Lagrangian scale decomposition via the graph Fourier transform

Theodore MacMillan and Nicholas T. Ouellette ⬤[*]

Department of Civil and Environmental Engineering, Stanford University, Stanford, California 94305, USA

Scale decomposition is ubiquitous in the analysis of complex fluid flows. Often with an eye towards reduced order modeling, considerable effort has been invested in the development of novel sets of basis functions for such decompositions. The most successful sets of basis functions have been developed in the Eulerian perspective, where traditional tools of calculus and newer methods from data science can be most easily leveraged. Here, we take advantage of recent interest in graph-based approaches to Lagrangian coherence as well as new methods for the scale analysis of complex networks to introduce a new scale decomposition that is instead fully Lagrangian and based on transport. To do so, we adapt a technique from network science know as the graph Fourier transform and develop a novel graph correlation function that allows us to quantitatively describe our Lagrangian decomposition as a function of scale. This method allows better interpretability of the dynamic consequences of kinematic coherence as well as the ability to perform traditional Eulerian tasks (such as decomposition, filtering, and compression) on Lagrangian quantities. We illustrate our techniques on examples drawn from coherent-structure analysis, ocean mixing, and cloud physics.

I. INTRODUCTION

The notions of scale and scale separation are firmly ingrained in many areas of modern physics. Since the pioneering work of Richardson and Kolmogorov, this has rung especially true in the study of turbulence, with the former famously musing that "big whirls have little whirls that feed on their velocity..." [1]. This cartoon was quickly formalized using Fourier transforms, a scale transformation that relies on the orthogonality of complex exponentials or sinusoids. Following this foundational work, theories of turbulence have relied heavily on the correspondence between Richardson's "whirls" and these orthogonal basis functions.

As the study of turbulence further matured, it became evident that many different orthogonal basis functions could form suitable scale-space representations, provided they still contained some metric by which to order them. The need for this approach was made acute as theories of turbulence expanded from homogeneous and isotropic flows towards flows bounded by walls or other symmetry-breaking features. An asset in some cases, the periodic nature of complex exponentials was a hindrance in others, so geometry-informed spectral transformations such as Chebyshev polynomials [2] or spherical harmonics [3] found applications. When a scale-local and space-local transformation was needed (noting that periodic basis functions often have the former quality but not the latter), researchers began to use wavelet transforms [4]. More recently, the explosion of data science has introduced general transformations that are not only geometry dependent but also flow dependent, and not necessarily ordered by spatial scale. The proper orthogonal decomposition (POD), for example, orders its modes from largest to smallest energy, while the dynamic mode

---

[*]nto@stanford.edu

124401-1

decomposition (DMD) orders its modes according to time-frequency oscillations [5]. What all of these decompositions share is a precise hierarchy in which their modes can be ordered.

Separately, but not entirely unrelatedly, so-called Lagrangian coherent structures (LCSs) have emerged as powerful tools for tackling problems of transport [6,7]. Very quickly, graph-theoretical approaches such as spectral clustering [8] and graph coloring [9] gained traction, and today are widely used [10,11]. These methods cast a group of trajectories as a dynamic network, where the connections between nodes (the trajectories) are determined by some metric that quantifies their kinematic similarity. Examples include a pure L1-norm average distance metric [8], a kinematic dissimilarity metric [9], a count of how many trajectories encounter one another [12], and many others. By studying the spectral signature of these trajectory graphs via an eigendecomposition, these methods elucidate the underlying structure of the trajectory space and the kinematic coherence within.

Methods for combining these two powerful approaches have so far eluded the broader fluid mechanics community. While it is natural to study Eulerian quantities through scale decompositions, multipoint Lagrangian quantities are typically not spatially differentiable and cannot be approached via the same set of tools. Some attempts have been made by spatially filtering Lagrangian trajectories [13] or by studying a frequency decomposition of the time history of Lagrangian trajectories [14], but few methods have emerged that respect the spatial organization of Lagrangian quantities and related structures. By contrast, coordinate transforms, which respect the spatial organization of Lagrangian trajectories [15,16], then have no corresponding notion of scale. For quantities that depend crucially on the Lagrangian history of parcels but also carry spatial signatures—such as the dynamically growing radius of a droplet in a cloud or the orientation of active swimmers in turbulence—this shortcoming prevents the kinds of scale arguments that are so often invoked for other quantities in fluid mechanics.

Here, we propose a way to surmount these problems by developing a scale decomposition based on the spectral properties of a Lagragian trajectory graph. Our approach builds on recent advances in graph signal processing and the related development of the graph Fourier transform (GFT) [17–19]. As we will show, tools from graph signal processing allow us to analyze functions defined on the nodes of a graph (i.e., dynamical Lagrangian quantities), an approach that is often more informative than studying the graph structure in isolation. We are then able to discuss the dynamical relevance of coherent structures rather than just their kinematic coherence—an important facet of any structure, but one that does not tell a complete story.

As we will demonstrate, one obstacle that has obstructed this development to date is that there is not a straightforward way to hierarchically order the eigenvectors of the graph Laplacian that corresponds to the physics or geometry of a given problem: the graph Fourier transform by itself does not reveal a quantitative spatial scale associated with each of its modes. We therefore begin by describing the GFT and develop a novel graph-based correlation function that allows us to order our scale-decomposition basis functions. Following this description, we revisit two well-studied LCS cases and demonstrate the added depth brought by our method. Finally, we describe various analyses of a turbulent flow now made possible with our transport-informed scale decomposition.

II. METHODS

Consider a graph $\mathcal{G} = (\mathcal{E}, \mathcal{N})$, where $\mathcal{N}$ is a set of N nodes and $\mathcal{E}$ is a set of weighted, undirected edges. This structure can be conveniently represented by the $N \times N$ adjacency matrix $\mathbf{W}$, where each entry $W_{i,j}$ represents the edge connecting nodes i and j. Because we confine our analysis to undirected graphs, the matrix $\mathbf{W}$ is symmetric. A function $f : \mathcal{V} \to \mathbb{R}$ takes a vertex of a graph as an argument and return a real-valued number, which applied to all the vertices on a graph can be written as a vector $\mathbf{f} \in \mathbb{R}^N$.

A. Graph Laplacian and its eigenvectors

As in many graph-based analyses, it is useful here to define a matrix known as the graph Laplacian. After first creating a degree matrix $\mathbf{D}$, where $D_{i,j} = \sum_{k=1}^{N} W_{i,k}$ when $i = j$ and 0 otherwise, we

can write the graph Laplacian as $\mathbf{L} = \mathbf{D} - \mathbf{W}$ and its normalized version as $\mathbf{L_{norm}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$. To motivate this definition, consider a graph comprised of nodes along a line in a periodic domain, each connected to only their direct neighbors and separated by distance Δx. This type of graph is called circulant. Constructing a weight matrix $W_{i,j} = |x(i) - x(j)|^{-2}$, we obtain

$$\mathbf{W} = \begin{pmatrix} 0 & \frac{1}{\Delta x^2} & 0 & \cdots & \frac{1}{\Delta x^2} \\ \frac{1}{\Delta x^2} & 0 & \frac{1}{\Delta x^2} & & \\ 0 & \frac{1}{\Delta x^2} & 0 & \frac{1}{\Delta x^2} & \\ \vdots & & \ddots & \ddots & \ddots \\ \frac{1}{\Delta x^2} & & 0 & \frac{1}{\Delta x^2} & 0 \end{pmatrix} \tag{1}$$

and the graph Laplacian is given by

$$\mathbf{L} = \mathbf{D} - \mathbf{W} = \begin{pmatrix} \frac{2}{\Delta x^2} & \frac{-1}{\Delta x^2} & 0 & \cdots & \frac{-1}{\Delta x^2} \\ \frac{-1}{\Delta x^2} & \frac{2}{\Delta x^2} & \frac{-1}{\Delta x^2} & & \\ 0 & \frac{-1}{\Delta x^2} & \frac{2}{\Delta x^2} & \frac{-1}{\Delta x^2} & \\ \vdots & & \ddots & \ddots & \ddots \\ \frac{-1}{\Delta x^2} & & 0 & \frac{-1}{\Delta x^2} & \frac{2}{\Delta x^2} \end{pmatrix}. \tag{2}$$

Now, consider the continuous Poisson equation $\nabla^2 v = -\lambda v$. On a one-dimensional periodic line, this equation becomes $\frac{d^2 v}{dx^2} = -\lambda v$. Discretizing the domain into points spaced a distance Δx apart, a first-order finite difference approximation to this equation is $\frac{v_{j+1} - 2v_j + v_{j-1}}{\Delta x^2} = -\lambda v_j$, or in matrix form $\mathbf{L}\mathbf{v} = -\lambda\mathbf{v}$, where $\mathbf{L}$ is exactly the matrix defined above. As it turns out, in periodic grid networks of any dimension, the graph Laplacian is exactly the finite-difference approximation of the continuous Laplace operator. So, the eigenfunctions of the continuous Laplace operator are equivalent to the eigenvectors of the graph Laplacian; that is, vectors $\mathbf{v}$ that satisfy the above equality. In the case of the one-dimensional circulant graph, these functions are simply $\cos(\omega_k x)$ and $\sin(\omega_k x)$, with corresponding eigenvalues $\lambda_k = \omega_k^2$.

