 doi:10.20944/preprints202011.0605.v1

Article

A parameter-free spectral clustering approach to coherent structure detection in geophysical flows

Margaux Filippi[1,*], Irina I. Rypina[2], Alireza Hadjighasem[1] and Thomas Peacock[1]

[1] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Physical Oceanography Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA
* Correspondence: margaux@mit.edu

Abstract: In Lagrangian dynamics, the detection of coherent clusters can help understand the organization of transport by identifying regions with coherent trajectory patterns. Many clustering algorithms, however, rely on user-input parameters, requiring *a priori* knowledge about the flow and making the outcome subjective. Building on the conventional spectral clustering method of Hadjighasem *et al* (2016), a new parameter-free spectral clustering approach is developed that automatically identifies parameters and does not require any user-input choices. A noise-based metric for quantifying the coherence of the resulting coherent clusters is also introduced. The parameter-free spectral clustering is applied to two benchmark analytical flows, the Bickley Jet and the asymmetric Duffing oscillator, and to a realistic, numerically-generated oceanic coastal flow. In the latter case, the identified model-based clusters are tested using observed trajectories of real drifters. In all examples, our approach succeeded in performing the partition of the domain into coherent clusters with minimal inter-cluster similarity and maximum intra-cluster similarity. For the coastal flow, the resulting coherent clusters are qualitatively similar over the same phase of the tide on different days and even different years, whereas coherent clusters for the opposite tidal phase are qualitatively different.

Keywords: parameter-free spectral clustering; Lagrangian Coherent Structures; clusters; geophysical flows; unsupervised machine learning

1. Introduction

In geophysical fluid flows, the Lagrangian approach, where one follows fluid parcels as they move through time and space, provides a natural perspective to study the dynamics of motion and the patterns of transport [10,35,36,41,48]. Even simple, time-periodic velocity fields can generate complicated trajectories and chaotic motion [4,46]. It thus comes as no surprise that Lagrangian transport in realistic oceanic flows, which are aperiodic, can be incredibly complex.

The Lagrangian structures that organize transport and govern coherent trajectory patterns are referred to as Lagrangian Coherent Structures (LCS), a term introduced by Haller and Yuan [26]. These structures act as the hidden skeleton of a flow [40,44] that can be uncovered using techniques from the dynamical systems theory [26,60,62]. Recent review papers of LCS detection methods include [2,5,24,28].

LCS methods can be classified into those looking to identify regions (two- and three-dimensional in 2D- and 3D-flows, respectively) with coherent Lagrangian motion, and those looking for the boundaries (one- and two-dimensional in 2D- and 3D-flows, respectively) between such regions. The majority of clustering methods, including the parameter-free spectral clustering presented in this paper, fall into the former category, whereas the latter category includes, for example, Finite-Time Lyapunov Exponent (FTLE) and Finite-Size Lyapunov Exponent (FSLE) ridges [7,9,26,57], geodesic and variational transport barriers [14,15,22,28], Lagrangian-Averaged Vorticity Deviation (LAVD) and Rotationally Coherent Lagrangian Vortices (RCLV) [16,17], and diffusive transport barriers [29].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

Cluster-based LCS methods take root in the field of unsupervised machine learning. Several algorithms have been adapted recently to LCS detection in geophysical flows [20,23,59]. A recurring concern with these methods is that the results rely on a set of user-input parameters, making the outcome subjective. Hadjighasem et al. [24] present a comparison of the number of parameters required by several clustering methods to construct a Lagrangian field and show how spectral clustering requires the least amount of user inputs. The authors describe how to "choose a reasonable set of parameters for each method" in order to yield "the most favorable outcome". They describe a robust outcome as an outcome where "small variations in the parameters do not lead to drastic changes in the outcome", which is a fair way to describe robustness. The reasonability of the parameters and the favorability of the outcome, however, suggest that, in the end, the user will have to make decisions about the outcomes of the analyses that could be arbitrary and subjective. In this paper, building on the conventional spectral clustering method of HA16 (explained in Sections 2.1-2.2), we introduce the new, improved spectral clustering analysis (explained in Section 2.4) that does not require any user-input parameters. We also introduce a noise-based metric for quantifying the coherence of the resulting coherent clusters (Section 2.4), which allows comparing the clusters to each other within the same flow, as well as inter-comparing the clusters between different flows. The parameter-free spectral clustering approach is then applied in section 3 to two idealized analytical flows, the Bickley Jet and the asymmetric Duffing oscillator, and to a realistic oceanic coastal flow. In the last example, the identified clusters are tested using observed trajectories of real drifters. The results are discussed in section 4.

2. Materials and Methods

Consider an unsteady flow that is known, either from measurements or simulations, over a finite time window from t_0 to t_f within a finite spatial domain. Our goal is to separate the domain into coherent clusters according to their Lagrangian behavior. Note that the time window and the domain size should be considered part of the problem set up rather than parameters of the cluster detection method.

Below we review the conventional spectral clustering method, discuss its shortcomings associated with the need for user-input parameters, and then introduce a new and improved parameter-free spectral clustering method. We also review the Finite-Time-Lyapunov Exponents and Poincaré section techniques, which will be used in section 3 for comparisons with the spectral clusters.

2.1. Conventional spectral clustering method

The spectral clustering approach to LCS detection by [23], hereinafter referred to as HA16, partitions the dataset according to similarity between trajectories, such that intra-cluster similarity is maximized and inter-cluster similarity is minimized. The similarity is determined from the pairwise distances between trajectories. Reviews of spectral clustering can be found in Shi and Malik [58] and von Luxburg [61]. Below we briefly summarize the method.

Let $x_k^i = x^i(t_k)$ and $x_k^j = x^j(t_k)$ be two trajectories in $\mathbb{R}^d$, where $d = [2,3]$ for two- or three-dimensional flows, respectively, with positions at n discrete times from t_0 to $t_{n-1} = T$. The distance r_{ij} between trajectories is

$$r_{ij} = \frac{1}{t_{n-1} - t_0} \sum_{k=0}^{n-2} \frac{t_{k+1} - t_k}{2} \left(\left| x_{k+1}^i - x_{k+1}^j \right| + \left| x_k^i - x_k^j \right| \right) \tag{1}$$

with $|\cdot|$ denoting the spatial Euclidean norm. Next, convert the pairwise distances r_{ij} into the similarity weights w_{ij}:

$$w_{ij} = \begin{cases} 1/r_{ij} & \forall i \neq j, \\ \text{constant offset } w = const & i = j. \end{cases} \tag{2}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

From the weights w_{ij}, build the similarity graph $G = (V, E, W)$, with nodes $V = \{x^1, x^2, ...\}$, edges $E = \{e_{ij}\} \in V \times V$ between nodes x^i and x^j, and similarity matrix $W = \{w_{ij}\} \in R^{n \times n}$ that associates the weights w_{ij} to the edges e_{ij}. Next, define the sparsification radius r and sparsify W by removing all weights

$$w_{ij} \leq 1/r. \tag{3}$$

The advantage of sparsifying W is two-fold: first, it saves computational efforts by removing a large amount of entries; second, it minimized influences of entries with insignificant similarities.

The partition of the domain into k subsets that maximize the intra-cluster similarity and minimize the inter-cluster similarity can be achieved by minimizing the Ncut function, $\text{Ncut}(V_1, ..., V_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{\sum_{j \in V_i, l \in \bar{V}_i} w_{j,l}}{\text{vol}(V_i)}$, where $\bar{V}_i$ is the complement of V_i and $\text{vol}(V) = \sum_{i \in V} \deg(x_i)$ is the volume of set V, with $\deg(x_i) = \sum_{j=1}^{n} w_{ij}$. For large datasets, however, minimizing Ncut is computationally expensive. Therefore, an approximate solution is often used, which can be efficiently constructed with the eigenvectors $\vec{v}_i$ of an eigenvalue problem for the graph Laplacian $L = D - W$ (see Shi and Malik [58], von Luxburg [61] for derivations of this result):

$$L\vec{v} = \lambda D\vec{v}, \tag{4}$$

where the diagonal degree matrix, D, contains degrees, $\deg(x_i)$, for all nodes x_i. From perturbation theory and the Davis-Kahan theorem, which bounds the differences between eigenspaces of symmetric matrices under perturbations, the eigenvectors $\vec{v}_i$'s are close to the vectors from the exact Ncut partition. The order of eigenvectors and eigenvalues is meaningful and the first k eigenvectors $\vec{v}_i$'s can be used as approximate solutions for the k connected components of the graph. The optimal number of eigenvectors that are kept is retrieved by the heuristic argument of the spectral eigengap, i.e. the largest gap between successive eigenvalues λ_i of L. This number k_{gap} equals the number of eigenvalues before the eigengap. The resulting partition separates the k_{gap} connected components from the rest of the domain, which is assigned to the "incoherent background" subset. The wider the eigengap, the closer the approximate solution is to the exact solution of the graph cut minimization problem and the better the partition of the domain will be.

To extract the coherent clusters within the domain, the K-Means algorithm ([38]), described below), is applied with $K = k_{gap} + 1$ to the set $U = (v_1, ..., v_{k_{gap}})$. The output is an assignment of the N trajectories into k_{gap} coherent structures and one incoherent background cluster corresponding to the mixing region filling the space around the coherent sets.

The K-Means clustering method, sometimes referred to as 'Hard C-Means' in contrast to the *soft* Fuzzy C-means clustering, partitions the data into K mutually exclusive clusters. Suppose N trajectories $x^i = \{x^i(t = 0), ..., x^i(t = T)\}, 1 \leq i \leq N$ are given at discrete times $t \in [0, T] \subset \mathbb{N}$. The Euclidean distance D between x^i and x^j is $D(x^i, x^j) = \sum_{t=0}^{T} ||x^i(t) - x^j(t)||^2 dt = ||x^i - x^j||^2$. Given K clusters with moving centers $C_k = \{c^k(t = 0), ..., c^k(t = T)\}, 1 \leq k \leq K$, K-Means assigns trajectories to clusters by minimizing the distances from trajectories to cluster centers, i.e., by minimizing

$$\sum_{k=1}^{K} \sum_{i=1}^{N} ||x^i - C_k||^2. \tag{5}$$

For a given number K of clusters, with K determined *a priori*, the algorithm is iterated as follows:

1. initialize by randomly assigning a *centroid* to each cluster;
2. calculate the distances between each trajectory and cluster centroids, following 5;
3. assign each trajectory to the cluster with the closest centroid;
4. calculate the moving cluster centers as the mean over trajectories in each cluster: $C_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x^i$;
5. reiterate step (2) to compute the trajectory-to-cluster distance 5 and evaluate whether the improvement is below a threshold; if yes, output the clusters; if no, repeat steps 2-4 until a threshold is reached.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

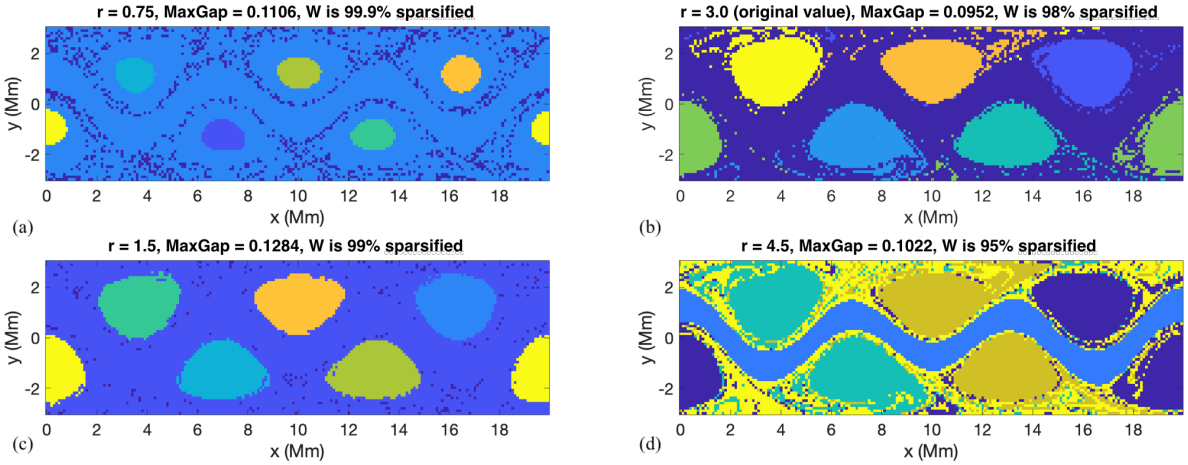


Figure 1. Example spectral clustering results for the Bickley Jet with parameters identical to those from HA16. The originally chosen value was $r = 3.0$ in panel (b). Three other values of r yielded a higher eigengap, however. Moreover, to follow the rule of thumb of 5–10% sparsification, $r = 4.5$ in panel (d) should be chosen instead. This value fails to detect the individual vortices, instead grouping them in pairs. It does, however, detects the meandering jet as an individual coherent structure. Looking at the maximum eigengap, the value of $r = 1.5$ in panel (c) should be chosen.

In the spectral clustering approach, the K-Means algorithm is not applied to the flow trajectories, but to the eigenvectors of the graph Laplacian L. The clustering of spectral vectors as opposed to a trajectory dataset is the key difference with other clustering methods for LCS detection, including the Fuzzy C-Means approach by [20].

2.2. Challenges of conventional spectral clustering

One important shortcoming of the conventional spectral clustering is the need for user-input parameters, including the sparsification radius r in equation (1) and the percentage of sparsification in W, both of which significantly affect the spectral clustering results. For r, HA16 advise to choose the r value which, first, sparsifies the weights matrix by 5-10% (so that 90-95% entries are removed during the sparsification step), and, second, the optimal r should correspond to the largest eigengap. As illustrated in Figure 1, however, the largest eigengap is often achieved at an r value that contradicts the 5-10% sparsification rule of thumb, and in many problems corresponds to the largest r considered. Thus, in the end, the choice of r is left entirely to the subjective judgement of the user and relies on the *a priori* knowledge of the system under study.

Another user input is the distance function and, subsequently, how it impacts the diagonal elements $w_{i,i}$ in W. [58] use a Gaussian kernel for the distance between nodes, a commonly-chosen function in spectral clustering ([42]), and choose an offset value for the diagonal of elements $w_{i,i}$ in W. HA16 use a different approach, where the distance function is described by equation (2) and the offset is added to W before computing D. This offset value in the diagonal elements is set to a large constant, which is analytically proven to be immaterial. In practice, however, the reader should be mindful of the step at which the diagonal offset is added, as the scalar chosen for this offset can impact the spectral clustering results. Very large values (towards 10^{16}) of the offset can introduce numerical errors due to the limits of double-precision. On the opposite, an offset value that is too small compared to the other w_{ij}'s entries of W can yield erroneous results. Unlike r, the offset is not a real physical parameter, but rather a numerical manipulation to avoid infinity in the diagonal. A convergence in the spectral decomposition is expected for large values of the offset, as long as these scalar values are numerically reasonable. 'Large' values, however, will depend on the flow: for this reason, the proposed approach verifies the convergence of the results and sets the offset value as a coefficient related to the maximum w_{ij} entry in W, as described in more detail below.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

2.3. Improved Parameter-Free Spectral Clustering

Here we propose an improvement to the conventional spectral clustering algorithm, which eliminates the need for user-input parameters. The optimal parameter choices are based on the notion of convergence (for $w_{i,i}$) and an argument involving the normalized spectral eigengap (for r). With $w_{i,i}$ and r defined in a user-independent fashion, sparsification number is no longer needed. With these modifications, the spectral clustering method becomes parameter-free.

