Evolutionary clustering of Lagrangian trajectories in turbulent Rayleigh-Bénard convection flows

Christiane Schneide,1 Philipp P. Vieweg,2 Jörg Schumacher,2,[a] and Kathrin Padberg-Gehle1,[b]

1) Institute of Mathematics and its Didactics, Leuphana Universität Lüneburg,
D-21335 Lüneburg, Germany
2) Institute of Thermodynamics and Fluid Mechanics, Technische Universität Ilmenau,
D-98684 Ilmenau, Germany

(Dated: 3 January 2022)

We explore the transport mechanisms of heat in two- and three-dimensional turbulent convection flows by means of the long-term evolution of Lagrangian coherent sets. They are obtained from the spectral clustering of trajectories of massless fluid tracers that are advected in the flow. Coherent sets result from trajectories that stay closely together under the dynamics of the turbulent flow. For longer times, they are always destroyed by the intrinsic turbulent dispersion of material transport. Here, this constraint is overcome by the application of evolutionary clustering algorithms that add a time memory to the coherent set detection and allow individual trajectories to leak in or out of evolving clusters. Evolutionary clustering thus also opens the possibility to monitor the splits and mergers of coherent sets. These rare dynamic events leave clear footprints in the evolving eigenvalue spectrum of the Laplacian matrix of the trajectory network in both convection flows. The Lagrangian trajectories reveal the individual pathways of convective heat transfer across the fluid layer. We identify the long-term coherent sets as those fluid flow regions that contribute least to heat transfer. Thus, our evolutionary framework defines a complementary perspective on the slow dynamics of turbulent superstructure patterns in convection flows that were recently discussed in the Eulerian frame of reference. The presented framework might be well suited for studies in natural flows which are typically based on sparse information from drifters and probes.

[a) Also at Tandon School of Engineering, New York University, New York, NY 11201, USA.
[b) Electronic mail: kathrin.padberg-gehle@leuphana.de]
The transfer of heat remains one of the central questions of turbulent convection. Our present understanding of its coupling to the material transport in fully turbulent flows over longer times is still incomplete. Here, we shed new light on this point by evolutionary spectral clustering in Lagrangian networks that are formed by trajectories of fluid particles which are advected in the flow. Our framework identifies the coherent sets in two- and three-dimensional Rayleigh-Bénard convection that stay together for longer times, detects their splits and mergers, and reveals their role as those spatial regions that contribute least to the global heat transfer from the bottom to the top. This paves the way for new sparse robust investigation methods of turbulent transport in natural flows that complement the standard analyses in the Eulerian frame of reference.

I. INTRODUCTION

Turbulent thermal convection is the essential mechanism by which heat is transported across extended layers or in closed vessels with numerous geophysical, astrophysical and technological applications. The description of convection in the Lagrangian frame of reference, which is always connected to the material transport, opens the perspective of a more precise quantification of the complex spatio-temporal pathways that heat takes through the flow and thus for a classification into spatial regions that contribute more or less to its transfer. Lagrangian transport and mixing processes have been intensively studied by means of mathematical and computational tools from dynamical systems theory. These methods have been successfully applied to the study of mass transport in several systems, such as marine ecosystems, cardiovascular flows, astrodynamical systems, or geophysical problems including predictions of oceanic pathways of pollution patterns and marine debris.