B. Graph Fourier transforms

The observant reader will notice that these functions closely resemble the basis functions used in the discrete Fourier transform (DFT) and its real-valued relative the discrete cosine transform (DCT) [20]. In recent years, researchers have taken advantage of this relationship to develop a version of Fourier transforms defined on graphs [17,19,21].

As the Laplacian matrix $\mathbf{L}$ is positive valued and symmetric, its eigenvectors $\mathbf{v}_{k \in 1, \dots, N}$ are real and form an orthogonal basis. The matrix $\mathbf{V}$ is a matrix whose columns consist of the eigenvectors of $\mathbf{L}$ and that satisfies the relationship $\mathbf{L}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix whose kth entry is λ_k, the eigenvalue corresponding to the kth eigenvector. Given a real-valued $1 \times N$ vector $\mathbf{f}$ defined on the nodes of $\mathcal{G}$, the graph Fourier transform (GFT) [21] is defined as

$$\hat{f}(\lambda_k) = \sum_{i=1}^{N} v_k(i)f(i), \tag{3}$$

or in matrix form as $\hat{\mathbf{f}} = \mathbf{V}^{\mathsf{T}}\mathbf{f}$, where $\hat{f}_k$ is the spectral graph coefficient corresponding to λ_k and $\mathbf{v}_k$. Similarly, we can define the inverse GFT as

$$f(i) = \sum_{k=1}^{N} v_k(i)\hat{f}(\lambda_k), \tag{4}$$

or in matrix form as $\mathbf{f} = \mathbf{V}\hat{\mathbf{f}}$.

C. Frequency on a graph

Now that we have defined the spectral coefficients of a graph, we would like to assign some sort of frequency or scale associated with them. In the simple case of the circulant graph described above, the eigenvectors of its graph Laplacian have a natural sense of frequency, with eigenvectors corresponding to smaller eigenvalues having fewer oscillations across the graph and eigenvectors corresponding to larger eigenvalues having many oscillations, though this relationship becomes less straightforward as more complicated graphs are considered [20]. For more general graphs such as the trajectory graphs we are interested in, however, it is typically the case that smaller eigenvalues correspond to lower-frequency eigenvectors. Many researchers [17,18,21,22] have shown that the total variation of the eigenvectors, defined as $\text{TV}(\mathbf{v}_k) = |\mathbf{v}_k - \mathbf{W}_{\text{norm}}\mathbf{v}_k|$ where $\mathbf{W}_{\text{norm}} = \frac{1}{\lambda_{\text{max}}}\mathbf{W}$, is a strictly increasing function of λ_k, or that $\text{TV}(\mathbf{v}_k) < \text{TV}(\mathbf{v}_{k+1})$ if the eigenvalues are ordered such that $\lambda_k < \lambda_{k+1}$. Connected to this result, Shuman *et al.* [21] showed that the number of zero-crossings monotonically increases for a collection of eigenvectors ordered in such a way.

This approach, however, fails to recover the clean eigenvalue-frequency relation of circulant graphs. To amend this shortcoming, we introduce a new measure of frequency on a graph by revisiting some canonical two-point statistical quantities in the graph setting.

The first step toward defining two-point statistics is to develop a definition of distance in the graph setting. If we compare to a typical one-dimensional autocovariance, which may be defined as $R(x, r) = f(x)f(x + r)$ where f is some function and x and r are absolute position and offset position, respectively, it is not immediately obvious what our equivalent offset position on a graph could be. In fact, part of the motivation behind the original development of the GFT was to implement graph convolutions (via multiplication in the graph spectral domain) without having to address this ambiguity directly. However, we can use several propositions and advances in recent years [23,24] to continue forward.

Here, we define a graph distance between two nodes i and j simply as

$$D_{i,j} = \min \sum_{(k,l)\in\Gamma_{ij}} W_{kl}^{-1},\tag{5}$$

where Γ_{ij} is the set of all paths connecting i and j (as in Ref. [24]). This is simply a shortest or least cost path metric, and in a grid setting is the same as the Manhattan distance. Because the graph space is fundamentally discrete, we cannot have a continuous covariance as before, so a graph autocovariance function is only defined at shortest path distances that exist in a given graph (any discrete autocovariance function is similarly constrained). So, we can define a graph autocovariance that begins at node i to be $r(i, D_{i,j}) = f(i)f(\arg_j D_{i,j})$, or as an average across all nodes i as $R(D_{i,j}) = \langle f(i)f(\arg_j D_{i,j})\rangle_{i\in\mathcal{N}}$. Similarly, a graph autocorrelation can be defined as $\rho(D_{i,j}) = \langle \frac{R(D_{i,j})}{f(i)^2}\rangle_{i\in\mathcal{N}}$. In plain language, our autocorrelation function is exactly the same as in the Euclidean setting, but the offset position is replaced by a shortest path distance. The motivation behind this endeavour is that now we can leverage standard measures of characteristic scale, which in the Euclidean setting have been tailored to this quantity. Common measures of the characteristic scale of a correlation function include the distance r at which $R(x, r)$ first crosses zero or the integral scale $\int_0^\infty R(x, r)\mathrm{d}r$.

To show how these definitions naturally now extend into the graph domain, we construct a circulant graph as before but in two dimensions. Figure 1 shows the graph and its associated spectrum along with some select eigenvectors. We use a 70×70 grid with $\Delta x = 1$ and use the L1-norm distance norm. Since the eigenvectors of the graph Laplacian are still just the eigenfunctions of the Laplace operator (this time two dimensional), we have an analytical form [25] that gives eigenvalues $\lambda = (k_x^2 + k_y^2)$ and eigenfunctions $\sin(k_x x)\sin(k_y y)$, $\sin(k_x x)\cos(k_y y)$, $\cos(k_x x)\sin(k_y y)$, and $\cos(k_x x)\cos(k_y y)$, a selection of which are shown in Fig. 1.

Now, for the first 50 eigenvectors, we compute the graph autocorrelation functions and plot them as a function of graph distance in Fig. 2. Although 50 lines are plotted, significantly fewer are visible due to the periodic nature of the domain and the particular graph chosen. Just as the eigenvalue
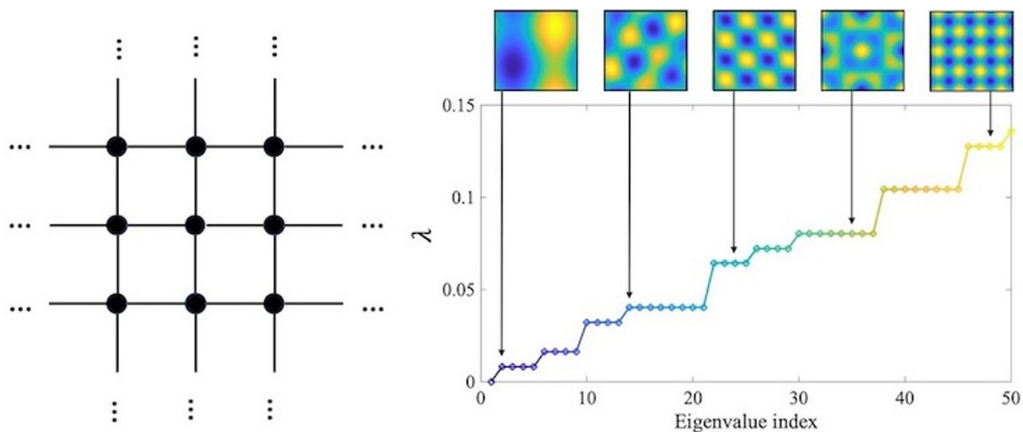
FIG. 1. (Left) Two-dimensional periodic grid. (Right) Spectrum of graph Laplacian with selected eigenvectors. The eigenvectors corresponding to larger eigenvalues vary more quickly across the network.

relationship has eigenvalues of multiplicity 4, each corresponding eigenvector is a shifted version of the others. Our approach captures the fact that shifts should not change either the autocorrelation or the length scales. Just as with the eigenvectors themselves, the autocorrelation functions that correspond to smaller eigenvalues vary slowly as a function of graph distance, either perfectly correlating or anticorrelating due to the periodic notion of the eigenvectors. Adopting the first-zero crossing approach to defining a characteristic scale, we compute the graph length scale of the first 50 autocorrelation functions and plot them as a function of eigenvalue in Fig. 3.