The sparsification radius r defines the largest allowed distance between trajectories, because all weights corresponding to larger distances are removed during sparsification. Thus, r can be thought of as the parameter defining the size of the resulting clusters. It is tempting to use the same eigengap argument to define the optimal r, which was used to define the optimal number of clusters k_{gap}, namely, that the optimal r should yield the largest eigengap. The eigenvalues of the graph Laplacian, and thus the corresponding eigengap value, however, typically decrease with r, as more and more entries are removed from the weight matrix during sparsification. Thus, the r value that yields the largest eigengap is often the largest r considered. To overcome this issue and use the spectral gap arguments for choosing the optimal r, we introduce the *normalized spectral eigengap* by referencing the largest eigengap to the full dynamic range of the spectral values. Specifically, we define the normalized eigengap as the absolute value of the maximum eigengap divided by the difference between the largest and smallest eigenvalues, $\lambda_{max} - \lambda_{min}$:

$$\text{Normalized eigengap} = \frac{\left|\text{maximum eigengap}\right|}{\left(\lambda_{max} - \lambda_{min}\right)} \tag{6}$$

This is similar in spirit to the recently-proposed Normalized Maximum Eigengap by [43], but in our approach the denominator is automatically calculated from the spectral decomposition and does not include any user-input coefficient. As normalization does not change the physical meaning of the eigengap, large values of the normalized eigengap still yield spectral clusters with high intra-cluster similarity and low inter-cluster similarity. The optimal r should result in clusters with higher/lower intra-/inter-cluster similarity than those for both locally smaller or larger r values. Thus, the optimal r values should correspond to local maxima in the normalized eigengap. If several local maxima are present, we suggest proceeding with identifying coherent clusters for all of the corresponding r values. This will ensure that in flows with more than one dominant length-scale, all coherent clusters of different sizes will be recovered. Since a reliable estimate of a coherent cluster requires at least several trajectories within it, the smallest r considered should be dictated by the resolution, i.e., by the number of trajectories released in the domain. As a rule of thumb, we propose starting the r search with a value 10 times larger than the grid spacing between neighboring trajectories. Finally, r values approaching the domain size will result in the nearly entire domain being assigned to one cluster, and while this partition has high intra- and low inter-cluster similarity, it is not physically relevant. Choosing r values that correspond to the local maxima in normalized eigengap (and not the largest value of the normalized eigengap) avoids this problem.

We now describe the strategy for eliminating the need for choosing the user-input parameter $w_{i,i}$. In practice, the offset value chosen for the diagonal of W impacts the spectral clustering results. Being the artificial parameter introduced for numerical stability considerations, convergence of results is expected for increasing values of w (HA16). However, while w should be substantially higher than the rest of the weights, excessively large w's result in numerical errors due to the limits of double-precision. Insufficiently large values of the offset will also yield erroneous results. Finally, the offset depends on the flow under study, and what is large for one system may be small for the other. For this reason, we propose to seek the optimal flow-dependent offset value of the form $\max(w_{ij}) \times 10^n$ with integer n, and pick n which corresponds to the smallest value at which convergence of the results, specifically, convergence of the normalized spectral eigengap, is achieved.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

With w and r determined based on the normalized eigengap arguments as outlined above, and the number of clusters K defined by the number of eigenvalues before the eigengap ($K = k_{gap} + 1$), the new spectral clustering protocol becomes parameter-free.

2.4. Noise-based cluster coherence metric

How coherent are the clusters that result from the spectral clustering partition? The eigenvalues of the graph Laplacian L contain information about coherence of eigenvectors but not of spectral clusters, because there is no one-to-one correspondence between former and latter. The spectral eigengap value, on the other hand, contains information about the overall skill of the partition, but not about coherence of individual clusters. In many applications, such as identifying an optimal field experiment design or a strategy for a search and rescue operation, it is the coherence of individual clusters that determines the drifter release strategy or the search pattern. To determine how coherent the clusters are, we define below one possible coherence metric that is based on the robustness of the clusters with respect to a small random perturbation. This applied numerical noise can be representative, for example, of uncertainties in ocean circulation models. Our metric is similar, in spirit, to the approach by [19].The robustness to noise is computed as follows. For trajectories starting at t_0 inside a given cluster, randomly perturb their final positions at time t_f, and advect trajectories backward in time to the start time t_0. The percentage of trajectories that return within the boundaries of their cluster is the proposed coherence metric. In all numerical simulations, we use the noise magnitude equal to $\frac{1}{100}$th of the trajectory grid spacing (i.e., distance between neighboring trajectories) but the results are similar for slightly larger or smaller perturbations. Essentially, our noise-based coherence metric favors coherent sets with large area-to-perimeter ratio and thus punishes small and filamentated clusters.

2.5. Parameter-free Spectral Clustering Algorithm Summary

The application of the parameter-free spectral clustering consists of 3 steps:

1. r-sweep: compute the normalized eigengap for variable sparsification radii r, from 10*trajectory spacing to the domain size, keeping the fixed offset value ($w = \max(w_{ij}) \times 10^7$ worked well in all examples), and identify all local maxima
2. w-sweep: for all local maxima, verify the convergence in the normalized eigengap by varying $w = \max(w_{ij}) \times 10^n, n = 2, ..., 10$
3. identify coherent sets corresponding to all local eigengap maxima and compute noise-based coherence metrics for the resulting clusters

2.6. Finite-Time Lyapunov Exponents and Poincaré sections

In the next section, we also use Finite-Time Lyapunov Exponents (FTLE) and Poincaré sections to provide insight about the flow and to compare with the coherent clusters obtained using parameter-free spectral clustering. For completeness, below we briefly describe both methods.

Poincaré section is a technique that allows mapping the chaotic and non-chaotic (regular) regions of the domain in time-periodic flows by stroboscopically sampling trajectories after each period of the perturbation. On the Poincaré section, regular trajectories appear as discretely-sampled smooth curves, and chaotic trajectories appear as clouds of dots covering finite areas.

FTLE approach ([26,28,57]) is probably the most widely used LCS detection method due to its intuitive nature and numerical robustness. FTLE fields can be computed by the differentiation of the flow map, $\vec{F}_{t_0}^{t} := \vec{x}_j(t_0, x_0; t)$, obtained from numerical trajectories $\vec{x}_j(x_0, t_0; t)$ that start at x_0 and advected by the flow over a finite integration time $t \in [t_0; t_1]$. Specifically, the gradient of the flow map is used to compute the Cauchy-Green strain tensor, $C_{t_0}^{t_1}(\vec{x_0}) = \left[\nabla \vec{F}_{t_0}^{t_1}(\vec{x_0})\right]^T \left[\nabla \vec{F}_{t_0}^{t_1}(\vec{x_0})\right]$, whose largest eigenvalue, $\lambda(\vec{x_0})$, quantifies the largest amount of stretching between neighboring trajectories and is then used to construct the FTLE field $\Lambda_{t_0}^{t_1}(\vec{x_0}) = \frac{1}{t_1 - t_0} \log \sqrt{\lambda(\vec{x_0})}$. Maximizing ridges of the FTLE field reveal the locally most repelling LCSs (provided that particle separation is dominated by strain and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

not shear; see Haller [28]), which are the finite-time counterparts of the invariant stable manifolds of hyperbolic trajectories in time-periodic flows. Similarly, maximizing ridges of the backward FTLEs reveal the most attracting LCSs.

3. Results

In this section, the new parameter-free spectral clustering method is applied to three flows with increasing complexity: the analytically-prescribed Bickley Jet (with coherent sets of the same size), the analytically-prescribed asymmetric Duffing oscillator (with coherent sets of different sizes), and a numerically-generated realistic coastal flow from the ocean circulation model MSEAS. In the latter case, spectral clusters from a model are compared against trajectories of real drifters.

3.1. Bickley Jet

The Bickley jet flow consists of a zonal jet on which 2 traveling Rossby waves are superimposed. Following [47], we use the streamfunction

$$\begin{aligned} &\psi(x,y,t) = \psi_0(x,y) + \psi_1(x,y,t) \quad \text{with} \\ &\psi_0(y) = -U_0 L \tanh\left(\frac{y}{L}\right); \quad \psi_1(x,y,t) = U_0 L \,\mathrm{sech}^2\left(\frac{y}{L}\right) \times \sum_{i=1}^{N=2} \epsilon_i \cos[k_i(x - c_i t)] \end{aligned} \tag{7}$$

where $U_0 = 62.66$ ms^{-1} is the velocity at the jet core; $L = 1770$ km is the characteristic jet width scale; $k_n = \frac{2n}{r_e}$ are wavenumbers with r_e the Earth radius, $c_1 = 0.205U_0$ and $c_2 = 0.461U_0$ are wave phase speeds; and $\epsilon_1 = 0.15$ and $\epsilon_2 = 0.30$ are wave amplitudes. In the reference frame moving with c_2, the flow consists of a steady background and a time-periodic perturbation:

$$\tilde{\psi} = c_2 y - U_0 L \tanh\left(\frac{y}{L}\right) + U_0 L \,\mathrm{sech}^2\left(\frac{y}{L}\right)\left[\epsilon_1 \cos(k_1 x - \sigma_1 t) + \epsilon_2 \cos(k_2 x)\right] \tag{8}$$

with $\sigma_1 = k_1(c_1 - c_2)$. The Poincaré map (computed over 1000 perturbation periods) and the forward-/backward- FTLEs (computed over 30 perturbation periods) for this flow are shown in Figure 2a. The regular (i.e., non-chaotic) meandering jet separates two chaotic zones to the north and south, with 3 regular vortices embedded into each chaotic zone. Between the vortices there exist 3 hyperbolic trajectories with stable and unstable manifolds that form heteroclinic tangles. HA16 also the Bickley Jet as a benchmark flow but with an extra (3rd) wave with $\epsilon_3 = 0.0075$ and $c_3 = c_2 + (\sqrt{5} - \frac{1}{2}) \times \frac{k1}{k3}(c_1 - c_2)$ in (13). We chose the 2-wave Bickley Jet for the benefit of comparing spectral clusters with the Poincaré section.

Following the proposed parameter-free spectral clustering approach, we performed sweeps in the sparsification radius r (from 0.5 to 4.25 with an incremental step of 0.25, and with finer steps around the detected peaks) and sweeps in the offset coefficient w ($w = \max\{w_{ij}\} \times 10^n$ with integer n from 1 to 11). The resulting normalized eigengap shows 3 local maxima in r (0.90, 1.25 and 2.0), yielding $K = 6$ clusters in all 3 cases. For all r, the normalized eigengap converges at $n = 7$. (See Figures A1a and A1b in Appendix A.1.) The coherent clusters, color-coded by their noise-based coherence metric, for all 3 peaks in r are shown in Figure 3. In all cases, the method successfully detected 6 vortices centered at each of the 6 regular islands in the Poincaré map. The sizes of the detected clusters grow with r, which is fully consistent with our interpretation of r as the parameter responsible for the cluster size. For $r = 0.9$ and 1.25, vortices are smaller and less filamented than for $r = 2$. In the latter case, each vortex contains one long and narrow filament. Note, however, that when advected to the final time t_f, identified coherent vortices do not get filamented any stronger than at t_0, whereas for an arbitrary subset, the degree of filamentation generally increases with time. The noise-based coherence metric suggests high coherence (>90%) for all of the detected coherent clusters, as well as for the incoherent background. The coherence is lower for smallest clusters (corresponding to $r = 0.9$) due to their smaller area-to-perimeter ratios. For $r = 0.9$ and 1.25, the coherence metric for the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

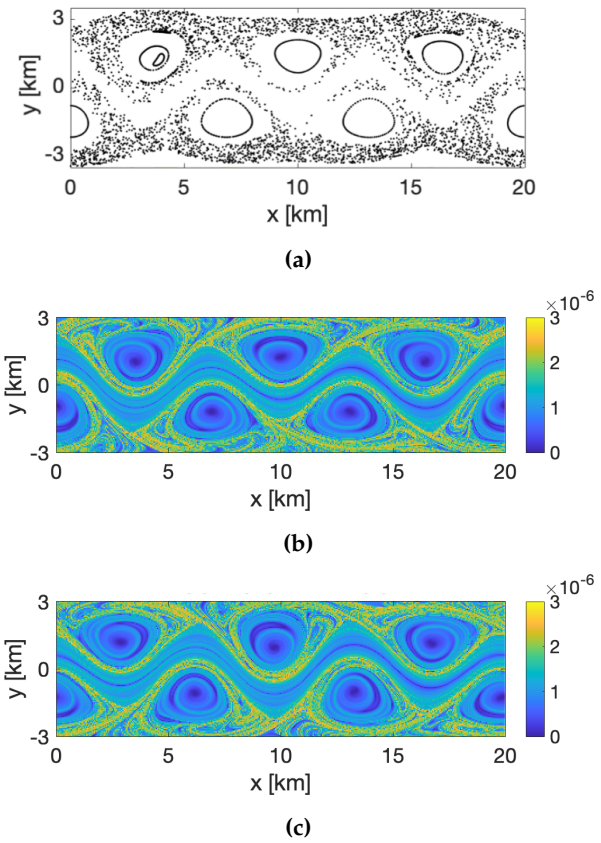


Figure 2. (a) Poincaré map for the periodic Bickley Jet flow, computed over 1000 perturbation periods. (b) Forward- and (c) Backward-FTLE field, computed over 30 peturbation periods.

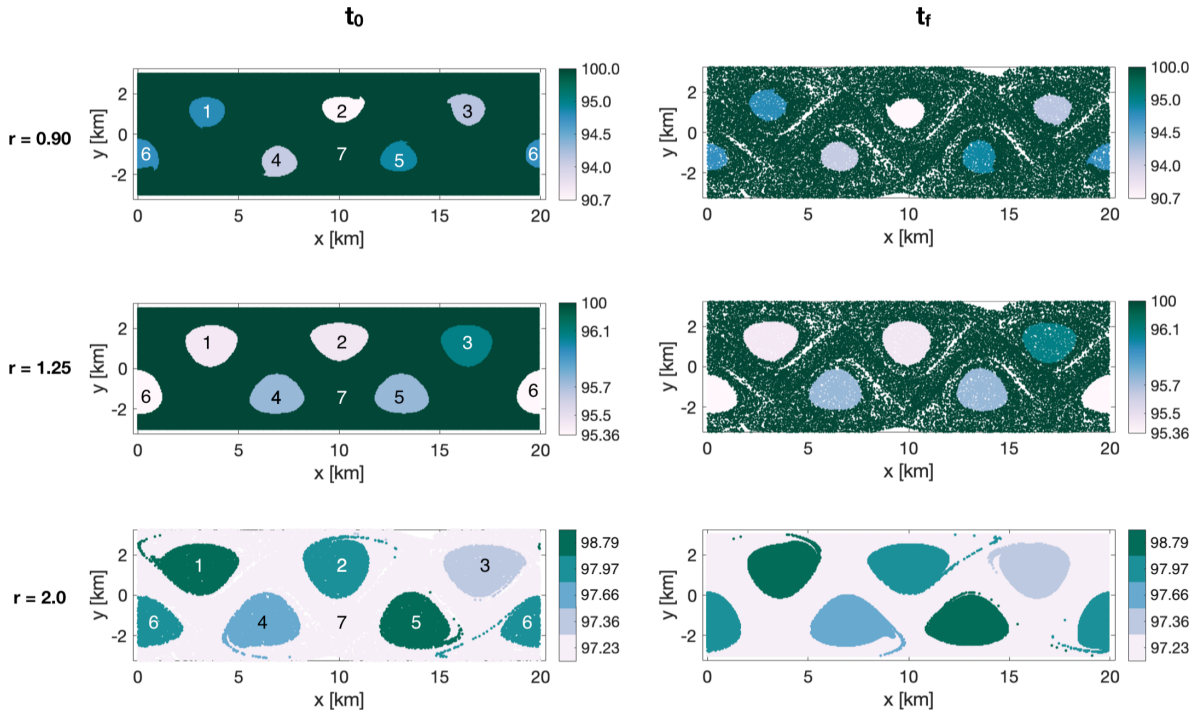


Figure 3. Coherent clusters color-coded by their coherence metrics resulting from the parameter-free spectral clustering for the Bickley Jet flow. The initial (t_0 - left) and final (t_f - right) positions are shown.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

incoherent background exceeds that of the 6 vortices and nearly reaches 100%, but at $r = 2.0$ it drops to a value that is below that for the vortices. This is likely due to a combination of the filamentation and smaller area-to-perimeter ratio for the incoherent background at $r = 2.0$. Advecting the clusters beyond the final time t_f of the integration window shows that the cluster with the lowest coherence metrics (cluster 3, 97.36%) starts deteriorating before the other clusters, as early as 40.5 days, whereas cluster 5, with the highest coherence metrics of 98.79%, starts deteriorating around 46.5 days.

