Supporting Information

Beron-Vera et al. 10.1073/pnas.1701392115

1. CLS Detection

The two methods used to detect CLSs are described here preceded by their mathematical setup.

1.1. Setup. Let $v(x,t)$ be a 2D velocity field, where x denotes position on the plane and t is time. Fluid particle motion then obeys the differential equation

$$\dot{x} = v(x,t). \qquad \textbf{[S1]}$$

The flow map

$$F_{t_0}^{t}(x_0) := x(t; x_0, t_0) \qquad \textbf{[S2]}$$

takes particle positions at time t_0 to their positions at a later time t.

1.2. The LAVD Method. Haller et al. (1) seek coherent Lagrangian eddies boundaries as locally maximal closed material curves along which all fluid particles experience the same bulk rotation relative to the mean rotation of the total fluid mass of interest. This relative bulk material rotation turns out to equal twice the LAVD field, defined as

$$\mathrm{LAVD}_{t_0}^{t}(x_0) := \int_{t_0}^{t} |\omega(F_{t_0}^{\tau}(x_0),\tau) - \bar{\omega}(\tau)|\,\mathrm{d}\tau, \qquad \textbf{[S3]}$$

where $\omega(x,t)$ is the vorticity and $\bar{\omega}(t)$ is its instantaneous average over the region $U(t)$ containing the fluid.

For an infinitesimal fluid region starting from x_0, $\mathrm{LAVD}_{t_0}^{t}(x_0)$ is an objective and dynamically consistent measure of bulk material rotation relative to the spatial mean rotation of the full fluid mass $U(t)$. Specifically, $\mathrm{LAVD}_{t_0}^{t}(x_0)$ is twice the intrinsic dynamic rotation angle generated by the relative rotation tensor $\Phi_{t_0}^{t}(x_0)$, which itself is obtained from the unique, dynamically consistent decomposition of the deformation gradient (2):

$$\mathrm{D}F_{t_0}^{t}(x_0) = \Phi_{t_0}^{t}(x_0)\Theta_{t_0}^{t}(x_0)M_{t_0}^{t}(x_0). \qquad \textbf{[S4]}$$

Here $\Theta_{t_0}^{t}(x_0)$ is the deformation gradient of a pure rigid-body rotation, and $M_{t_0}^{t}(x_0)$ the deformation gradient of a unique, purely straining flow; cf. ref. 2 for details.

Over the finite time interval $[t_0, t_1]$, Haller et al. (1) define a rotationally coherent Lagrangian eddy as an evolving fluid region $\mathcal{R}(t)$ such that $\mathcal{R}(t_0)$ is filled with a nested family of level curves of $\mathrm{LAVD}_{t_0}^{t_1}(x_0)$ with outward-increasing values. The boundary $\mathcal{B}(t)$ of $\mathcal{R}(t)$ is a material loop such that $\mathcal{B}(t_0)$ is the outermost level curve of $\mathrm{LAVD}_{t_0}^{t_1}(x_0)$ in $\mathcal{R}(t_0)$.

We note that rotationally coherent Lagrangian eddies are allowed to experience filamentation. However, if the boundary is initially convex, any filamented piece is guaranteed to rotate together with the rotationally coherent Lagrangian structure without global breakaway.

The numerical implementation of the LAVD method is quite simple. It involves integrating Eq. **S1** from some dense grid of initial fluid particle positions, then evaluating the vorticity along these fluid trajectories, and finally computing the LAVD field in Eq. **S3**. Once this computation is carried out, one can identify coherent Lagrangian eddy boundaries as outermost convex level curves of the LAVD field encircling its local maxima. As a practical way to deduce the convexity of these curves, one requires their convexity deficiency to be less than a sufficiently small bound. Convexity deficiency of a closed curve on a plane is defined as the ratio of the area between the curve and its convex hull to the area enclosed by the curve.

We have carried out all integrations using a step-adapting fourth/fifth-order Runge–Kutta method with interpolations done with a cubic method. The width of the initial position grid was set to 0.1 km, which is about 100 times smaller than the smallest spatial scale that the NCOM model can effectively resolve. Using smaller widths makes computations only more expensive, while using larger widths typically results in coarser eddy boundaries which tend to filament too quickly. The vorticity was computed from the model's multiscale velocity output using central differences. No smoothing was applied on the vorticity as this was found to produce smoother eddy boundaries that also tended to filament quickly. The convexity deficiency tolerances used were set as small as the multiscale data allowed us. For the mesoscale eddy boundary extraction, we were able to set the tolerance to about 0.25. No boundary was found using a larger value, while smaller values resulted in boundaries that enclosed areas with mean radii much smaller than the scale of interest. For the submesoscale extractions, the tolerance was possible to decrease to 0.025. Similarly, we could not find any boundaries using larger values, and smaller values produced too small eddies.