As expected, smaller eigenvalues are associated with larger length scales (smaller spatial frequencies). In this case, we also have an exact relation with which to compare: for circulant graphs, the eigenvalues should go as $\lambda \sim (\frac{1}{L})^2$ or $L \sim \frac{1}{\sqrt{\lambda}}$, so $\ln(L) \sim -\frac{1}{2}\ln(\lambda)$. Plotted in log space, we measure a slope of -0.5070, suggesting that our method for extracting a scale for each eigenvalue is nearly exact. Having thus defined a notion of graph frequency, we can begin to analyze trajectory graphs in graph spectral space and assign scales to the structures we discover.

D. Graph-based Lagrangian coherent structures

The actual weighted matrices that we want to analyze are those built from applying similarity measures to Lagrangian trajectories. Given a set of N_p trajectories $X_i(t)$, we can construct
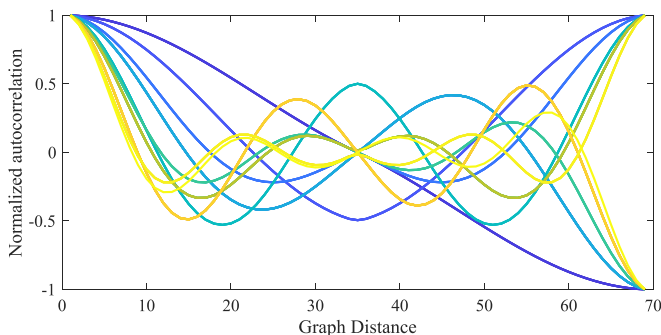

FIG. 2. Graph autocorrelation function for the first 50 eigenvectors. Colors correspond to the spectrum colors in Fig. 1.
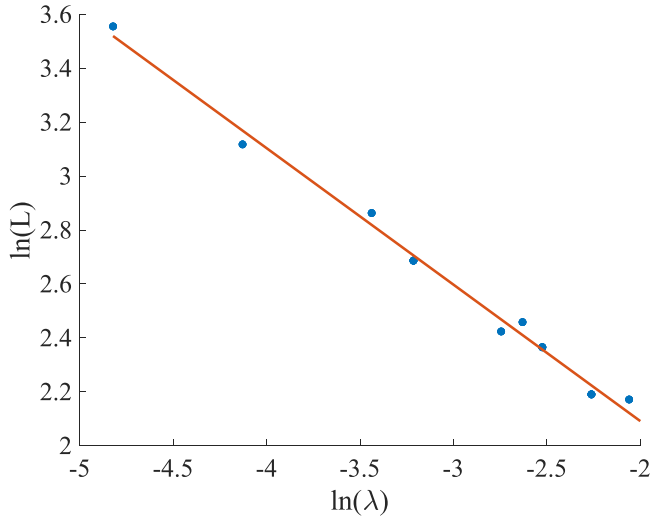
FIG. 3. Measured correlation length scale L as a function of eigenvalue λ. The solid line is a best-fit power law. Fewer than 50 points appear because eigenvectors with identical eigenvalues have identical graph autocorrelation functions.

a weighted adjacency matrix $\mathbf{W}$, where each entry corresponds to some measure of similarity between trajectories X_i and X_j. First, as with any spectral method, we expect some sensitivity to the number of trajectories chosen (see Ref. [9] for a more thorough discussion). Further, the choice of the relationship between X_i and X_j is itself a significant choice and will impact the results of subsequent calculations that employ the adjacency matrix $\mathbf{W}$, and so many different options have been developed and studied. Schlueter-Kuck and Dabiri [9], for example, used the metric $W_{ij} = \frac{1}{\overline{r_{ij}}T^{1/2}}[\sum_{k=0}^{T-1}(\overline{r_{ij}} - r_{ij}(t_k))^2]^{1/2}$ to penalize the kinematic dissimilarity between particles rather than to reward particles staying close over time. Their results are therefore agnostic to the actual location of particles, and the structures extracted from their adjacency matrix are not necessarily local in space. Other approaches construct the graph using some sort of distance-based metric [8]. Common methods include using either the L1 norm, $W_{ij} = [\sum_{k=0}^{T-1}(X_i(t_k) - X_j(t_k))^2]^{-1/2}$, the L2- norm, $W_{ij} = [\sum_{k=0}^{T-1}(X_i(t_k) - X_j(t_k))^2]^{-1}$, or some more involved distance metric. There are other methods still to generate the adjacency matrix, including that of Schneide *et al.* [12], where entries in $\mathbf{W}$ are set to 1 if two trajectories come within a certain encounter threshold ϵ and 0 otherwise. Although this list is not exhaustive, all graph-based formulations of coherent structures analysis aim to generate a large entry in W_{ij} when trajectories X_i and X_j are somehow more similar or dissimilar to one another over some time frame. The specific notion of similarity or dissimilarity will vary from method to method.

Standard practice then considers the eigenvectors corresponding to the k smallest eigenvalues and uses their information to infer something about coherence. The value of k is typically found via the eigengap heuristic, where eigenvalues close to 0 are assumed to indicate approximately coherent sets. It is not always the case that the eigenvectors are used directly, as in Froyland *et al.* [26] where the eigenvectors are used just to identify an appropriate subspace.

It is not straightforward which distance metric or form of the Laplacian to use when carrying out such an analysis [27], although it is fairly common practice to use the L1 norm and the generalized eigenvectors of Laplacian. In the following examples, we employ the L1 norm and the eigenvectors of the normalized graph Laplacian, which appear to give the most robust results. In contrast to typical graph-based studies of coherent structures, however, we keep all of the eigenvectors as a projection basis.
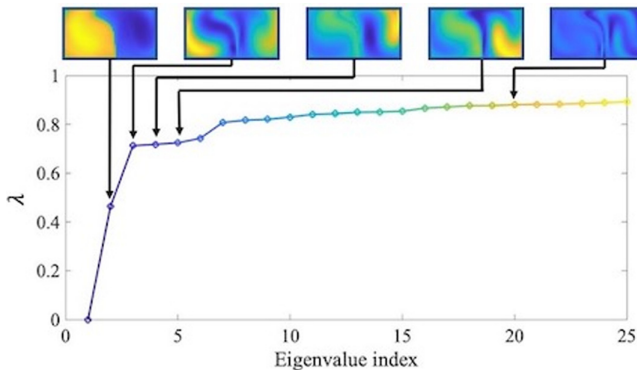
FIG. 4. Eigenvalue spectrum of the double-gyre trajectory graph along with a selection of eigenvectors.

III. DYNAMICALLY RELEVANT COHERENT STRUCTURES

In this section, we revisit two frequently considered LCS test cases, using the GFT to identify dynamically relevant coherent structures as well as specific length scales of interest. We first study a toy case known as the double gyre with two sets of initial conditions to show how kinematic coherence can be more or less important depending on what exactly is being transported [7]. Then, we explore the real-world transport of temperature, salinity, and vorticity in the Agulhas leakage.

A. Two initial conditions

The double gyre has been the subject of many LCS analyses [11] due to its prominent counterrotating vortices and complex transport dynamics. The flow is described by the set of coupled differential equations [28]

$$u = -\pi A \sin(\pi f) \cos(\pi y) \tag{6}$$

$$v = -\pi A \cos(\pi f) \cos(\pi y) \frac{df}{dx} \tag{7}$$

$$f = \epsilon \sin(\omega t) x^2 + [1 - 2\epsilon \sin(\omega t)]x, \tag{8}$$

where $A = 0.1$, $\epsilon = 0.25$, $\omega = 3\pi/5$, u, and v are velocities associated with the x and y directions, and the domain is defined on $x \in [0, 2]$, $y \in [0, 1]$. Here, we also consider the dynamics of an advected scalar field ϕ that satisfies the advection-diffusion equation

$$\frac{\partial \phi}{\partial t} = -\mathbf{u} \cdot \nabla \phi + \mu \nabla^2 \phi, \tag{9}$$

where $\mu = 10^{-3}$. Employing the L1 norm for our trajectory graph, we integrate an evenly spaced grid of 2500 particles for $T = 10$ s and compute the normalized graph Laplacian. Figure 4 shows the spectrum along with an illustrative selection of eigenvectors. As before, we also plot the graph autocorrelation functions for the first 25 eigenvectors in Fig. 5.

