^* = [8.0, 10.2]$, using 30 500 particles.

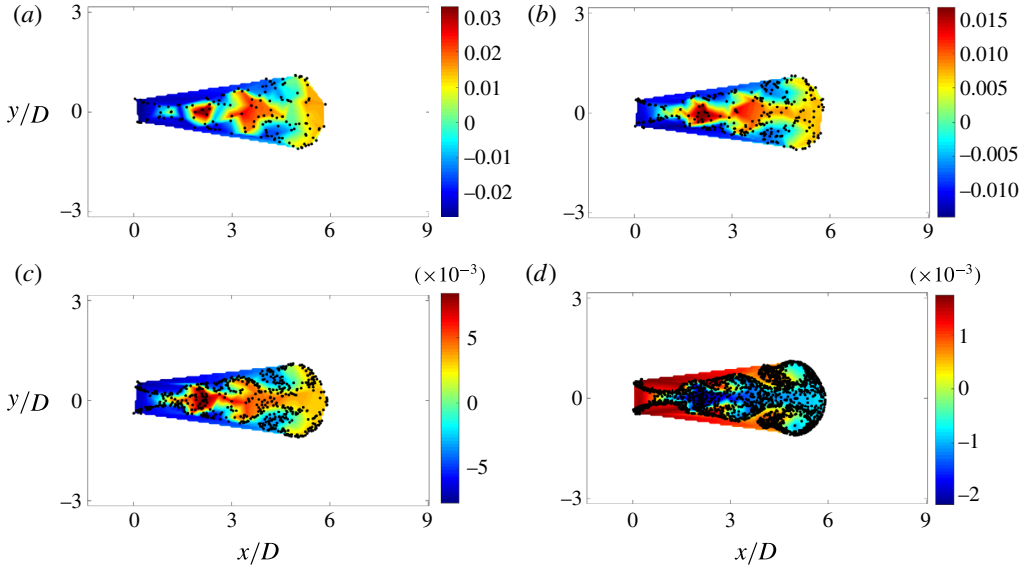

FIGURE 10. Vortex ring CSC, for particles introduced at nozzle exit plane while vortex ring is forming, calculated over the time interval $t^* = [8.0, 10.2]$: (*a*) 150 particles, (*b*) 300 particles, (*c*) 600 particles and (*d*) 2400 particles.

The corresponding backward FTLE field, computed using 30 500 particles, is shown in figure 9(*b*). Figure 10 shows the CSC calculated using a total of 150, 300, 600 and 2400 particles initiated at the nozzle exit plane near the left of the frame between $t^* = 0.04$ and $t^* = 8.4$. The CSC algorithm identifies all three vortex rings, which is in qualitative agreement with the dark ridges of the FTLE field. While the resolution of the CSC contours increases with the number of particles, it is clear that 300 particles is sufficient to obtain a qualitatively similar result to the case with eight times as many particles, and to the FTLE calculation based on 30 500 particles.

The sensitivity of the CSC method to the size of the domain of particles and the time of release was also characterized using the vortex ring PIV data. In figure 11(*a*), the entire flow field was initialized with randomly located particles at $t^* = 0$, and subsequently 1200 additional particles were added between $t^* = 0.04$ and $t^* = 8.4$ at the nozzle exit plane. Because the CSC algorithm groups regions with a low

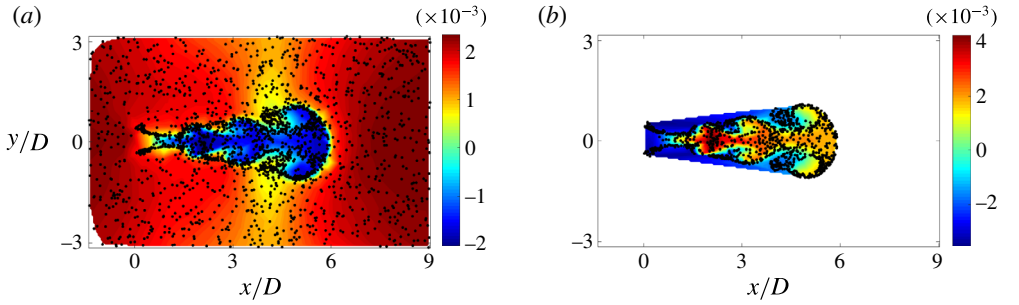

FIGURE 11. Vortex ring CSC, calculated over the time interval $t^* = [8.0, 10.2]$: (*a*) 1200 particles initiated randomly in full domain at $t^* = 0$ and 1200 particles introduced at nozzle exit plane during vortex ring formation time, $t^* = [0.04, 8.4]$; and (*b*) 1200 particles introduced at nozzle exit plane during vortex ring formation time, $t^* = [0.04, 8.4]$.