3.2. Asymmetric Duffing oscillator

To test the method on a system with coherent sets of different sizes, we use an asymmetric Duffing Oscillator. The Duffing oscillator is another commonly-used benchmark flow that consists of two gyres with the same sign of rotation located on opposite sides of the hyperbolic point at the origin [55]. The classic Duffing Oscillator has two identical gyres of same size. Here, we use a longitudinal asymmetry to generate a right left gyre that is smaller than the left gyre. When the gyres oscillate around their mean position periodically with time, the stable and unstable manifolds of the hyperbolic trajectory form a homoclinic tangle that induces chaos in a figure-eight-shaped chaotic region around the gyres. The asymmetric Duffing oscillator velocity is two-dimensional, incompressible and is described by a streamfunction

$$\psi(x, y, t) = [1 - \epsilon \cos(\omega t + \phi)]\left(\frac{x^2}{2} - \frac{ax^4}{4}\right) - \frac{y^2}{2} \tag{9}$$

with

$$\epsilon = 0.1 \times \frac{x + 4.5}{2}.$$

Here, $\epsilon = 0.1$, $\omega = \frac{3\pi}{2}$, $\phi = \frac{\pi}{4}$ and $a = 0.5$. The corresponding Poincaré map and FTLE fields are shown in Figure 4.

To identify coherent sets we again followed the parameter-free spectral clustering algorithm describes in Section 2. The analysis revealed one peak in the normalized eigengap at $r = [1.0]$ (Figure A2a in Appendix A.2) with $k = 2$ eigenvalues before the eigengap. Convergence of the normalized eigengap with respect to the increasing values of the offset parameter is achieved at $w = \max\{w_{ij}\} \times 10^7$ (Figure A2b in Appendix A.2), the same value of $n - 7$ as for the Bickley Jet. The resulting two clusters plus the incoherent background, color-coded by their noise-based coherence metrics, are shown in Figure 5. As expected, the clusters correspond to the two gyres, the larger gyre on the left and smaller on the right. Both coherent clusters, but especially the right cluster, in this example are more filamented than for the Bickley Jet flow. Nevertheless, the cluster coherence is still high, about 92 and 97%, which is above that of the incoherent background (about 86%). Interestingly, at initial time t_0,the filaments are approximately aligned with the stable manifolds (black curve in Figure 5, left), outlining the turnstyle lobes in such a manner that, when advected forward to t_f, filamentation of the clusters does not increase. Overall, in this example the parameter-free spectral clustering performed well at identifying coherent clusters of slightly different sizes and partitioning the domain into subsets with high intra- and low inter-cluster similarity. The application of the parameter-free spectral clustering to the flow where cluster sizes are very different is shown in Appendix B.

3.3. Geophysical example: flow around No Man's Land island

In 2017 and 2018, two field experiments took place in coastal waters south of Martha's Vineyard, MA as part of the NSF-funded Advanced Lagrangian Predictions for Hazards Assessments (ALPHA) project. The goal of the project was to develop and test Lagrangian methods for the prediction, mitigation and response to environmental hazards (for example, oil spills), as well as to aid in the search and rescue operations. The experiments consisted of releasing surface drifters targeting the LCS that were predicted in near-real time based on velocity forecasts from a high-resolution ocean circulation model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

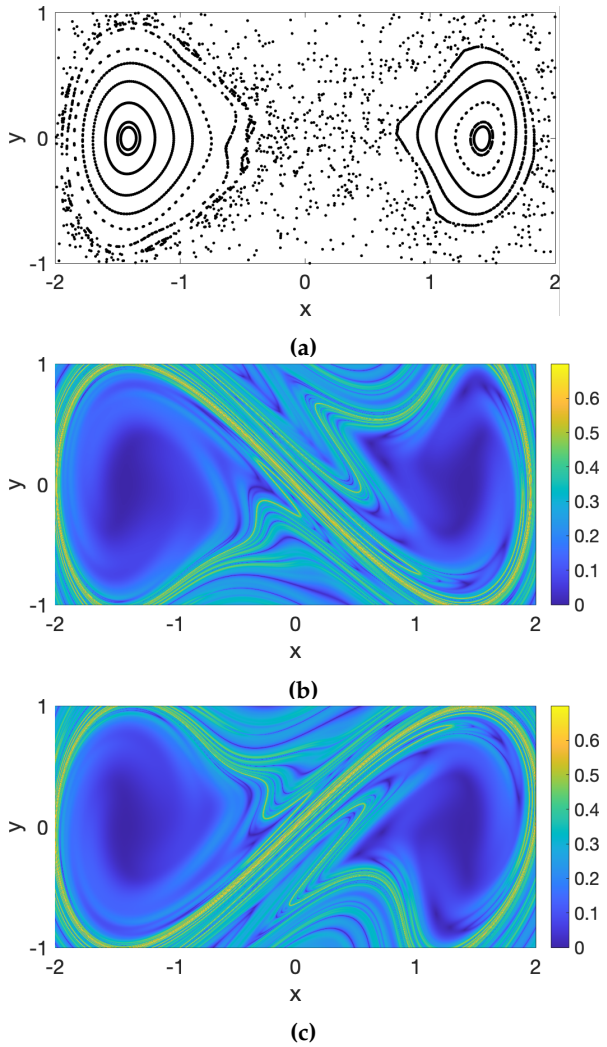


Figure 4. The asymmetric Duffing oscillator. (a) Poincaré map with 20 periods of perturbation T_{pert}. (b) Forward- and (c) Backward- FTLE for $10T_{pert}$.

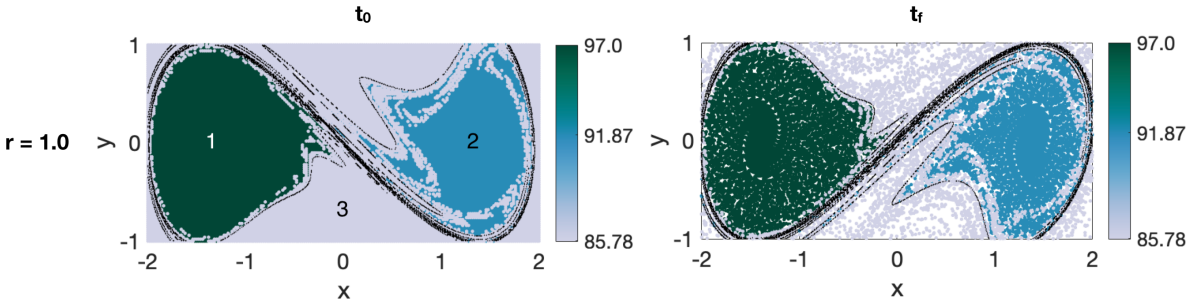


Figure 5. Coherent clusters, color-coded by their coherence metrics, resulting from the parameter-free spectral clustering for the asymmetric Duffing Oscillator flow. The initial (t_0 - left) and final (t_f - right) positions are shown. For comparison, black curves show the FTLE ridges in forward-time at t_0 and in backward time at t_f. The spectral clustering was done for $30T_{pert}$; FTLE ridges were computed for $10T_{pert}$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

Our specific area of interest was the region around a small uninhabited No Man's Land island located approximately 5 km south of the western end of Martha's Vineyard (Figure 6). The depth of the channel between Martha's Vineyard and No Man's Land is about 10 meters on average, with steep gradient on each side. East and west of No Man's Land, the depths increase to 25 m over a few kilometers. The flow south of Martha's Vineyard is strongly affected by wind and tides [52,53].

The drifters used in our study were the same as in [52], i.e., the Coastal Dynamics Experiment (CODE) type, also called Davis type, developed by [11] and manufactured by metOcean telematics. The model used was the primitive equation Multidisciplinary Simulation, Estimation, and Assimilation Systems (MSEAS) model developed at MIT [30,31]. MSEAS was configured with a two-way nesting: the domain around the continental shelf had a 600-meter resolution and the domain around Martha's Vineyard had a 200-meter resolution. The model was forced by the atmospheric NCEP flux forecasts and tidal forcing from the Oregon State University barotropic tide model adapted to the bathymetry around Martha's Vineyard. MSEAS was initialized with historical data from the National Marine Fisheries Service (NMFS) for conductivity, temperature and depth (CTD), and the sea surface temperature data provided by the Johns Hopkins University's Applied Physics Lab (JHU APL). About a week prior to each field experiment, hydrographic surveys were conducted in the area of interest south of Martha's Vineyard, and the collected CTD measurements were used to adjust the model boundary conditions with the observed ocean conditions. Details about surface forcing, and initial and boundary conditions for these runs can be found on the MSEAS website http://mseas.mit.edu/Sea_exercises/NSF_ALPHA.

MSEAS was re-run in a hindcast mode using the observed NCEP reanalysis wind forcing (instead of forecasted winds) after the end of the field experiments. The hindcast runs are qualitatively similar to the forecast runs in all respects. Quantitatively, MSEAS hindcasts for both 2017 and 2018 generally agree better with the observed real drifter motion than the forecasts, although the short-term strong-wind events (discussed in more detail below) that occurred during the 2018 field experiment and that were not well-represented by the NCEP reanalysis, led to the poorer agreement between model and drifters in 2018 compared to the 2017 experiment. For brevity, in this paper we only present the analyses of the MSEAS hindcasts, but the forecast-based results obtained in real time during the field experiments are qualitatively similar.

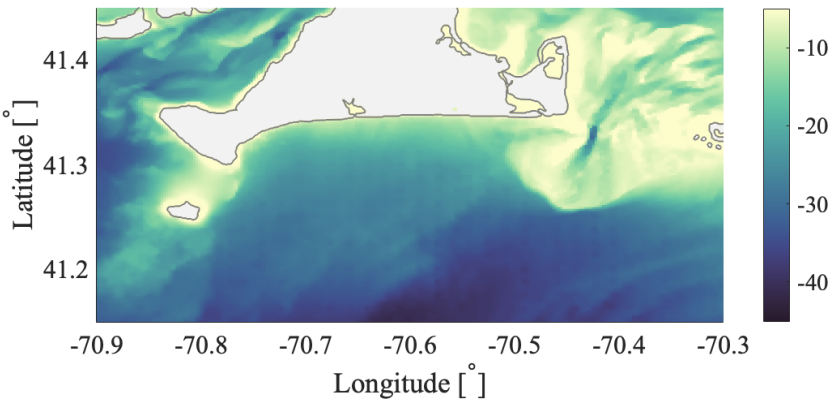


Figure 6. Bathymetry of the 200-m resolution model domains around Martha's Vineyard, extending south towards the continental shelf break. Depths in meters. Data from MSEAS. Large island is Martha's Vineyard; small island near 70.82W and 41.25N is No Man's Land.

For ease and speed of Lagrangian calculations, the TRACE web-based gateway [3] (http://transport.me.berkeley.edu/thredds/catalog/public/MIT-MSEAS/catalog.html) was used to compute trajectories of simulated drifters advected by MSEAS currents, as well as the FTLE fields. The trajectory integration time in this study was 6 hours, which is an important timescale for a flow with a strong M2 tidal component, and also represents the upper time limit in the real person-lost-at-sea search-and-rescue operations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

3.3.1. 2017 experiment

On August 14, 2017, we released 14 drifters at 10 locations (i.e., triplets were released at 2 locations - station #4 from the west and the southernmost station in Figure 7.a) in a circumvent pattern around No Man's Land. Deployment took 1.5 hours, with the last drifter hitting water at 15:51. The deployment strategy targeted different coherent regions delineated by the separating LCSs. The 360-degree circumvent pattern around the island ensured that drifters would be deployed on opposite sides of any LCS that protrudes from any point on the island in any direction, thus mitigating uncertainties in the exact positions of model-based LCSs. Figure 7 shows the drifter deployment locations and 6-hour long drifter trajectories, as well as the forward- and backward-FTLE fields and the coherent clusters identified using the parameter-free spectral clustering method.

For a unidirectional flow that impinges on an island from any direction, a hyperbolic trajectory is expected to form on the side of the island facing the inflow and on the opposite side facing the outflow. From the former hyperbolic trajectory, a stable manifold is expected to protrude into the incoming flow, separating trajectories passing the island on opposite sides. The real oceanic flow is not unidirectional and, having a strong tidal component, changes significantly over the 6-hour analysis window. Nevertheless, consistent with the expectations described above, forward FTLEs show a strong repelling ridge extending from No Man's Land across the eastern side of the channel toward Martha's Vineyard (Figure 7.a). (Some indication of multiple backward FTLE ridges is also seen on the western side of the No Man's Land (Figure 7.b) but these are not as strong as the above-mentioned forward FTLE ridge.) As we go from west to east across the red forward FTLE ridge in Figure 7.a, 4 deployment locations (black dots) lie to the west of it and 3 to the east, with the remaining 3 deployment locations further away from it on the southern side of the island. The drifters on opposite sides of this ridge are expected to have qualitatively different Lagrangian motion, and indeed, trajectories of the three drifters on the eastern side of this repelling ridge are significantly different from the rest in that they barely move over the subsequent 6 hours and are by far the shortest. This is also consistent with the smaller FTLE values in this region (fainter red) indicating a more quiescent character of the flow. Multiple additional forward FTLE ridges are seen west and north of the No Man's Land, which lead to the complicated behavior of drifters in this quadrant. Specifically, out of the 6 drifters deployed there, the westernmost drifter did not move much, the 2 northernmost drifters strayed far north, and the easternmost triplet of drifters looped around and came back.