As is evident from both the spectrum and the plot of the graph autocorrelations, the largest-scale structure is given by the first nontrivial eigenvalue and corresponds to the two counterrotating vortices. The largest-scale structure, however, is not always the most dynamically relevant. Consider the two initial conditions given by $\phi_A(x, y, t = 0) = \sin(2\pi y)$ (condition A) and $\phi_B(x, y, t = 0) = \sin(\pi x)$ (condition B). After evolving both initial conditions under Eq. (9), we compute the GFT of the Lagrangian-averaged scalar field on the time interval $T = [0, 10]$ s using the eigenvectors of the normalized graph Laplacian according to Eq. (3).
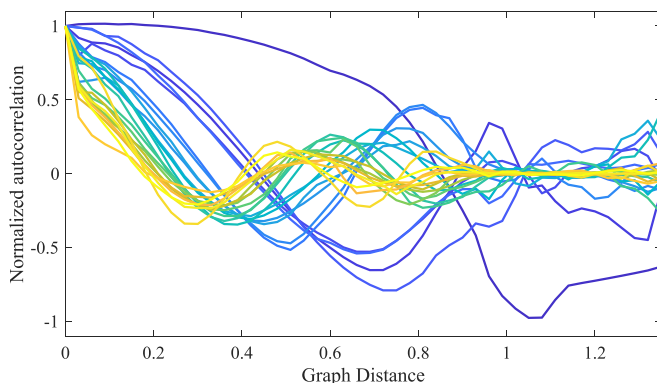
FIG. 5. Graph autocorrelation function for the first 25 eigenvectors of the double-gyre trajectory graph. Colors correspond to those shown in Fig. 4.

Although the kinematic behavior of the double gyre is identical in both scenarios, resulting in identical eigenvectors, the spectral content of each scalar field is substantially different. Figure 6 shows the Lagrangian-averaged scalar field and its corresponding spectral representation. Overlaid over each field is a bar corresponding to the length scale of the dominant eigenvector (see Fig. 5). For condition A, the second eigenvector is the most dynamically important for reconstructing the signal, and carries considerably more weight in spectral space than any other eigenvector. This result is in agreement with traditional intuition about the system, which would argue that the two vortices separated by a transport barrier dominate the transport dynamics. Observing the spectrum for condition B, however, it appears that this large-scale structure is no longer the most important for describing the dynamics. Instead, it is the higher harmonics that occur beyond the obvious spectral gap that dominate the transport. Because condition B is vertically inhomogeneous but well mixed
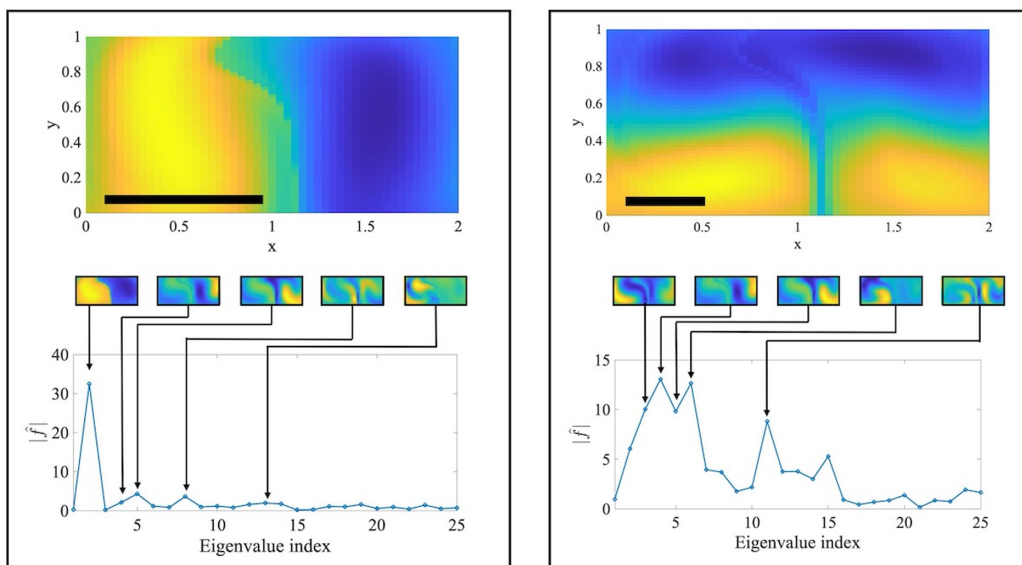

FIG. 6. Lagrangian-averaged scalar fields for condition A (left) and condition B (right), with corresponding spectra and the five most dominant eigenvectors. The bar on the scalar fields shows the length scale of the eigenvector with the most spectral weight.
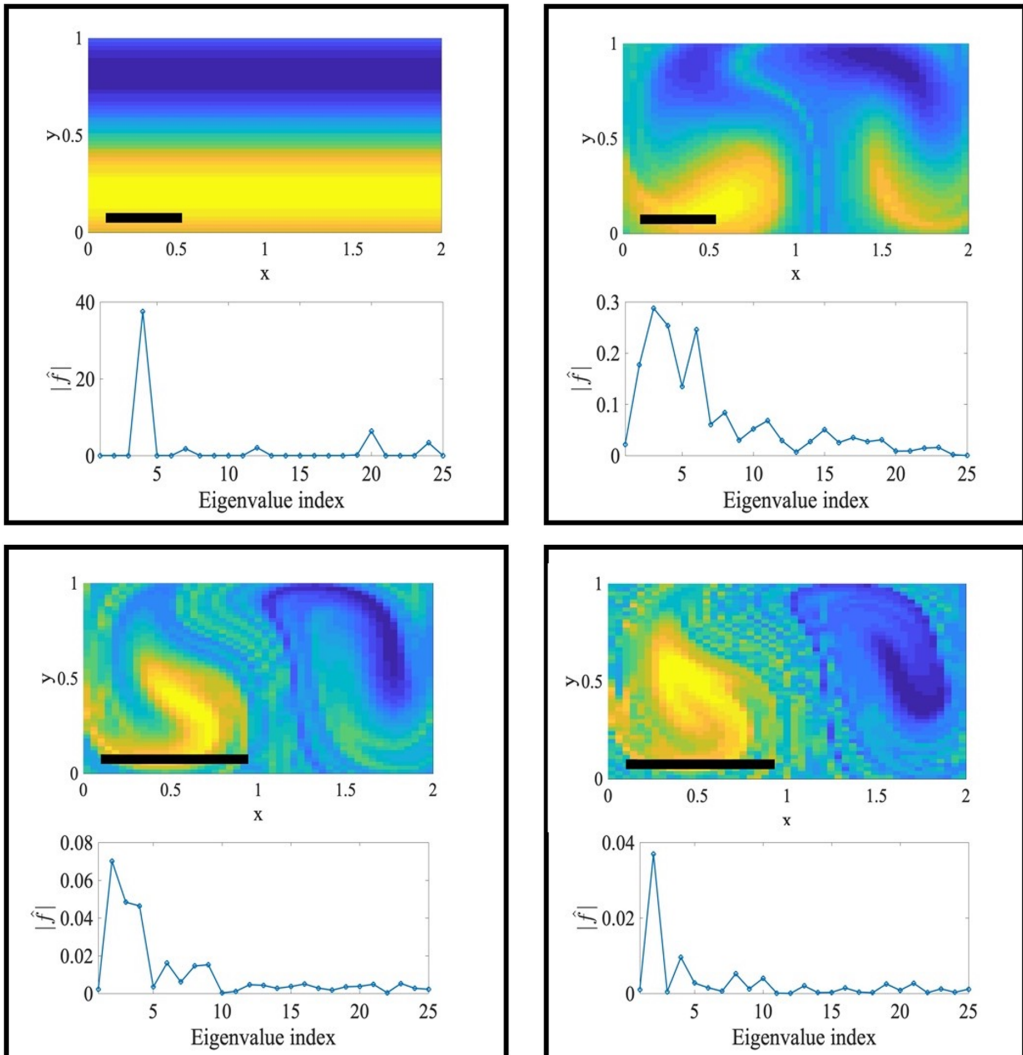
FIG. 7. Final time scalar fields corresponding to condition B with their spectra for times $T = 0$ s (top left), $T = 10$ s (top right), $T = 20$ s (bottom left), and $T = 30$ s (bottom right). As before, the bar on the scalar fields shows the length scale of the eigenvector with the most spectral weight.

in the horizontal direction, advection by the double-gyre structure (a primarily horizontal structure) is not as relevant.