1.3. The Spectral Clustering Method. Hadjighasem et al. (3) seek coherent Lagrangian eddies as distinguished sets of fluid particles that maintain short distances among themselves relative to their distances to particles in other regions over the finite time interval $[t_0, t_1]$.

To implement the spectral clustering approach, the first step is to construct a trajectory array $X \in \mathbb{R}^{n\times m\times d}$ whose rows $(X_i)_{i=1,\dots,n}$ contain positions of n Lagrangian particles over m discrete time intervals in d-dimensional space; that is, $X_i = \{x_i(t_0), \dots, x_i(t_1)\}$. The maximum dynamic distance r_{ij} between particle trajectories X_i and X_j is then defined as

$$r_{ij} = \max_t |x_i(t) - x_j(t)|, \qquad \textbf{[S5]}$$

where $|\cdot|$ denotes the spatial Euclidean norm.

The second step is to construct a similarity graph $G = (V, E, W)$, which is defined by the set of its nodes $V = \{v_1, \cdots, v_n\}$, the set of its edges $E \subseteq V \times V$ connecting nodes, and a similarity matrix $W \in \mathbb{R}^{n\times n}$ which associates weights to the edge set. In this context, each graph node v_i represents a fluid particle, and each element of the similarity matrix w_{ij} equals the inverse of maximum dynamic distance between the particle i and particle j. However, only those elements from the similarity matrix w_{ij} whose values exceed a specified threshold ϵ are retained. All other w_{ij} entries are otherwise set to zero and hence require no storage. The reason for sparsifying the similarity matrix is that, when the dynamic distance between the particle i and particle j is too large, then the chance that these two particles can be in the same coherent region is too low. Hence, the edge between the corresponding graph nodes v_i and v_j can be safely ignored, and the corresponding edge weight w_{ij} can be set to zero.

With the similarity weights at hand, the degree matrix D is defined as a diagonal matrix, which contains the row/column sums of the adjacency matrix along the diagonal, as

$$d_{ij} = \begin{cases} \sum_{j=1}^{n} w_{ij}, & i = j \\ 0, & i \neq j. \end{cases} \qquad \textbf{[S6]}$$

With the notation developed so far, the problem of identifying k coherent fluid regions can now be posed in terms of a normalized graph cut problem: Given a similarity graph $G = (V, E, W)$,

partition the graph nodes V into k clusters $A_1, A_2, \cdots, A_k$ such that the following (dis)similarity conditions hold. First, nodes within the same cluster have large edge weights between themselves; i.e., particles in a coherent fluid region have mutually short dynamical distances. Second, nodes in different clusters have small edge weights between themselves. In other words, particles in a coherent fluid region have long dynamic distances from the rest of the particles, particularly those located in the mixing region (i.e., noise cluster) that fills the space between the coherent fluid regions.

The normalized cut that implements the above (dis)similarity conditions can now be formulated in terms of a minimization problem as

$$\mathrm{Ncut}(A_i, \cdots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \bar{A}_i)}{\mathrm{vol}(A_i)}, \qquad \textbf{[S7]}$$

$$\mathrm{cut}(A_1, \cdots, A_k) = \frac{1}{2}\sum_{i=1}^{k} W(A_i, \bar{A}_i), \qquad \textbf{[S8]}$$

$$\mathrm{vol}(A_i) = \sum_{i \in A_i} \deg(v_i), \qquad \textbf{[S9]}$$

where $\bar{A}$ denotes the complement of set A in V. The minimization of the normalized cut exactly is an NP-complete problem, even for $k = 2$ clusters (64). The optimal relax solutions of the Ncut problem, however, can be approximated from the first k eigenvectors $u_1, \cdots, u_k$ of the following eigensystem (4):

$$(D - W)u = \lambda D u. \qquad \textbf{[S10]}$$

Here, k equals the number of eigenvalues that precede the largest gap in the eigenvalue sequence. The first k generalized eigenvectors $u_1, \cdots, u_k$ provide a reduced space in which k coherent fluid regions can be extracted using a simple *K*-means algorithm (5).