The coherent clusters obtained using the parameter-free spectral clustering method are shown in Figure 7(c-d). The sweep with increasing r (Figure A3 in Appendix A.3) revealed 2 peaks in the normalized eigengap - at 0.0125 and 0.0425 deg (1.4 and 4.7 km). With respect to the offset, convergence in normalized eigengap was again achieved at $w = \max\{w_{ij}\} \times 10^7$ for all r. The first peak at $r = 1.4$ km yielded two coherent clusters: a less coherent purple one (78.6% coherence metric) located in the dynamically-active region inside the channel, and a more coherent blue one (97.8% coherence metric) in the most quiescent southeastern corner of the domain; the rest of the domain was assigned to the green incoherent background (which was in fact very coherent with >99% coherence metric). The purple cluster is roughly delineated by the 2 forward FTLE ridges - a stronger one on the east and a weaker one on the west. When advected forward in time for 6 hours, the purple cluster moves northward and shrinks in area, suggesting that convergence of the surface flow could be an important physical mechanism responsible for the existence of this coherent cluster. The second peak, at $r = 4.7$ km, yielded one large coherent cluster (97.5% coherence) occupying the majority of the quiescent eastern part of the domain, with the rest of the domain assigned to the incoherent background (again with >%99 coherence).

Comparing the motion of real drifters against the model-based spectral clusters, the purple cluster in the channel did not hold the triplet of drifters that were released there. After initially heading northwest with the shrinking purple cluster, all 3 drifters looped around and headed back southeast, ending up in the green incoherent background not far from their release location after 6 hours. This is most likely due to a combination of the uncertainties in the numerically-simulated ocean currents

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

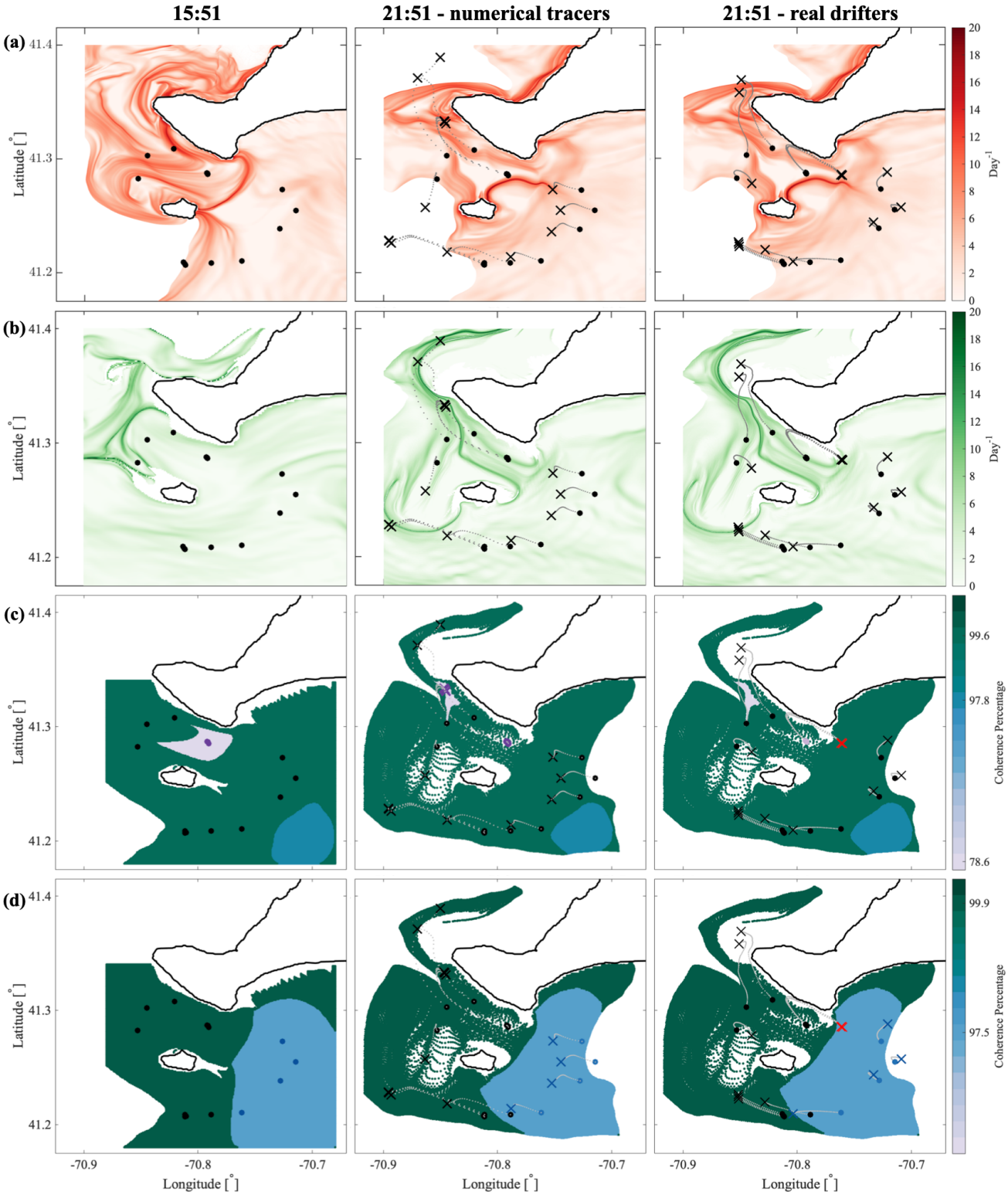


Figure 7. LCS, numerical tracers and experimental drifter positions at the start (15:51 UTC) and end (21:51 UTC) of the 2017 experiment on August 14. (a) Forward and (b) backward FTLE. (c-d) Spectral clusters with coherence metric. The drifters in c-d are color-coded according to the cluster to which they belong, with red crosses if their final positions was outside of their initial cluster.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

(see Figure A10 in Appendix C.1 for the differences between the real and simulated drifters), the most dynamic character of the flow in this region with high sensitivity to the drifters deployment locations, and the fact that the purple cluster was the least coherent out of all clusters (only 78% coherent) and thus is most sensitive to noise. All other drifters were deployed in the green region in Figure 7.c, and all stayed there after 6 hours. With respect to the clusters in Figure 7.d, 4 drifters were released within the blue coherent cluster and all 4 stayed there. The rest of the drifters were deployed in the green background region; 7 stayed there, whereas one triplet switched from green to blue.

3.3.2. 2018 experiment

The 2018 No Man's Land experiment took place on August 7. A similar circumvent deployment route around No Man's Land was used, but with 18 drifters, instead of 14 as in 2017. The deployment was completed at 16:00 so that over the subsequent 6 hours the drifters would sample the same part of the tidal cycle as in 2017.

Intermittent eastward wind gusts of up to 25 m/s started to occur shortly after the drifter deployment, from 16:00 until 20:00. (Measured wind time series for the Martha's Vineyard Airport station for 7/8/2018 can be found on The Weather Underground website). These wind gusts were absent in the NCEP wind forecasts and the corresponding MSEAS model forecasts, and were under-represented in the NCEP hindcast wind reanalysis (due to temporal intermittency and spatial variation of winds around Martha's Vineyard and No Man's Land) and the corresponding MSEAS hindcast currents. The wind gusts strongly affected the drifter trajectories, caused beaching of 2 drifters, and led to the generally poor agreement between the real and simulated drifter trajectories over the first 6 hours of the experiment (Figure A11 in Appendix C.2). Agreement between the real and simulated drifters had improved after the end of the wind gust period (Figure A12), but the 6-hour time interval from 20:00 on August 7 until 02:00 on August 8 corresponds to a different part of the tidal cycle and thus cannot be directly compared to the 2017 results. Analysis of the same tidal phase of the next tidal cycle, i.e., 04:00-10:00 on August 8, is possible but by then all drifters had moved east of No Man's Land (Figure A13), and the initial elliptical deployment pattern had been distorted into a geometrical configuration that was sub-optimal for testing the agreement with spectral clusters. Thus, the comparison of the 2018 spectral clusters with real drifters and with the results from the 2017 experiment is challenging, no matter what time window is chosen. Here, we present the analysis for 3 different time windows, and we learn something from each.

The results for the 3 different 6-hour time windows, 16:00-22:00 on August 7; 20:00 on August 7 - 02:00 on August 8; and 04:00-10:00 on August 8, are presented in Figures 8 - 10. The format of each figure is akin to the 2017 results in Figure 7: left and right panels correspond to the start and end of each time window, the top two rows show forward and backward FTLEs, and the bottom rows show spectral clusters.

For the 16:00-22:00 and 04:00-10:00 time windows (8 and 10), the FTLE fields are qualitatively similar to each other and, in many respects, to those from 2017 (Figure 7). A strong repelling forward FTLE ridge is seen extending from No Man's Land to Martha's Vineyard across the eastern part of the channel (panel a-left). Additional ridges and the overall elevated FTLE values are seen in the northwestern quadrant of the domain, reflecting on the complex Lagrangian behavior and strong sensitivity to deployment locations in this region. The opposite southeastward quadrant shows the smallest FTLEs indicative of the more quiescent character of the flow with lesser drifter separation. For this tidal phase in 2018, similar to 2017, the parameter-free spectral clustering results have 2 peaks (at $r = [0.0225, 0.0425]$ for 16:00-22:00 and $r = [0.03125, 0.04625]$ for 04:00-10:00, with convergence in w at the same $n = 7$ coefficient as in all other examples – see supplementary Figures A4– A6).

At 16:00, the first peak revealed 3 clusters – a cluster in the channel and two clusters to the south-southwest of No Man's (Figure 8). The channel cluster is qualitatively similar to the 2017 purple cluster. The rest of the domain is assigned to the green around-cluster background. When advected forward over 6 hours, the clusters to the south-southwest of No Man's Land moves anticyclonically

 doi:10.20944/preprints202011.0605.v1

around the island and into the channel, whereas the purple channel cluster moves northward, hugging the tip of Martha's Vineyard and shrinking in size, which is again qualitatively similar to 2017. The second peak revealed a large cluster in the center-eastern part of the domain that encompasses the dark green and turquoise coherent clusters from the first peak, No Man's Land and parts of the purple cluster. Comparing the spectral clusters for the 16:00-22:00 time window against real drifters, the agreement is poor as all drifters released within a cluster ended up in the incoherent background, very likely due to the unresolved wind gusts in the model.

The 04:00-10:00 time window (Figure 8) corresponded to a similar phase of the tidal cycle and each peak in r yielded results that had many qualitative similarities to the 16:00-22:00 time window. The first peak again revealed a purple channel cluster and a second, turquoise, cluster south of No Man's Land. After 6 hours, the channel cluster moves northward around Martha's Vineyard and the turquoise cluster moves anticyclonically around No Man's Land and into the channel. The second peak revealed a large, highly coherent cluster shown in dark teal that encompasses parts of both clusters from the first peak. In both cases, l7 drifters were initially located in the green inter-cluster background and 1 drifter started in the channel, in the purple cluster in (c) and the dark teal cluster in (d). All drifters stayed within their respective cluster, but since only 1 drifter was initially inside the purple cluster, the comparison might not be statistically robust.

Looking at the result over the different phase of the tide, from 20:00 to 04:00 UTC in Figure 9, we see a qualitatively different geometry both in the FTLE fields and spectral clusters. The parameter-free spectral clustering has 2 peaks in the normalized eigengap - at $r = [0.0325, 0.0375]$. Both have very large number of clusters (17 and 18, respectively), and both do not identify the inter-cluster background. This large number of clusters suggests that over this tidal period, no part of the domain is significantly different from the rest of the domain in terms of Lagrangian connectivity, unlike in all 3 other Martha's Vineyard cases that consistently show 1 to 3 coherent clusters embedded into the inter-cluster background. For the 20:00-04:00 time window, the comparison with real drifters showed mixed results. The four drifters released within the channel clusters in Figure 9(a-d) stayed within their assigned channel clusters, but otherwise, about half of the drifters left their assigned clusters by 04:00.

4. Discussion

A reliable method for identifying coherent clusters in oceanic flows is important for a variety of applications – from understanding of mixing and exchanges of biogeophysical tracers, to hazard mitigation and search-and-rescue operations. In studies of exchange processes, coherent clusters can help visualizing tracers that stay coherent over time. In hazard mitigation applications, clusters are useful for identifying areas where mitigation assets should be focused, and in the search-and-rescue operations, cluster geometries help narrow down on the optimal search and rescue pattern. The conventional spectral clustering algorithm has many of the desired qualities, yielding clusters that maximize inter-cluster similarity and intra-cluster differences of the underlying Lagrangian motion. Spectral clustering also has advantages over other techniques for LCS identification, such as FTLEs, in applications with sparse Lagrangian data. The major drawback of the conventional spectral clustering is the need for subjective user-defined choices for several method parameters, which requires *a priori* knowledge about the flow. The parameter choices vary between different flows and depend on flow characteristics such as intensity, spatial variability, and dominant spatio-temporal scales of the currents, which might not be known in real applications. Different parameter choices generally yield substantially different coherent clusters, which drastically complicates the applicability of this method.

In this paper, we describe a new parameter-free spectral clustering algorithm, which automatically identifies optimal parameters for a given flow and does not require any user-input choices. The specific parameters of interest are the sparsification radius r, which is responsible for the sizes of the identified clusters, and the offset parameter $w_{i,i}$, which is responsible for the numerical stability of the algorithm. Automated selection of optimal r and $w_{i,i}$ requires modifications of the conventional spectral clustering.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

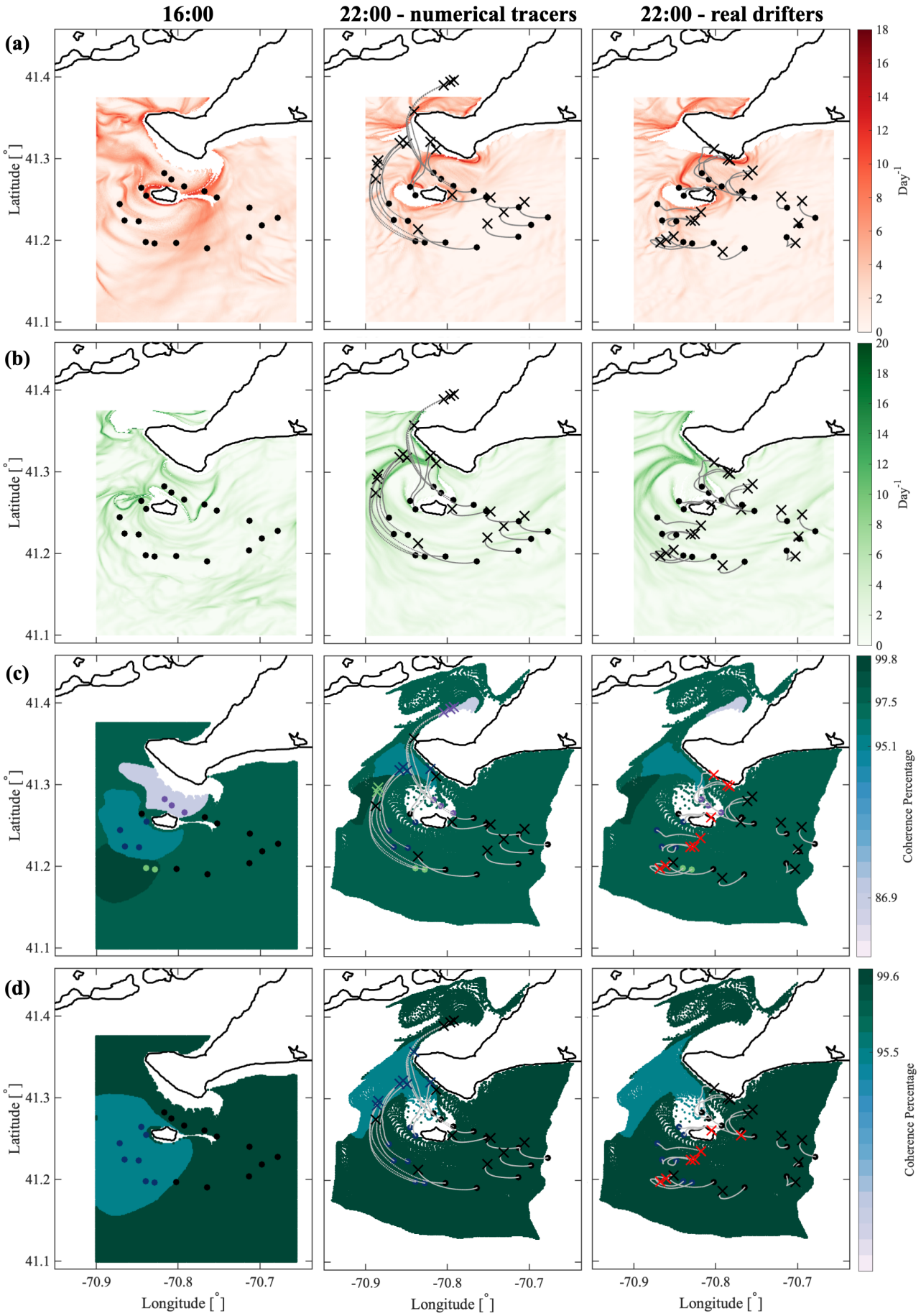


Figure 8. LCS, numerical tracers and experimental drifter positions at the start (16:00 UTC) and end (22:00 UTC) of the 2018 experiment on August 7. (a) Forward and (b) backward FTLE. (c-d) Spectral clusters with coherence metric. The drifters are color-coded according to the cluster to which they belong, with red crosses if their final positions was outside their initial cluster.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