This conclusion, however, is only valid for the $T = 10$ s time interval studied. Further, it is not entirely clear what part of these length scales are a result of the initial condition as opposed to the dominant kinematic length scales. One would expect the $T = 0$ s spectrum to be dominated completely by the length scale of the initial condition, whereas in the long-time limit the scalar would diffuse to the domain scale. Somewhere in between, the most relevant length scales would likely be those set by the kinematics of the flow. To tease apart this relationship, we plot four different spectra at progressively later times for condition B in Fig. 7. To avoid the spectrum being dominated by the large initial scalar gradients, we compute the spectrum of the final Lagrangian
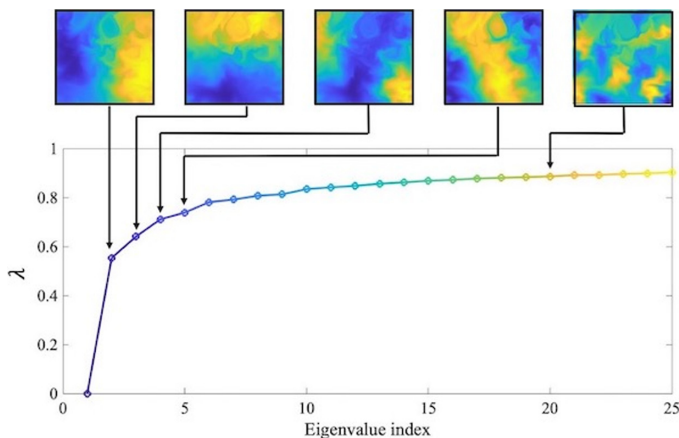
FIG. 8. Eigenvalue spectrum of the Agulhas leakage trajectory graph along with a selection of eigenvectors.

scalar values (as opposed to Lagrangian-averaged values) and show each trajectory at its initial position in the flow.

Initially, the spectrum is centered on the fourth eigenvector, the natural length scale of the initial condition. As the flow is advected, the dominant length scale becomes that of the two gyres, supporting previous conclusions that minimally diffusive barriers should be independent of initial conditions [29]. Although one might expect that advection would introduce smaller length scales as the flow evolves, this folding and stretching action of the flow is contained entirely in the eigenvectors themselves. Finally, we note that the combination of advection and diffusion in this system seems to operate as a shift operator in the graph frequency space. Conclusions from this are limited, however, as the eigenvectors themselves are also evolving with the flow.

B. Ocean transport

Using the same tools, we now turn to the real-world transport of temperature, salinity, and vorticity in the Agulhas leakage. The Agulhas leakage has been the focus of many studies that analyze its coherent structures [30,31], and separately work that considers its importance to ocean transport and climate [32]. Unifying these two threads of research has, however, proven difficult.

We numerically integrated a 100×100 grid of ocean parcels in the Agulhas leakage between latitudes $[-48 \ -41]$ and longitudes $[10 \ 20]$ over the span of 30 days [May 1, 2020, May 31, 2020] using the open-source COPERNICUS database, and computed the normalized Laplacian using the L1 norm. The eigenvalue spectrum with selected eigenvectors is shown below in Fig. 8, and the corresponding graph autocorrelation function is shown in Fig. 9. In contrast to the double gyre, we no longer see the large coherent vortices as the leading eigenvector. Instead, although close inspection does show the footprint of vortices, the leading eigenvectors look much more like the sinusoids of our first example. The graph autocorrelation function oscillates as well, suggesting that the parcels have not moved considerably over the time frame considered.

To begin our study of transport, we compute the Lagrangian-averaged temperature, salinity, and vorticity deviation (in the manner of Ref. [30]) and plot the related spectra in Fig. 10. At a first glance, coherent transport in the Agulhas leakage seems very relevant, as the temperature and salinity fields reveal a circular eddy with significantly larger temperature and salinity than its neighbors. A GFT puts this transport into more perspective. For both fields, almost all of the variation in the signal is held in the lowest-frequency eigenvectors—those shown in Fig. 8 to be varying north to south (the second nontrivial eigenvector for the temperature field) and other additional slowly varying modes (for the salinity field).
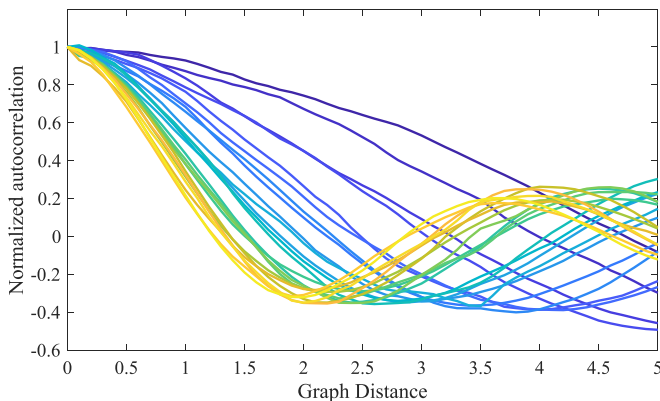
FIG. 9. Graph autocorrelation function for first 25 eigenvectors of the Agulhas leakage trajectory graph. Colors correspond to those shown in Fig. 4.

We can see that the situation is different for the vorticity deviation field. For the case of vorticity transport, the coherent vortices have a dominant role and the field is significantly more broadband. To investigate this result further, we compute the length scales of the leading 500 eigenvectors via our graph autocorrelation method and plot the spectrum as a function of length scale. Then, to test whether the length scales computed are independent of domain size, we perform the same computation but for a 100×100 grid of ocean parcels seeded in the subdomain between latitudes $[-43.5 \ -41]$ and longitudes $[14 \ 16.5]$. These two domains are shown in Fig. 11.

For the case on the larger domain, it appears that most of the signal is captured at the domain scale (approximately 4 degrees) but has broadband signatures at the vortex scale (0.5–2 degrees). Comparing this spectrum to the smaller case, we find many of the same approximate signatures, but this time the domain is almost entirely centered around a large vortex, and as expected the signal is most prominent at 2 degrees (the length scale associated with the largest vortex) and 0.5–1 degree (the length scale associated with the smaller vortex).
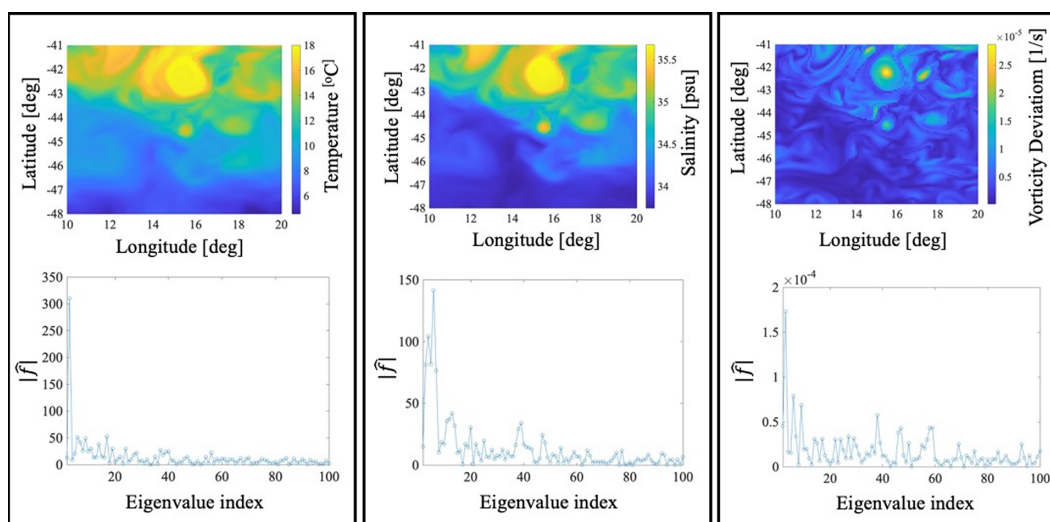