At the core of Lagrangian approaches is the concept of a Lagrangian coherent set, representing a regularly shaped region in the fluid volume that only weakly mixes with its surrounding. The boundaries of such regions can be identified within a geometric approach, where Lagrangian coherent structures (LCS) represent minimal curves or surfaces that enclose coherent sets. The LCS framework makes use of local stretching properties of the underlying flow such as measured by finite-time Lyapunov exponents (FTLE) and
has been recently extended to the case of weakly diffusive transport and to transport of other quantities. Coherent sets were originally introduced within a probabilistic approach based on transfer operators and they have been characterized as sets which possess a minimal boundary-to-volume ratio for the dynamics by means of spectral properties of a related dynamic Laplacian operator. An extension of this framework allows for studying the emergence and decay of coherent sets. Recent approaches make use of spatio-temporal clustering algorithms applied to Lagrangian trajectory data. These unsupervised machine learning algorithms aim at identifying coherent sets as groups of trajectories that remain close to each other in the time interval under investigation and can deal with sparse data and gaps in the observation. Coherent sets are also connected to the turbulent superstructures of convection, a gradually evolving large-scale skeleton detected in fully turbulent flows in the complementary Eulerian frame of reference, forming the backbone of turbulent heat transport.

Conceptually, all Lagrangian approaches are tailored to identify material structures characterized by minimal mass transport. Such a property is not expected to hold for fully turbulent flows. Turbulent dispersion dissolves larger coherent sets quickly. Coherence in this material sense rather holds for smaller-scale features such as intense vortices, which all of these approaches will preferentially identify. Moreover, the trajectory-based clustering approaches typically rely on a spatio-temporal similarity measure, which aggregates the spatio-temporal relationship between two trajectories into a single number. These similarity measures quickly become meaningless due to turbulent dispersion, causing Lagrangian approaches to fail for long-term investigations in fully turbulent flows. One first attempt to overcome this problem is to follow coherent sets from short-time computations by means of a multiple object tracking method. However, such procedure strongly relies on the capability of the underlying clustering algorithm to identify reliable structures. This sets the major motivation for the present work.

Here, we extend the trajectory-based network approach to an evolutionary spectral clustering framework which blends instantaneous analyses with a dynamic short-term memory. In this way, the formerly required material property of the sets of interest is relaxed thus opening doors to study the long-term evolution of Lagrangian coherent sets in two- and three-dimensional turbulent flows. The result are persistent sets with altering components, which we extract via a sparse eigenbasis approximation methodology from the network
Laplacian matrix. Thereby, we are able to restrict the clustering to sets which are at the center regions of the convection rolls in turbulent convection. We can then study the long-term evolution of these structures including mergers and splits - which significantly extends previous results\cite{26,27} - and confirm that the coherent sets contribute least to the turbulent heat transport\cite{27}. Additionally, these splits and mergers can be clearly identified by fast switches of the gaps in the spectrum of the network Laplacian. We thus present a method to study the slow dynamics of turbulent superstructures in convection from a Lagrangian view.

The outline of the manuscript is as follows. Section II introduces the numerical simulation model of the convection flow together with important dimensionless parameters that characterize the turbulent transport. Sections III and IV discuss in brief the concept of Lagrangian coherent sets and the evolutionary clustering, respectively. Sections V and VI summarize the results for the two- and three-dimensional cases. We conclude the work in Section VII and provide further trajectory analysis details in two appendices.

\section{II. TWO- AND THREE-DIMENSIONAL RAYLEIGH-BÉNARD CONVECTION MODELS}

We study Lagrangian coherent sets in thermal convection flows in a two-dimensional (2d) or three-dimensional (3d) Cartesian domain, which is uniformly heated from below and cooled from above. The horizontal extension $L$ of the domains is in both cases significantly larger than their vertical extension $H$, which is quantified by the aspect ratio $\Gamma = L/H$. The two-dimensional (2d) flow case is studied in a closed domain $x \in [-4, 4]$ and $z \in [0, 1]$. In the three-dimensional (3d) case, we consider $x, y \in [-8, 8]$ and the vertical coordinate $z$ in the same range as in the 2d case. All coordinates are given in units of the layer height $H$.