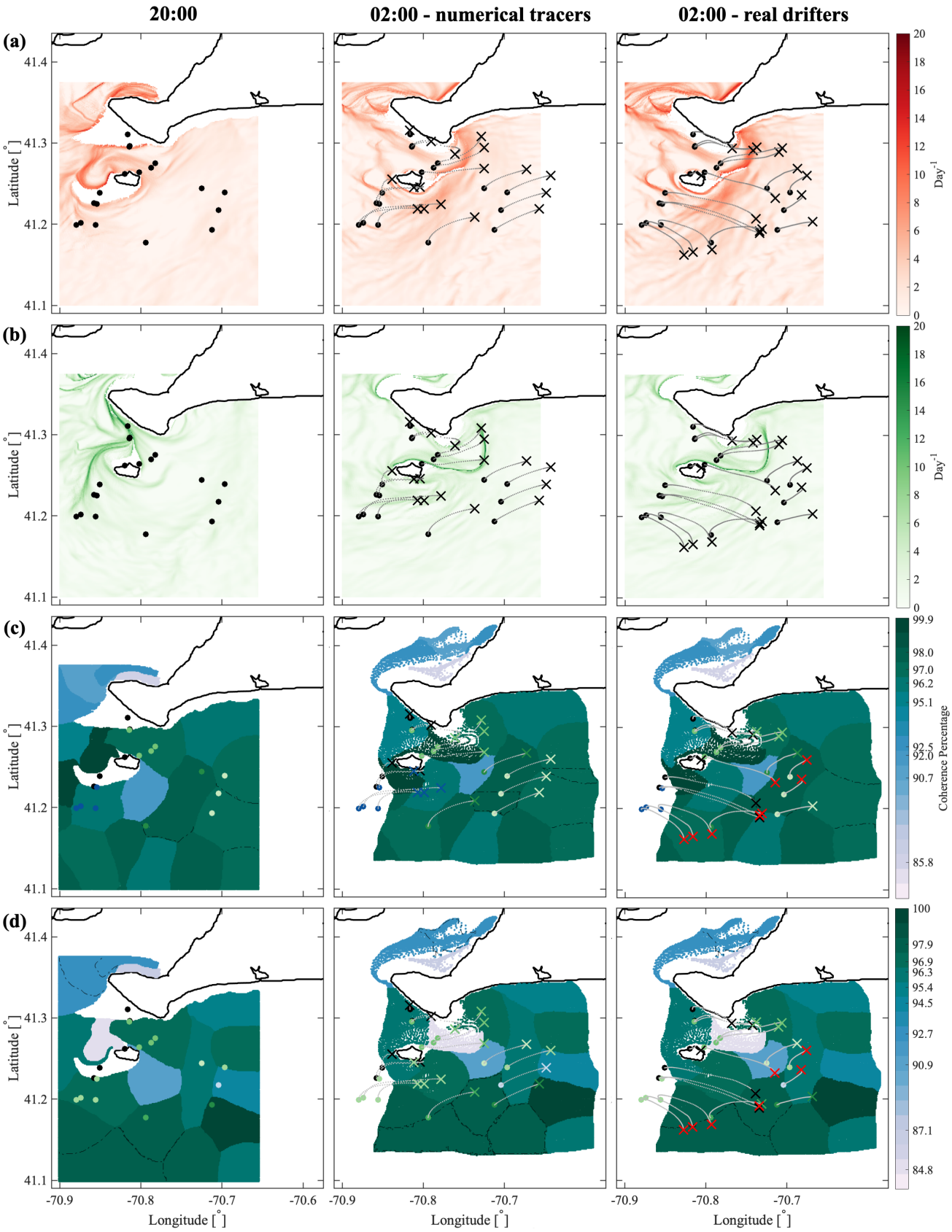


Figure 9. LCS, numerical tracers and drifter positions at the start (20:00 UTC) and end (02:00 UTC) of the 2018 experiment on August 7-8. (a) Forward and (b) backward FTLE. (c-d) Spectral clusters with coherence metric. The drifters are color-coded according to the cluster to which they belong, with red crosses if their final positions was outside their initial cluster.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

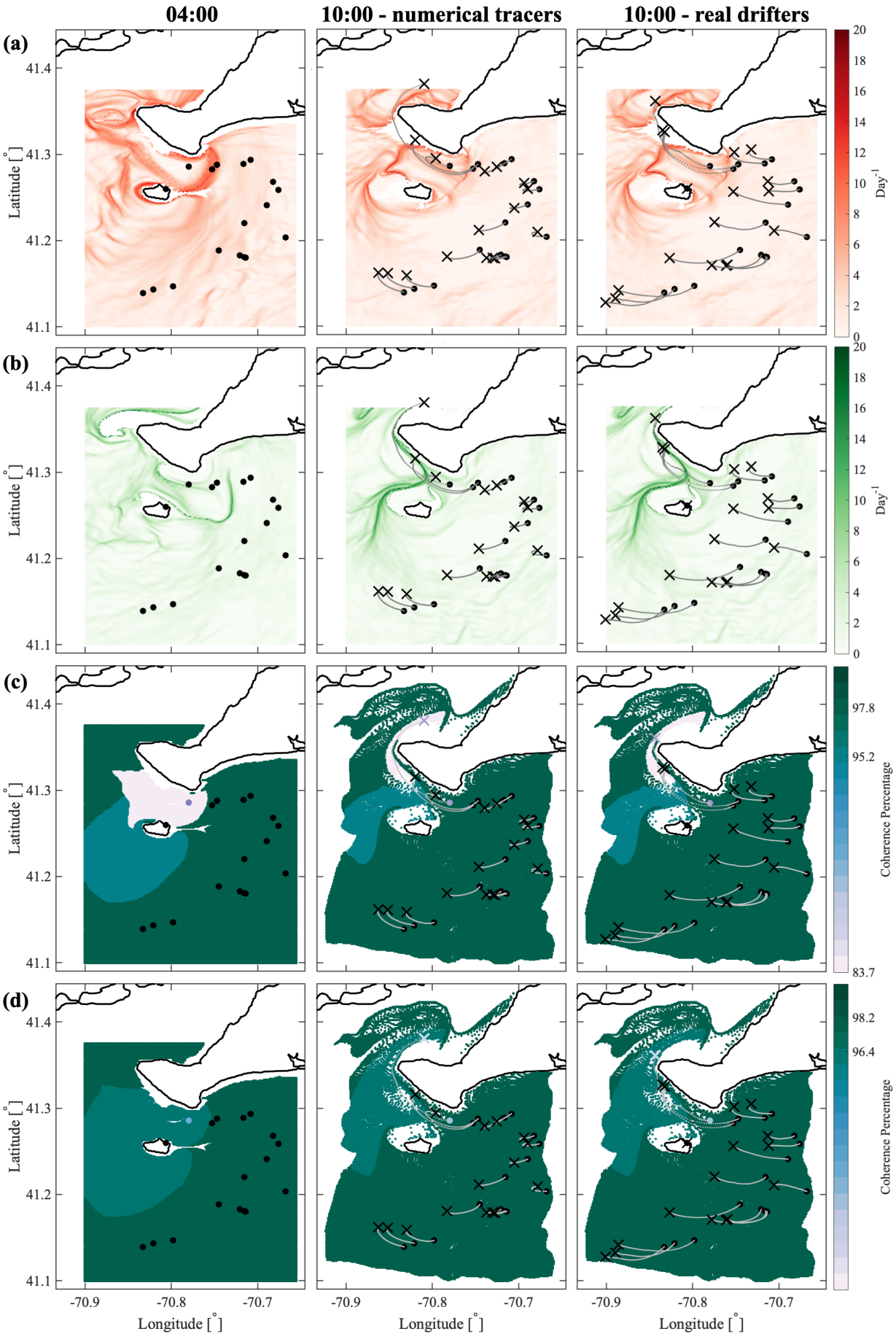


Figure 10. LCS, numerical tracers and drifter positions at the start (04:00 UTC) and end (10:00 UTC) of the 2018 experiment on August 8. (a) Forward and (b) backward FTLE. (c-d) Spectral clusters with coherence metric. The drifters are color-coded according to the cluster to which they belong, with red crosses if their final positions was outside their initial cluster.

 doi:10.20944/preprints202011.0605.v1

Specifically, we introduce a normalized spectral eigengap, which allows inter-comparison between different r-choices, instead of absolute eigengap used in the conventional method, which typically decreases with r. (As in the conventional spectral clustering, the optimal number of clusters is defined as the number of eigenvalues before the eigengap.) The optimal r value(s) is then identified as the value(s) that corresponds to the local maximum (or maxima) of the normalized eigengap. Specifying r in this manner ensures that for clusters of that particular size the inter-/intra-cluster similarity is largest/smallest compared to clusters that are either smaller and larger. Since many geophysical flows operate on multiple spatio-temporal scales, clusters of very different sizes could simultaneously exist within the same flow, and looking at all peaks in r ensures that we identify all of them. The second method parameter, $w_{i,i}$, is chosen as the smallest $w_{i,i}$ at which convergence of the normalized eigengap is achieved. Finally, to evaluate the coherence of the clusters, we introduce a noise-based coherence metric that allows comparing clusters within the same flow or between different flows. This metric quantifies the relative robustness of the resulting clusters with respect to applied numerical noise.

The automatic detection of multi-scale features is a well-known problem for conventional graph cut methods, including the normalized cut (which lies at the core of the spectral clustering) that favors clusters of equal size [42]. When clusters are well separated from each other and their sizes are only slightly different, such as in the periodically forced pendulum example in HA16 or in our slightly-asymmetric Duffing Oscillator example, all clusters can be identified using a single r value. However, when the clusters are sufficiently different in size, such as in our stongly-asymmetric Duffing Oscillator example, different r values are needed to correctly identify all the coherent sets. Thus, in general, no single r can be guaranteed to identify all of the clusters. By doing sweeps in r and identifying all of the r values that correspond to the local maxima in the eigengap ratio, we are able to overcome this challenge and identify all of the underlying coherent sets across a wide range of spatial scales. This is particularly important in oceanic applications, in which very little is known *a priori* about the flow and the scales of the coherent features.

The optimality of the parameter choices, as well as the specific noise-based metric of coherence that we have proposed here, present one possible way to eliminate the user-input parameter choices and quantify the coherence of the resulting clusters. We explained the logical arguments underlying our proposed parameter-free algorithm, and we illustrated that our method reliably yields parameters and clusters that are optimal according to our specific definition. The strength of our parameter-free spectral clustering method is that it allows applying the method to any flow without any prior knowledge about it, and that the method automatically identifies clusters without any user input, along with a coherence value for each cluster.

We have tested the parameter-free spectral clustering in 2 commonly-used benchmark analytic flows – Bickley Jet and Duffing Oscillator – and in a real-life oceanic application to a high-resolution model-based surface flow in the coastal region south of Martha's Vineyard, MA. The results are encouraging in that our method identified, without any user-input parameters, all known coherent clusters in both benchmark flows, and it identified reasonably-looking clusters in the realistic oceanic example.

In The Martha's Vineyard case study, since the flow south of the Vineyard has a strong tidal component, the resulting coherent clusters are qualitatively similar over the same phase of the tide on different days and even different years, whereas coherent clusters for the opposite tidal phase are qualitatively different. The strong dependence of the cluster geometry on the tidal phase has been also observed in tidally-driven flow over an experiment in Scott Reef, Australia [18]. Comparing the model-based clusters south of Martha's Vineyard to the motion of real drifters, for most of the identified clusters, drifters deployed within a cluster stayed within the same cluster for 6 hours, which was the time interval chosen for our analysis. Exceptions, i.e., clusters from which real drifters escaped in less than 6 hours, include small clusters with lower coherence metric or clusters identified over the time interval with strong wind gusts that were not realistically represented in the numerical model of oceanic currents.

 doi:10.20944/preprints202011.0605.v1

Comparing coherent clusters with FTLE fields reveals that cluster boundaries often have similarities with the forward-time FTLE ridges. This qualitative similarity is due to the repelling character of the forward FTLE ridges, which maximize separations between neighboring drifters and delineate regions of qualitatively different Lagrangian motion. In other cases, coherent clusters were identified in areas that shrink significantly over the subsequent 6 hours, suggesting that these clusters might be dominated by the surface flow convergence. Finally, some of the coherent clusters were identified over the most quiescent regions with lowest FTLEs, which encompass groups of drifters that moved (and separated) less than others. Note, however, that because the spectral clustering method identifies clusters based on a specific criterion, i.e., by maximizing intra-/inter-cluster similarities/differences, one should not expect a one-to-one agreement between spectral clusters boundaries and FTLEs or other LCSs that utilize another underlying flow property. Thus, various complementary LCS techniques could be applied in concert to provide the most comprehensive view of the underlying Lagrangian transport.

Author Contributions: Conceptualization, M.F., I.R. and T.P.; methodology, M.F., A.H. and I.R.; software, M.F. and A.H.; validation, M.F.; investigation, M.F.; resources, I.R. and T.P.; data curation, M.F.; writing–original draft preparation, M.F.; visualization, M.F.; supervision, A.H., I.R. and T.P.; project administration, I.R. and T.P.; funding acquisition, I.R. and T.P. All authors contributed to the formal analysis and to the writing–review and editing.

Funding: This research was funded by the Woods Hole Oceanographic Institution. Field work was supported by the U.S. National Science Foundation (NSF) under Grant Number AGS 1520825 (Hazards SEES: Advanced Lagrangian Methods for Prediction, Mitigation and Response to Environmental Flow Hazards).