normalized standard deviation in relative particle separation, the nominally quiescent background flow was assigned a CSC value that contrasts most sharply with the entire starting jet flow. Consequently, details of the internal structure of the vortex ring and trailing jet are lost. However, if we recursively apply the CSC algorithm only to the flow trajectories in the starting jet, as shown in figure 11(*b*), we see that the algorithm is able to detect the structure of the primary, secondary and tertiary vortex rings in greater detail, despite the fewer number of total particles.

4. Comparison with other graph-theory-based methods

A related method for coherent structure detection that is also based on graph theory uses the concept of an eigen-gap heuristic to determine the number of coherent structures present in the flow (Hadjighasem *et al.* 2016). For this method, the weights assigned to the edges of the graph are equal to the reciprocal of the average distance between particle pairs. The generalized eigenvalue problem solved in this method is $\boldsymbol{L}x = \lambda \boldsymbol{D}x$, where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ is the graph Laplacian, and $\boldsymbol{D}$ is the diagonal degree matrix, where d_{ii} is equal to the sum of the elements in row i of the adjacency matrix $\boldsymbol{A}$. This method assumes that, by examining the smallest eigenvalues of the generalized eigenvalue problem, the number of coherent structures in a flow can be determined by locating the largest numerical gap between successive eigenvalues; the number of eigenvalues before the gap is assumed to correspond to the number of coherent regions in the flow.

Here we present an analysis of the aforementioned technique and its response to several input variables, comparing its robustness to the method introduced in this paper. This analysis again uses the Bickley jet described by (3.4) and (3.5) and the values of the parameters listed in § 3.2. The response of the eigenvalue spectrum for the Bickley jet flow to changes in the initial position of tracer particles and to the value of the sparsification parameter ϵ are shown in figure 12 for 1920 particles and in figure 13 for 480 particles. For the case with 480 particles initialized randomly in the domain, exactly the same particle trajectories were used as in the analysis in § 3.2 to allow for a direct comparison between the two methods.

Based on prior knowledge of the Bickley jet flow, we expect to resolve six coherent structures: the five full vortices flanking the meandering jet, and due to the periodic nature of the flow in the x-direction, and two half vortices identified

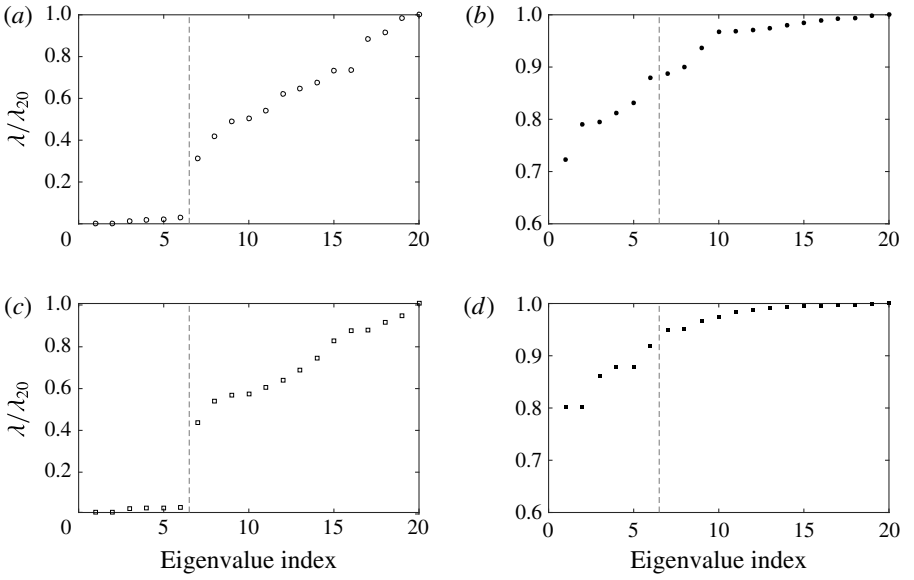

FIGURE 12. Bickley jet eigenvalue spectra for 1920 particles: (*a*) randomized particle initialization, sparsification of adjacency matrix for average particle pairwise distances greater than 3×10^6 m; (*b*) randomized particle initialization, no sparsification of adjacency matrix; (*c*) gridded particle initialization, sparsification of adjacency matrix for average particle pairwise distances greater than 3×10^6 m; and (*d*) gridded particle initialization, no sparsification of adjacency matrix. Dotted vertical lines indicate the expected location of the eigen-gap based on six coherent structures.