FIG. 10. Three different scalar fields [temperature (left), salinity (center), and vorticity (right)] for the Agulhas leakage and their associated GFT spectra.
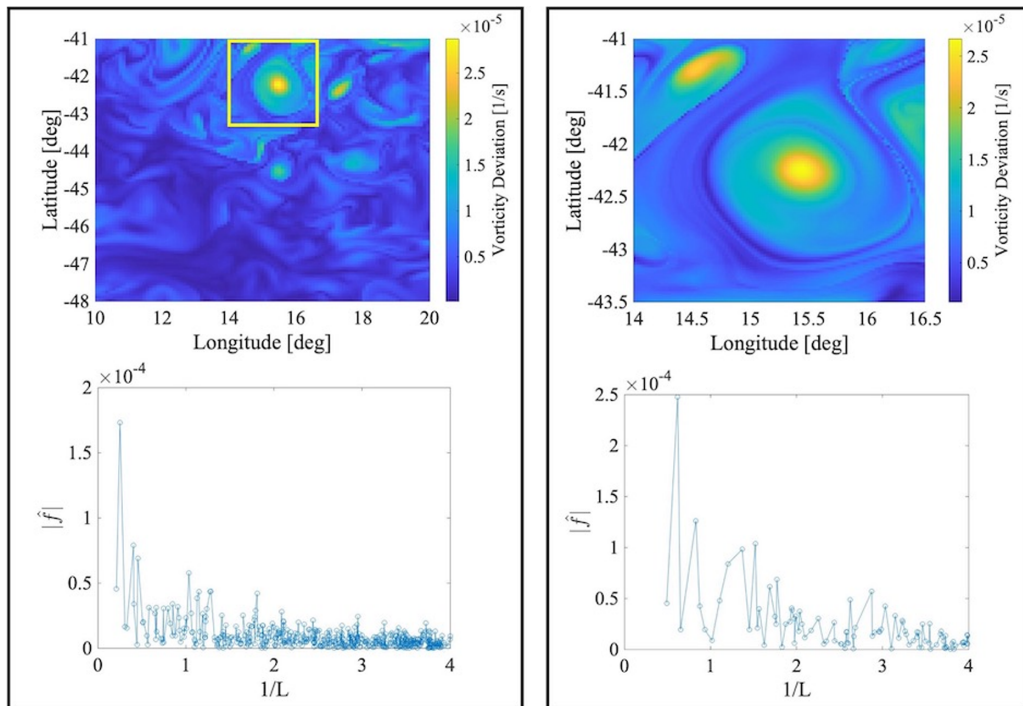
FIG. 11. Vorticity deviation fields for the Agulhas leakage with associated spectra computed at two different domain scales. The domain for the right-hand plot is indicated on the left-hand plot.

IV. TRANSPORT-BASED SCALE ANALYSIS

Augmenting LCS analysis with a precise notion of scale, as discussed above, suggests new possibilities for assessing the dynamical relevance of coherent structures for particular problems of interest. This importance of scale in describing dynamics, while relatively new for LCS analysis, has a longstanding and foundational role in the study of turbulent flows, albeit from an Eulerian perspective. Many problems in turbulence, however, are fundamentally linked with transport, and so a fully Lagrangian scale-based analysis such as the one we have developed would have a wide range of applicability.

Here, to illustrate the utility of our methods, we consider the case of droplets suspended in a turbulent flow and growing according to a growth law $\frac{dr}{dt} = \phi/r$, where r is the droplet radius and ϕ is a scalar field (for example, the water vapor saturation field) governed by Eq. (9). This problem has clear relevance to cloud physics and atmospheric science [33], but also shows how fundamentally Lagrangian quantities may be studied with our transport-based approach.

To begin, we solve the incompressible Navier-Stokes equations

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \nu \frac{\partial u_i}{\partial x_j \partial x_j} \tag{10}$$

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{11}$$

in two dimensions on the periodic domain $[0\ \pi] \times [0\ \pi]$. The details of the solver are given in Ref. [34]. We treat the droplets as passive, point particles and advect them in this flow field initially seeded on a 100×100 grid and with an initial radius of $r = 1\ \mu$m for a total time of $T = 5$ s. The initial scalar field is given by $\phi = 1 + \sin(2x)$. After integration of the droplet trajectories
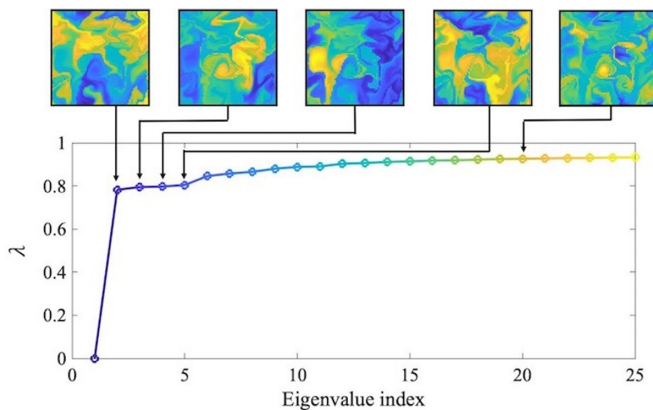
FIG. 12. Eigenvalue spectrum of the turbulence trajectory graph along with a selection of eigenvectors.

we compute the trajectory graph and in Figs. 12 and 13 show the spectrum and a selection of eigenvectors along with the graph autocorrelation function for the first 25 eigenvectors.

A. Comparison to Fourier transform

To compare the results of our method directly with what one would find from a Fourier transform, it is worth considering what field exactly is being scale decomposed. In the graph sense, the droplets can be described as nodes that exist in a high-dimensional trajectory space, and the GFT eigenvectors are defined precisely in this space on these nodes. Direct comparison with a Fourier transform is thus difficult: the modes of the Fourier transform are defined in physical space, so it is unclear whether we should be transforming the radius field based on the initial grid of droplets or based on the locations where they are advected to. To illustrate the differences between the two approaches, Figure 14 shows the two droplet radius fields. Both are colored according to droplet radius, but the field on the left is shown based on the final locations of the droplets, while the field on the right is based on the locations of the droplets pulled back to where they were initially seeded. The latter approach essentially eliminates the effects of advection, so can be thought of as a Lagrangian field.
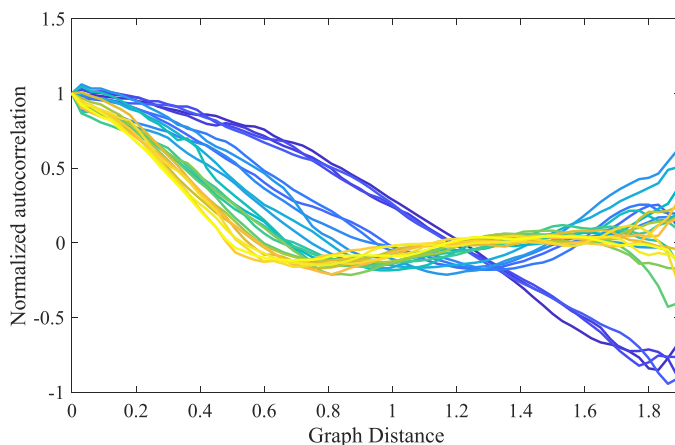

FIG. 13. Graph autocorrelation function for first 25 eigenvectors of the turbulence trajectory graph. Colors correspond to those shown in Fig. 12.
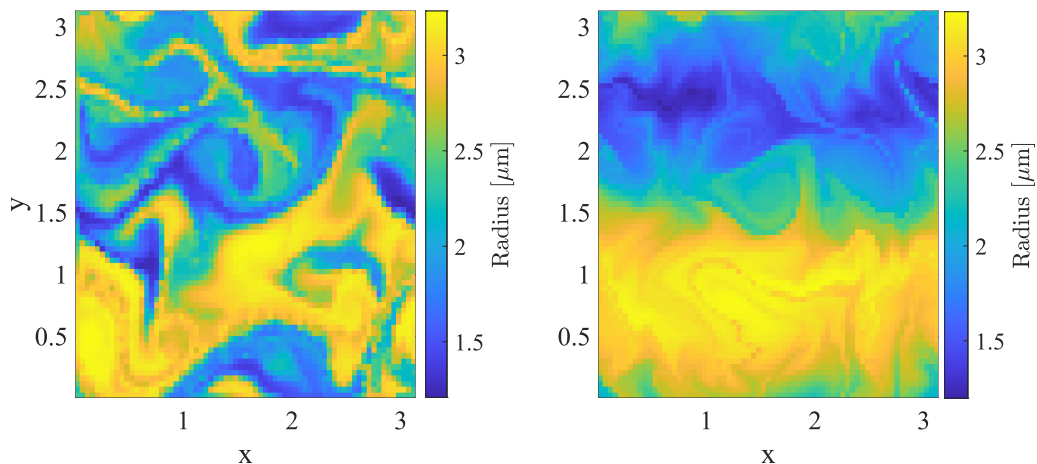
FIG. 14. (Left) Droplet radius field without pullback (that is, plotted at the final droplet locations). (Right) Droplet radius field with pullback (that is, plotted at the initial droplet locations).

To compare our results to a Fourier transform fairly, we therefore consider the Fourier transform of both fields—the one without pullback, which we will call Eulerian, and the one with pullback, which we will call Lagrangian. Another way of understanding the pullback operation is that we consider the Eulerian field but also advect the Fourier modes as if they were nondiffusing scalar fields. A more detailed treatment of this distinction is provided in [35].

Once we compute the Fourier spectrum with and without pullback as well as the GFT of the droplet radii, shown in Fig. 15, it is immediately clear that both the pullback DFT and the GFT are considerably more spectrally compact than the traditional DFT. This is not surprising, as (nonlinear) advection can create fronts that have very large spatial frequency components. To make this comparison quantitative, we reconstruct the radius field using the top N most contributing components and measure the error as a function of N, as shown in Fig. 16. This computation reveals that the GFT is much more compact than even the DFT with pullback, in that we can achieve a much more faithful reconstruction with considerably fewer modes. The tradeoff is that more computational time is required for computing the GFT modes.