Acknowledgments: The authors acknowledge the contributions from the MSEAS team, who provided the ocean circulation models and the output numerical velocity fields, and Siavash Ameli for providing the TRACE web-based gateway. The authors thank the ALPHA project team for their help with field work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

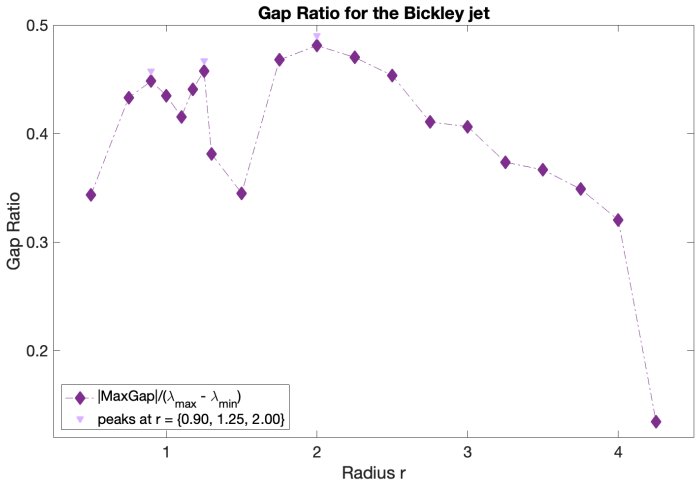
The following abbreviations are used in this manuscript:

FTLE	Finite-Time Lyapunov Exponent
FCM	Fuzzy C-Means
LCS	Lagrangian Coherent Structure

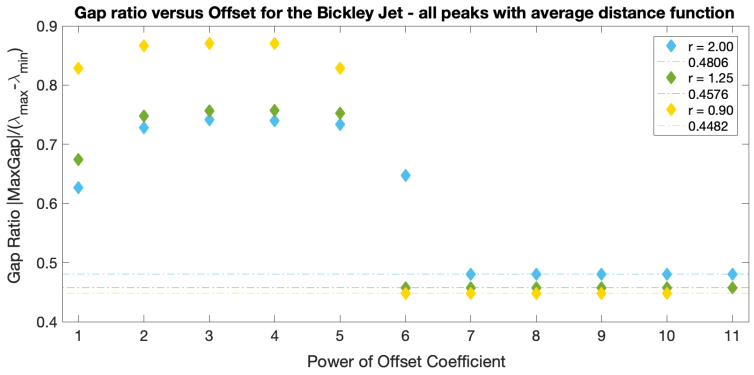
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

Appendix A Verification of parameter convergence

Appendix A.1 Bickley Jet



(a)

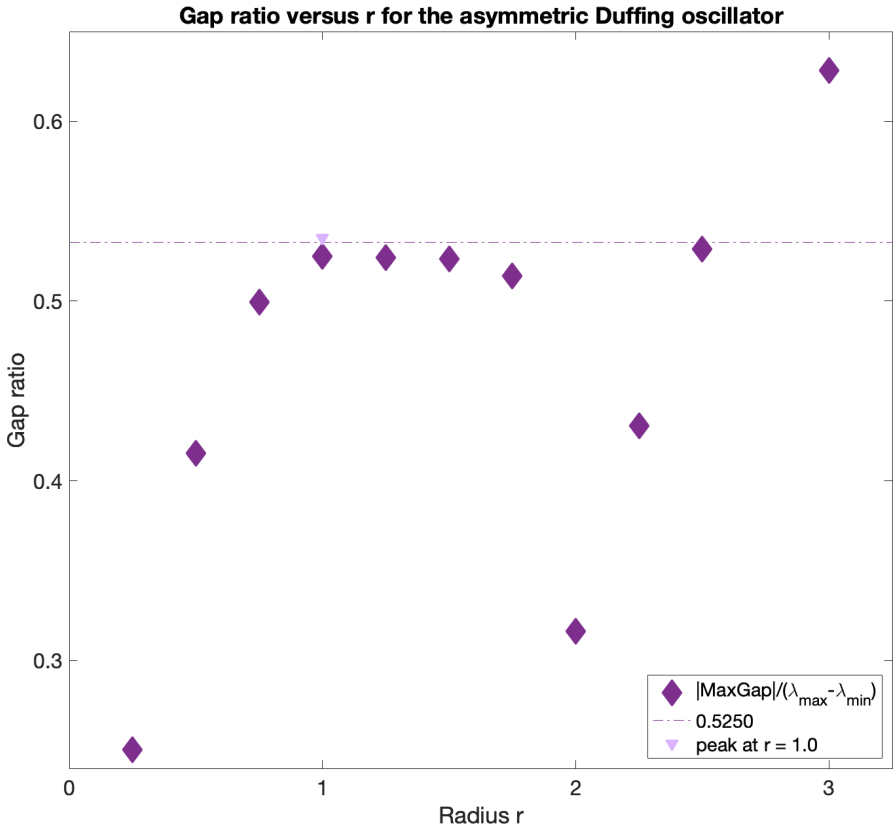


(b)

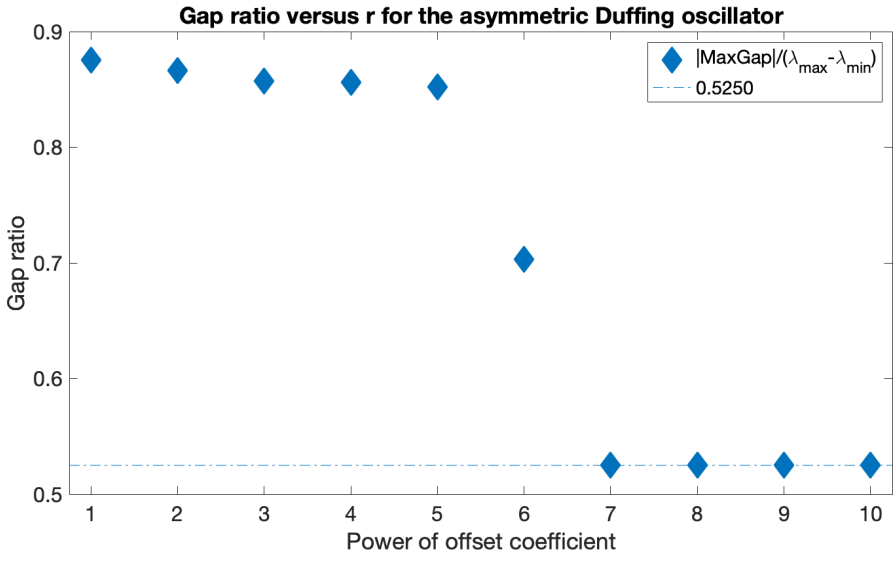
Figure A1. (a) Step 1 of the spectral clustering protocol for the Bickley Jet example: gap ratio as a function of r with the average distance function. (b) Step 2 of the spectral clustering protocol for the Bickley jet: sweep of offset coefficients 10^n for the gap ratio peaks in (a) at $r = 0.90$, $r = 1.25$ and $r = 2.00$ with the average distance function.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

Appendix A.2 Asymmetric Duffing oscillator



(a)

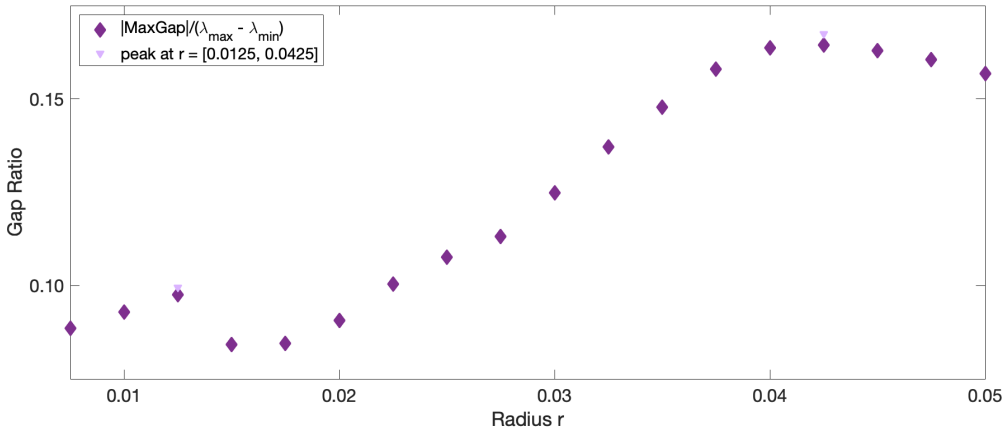


(b)

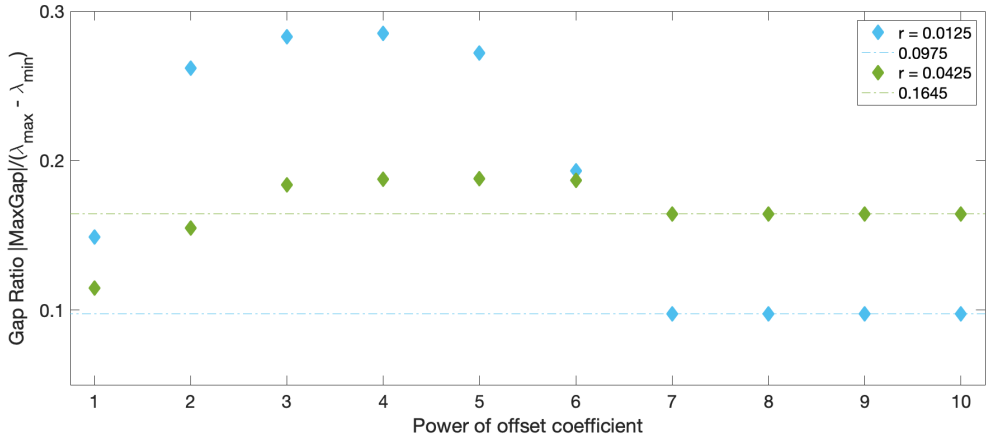
Figure A2. Steps 1 and 2 of the spectral clustering protocol for the asymmetric Duffing oscillator. (Top) Sweep of r parameters with offset coefficient 10^7 for the average distance function. (Bottom) Sweep of offset coefficients 10^n for average distance function and the gap ratio peak at $r = 1.0$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

Appendix A.3 2017 No Man's Land experiment



(a)

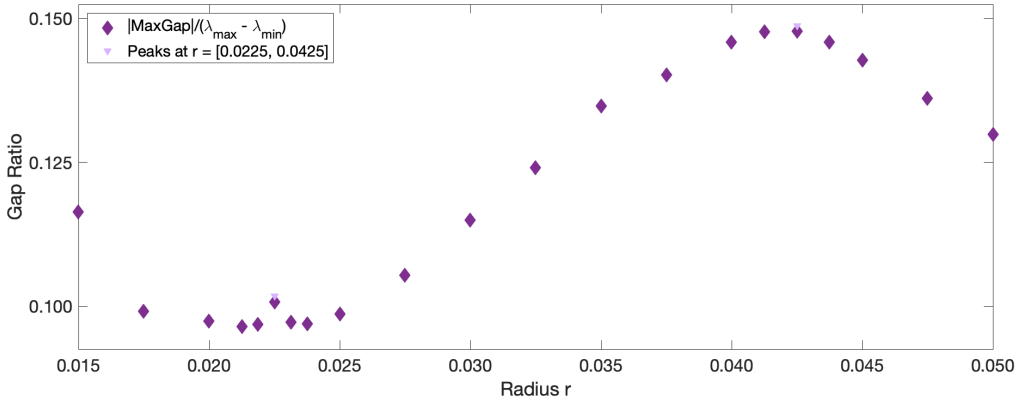


(b)

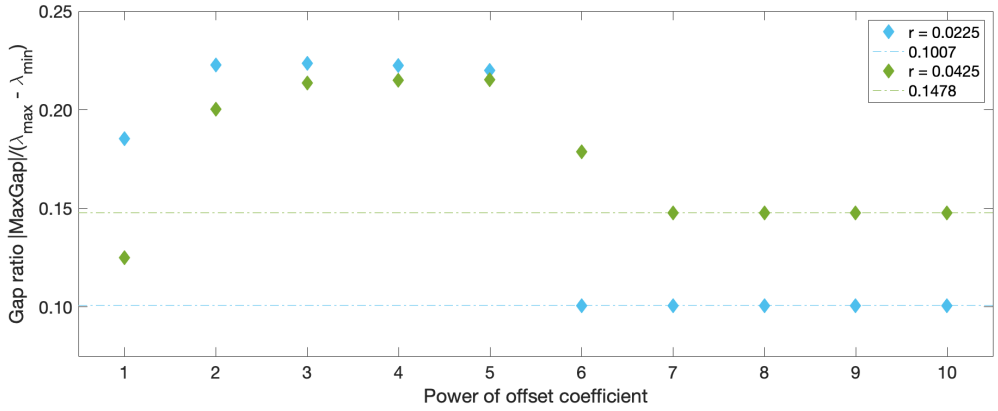
Figure A3. Steps 1 and 2 of the spectral clustering protocol for the 2017 No Man's Land experiment for the 15:51–21:51 window. (Top) Sweep of r parameters with offset coefficient 10^7 for the average distance function. (Bottom) Sweep of offset coefficients 10^n for average distance function and the gap ratio peaks in (a).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

Appendix A.4 2018 No Man's Land experiments



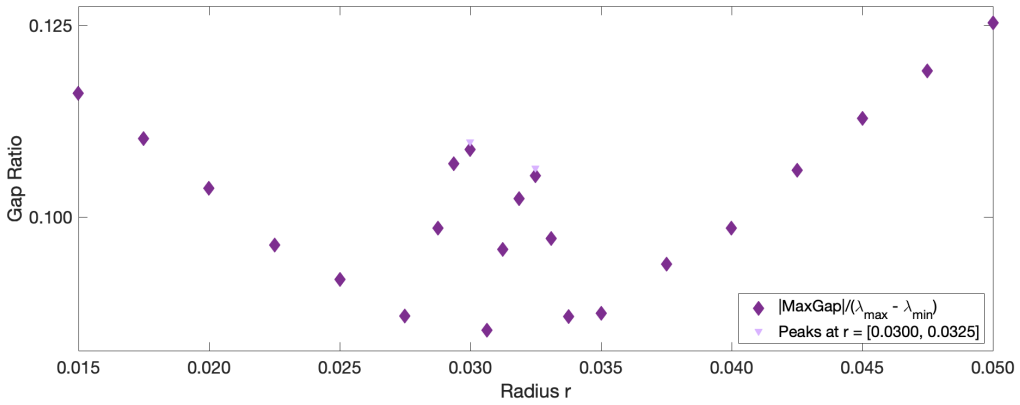
(a)



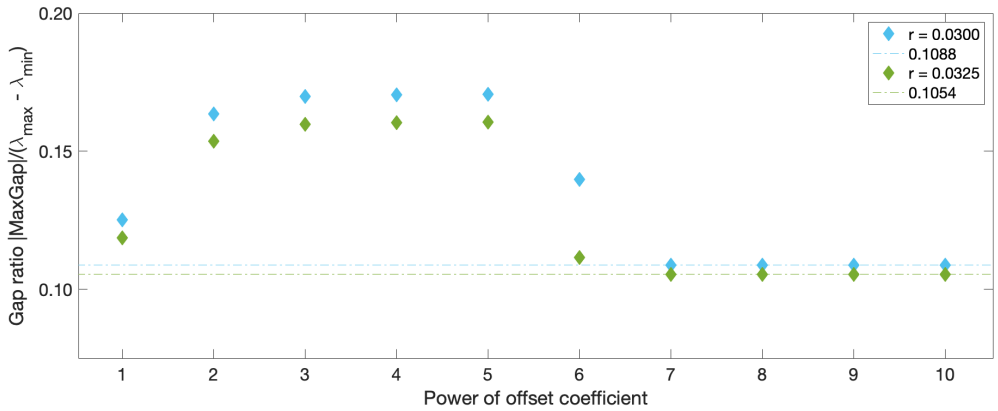
(b)