together as a sixth coherent structure. Thus, the eigen-gap heuristic should predict a numerical gap between the sixth and the seventh eigenvalues. From figure 12(*a*,*c*) we can see that for 1920 particles, regardless of whether particles are initialized on a Cartesian grid or randomly throughout the domain, the largest gap in the smallest 20 eigenvalues is between the sixth and seventh eigenvalues, as expected. However, when the adjacency matrix is not sparsified to remove weights corresponding to particle pairs with an average distance greater than 3×10^6, as shown in figure 12(*b*,*d*), the eigen-gap is located between the first and second eigenvalues for the random particle initialization and between the second and third eigenvalues for the gridded particle initialization. These results indicate that the eigen-gap heuristic is sensitive to the level of sparsification used in the adjacency matrix. Consequently, without prior knowledge of the size of the coherent structures to inform an appropriate level of sparsification, this method is not able to robustly determine the number of coherent structures in the flow.

The analysis of the eigenvalue spectrum for 480 particles, the same number as used in the analysis of CSC for the Bickley jet flow in § 3.2, is shown in figure 13. Here we see in figure 13(*a*) that, for random particle initialization with sparsification, the eigen-gap is between the ninth and tenth eigenvalues. Additionally, if the particles are initialized on a grid (figure 13*c*), or sparsification is not used (figure 13*b*), or both (figure 13*d*), the eigen-gap is also not correctly identified. Based on this analysis, we can conclude that for small numbers of tracer particles, on the order of 10^2–10^3, use of the eigen-gap heuristic to determine the number of coherent structures in the flow is ineffective based on the lack of robustness to initial tracer locations (often

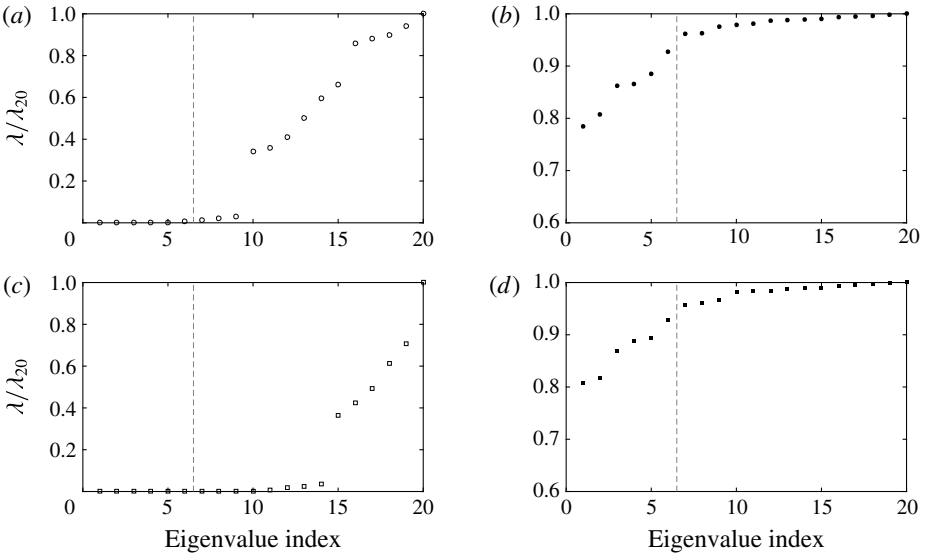

FIGURE 13. Bickley jet eigenvalue spectra for 480 particles: (*a*) randomized particle initialization, sparsification of adjacency matrix for average particle pairwise distances greater than 3×10^6 m; (*b*) randomized particle initialization, no sparsification of adjacency matrix; (*c*) gridded particle initialization, sparsification of adjacency matrix for average particle pairwise distances greater than 3×10^6 m; and (*d*) gridded particle initialization, no sparsification of adjacency matrix. Dotted vertical lines indicate the expected location of the eigen-gap based on six coherent structures.