The underlying equations of motion in the Boussinesq approximation are solved in both cases by a direct numerical simulation applying the spectral element method Nek5000\cite{38}. The equations are made dimensionless by the layer height $H$, wall-to-wall temperature difference $\Delta T$, and free-fall velocity $U_f = \sqrt{g \alpha \Delta TH}$ with the acceleration due to gravity $g$ and the isobaric expansion coefficient $\alpha$. Consequently times are measured by the free-fall time $T_f$.
which is given by

$$T_f = \frac{H}{U_f} = \sqrt{\frac{H}{g\alpha\Delta T}}.$$  \hfill (1)

The equations of motion in dimensionless form follow to

$$\nabla \cdot \mathbf{u} = 0,$$  \hfill (2)

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \sqrt{\frac{Pr}{Ra}} \nabla^2 \mathbf{u} + T \mathbf{e}_z,$$  \hfill (3)

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla) T = \frac{1}{\sqrt{Ra Pr}} \nabla^2 T.$$  \hfill (4)

The variables $\mathbf{u}(x, t), T(x, t)$ and $p(x, t)$ represent the non-dimensional velocity, temperature, and pressure fields, respectively, with $x = (x, y, z)^T$. The Rayleigh number $Ra$, which is given by

$$Ra = \frac{\alpha g \Delta T H^3}{\nu \kappa},$$  \hfill (5)

describes the strength of thermal driving. The Prandtl number $Pr$, which is given by

$$Pr = \frac{\nu}{\kappa},$$  \hfill (6)

measures the ratio of kinematic viscosity $\nu$ to temperature diffusivity $\kappa$, thus being a working fluid parameter. Table I summarizes all important parameters and quantities for both, the two-dimensional and three-dimensional simulation. Both simulations are for the turbulent flow case. Figure 1 shows an iso-surface snapshot of the 3d turbulent temperature field at the isolevel $T = 0.7$ which illustrates clearly the rising thermal plume structures. While the spectral element code works in the Eulerian frame of reference, the advection of $N$ massless tracer particles is an analysis in the Lagrangian frame of reference. It is conducted together with the time advancement of the Boussinesq equations. The trajectories of the $N$ tracers follow

$$\dot{\mathbf{X}}_i = \mathbf{u}(\mathbf{X}_i, t) \quad \text{for} \quad i = 1, \ldots, N.$$  \hfill (7)

The velocity field components in (7) are obtained by spectral interpolation to the tracer position $\mathbf{X}(t)$. The spectral resolution of the simulations is governed by the number of spectral elements $N_e$ together with the polynomial order $N_p$ on each spectral element and in each spatial dimension. The temporal advancement of the simulations exploits a second order backwards difference formula.
FIG. 1. Perspective view of the instantaneous iso-surface at $T(x,y,z) = 0.7$ in turbulent Rayleigh-Bénard convection. The system parameters $Pr = 0.7$, $Ra = 10^5$ and $\Gamma = 16$. This iso-surface shows regions of up-welling fluid flow, indicating regions of large convective heat flux. (Leaking) Lagrangian coherent sets can be found between this hot iso-surface and its cold counterpart (down-welling thermal plumes, not shown), thus reducing the overall heat transport.

The initial conditions (ICs) of the simulations are given by a linear vertical temperature profile at equilibrium subject to a small random perturbation $\Upsilon$ of maximum order $O(10^{-3})$ together with a fluid at rest, i.e., $T(x,t = 0) = 1 - z + \Upsilon$ and $u(x,t = 0) = 0$. The initial transient evolution of the dynamical system towards a statistically stationary state is ignored for the fields; the analysis starts once the tracer particles are homogeneously distributed across the simulation domain. The boundary conditions (BCs) are as follows. The flows are driven by uniform temperatures (Dirichlet-type BCs) at the bottom and top plate, i.e. $T(x,y,z = 0,t) = 1$ and $T(x,y,z = 1,t) = 0$, respectively, whereas the lateral walls are thermally insulating (Neumann-type BC), i.e. $\partial T/\partial n = 0$ where $n$ is the normal vector of the corresponding lateral wall. The simulation domains are closed in all directions and exhibit – just as in a real laboratory experiment – no-slip BCs, i.e. $u(x = x_b,t) = 0$ with domain boundaries located at $x_b$. The aspect ratios are chosen large enough such that finite-domain-size effects are negligible.