It is also helpful to view the reconstructed fields themselves to qualitatively assess what features the GFT captures best. So, in Fig. 17 we show four reconstructed Lagrangian radius fields. In contrast to a traditional DFT, we can see that certain gradients in the flow are accurately reconstructed
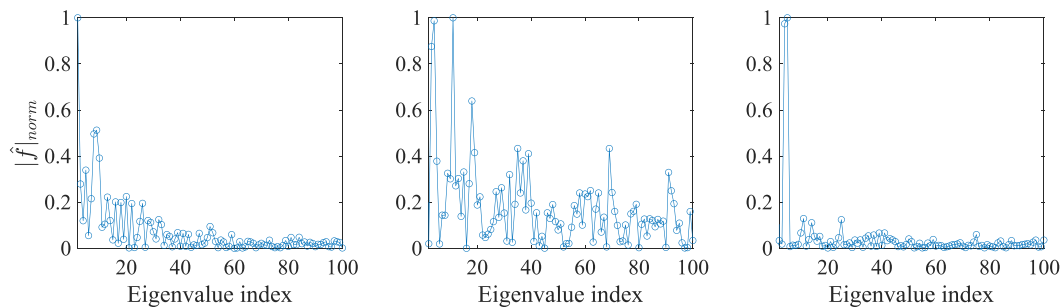

FIG. 15. (Left) GFT of the droplet radius field. (Center) DFT of the droplet radius field, with no pullback. (Right) DFT of the droplet radius field, with pullback.
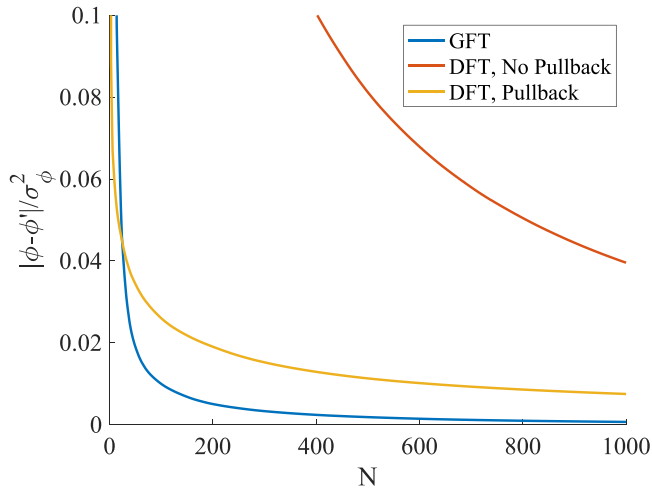
FIG. 16. Reconstruction error of the droplet radius field (see text) using only the N the most contributing modes as a function of N for the GFT, DFT without pullback, and DFT with pullback.

using only two eigenvectors. This is because the eigenvectors themselves, while smooth in trajectory space, have high-frequency components in physical space (see example eigenvectors in Fig. 12). We lose, however, some large-scale spatial structure that would likely be captured with a traditional
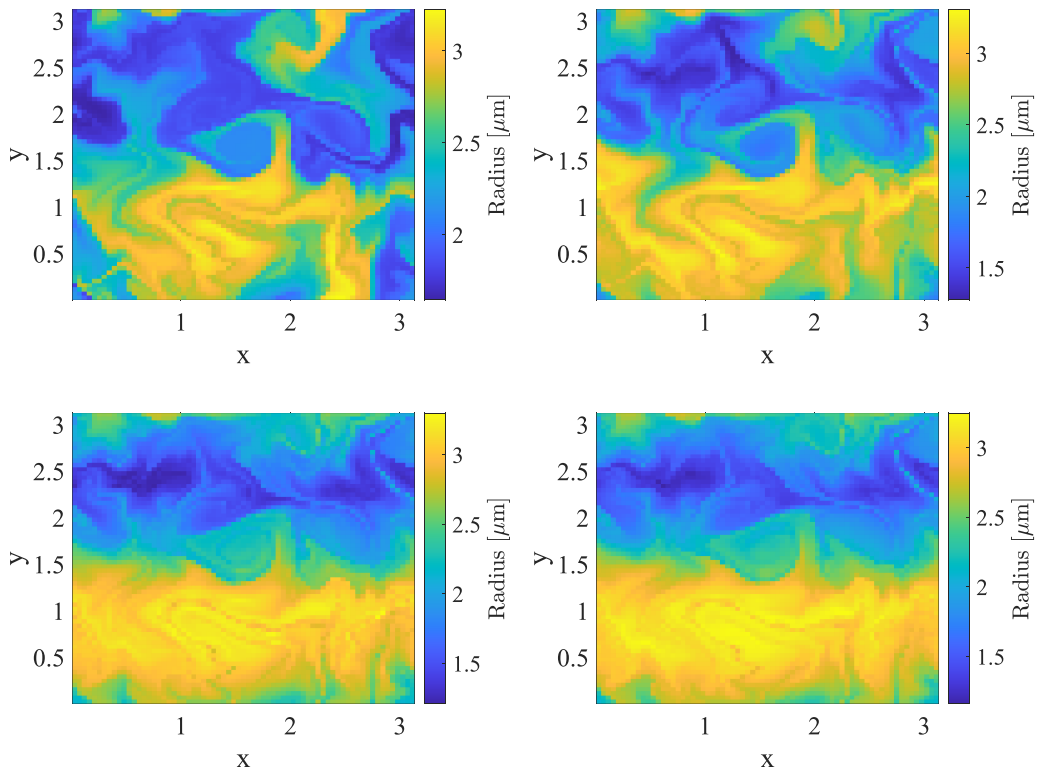

FIG. 17. Reconstructed Lagrangian field using $N = 2$ (top left), $N = 10$ (top right), $N = 100$ (bottom left), and $N = 500$ (bottom right) modes.
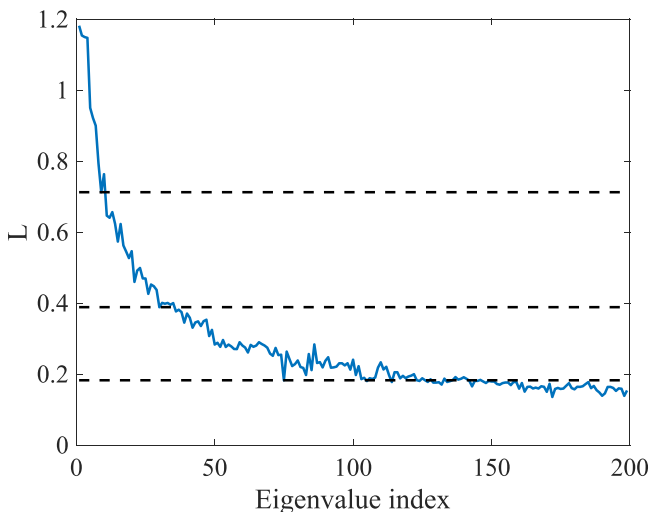
FIG. 18. Length scale of the leading 200 eigenvectors for the turbulence trajectory graph. Dashed lines show the selected filtering scales ($L = \pi/4$, $L = \pi/8$, and $L = \pi/16$).

DFT. This is likely why the DFT with pullback initially outperforms the GFT for a very small number of modes.

B. Filtering

This last result suggests that it is also possible to use our transport-based scale decomposition to filter Lagrangian quantities according to their Lagrangian footprint. To illustrate this idea, we compute the length scale of the leading 200 eigenvectors and filter at three different spatial cutoffs, shown in Fig. 18.

Next, in Fig. 19 we show the true probability density function (PDF) of the droplet radius along with the PDF of droplet radius reconstructed using only modes larger than various cutoff length scales. By using only larger-scale spatial modes, we retain the Lagrangian signal that is contained
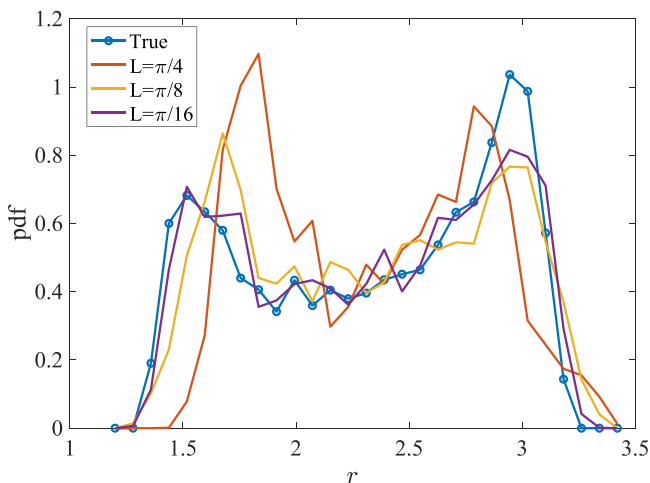

FIG. 19. True PDF of droplet radius as well as PDFs of the droplet radius filtered at different scales.

in the largest-scale Lagrangian structures. In this case, we can see that by only considering the largest Lagrangian modes, we capture the bimodal distribution of droplet radius but not the more detailed behavior. As smaller length scales are included, the PDFs approach the true distribution. Using such an approach, it is possible to uncover exactly which turbulent scales (in the Lagrangian sense) contribute to higher-order Lagrangian statistics.