If a spectral gap does not exist in a given flow, then no coherent structures can be identified based on the trajectory-distance metric. This outcome, however, appears to be atypical in real-life unsteady flows. Indeed, on several challenging fluid flow examples discussed by Hadjighasem et al. (6), spectral clustering with the spectral gap condition has consistently outperformed other clustering-type methods, such as the transfer operator and C-means clustering approaches reviewed by Hadjighasem et al. (6). As with any open condition formulated for trajectories, pronounced spectral gaps remain robust under small perturbations and uncertainties due to the continuous dependence of trajectories on parameters over finite time intervals. The behavior of spectral gaps under larger perturbations is generally unknown.

2. The NCOM Simulation

The NCOM simulation uses assimilation and nowcast analyses from Navy Coupled Ocean Data Assimilation (NCODA) (7). Forecasts are generated by systems linking NCODA with regional implementations (8) of NCOM (9). The model has 1-km horizontal resolution and was initiated on May 15, 2012, from the then operational global ocean model Global Ocean Forecast System (GOFS) 2.6 (10). Daily boundary conditions are received from the current operational GOFS using the Hybrid Coordinate Ocean Model (HYCOM) (11). The vertical grid is composed of 49 total levels: 34 terrain-following σ levels above 550 m and 15 lower z levels. The σ-coordinate structure has higher resolution near the surface with the surface layer having 0.5-m thickness. The simulation uses atmospheric forcing at the sea surface from the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) (12) to generate forecasts of ocean state out to 72 h in 3-h increments. The observational data assimilated in these studies are provided by the Naval Oceanographic Office (NAVOCEANO) and introduced into NCODA via its ocean data quality control process. Observations are 3D variational (3D-Var) assimilated (13) in a 24-h update cycle with the first guess from the prior day NCOM forecast.

1. Haller G, Hadjighasem A, Farazmand M, Huhn F (2016) Defining coherent vortices objectively from the vorticity. *J Fluid Mech* 795:136–173.
2. Haller G (2016) Dynamically consistent rotation and stretch tensors from a dynamic polar decomposition. *J Mech Phys Sol* 86:70–93.
3. Hadjighasem A, Karrasch D, Teramoto H, Haller G (2016) Spectral clustering approach to Lagrangian vortex detection. *Phys Rev E* 93:063107.
4. Shi JB, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Machine Intell* 22:888–905.
5. Lloyd SP (1982) Least squares quantization in PCM. *IEEE Trans Inf Theor* 28:129–137.
6. Hadjighasem A, Farazmand M, Blazevski D, Froyland G, Haller G (2017) A critical comparison of Lagrangian methods for coherent structure detection. *Chaos* 27:053104.
7. Cummings JA (2005) Operational multivariate ocean data assimilation. *Q J R Meteorol Soc* 131:3583–3604.
8. Rowley C, Mask A (2014) Regional and coastal prediction with the relocatable ocean nowcast/forecast system. *Oceanography* 27:44–55.
9. Barron CN, Kara AB, Martin PJ, Rhodes RC, Smedstad LF (2006) Formulation, implementation and examination of vertical coordinate choices in the global Navy Coastal Ocean Model (NCOM). *Ocean Modell* 11:347–375.
10. Barron CN, Smedstad LF, Dastugue JM, Smedstad OM (2007) Evaluation of ocean models using observed and simulated drifter trajectories: Impact of sea surface height on synthetic profiles for data assimilation. *J Geophys Res* 112:C07019.
11. Metzger EJ, Smedstad OM, Carroll SN (2009) *User's Manual for Global Ocean Forecast System (GOFS) Version 3.0 (V3.0)* (Naval Research Laboratory, Stennis Space Center, MS), NRL Memorandum Report NRL/MR/7320-09-9175.
12. Hodur RM (1997) The Naval research Laboratory's Coupled Ocean/Atmosphere mesoscale prediction system (COAMPS). *Mon Weather Rev* 125:1414–1430.
13. Smith SR, et al. (2011) *Validation Test Report for the Navy Coupled Ocean Data Assimilation 3D Variational Analysis (NCODA-VAR) System, Version 3.43* (Naval Research Laboratory, Stennis Space Center, MS), NRL Memorandum Report NRL/MR/7320-11-9363.