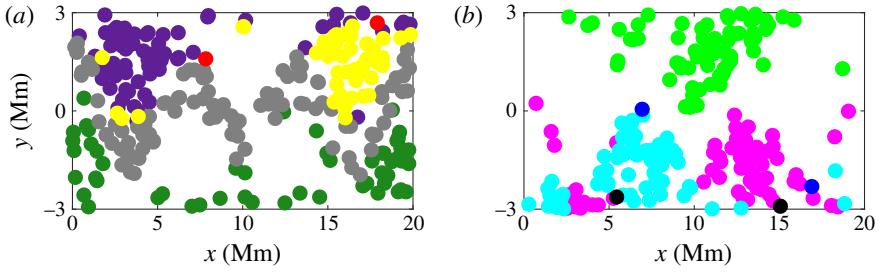

FIGURE 14. *K*-means clustering of the Bickley jet with 480 randomly initialized particles; five of the 10 total clusters identified are shown in (*a*) and the remaining five in (*b*).

beyond the control of the investigator for experimental applications) and to the level of sparsification of the adjacency matrix (which requires *a priori* knowledge of the size of the coherent structures, if sparsification is to be used effectively).

Despite an incorrect identification of the eigen-gap, the coherent structures can theoretically still be identified using a *K*-means clustering of the eigenvectors associated with the eigenvalues below the eigen-gap according to this method. To be sure, if the eigen-gap heuristic overestimates the number of structures, some structures might be split into several, and if the number of structures is underestimated, several structures might be merged together. Thus, for the case where 480 particles were randomly initialized in the domain and sparsification was used in the analysis (i.e. figure 13*a*), the results of the *K*-means clustering is shown in figure 14,

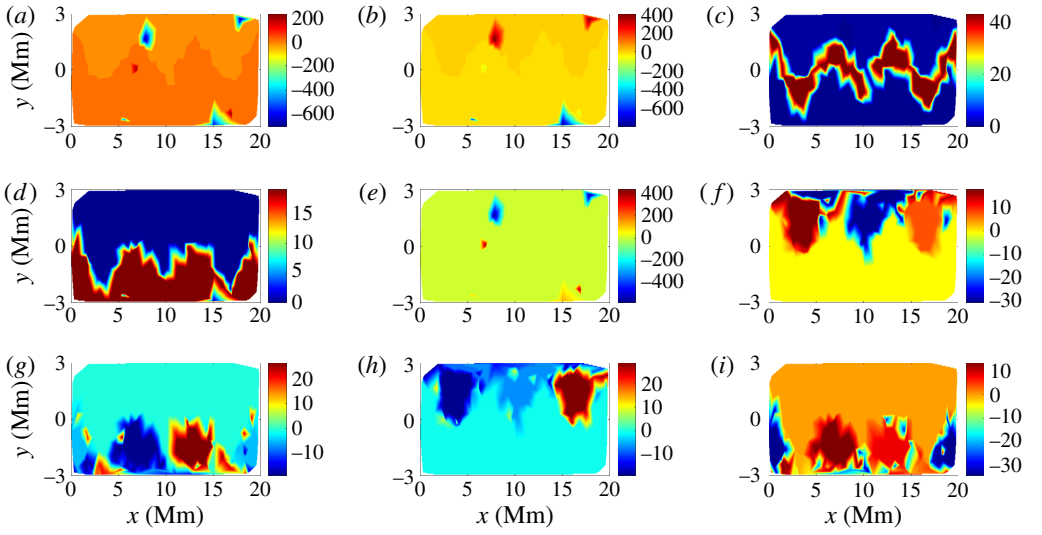

FIGURE 15. Eigenvectors corresponding to the eigenvalues below the eigen-gap for the Bickley jet flow with 480 randomly initialized particles.