Figure A4. Steps 1 and 2 of the spectral clustering protocol for the 2018 No Man's Land 16:00 - 22:00 experiment. (Top) Sweep of r parameters with offset coefficient 10^7 for the average distance function. (Bottom) Sweep of offset coefficients 10^n for average distance function and the gap ratio peaks from (a).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1



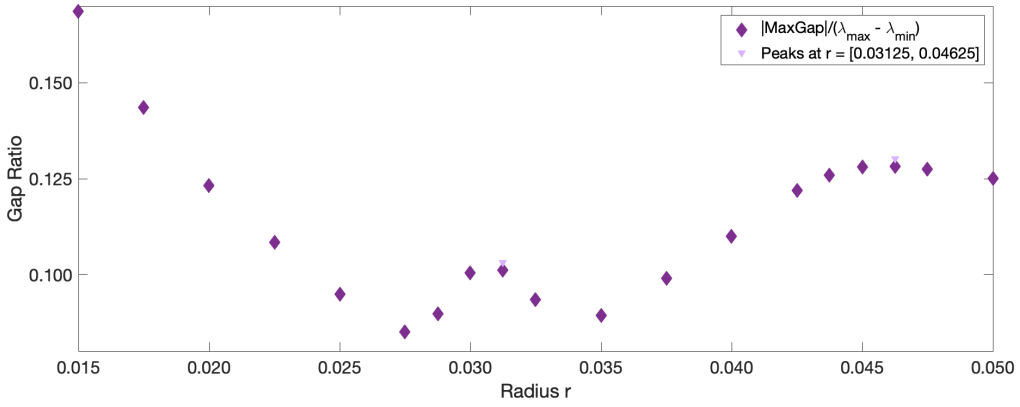
(a)



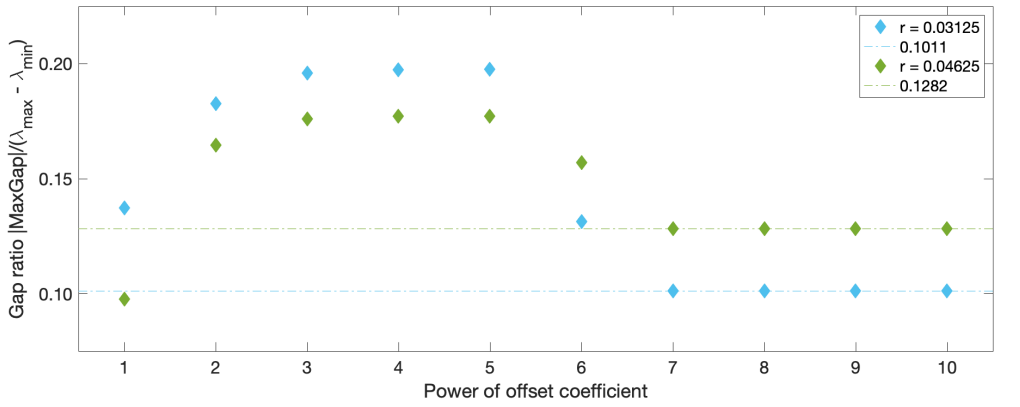
(b)

Figure A5. Steps 1 and 2 of the spectral clustering protocol for the 2018 No Man's Land 20:00 - 02:00 experiment. (Top) Sweep of r parameters with offset coefficient 10^7 for the average distance function. (Bottom) Sweep of offset coefficients 10^n for average distance function and the gap ratio peaks from (a).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1



(a)



(b)

Figure A6. Steps 1 and 2 of the spectral clustering protocol for the 2018 No Man's Land 04:00 - 10:00 experiment. (Top) Sweep of r parameters with offset coefficient 10^7 for the average distance function. (Bottom) Sweep of offset coefficients 10^n for average distance function and the gap ratio peaks from (a).

Appendix B Duffing oscillator with increased asymmetry

When the asymmetry of the Duffing Oscillator is increased, such as in the example presented in Figure A7 the sizes of the left and right gyres become increasingly different from each other and no single r value is able to successfully identify both of them at the same time.

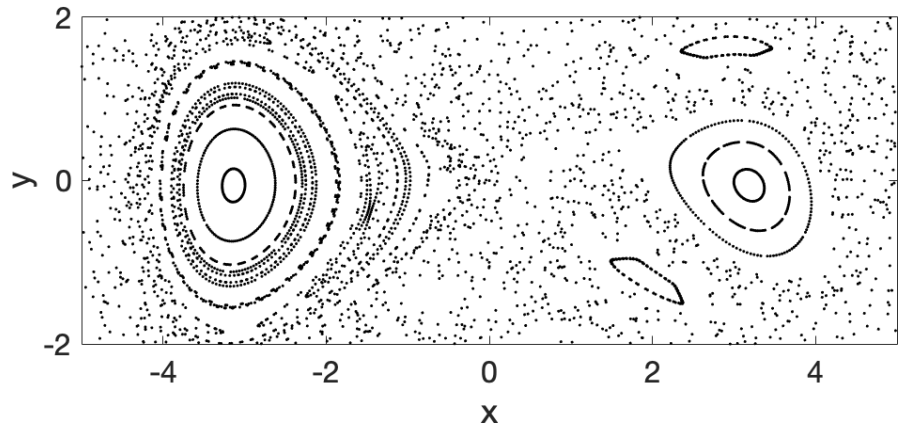


Figure A7. Poincaré map for a second example of the asymmetric Duffing oscillator with 100 periods of perturbation T_{pert}.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

In this case, the r-sweep, shown in Figure A8, shows three local maxima in the normalized eigengap. The resulting coherent clusters, color-coded by their noise-based coherence metric are shown in Figure A9. Each peak in r shows 2 coherent clusters centered at the left and right gyres, and an incoherent background filling the space between and around them. The sizes of the clusters are different for different r (consistent with our interpretation of r as the parameter defining the size of the resulting clusters), as is the amount of cluster filamentation. The peak at the smallest $r = 1.125$ correctly identifies the smaller right gyre, but pairs it with a left cluster that is smaller than the full extent of the left gyre. The peaks at $r = 2.0$ correctly identifies the full extent of the larger left gyre, but pair it with a right cluster that is very filamented. The peak at the largest $r = 3.75$ shows two clusters which are both too large and strongly filamented. The noise-based coherence metric can then be used to choose between the co-located clusters to pick one (out of three in this case) with the higher coherence, correctly yielding two clusters (the grey right cluster at $r = 1.125$ and the green left cluster at $r = 2$) corresponding to the full extent of the left and right gyres.

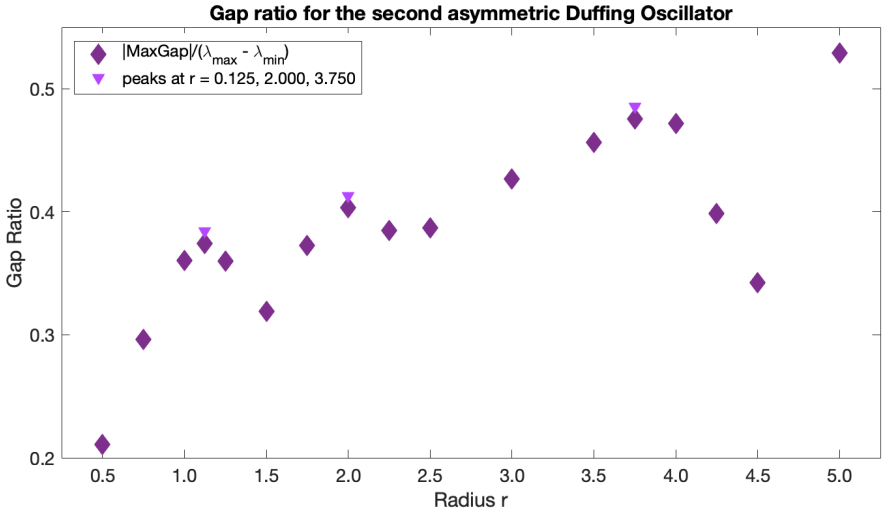


Figure A8. Step 1 of the spectral clustering protocol for the asymmetric Duffing oscillator shown in Figure A7: sweep of r parameters with offset coefficient 10^7 for the average distance function.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

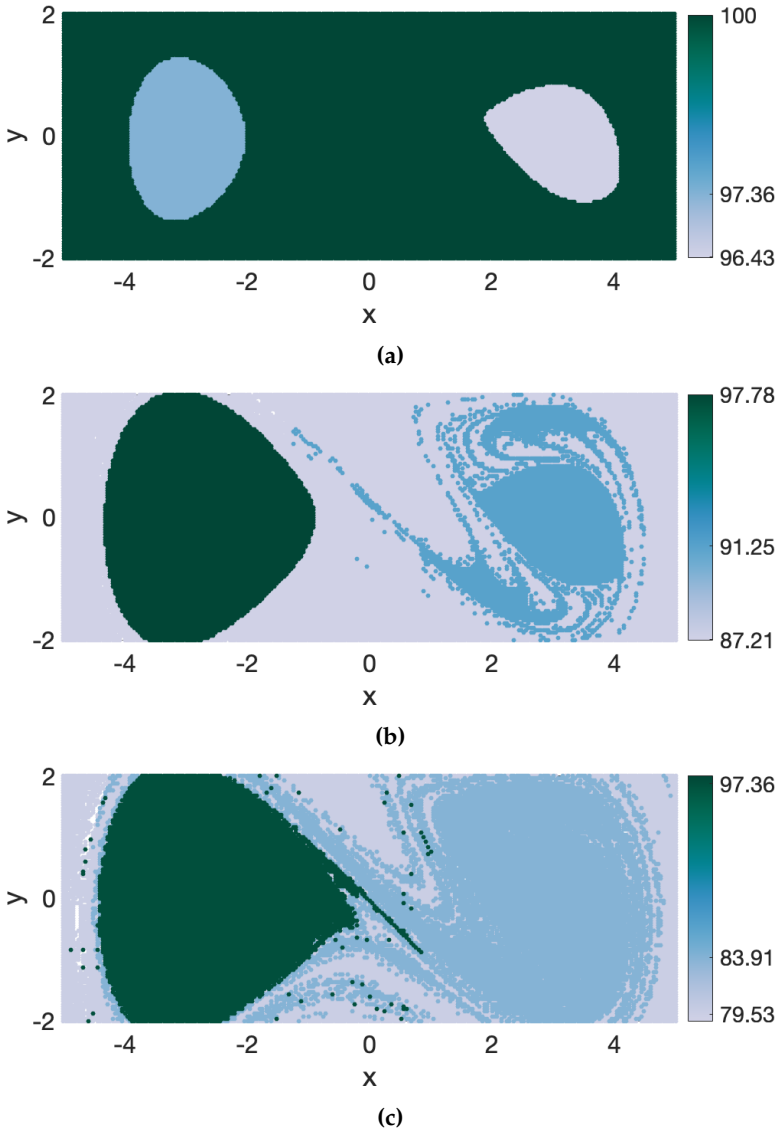


Figure A9. Step 3 of the spectral clustering protocol for the asymmetric Duffing oscillator shown in Figure A7 for (a) $r = 1.125$, (b) $r = 2.0$ and (c) $r = 3.75$. All individual clusters are color-coded by their coherence metrics.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

Appendix C Comparisons between real and simulated drifter trajectories

Appendix C.1 2017 No Man's Land experiment

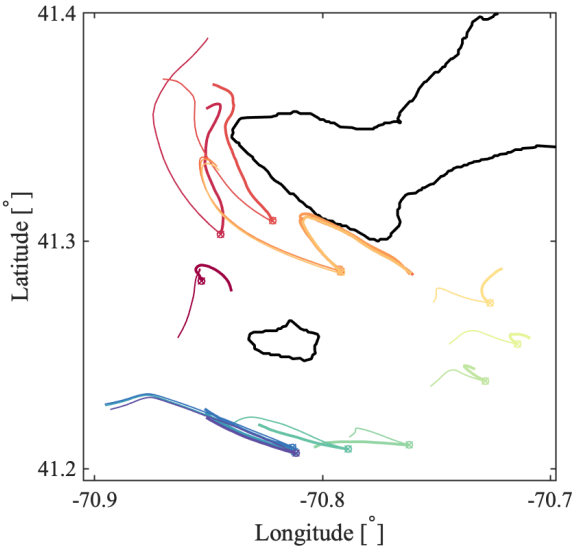


Figure A10. Comparison between real and simulated drifter trajectories for the 2017 No Man's Land experiment. The trajectories of CODE drifters are plotted with thick lines. The corresponding numerical trajectories, simulated with the same initial conditions as their corresponding CODE drifter, are plotted with thin lines.

Appendix C.2 2018 No Man's Land experiments

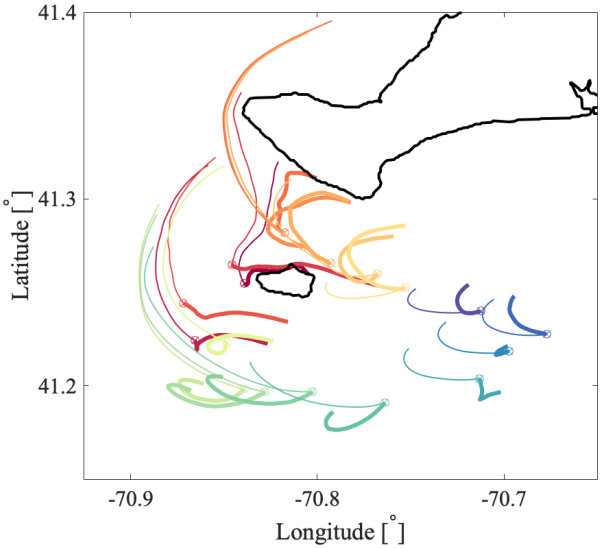


Figure A11. Comparison between real and simulated drifter trajectories for the 2018 No Man's Land experiment between 16:00 and 22:00 UTC. The trajectories of CODE drifters are plotted with thick lines. The corresponding numerical trajectories, simulated with the same initial conditions as their corresponding CODE drifter, are plotted with thin lines.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2020 doi:10.20944/preprints202011.0605.v1

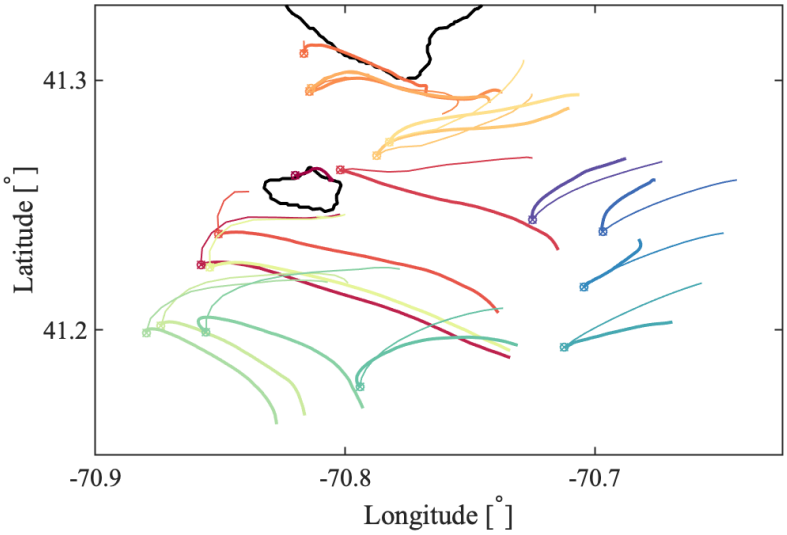


Figure A12. Comparison between real and simulated drifter trajectories for the 2018 No Man's Land experiment between 20:00 and 02:00 UTC. The trajectories of CODE drifters are plotted with thick lines. The corresponding numerical trajectories, simulated with the same initial conditions as their corresponding CODE drifter, are plotted with thin lines.