The Nusselt number (in the Eulerian frame of reference) $\text{Nu}^E$ is given by

$$
\text{Nu}^E = -\left\langle \frac{\partial T}{\partial z} \right\rangle_{z=0}^{\partial \Omega_{z,t}} = -\left\langle \frac{\partial T}{\partial z} \right\rangle_{z=1}^{\partial \Omega_{z,t}}.
$$

(8)

It is the global measure of mean turbulent heat transfer from the bottom to the top with
Here, $\langle \cdot \rangle_{\partial \Omega, t}$ is either a combined average with respect to time $t$ and a line at constant $z$ in 2d or time and a horizontal cross-section area $A$ in 3d. The Reynolds number $Re$, which is given by

$$Re = \sqrt{\frac{Ra}{Pr} \langle u^2 \rangle_{\Omega, t}},$$

(9)
is the global measure of turbulent momentum transport. Here, $\langle \cdot \rangle_{\Omega, t}$ is either a combined average with respect to time $t$ and simulation plane in 2d or time and simulation volume in 3d. In addition to the global measure of turbulent heat transport (see above), we compute for each individual tracer particle $i$ a local Nusselt number $Nu_{\text{local}}$ in the Lagrangian frame of reference along its trajectory $X_i(t)$. This Nusselt number is given by

$$Nu_{\text{local}}[X_i(t)] = \sqrt{RaPr} u_z T \bigg|_{X_i(t)} \frac{\partial T}{\partial z} \bigg|_{X_i(t)}.$$  

(10)

We verified that the Lagrangian ensemble average of the time-averaged local Nusselt numbers coincides with the mean value $Nu^E$ in the Eulerian frame of reference. This is documented in Table I. Here, the time average is taken over the whole time interval $T$ and the Lagrangian ensemble average over all tracers ($L$). This results to

$$\langle Nu_{\text{local}} \rangle_{T, L} = \frac{1}{N} \sum_{i=1}^{N} \langle Nu_{\text{local}}[X_i(t)] \rangle_T \approx Nu^E.$$  

(11)

From the table we note that the standard deviations of the Eulerian and Lagrangian Nusselt numbers differ considerably. This is caused by the ways they are computed. The comparatively large standard deviation in the Lagrangian frame represents different contributions of individual trajectories; the smaller Eulerian value represents temporal changes in the line or area average $\langle \cdot \rangle_{\partial \Omega}$.  

III. TRAJECTORY-BASED LAGRANGIAN COHERENT SETS

Recently, complex network approaches have been successfully applied to the study of turbulent flows. In this work, we extend the trajectory-based network framework for the identification of coherent sets proposed in Ref. 25, where an unweighted network was considered. Given $N$ tracer trajectories $X_i(t)$ at discrete time instances $t \in T = \{0, 1, \ldots, \Theta\}$ and a threshold $\epsilon > 0$, we set up instantaneous adjacency matrices $A_t \in \{0, 1\}^N$ where $A_{ij,t} = 1$ if $\|X_i(t) - X_j(t)\| < \epsilon$ for $i \neq j$ and $A_{ij,t} = 0$ otherwise. From this we obtain a network weight
TABLE I. Parameters and statistical measures of the numerical models. The dimension $d$ of the spatial domain, the aspect ratio $\Gamma$, the Rayleigh number $Ra$, the Prandtl number $Pr$, the number of spectral elements $Ne$, the polynomial order $Np$ on each spectral element, the total time of Lagrangian analysis $\Delta t^L$, the number $N$ of massless Lagrangian tracer particles, the global Nusselt number $Nu^E$ in the Eulerian frame of reference, its Lagrangian trajectory-based counterpart $\langle Nu_{\text{local}} \rangle_{T,L}$, as well as the global Reynolds number $Re$ are listed. The latter three quantities are given with respect to their mean value and the corresponding standard deviation.