and the nine eigenvectors used in the clustering analysis are shown in figure 15. The clustering analysis searched for 10 groups corresponding to the nine coherent structures expected from the eigen-gap heuristic, in addition to one structure for the incoherent background flow. In figure 14, these 10 clusters are plotted between two separate panels to aid in visualization of the individual clusters. From the clustering, we can see that the grey cluster roughly corresponds to the meandering jet, while the flanking vortices are somewhat consistent with the purple, yellow and light green clusters on the top and the dark green, cyan and magenta clusters on the bottom. The black, red and blue clusters identify only a few seemingly random particles each. Even if the three smallest clusters are ignored, the boundaries of the clusters corresponding to the vortices are significantly different from the boundaries of the vortices themselves, as observed in the FTLE analysis in figure 4(*b*). From observing the eigenvectors used for this analysis, it is evident that the first, second and fifth eigenvectors (figure 15*a*,*b*,*e*) are responsible for the small clusters identified by the *K*-means analysis, while the remaining six eigenvectors roughly correspond to the boundaries of groups of the flanking vortices and meandering jet. Although not shown, if *K*-means clustering is performed using the third and fifth to ninth eigenvectors to identify seven clusters, the clusters identified are almost identical to the clusters in figure 14 excluding the small blue, red and black clusters. However, the boundaries of these clusters are still not consistent with the boundaries of the jet and the vortices.

5. Conclusions

This paper presents an algorithm for detecting coherence in flows where only sparse velocity data are available, as is often the case in particle tracking velocimetry, or oceanographic tracking of surface floats. In this regime, alternative methods of evaluating coherence either require knowledge of the number of coherent structures *a priori*, or break down due to the sparsity of the data. The present method, based on the concepts of graph colouring and spectral graph drawing, examines the kinematic

dissimilarity of every pair of trajectories, and organizes these data into a weighted adjacency matrix. As such, we consider a different definition of coherence from other coherent structure detection methods, which only consider groups of particles that remain close as time progresses without mixing with the surrounding fluid to be coherent structures. The eigenvector associated with the maximum eigenvalue of the generalized eigenvalue problem $\boldsymbol{LX} = \lambda \boldsymbol{DX}$ assigns a value of CSC to each particle, such that similar CSC values indicate coherence in the flow. This algorithm has several inherent strengths, including that the number of coherent structures does not need to be known *a priori*. Additionally, information about all coherent structures in the flow is contained in a single eigenvector of the generalized eigenvalue problem associated with the graph Laplacian. The algorithm is also capable of detecting coherent structures with the small number of trajectories associated with many PTV and ocean drifter datasets, and was shown to be robust to different types of data loss common in particle/drifter tracking applications. Although only two-dimensional datasets were examined here, the kinematic dissimilarity metric in (2.1) and the subsequent maximization problem can both be extended to higher dimensions without loss of generality and with limited additional computational cost, since the adjacency matrix scales with the square of the number of particles, independent of the dimensionality of their trajectories. The CSC method has the potential to be extended to analyse other properties of fluid flows in addition to coherence, and it may also find application in other data analysis problems for which coherent structure identification remains a challenge.

A MATLAB implementation of the CSC algorithm is available for free download at http://dabirilab.com/software.

Acknowledgements

This work was supported by the US National Science Foundation and by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

REFERENCES

ALLSHOUSE, M. & PEACOCK, T. 2015 Lagrangian based methods for coherent structure detection. *Chaos* **25**, 097617.

ALLSHOUSE, M. & THIFFEAULT, J.-L. 2012 Detecting coherent structures using braids. *Physica* D **241**, 95–105.

BANISCH, R. & KOLTAI, P. 2016 Understanding the geometry of transport: diffusion maps for Lagrangian trajectory data unravel coherent sets. arXiv:1603:04709v2 [math.DS].

BEZDEK, J., EHRLICH, R. & FULL, W. 1984 FCM: the fuzzy c-means clustering algorithm. *Comput. Geosci.* **10**, 191–203.

CHANG, T., WILCOX, N. & TATTERSON, G. 1984 Application of image processing to the analysis of three-dimensional flow fields. *Opt. Engng* **23** (3), 283–287.

DAVIS, R. 1991 Lagrangian ocean studies. *Annu. Rev. Fluid Mech.* **23**, 43–64.

ELSINGA, G., SCARANO, F., WIENEKE, B. & VAN OUDHEUSDEN, B. 2006 Tomographic particle image velocimetry. *Exp. Fluids* **41**, 933–947.

FROYLAND, G. & PADBERG-GEHLE, K. 2015 A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data. *Chaos* **25**, 087406.

GHARIB, M., RAMBOD, E. & SHARIFF, K. 1998 A universal time scale for vortex ring formation. *J. Fluid Mech.* **360**, 121–140.