V. CONCLUSION

We have adapted recent advances in graph signal processing to introduce a novel scale decomposition based on the transport properties of given flow. In contrast to previous well-established frequency-space transforms, our method is purely Lagrangian. We then demonstrate that this new approach has value in reassessing and quantifying the contribution of Lagrangian coherent structures to the dynamical behavior of systems, and also that it expands the range of problems one can approach from a Lagrangian perspective. Common procedures in flow analysis such as filter-based techniques can now be understood through a perspective grounded in the flow's transport properties.

There are many future avenues worth pursuing in applying the graph Fourier transform in the context of fluid dynamics. Resolution upscaling and data cleaning, for example, could be used for analyzing ocean data, where kinematic data such as velocity may be known at a high resolution (say, from satellite wind data) but dynamic data such as salinity may be known only at a lower resolution (say, from Argo floats). A Lagrangian-based scale decomposition offers a framework in which data acquired at these two resolutions can be integrated. As we have demonstrated that a GFT represents Lagrangian data in a storage-efficient manner, there is also the possibility of alleviating the bottleneck of storing large quantities of climate data from the ocean and the atmosphere. In addition, we speculate that the GFT is a promising route for approaching the problem of closure models for Lagrangian quantities such as droplet radius. Overcoming hurdles such as the computational burden of finding the graph eigenvectors and a strong physics-based interpretation of what exactly the modes represent will be key for a broader utilization of transport-based scale decompositions. Finally, we note that scale decompositions based on the eigenvectors of the dynamic Laplacian [25], while not graph based, could have similar desirable properties to those we have demonstrated here.

---

[1] L. Richardson, *Weather Prediction by Numerical Process* (Cambridge University Press, Cambridge, 1922), p. 66.

[2] P. Le Quere, and T. A. De Roquefortt, Computation of natural convection in two-dimensional cavities with chebyshev polynomials, J. Comput. Phys. **57**, 210 (1985).

[3] T. DelSole and M. K. Tippett, Laplacian eigenfunctions for climate analysis, J. Clim. **28**, 7420 (2015).

[4] M. Farge, Wavelet transforms and their applications to turbulence, Annu. Rev. Fluid Mech. **24**, 395 (1992).

[5] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. **52**, 477 (2020).

[6] G. Haller, Lagrangian coherent structures, Annu. Rev. Fluid Mech. **47**, 137 (2015).

[7] S. Balasuriya, N. T. Ouellette, and I. I. Rypina, Generalized lagrangian coherent structures, Physica D **372**, 31 (2018).

[8] A. Hadjighasem, D. Karrasch, H. Teramoto, and G. Haller, A spectral clustering approach to lagrangian vortex detection, Phys. Rev. E **93**, 063107(2016).

[9] K. L. Schlueter-Kuck and J. O. Dabiri, Coherent structure colouring: Identification of coherent structures from sparse data using graph theory, J. Fluid Mech. **811**, 468 (2017).

[10] A. Hadjighasem, M. Farazmand, D. Blazevski, G. Froyland, and G. Haller, A critical comparison of lagrangian methods for coherent structure detection, Chaos **27**, 053104 (2017).

[11] M. R. Allshouse and T. Peacock, Lagrangian based methods for coherent structure detection, Chaos **25**, 977617 (2015).

[12] C. Schneide, P. P. Vieweg, J. Schumacher, and K. Padberg-Gehle, Evolutionary clustering of Lagrangian trajectories in turbulent Rayleigh-Bénard convection flows, Chaos **32**, 013123 (2022).

[13] G. W. He, H. P. Ansari, G. D. Jin, D. Li, and A. Mani, A Lagrangian filtering approach for large-eddy simulation of particle-laden turbulence, *Proceeding of the Summer Program 2012* (Center for Turbulence Research, Stanford University, Stanford, CA, 2012), pp. 375–383.

[14] A. El Aouni, H. Yahia, K. Daoudi, and K. Minaoui, A fourier approach to Lagrangian vortex detection, Chaos **29**, 093106 (2019).

[15] T. MacMillan and D. H. Richter, The most robust representations of flow trajectories are Lagrangian coherent structures, J. Fluid Mech. **927**, A26 (2021).

[16] R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, Discovering Physical Concepts with Neural Networks, Phys. Rev. Lett. **124**, 010508 (2020).

[17] A. Sandryhaila and J. M. Moura, Discrete signal processing on graphs: Frequency analysis, IEEE Trans. Signal Proc. **62**, 3042 (2014).

[18] B. Ricaud, P. Borgnat, N. Tremblay, P. Gonçalves, and P. Vandergheynst, Fourier could be a data scientist: From graph Fourier transform to signal processing on graphs, C. R. Phys. **20**, 474 (2019).

[19] G. Cheung, E. Magli, Y. Tanaka, and M. K. Ng, Graph spectral image processing, Proc. IEEE **106**, 907 (2018).

[20] N. Saito, How can we naturally order and organize graph laplacian eigenvectors? *2018 IEEE Statistical Signal Processing Workshop (SSP)* (IEEE, Piscataway, NJ, 2018).

[21] D. I. Shuman, B. Ricaud, and P. Vandergheynst, Vertex-frequency analysis on graphs, Appl. Comput. Harm. Anal. **40**, 260 (2016).

[22] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Proc. Mag. **30**, 83 (2013).

[23] F. Iannelli, A. Koher, D. Brockmann, P. Hövel, and I. M. Sokolov, Effective distances for epidemics spreading on complex networks, Phys. Rev. E **95**, 012313 (2017).

[24] F. Iannelli and I. M. Sokolov, Path-integral formulation of spreading processes in complex networks, Eur. Phys. J.: Spec. Top. **230**, 2793 (2021).

[25] G. Froyland, Dynamic isoperimetry and the geometry of Lagrangian coherent structures, Nonlinearity **28**, 3587 (2015).

[26] G. Froyland, C. P. Rock, and K. Sakellariou, Sparse eigenbasis approximation: Multiple feature extraction across spatiotemporal scales with application to coherent set identification, Commun. Nonlin. Sci. Numer. Simul. **77**, 81 (2019).

[27] U. von Luxburg, A Tutorial on spectral clustering, arXiv:0711.0189.

[28] S. C. Shadden, F. Lekien, and J. E. Marsden, Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows, Physica D **212**, 271 (2005).

[29] G. Haller, D. Karrasch, and F. Kogelbauer, Material barriers to diffusive and stochastic transport, Proc. Natl. Acad. Sci. USA **115**, 9074 (2018).

[30] G. Haller, A. Hadjighasem, M. Farazmand, and F. Huhn, Defining coherent vortices objectively from the vorticity, J. Fluid Mech. **795**, 136 (2016).

[31] T. MacMillan, N. T. Ouellette, and D. H. Richter, Detection of evolving Lagrangian coherent structures : A multiple object tracking approach, Phys. Rev. Fluids **5**, 124401 (2020).

[32] L. M. Beal, W. P. De Ruijter, A. Biastoch, R. Zahn, M. Cronin, J. Hermes, J. Lutjeharms, G. Quartly, T. Tozuka, S. Baker-Yeoboah, T. Bornman, P. Cipollini, H. Dijkstra, I. Hall, W. Park, F. Peeters, P. Penven, H. Ridderinkhof, and J. Zinke, On the role of the Agulhas system in ocean circulation and climate, Nature (London) **472**, 429 (2011).

[33] R. A. Shaw, Particle-turbulence interactions in atmospheric clouds, Annu. Rev. Fluid Mech. **35**, 183 (2003).

[34] D. Karrasch and N. Schilling, Fast and robust computation of coherent Lagrangian vortices on very large two-dimensional domains, SMAI J. Comput. Math. **6**, 101 (2020).

[35] D. Karrasch and J. Keller, A geometric heat-flow theory of Lagrangian coherent structures, J. Nonlinear Sci. **30**, 1849 (2020).