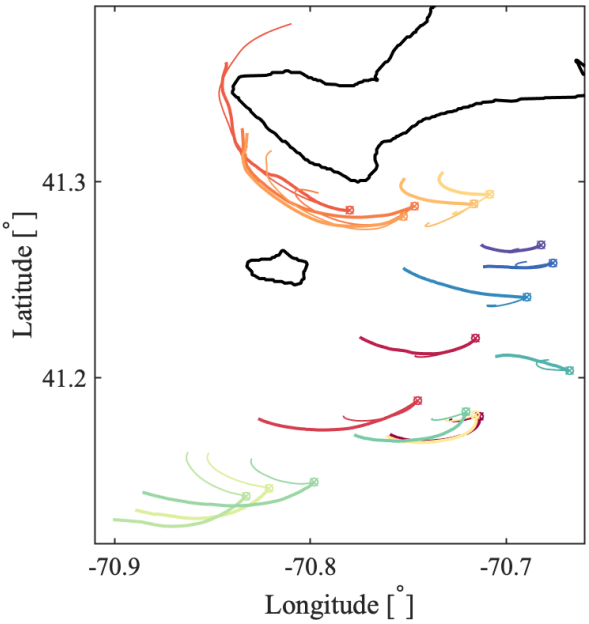


Figure A13. Comparison between real and simulated drifter trajectories for the 2018 No Man's Land experiment between 04:00 and 10:00 UTC. The trajectories of CODE drifters are plotted with thick lines. The corresponding numerical trajectories, simulated with the same initial conditions as their corresponding CODE drifter, are plotted with thin lines.

References

1. Aggarwal, C.C.; Hinneburg, A.; Keim, D.A. On the Surprising Behavior of Distance Metrics in High Dimensional Space. In *Database Theory — ICDT 2001*; Van den Bussche, J., Vianu, V., Eds., Springer Berlin Heidelberg: Berlin, Germany, 2001; pp. 420–434.
2. Allshouse, M.R.; Peacock, T. Lagrangian based methods for coherent structure detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2015**, *25*, (9)097617.
3. Ameli, S.; Shadden, S.C. A Transport Method for Restoring Incomplete Ocean Current Measurements. *Journal of Geophysical Research: Oceans* **2019**, *124*, (1)227–242.
4. Aref, H. Stirring by chaotic advection. *Journal of Fluid Mechanics* **1984**, *143*, 1–21.
5. Balasuriya, S; Ouellete, N.; Rypina, I.I. Generalized Lagrangian coherent structures *Physica D: Nonlinear Phenomena* **2018**, *372*, 31–51.
6. Beron–Vera, F.J.; Hadjighasem, A.; Xia, Q,; Olascoaga, M.J.; Haller, G. Coherent Lagrangian swirls among submesoscale motions. *Proceedings of the National Academy of Sciences* **2018**, *116*, (37):18251–18256.
7. Bettencourt, J.H.; López, C.; Hernández-García, E.Characterization of coherent structures in three-dimensional turbulent flows using the finite-size Lyapunov exponent. *Journal of Physics A: Mathematical and Theoretical* **2013**, *46*, 254022.
8. Cebeci, Z.; Yildiz, F. Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *Journal of Agricultural Informatics* **2015**, *6*, (3):13–23.
9. Cencini, M.; Vulpiani, A. Finite size Lyapunov exponent: review on applications. *Journal of Physics A: Mathematical and Theoretical* **2013**, *46*, 254019.
10. Davis, R.E. Oceanic property transport, Lagrangian particle statistics, and their prediction. *Journal of Marine Research* **1983**, *41*, 163–194.
11. Davis, R.E. Drifter observations of coastal surface currents during CODE: The method and descriptive view. *Journal of Geophysical Research: Oceans* **1985**, *90*, (C3):4741–4755.
12. Domingos, , E.; Pierre F. J. Lermusiaux, P.F.J. Multi-scale modeling: nested-grid and unstructured-mesh approaches. *Ocean Dynamics* **2008**, *58*, 335–336.
13. Domingos, P. A Few Useful Things to Know About Machine Learning. *Communications of the ACM* **2012**, *55*, (10):78–87.
14. Farazmand, M.; Haller, G. Computing Lagrangian Coherent Structures from their variational theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2012**, *22*, (1):013128.
15. Farazmand, M.; Blazevski, D.; Haller, G. Shearless transport barriers in unsteady two-dimensional flows and maps. *Physica D: Nonlinear Phenomena* **2014**, *278–279*, 44–57.
16. Farazmand, M.; Haller, G. Polar rotation angle identifies elliptic islands in unsteady dynamical systems. *Physica D* **2016**, *315*, 1–12.
17. Farazmand, M.; Haller, G.; Huhn, F. Defining coherent vortices objectively from the vorticity. *Journal of Fluid Mechanics* **2016**, *795*, 136–173.
18. Filippi, M. Advancing the theory and applications of Lagrangian Coherent Structures methods for oceanic surface flows. *Doctoral dissertations*, Massachusetts Institute of Technology.
19. Froyland, G. An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems. *CPhysica D: Nonlinear Phenomena* **2013**, *250*, 1–19.
20. Froyland, G.; Padberg–Gehle, K. 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* **2015**, *25*, (8):087406.
21. Ghosh, S.; Dubey, S.K. Comparative Analysis of K-Means and Fuzzy CMeans Algorithms. *International Journal of Advanced Computer Science and Applications* **2013**, *4*, (4):35–39.
22. Hadjighasem, A.; Haller, G. Geodesic Transport Barriers in Jupiter's Atmosphere: A Video-Based Analysis. *SIAM Review* **2016**, *58*, (1):69–89.
23. Hadjighasem, A.; Karrasch, D.; Teramoto, H.; Haller, G. Spectral-clustering approach to Lagrangian vortex detection. *Phys. Rev. E* **2016**, *93*, 063107.
24. Hadjighasem, A.; Farazmand, M.; Froyland, G.; Haller, G. A critical comparison of Lagrangian methods for coherent structure detection. *CHAOS* **2017**, *27*, 053104.

25. Haller, G.; Poje, A.C. Finite time transport in aperiodic flows. *Physica D: Nonlinear Phenomena* **1998**, *119*, (3):352–380.
26. Haller, G.; Yuan, G. Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica D: Nonlinear Phenomena* **2000**, *147*, (3):352–370.
27. Haller, G.; Beron–Vera, F.J. Coherent Lagrangian vortices: The black holes of turbulence. *Journal of Fluid Mechanics* **2013**, *731*, R4.
28. Haller, G. Lagrangian Coherent Structures. *Annu. Rev. Fluid Mech.* **2015**, *47*, (1):137–162.
29. Haller, G.; Karrasch, D.; Kogelbauer, F. Material barriers to diffusive and stochastic transport. *Proceedings of the National Academy of Sciences* **2018**, *115*, (37):9074–9078.
30. Haley, P.J.; Lermusiaux, P.F.J. Multiscale two-way embedding schemes for free-surface primitive equations in the "Multidisciplinary Simulation, Estimation and Assimilation System". *Ocean Dynamics* **2010**, *60*, (6):1497–1537.
31. Haley, P.J.; Agarwal, A.; Lermusiaux, P.F.J. Optimizing velocities and transports for complex coastal regions and archipelagos. *Ocean Modelling* **2015**, *89*, 1–28.
32. Haza, A.; Griffa, A.; Martin, P.; Molcard, A.; Özgökmen, T.M.; Poje, A.; Barbanti, R.; Book, J.; Poulain, P.-M.; Rixen, M.; Zanasca, P. Model-based directed drifter launches in the adriatic sea: Results from the DART experiment. *Geophys. Res. Lett.* **2007**, *34*, L10605.
33. Haza, A.; Özgökmen, T.M.; Griffa, A.; Molcard, A.; Poulain, P.M.; Peggion, G. Transport properties in small-scale coastal flows: relative dispersion from VHF radar measurements in the Gulf of La Spezia. *Ocean Dynamics* **2010**, *60*, 861–882
34. Kirincich, A. Remote Sensing of the Surface Wind Field over the Coastal Ocean via Direct Calibration of HF Radar Backscatter Power. *Journal of Atmospheric and Oceanic Technology* **2016**, *33*, (7):1377–1392.
35. LaCasce, J.H. Statistics from Lagrangian observations. *Progress in Oceanography 2008*, *77*, (1):1–29.
36. Lehahn, Y.; d'Ovidio, F.; Koren, I. A Satellite-Based Lagrangian View on Phytoplankton Dynamics. *Annual Review of Marine Science* **2018**, *10*, (1):99–119.
37. Lermusiaux, P.F.J.; Schröter, J.; Danilov, S.; Iskandarani, M.; Pinardi, N.; Westerink, J.J. Multiscale Modeling of Coastal, Shelf and Global Ocean Dynamics. *Ocean Dynamics* **2013**, *63*, 1341–1344.
38. Lloyd, S.P. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* **1982**, *28*, 129–137.
39. McWilliams, J.C. Maps from the Mid-Ocean Dynamics Experiment: Part I. Geostrophic Streamfunction. *Journal of Physical Oceanography* **1976**, *6*, (6):810–827.
40. Mathur, M.; Haller, G.; Peacock, T.; Ruppert-Felsot, J.E.; Swinney, H.L. Uncovering the Lagrangian Skeleton of Turbulence. *Phys. Review Lett.* **2007**, *98*, (14):144502.
41. Mendoza, C.; Mancho, A.M. The Lagrangian description of aperiodic flows: a case study of the Kuroshio Current. *Nonlinear Processes in Geophysics* **2012**, *4*, 449–472.
42. Nadler, B.; Galun, M. Fundamental Limitations of Spectral Clustering. In *Advances in Neural Information Processing Systems: Proceedings of the 2006 Conference*; Schölkopf, B, Platt, J., Hofmann, T., Eds.; MIT Press, Cambridge, U.S.A., 2007, pp.1017-1024.
43. Park, T.J; Han, K.J.; Kumar, M.; Narayanan, S. Auto-tuning spectral clustering for speaker diarization using normalized maximum eigengap. *IEEE Signal Processing Letters* **2019**, *27*, 381–385.
44. Peacock, T.; Haller, G. Lagrangian coherent structures: The hidden skeleton of fluid flows. *Physics Today* **2013**, *66* (2):41–47.
45. Rayson, M.D.; Ivey, G.N.; Jones, N.L.; Fringer, O.B. Resolving high-frequency internal waves generated at an isolated coral atoll using an unstructured grid ocean model. *Ocean Modelling* **2018**, *122*, 67–84.
46. Rypina, I.I. Lagrangian Coherent Structures and Transport in Two-Dimensional Incompressible Flows with Oceanographic and Atmospheric Applications. *University of Miami PhD thesis* **2007**.
47. Rypina, I.I.; Brown, M.G.; Beron-Vera, F.J.; Koçak, H.; Olascoaga, M.J.; Udovydchenkov, I.A. On the Lagrangian Dynamics of Atmospheric Zonal Jets and the Permeability of the Stratospheric Polar Vortex. *Journal of the Atmospheric Sciences* **2007**, *64*, (10):3595–3610.
48. Rypina, I.I.; Pratt, L.J.; Pullen, J.; Levin, J.; Gordon, A.L. Chaotic Advection in an Archipelago. *Journal Physical Oceanography* **2010**, *40*, 1988–2006,
49. Rypina, I.I.; Scott, S.E.; Pratt, L.J.; Brown, M.G. Investigating the connection between complexity of isolated trajectories and Lagrangian Coherent Structures. *Nonlinear Processes in Geophysics* **2011**, *18*, (6):977–987.

50. Rypina, I.I.; Jayne, S.R.; Yoshida, S.; Macdonald, A.M.; Douglass, E.; Buesseler, K. Short-term dispersal of Fukushima-derived radionuclides off Japan: modeling efforts and model-data intercomparison. *Biogeosciences* **2013**, *10*, (7):4973–4990.
51. Rypina, I.I.; Jayne, S.R.; Yoshida, S.; Macdonald, A.M.; Buesseler, K. Drifter-based estimate of the 5 year dispersal of Fukushima-derived radionuclides. *Journal of Geophysical Research: Oceans* **2014**, *119*, (11):8177–8193.
52. Rypina, I.I.; Kirincich, A.R.; Limeburner, R.; Udovydchenkov, I.A. Eulerian and Lagrangian Correspondence of High-Frequency Radar and Surface Drifter Data: Effects of Radar Resolution and Flow Components. *Journal of Atmospheric and Oceanic Technology* **2014**, *31*, (4):945–966.
53. Rypina, I.I.; Kirincich, A.; Lentz, S.; Sundermeyer, M.A. Investigating the Eddy Diffusivity Concept in the Coastal Ocean. *Journal of Physical Oceanography* **2016**, *46*, (7):2201–2218.
54. Rypina, I.I.; Fertitta, D.; Macdonald, A.; Yoshida, S.; Jayne, S. Multi-Iteration Approach to Studying Tracer Spreading Using Drifter Data. *Journal of Physical Oceanography* **2017**, *47*, (2):339–351.
55. Rypina, I.I.; and Pratt, L.J. Trajectory encounter volume as a diagnostic of mixing potential in fluid flows. *Nonlinear Processes Geophysics* **2017**, *24*, 189–202.
56. Rypina, I.I.; Smith, L.; Stefan, G.; Pratt, L.J. Connection between encounter volume and diffusivity in geophysical flows. *Nonlinear Processes in Geophysics* **2018**, *25*, (2):267–278.
57. Shadden, S.C.; Lekien, F.; Marsden, J.E. Definition and properties of Lagrangian Coherent Structures from Finite-Time Lyapunov Exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena* **2005**, *212*, (3):271–304.
58. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2000**, *22*, (8):888–905.
59. Vieira, G.S.; Allshouse, M.R. Internal wave boluses as coherent structures in a continuously stratified fluid. *J. Fluid Mech.* **2020**, *885*, A35.
60. Villermaux, E. Mixing Versus Stirring. *Annual Review of Fluid Mechanics* **2019**, *51*, (1):245–273.
61. von Luxburg, U. A tutorial on spectral clustering. *Statistics and Computing* **2007**, *17*, (4):395–416.
62. Wiggins, S. The dynamical system approach to Lagrangian transport in oceanic flows. *Annual Review of Fluid Mechanics* **2005**, *37*, 295–328.

Sample Availability: Samples of the compounds ..... are available from the authors.