HADJIGHASEM, A., KARRASCH, D., TERAMOTO, H. & HALLER, G. 2016 Spectral-clustering approach to Lagrangian vortex detection. *Phys. Rev.* E **93**, 063107.

HALL, K. 1970 An r-dimensional quadratic placement algorithm. *Manage. Sci.* **17** (3), 219–229.

HALLER, G. 2005 An objective definition of a vortex. *J. Fluid Mech.* **525**, 1–26.

HALLER, G. 2015 Lagrangian coherent structures. *Annu. Rev. Fluid Mech.* **47**, 137–162.

HALLER, G. & YUAN, G. 2000 Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica* D **147**, 352–370.

HO, J., TOLOUI, M., CHAMORRO, L., GUALA, M., HOWARD, K., RILEY, S., TUCKER, J. & SOTIROPOULOS, F. 2014 Natural snowfall reveals large-scale flow structures in the wake of a 2.5-MW wind turbine. *Nat. Commun.* **5**, 4216.

HUHN, F., VON KAMEKE, A., PÉREZ-MUÑUZURI, V., OLASCOAGA, M. & BERON-VERA, F. 2012 The impact of advective transport by the South Indian Ocean Countercurrent on the Madagascar plankton bloom. *Geophys. Res. Lett.* **39**, L06602.

HUNT, J., WRAY, A. & MOIN, P. 1988 Eddies, streams, and convergence zones in turbulent flows. *Center for Turbulence Research Report* CTR-S88, 193–208.

JEONG, J. & HUSSAIN, F. 1995 On the identification of a vortex. *J. Fluid Mech.* **285**, 69–94.

KATIJA, K. & DABIRI, J. 2008 *In situ* field measurements of aquatic animal-fluid interactions using a Self-Contained Underwater Velocimetry Apparatus (SCUVA). *Limnol. Oceanogr.* **6**, 162–171.

KIM, D., HUSSAIN, F. & GHARIB, M. 2013 Vortex dynamics of clapping plates. *J. Fluid Mech.* **714**, 5–23.

KOREN, Y. 2005 Drawing graphs by eigenvectors: theory and practice. *Comput. Maths Applics.* **49**, 1867–1888.

LI, D., ZHANG, Y., SUN, Y. & YAN, W. 2008 A multi-frame particle tracking algorithm robust against input noise. *Meas. Sci. Technol.* **19**, 105401.

LÜTHI, B., TSINOBER, A. & KINZELBACH, W. 2005 Lagrangian measurement of vorticity dynamics in turbulent flow. *J. Fluid Mech.* **528**, 87–118.

MUÑOZ, S., ORTUÑO, M. T., RAMÍREZ, J & YÁÑEZ, J. 2005 Coloring fuzzy graphs. *Omega* **33**, 211–221.

MURAI, Y., NAKADA, T., SUZUKI, T. & YAMAMOTO, F. 2007 Particle tracking velocimetry applied to estimate the pressure field around a Savonius turbine. *Meas. Sci. Technol.* **18**, 2491–2503.

OKUBO, A. 1970 Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences. *Deep-Sea Res.* **17**, 445–454.

RACCA, R. & DEWEY, J. 1988 A method for automatic particle tracking in a three-dimensional flow field. *Exp. Fluids* **6**, 25–32.

RYPINA, I., BROWN, M. & BERON-VERA, F. 2007 On the Lagrangian dynamics of atmospheric zonal jets and the permeability of the stratospheric polar vortex. *J. Atmos. Sci.* **64**, 3595–3610.

SCHANZ, D., GESEMANN, S. & SCHRÖDER, A. 2016 Shake-the-box: Lagrangian particle tracking at high particle image densities. *Exp. Fluids* **57** (5), 70.

SCHLUETER-KUCK, K. & DABIRI, J. 2016 Pressure evolution in the shear layer of forming vortex rings. *Phys. Rev. Fluids* **1**, 012501.

SHADDEN, S., LEKIEN, F. & MARSDEN, J. 2005 Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. *Physica* D **212**, 271–304.

VIRANT, M. & DRACOS, T. 1997 3D PTV and its application on Lagrangian motion. *Meas. Sci. Technol.* **8**, 1539–1552.

WEISS, J. 1991 The dynamics of enstrophy transfer in two-dimensional hydrodynamics. *Physica* D **48**, 273–294.