| $d$ | $\Gamma$ | Ra | Pr | $Ne$ | $Np$ | $\Delta t^L$ | $N$ | $Nu^E$ | $\langle Nu_{\text{local}} \rangle_{T,L}$ | Re |
|-----|----------|----|----|------|------|-------------|----|--------|-----------------|----|
| 2   | 8        | $10^8$ | 7.0 | 384  | 11   | 1500  | 16,384 | 26.4 ± 0.9 | 26.2 ± 9.6       | 567.3 ± 10.7 |
| 3   | 16       | $10^5$ | 0.7 | 441,000 | 5   | 158   | 65,536 | 4.1 ± 0.1  | 4.3 ± 1.0        | 91.1 ± 0.5     |

FIG. 2. Sketch of Lagrangian tracers moving in a convection cell. The black dashed ellipse represents the $\epsilon$-neighborhood of the yellow particle. Consecutive weighted networks computed over the first and second snapshot (second and third, respectively) are shown below.

The matrix $W = \sum_{t \in \mathbb{T}} A_t$ with components $W_{ij}$ that encode the number of $\epsilon$-close encounters of tracers $X_i$ and $X_j$ over the entire time span. Hence, $W_{ij}$ is large when the trajectories $X_i$ and $X_j$ are close. Classical network measures can be employed, such as the weighted and unweighted node degree which are given by $d_{w,i} = \sum_{j=1}^{N} W_{ij}$ and $d_{u,i} = \sum_{j=1}^{N} W_{ij}$, respectively. Here, $\overline{W} \in \{0,1\}^N$ is the adjacency matrix of the unweighted network. The latter measures how many different trajectories interact with $X_i$ and can be linked to FTLE$^{[10]}$. In our context, the normalized node degree, which is given by $d_i = d_{u,i} / d_{w,i}$, measures the
number of different links of node \( X_i \) in relation to the total links weights. It is expected to be large in highly mixing regions, where each particle interacts with many different other ones. In order to analyze regions with different dynamical behavior, we seek to create a hard clustering of the trajectories. To this end, for example, one could define a threshold on the normalized node degree. Another approach is the application of spectral graph partitioning techniques, i.e. the partitioning of a data set that is represented by a graph based on spectral properties of a corresponding similarity matrix. Here, we use the weight matrix \( W \) as similarity matrix and optimize the normalized cut (NCut), originally proposed in Ref. [41] to determine the hard clustering. The cost function of the k-way normalized cut \(^{42}\), i.e., a partitioning into \( k \) sets, to be minimized is given by

\[
\text{Cost}_{\text{NCut}} = k - \text{Tr} \left[ X^T \left( D^{-1/2} W D^{-1/2} \right) X \right],
\]

where \( D \in \mathbb{R}^{N,N} \) denotes the degree matrix with \( D_{ii} = d_{w,i} \). Therefore, the entry \( i,j \) of the matrix \( D^{-1/2} W D^{-1/2} \) is given by \( W_{ij} / \sqrt{d_{w,i} d_{w,j}} \). The normalized weight matrix resembles a random walk matrix and has eigenvalues \( \lambda_i \leq 1 \). The cost function \( \text{Cost}_{\text{NCut}} \) is minimized by the matrix \( X \in \mathbb{R}^{N,k} \) if the columns of \( X \) are composed of the eigenvectors associated to the \( k \) largest eigenvalues of the matrix \( D^{-1/2} W D^{-1/2} \), where \( k \) can be determined using a spectral gap criterion. The clustering is subsequently produced in a post-processing step using the eigenvectors of the matrix \( X \). To this end, we use the sparse eigenbasis approximation (SEBA)\(^{37}\). From this, we derive a cluster membership indicator \( S_{\text{max}} \), which highlights particles in coherent regions (see Appendix A for mathematical details).

IV. EVOLUTIONARY SPECTRAL CLUSTERING

Building on the trajectory-based network representation we propose two evolutionary frameworks that allow us to study the evolution of large-scale coherent sets in turbulent flows in the Lagrangian frame of reference over long time spans.

A. Evolving network approach

First, we consider time-dependent weight matrices \( W_t \) that are computed over a shorter time period \( \tau \ll \Theta \), centered at \( t \), i.e., \( T_{t,\tau} = \{ t - \frac{\tau}{2}, \ldots, t + \frac{\tau}{2} \} \), via \( W_t = \sum_{s \in T_{t,\tau}} A_s \).
FIG. 3. Long-term evolution of coherent sets using the network evolution approach displays two merging events. (a) Postprocessing of the spectral information outputs a likelihood $S_{\text{max}}(X_i(t))$ for each particle trajectory $X_i$ to belong to a cluster over the time span $t \pm \frac{T}{2}$ (multimedia view), see Appendix A. (b) The normalized node degree $d(X_i(t))$ gives a similar distinction of clusters and mixing regions (multimedia view). Similarly to panel (a), the whole simulation domain is visualized at 6 time instances. (c) The 10 leading eigenvalues of the matrix $D_t^{-1/2}W_tD_t^{-1/2}$ plotted versus time highlight the existence of eight coherent sets at the beginning which merge into seven and then six at about $t = 480T_f$ and $1150T_f$, respectively; see Appendix B for more details on the evolution of the spectrum.

Obviously, the weight matrix $W_t$ is constructed from $W_{t-1}$ by adding future information $A_{t+\tau/2}$ and dropping the most historical information $A_{t-\tau/2}$, see Figure 2 for an illustration. When $\tau$ is sufficiently large, $W_t$ and thus also the matrix of the leading $k$ eigenvectors $X_t$ from the relaxed normalized cut problem can be expected to vary “smoothly” in time.

This construction is an adoption of the sliding window approaches in the transfer operator framework\cite{43,44} to the trajectory-based context and thus considered as a variation of the evolutionary spectral clustering approach preserving cluster quality (PCQ)\cite{36}. Notably, by this evolutionary construction the material property of coherent sets over the whole time
span $\mathbb{T}$ is relaxed and holds strictly only over smaller time spans of length $\tau$ used for setting up $W_t$. Significant changes in the dynamics such as mergers or splits of coherent sets will leave their footprints in the time evolution of the spectral properties of $D_t^{-1/2}W_tD_t^{-1/2}$, which can be highlighted by means of the cluster membership certainty $S_{\text{max}}$ or local network measures such as the normalized node degree $d$. By $S_{\text{max}}(X_i(t))$ as well as $d(X_i(t))$ we denote the respective time-dependent quantities for trajectories $X_i$ over the time span $t \pm \frac{\tau}{2}$ centered at time $t$.

B. Evolving cluster approach

The evolving network approach works well when sufficiently strong spectral gaps exist such that short-term variations due to systematic errors or noise do not cause the recent clustering to deviate strongly from the preceding one, such as also required in Ref. 35. In case of a simultaneous monitoring of the evolution of several coherent sets or of sub-dominant coherent sets, one can consider a second evolutionary framework, the preserving cluster membership (PCM)\textsuperscript{36}, which directly compares the recent clustering to the preceding clustering. This concept translates into a normalized cut problem of the form\textsuperscript{36}

\[
\text{Cost}_{\text{NCut},t} = k - \text{Tr} \left[ X_t^T \hat{W}_t X_t \right],
\]

which is solved by $X_t \in \mathbb{R}^{N,k}$ if the columns of $X_t$ are composed of the eigenvectors associated to the $k$ leading largest eigenvalues of the matrix

\[
\hat{W}_t = \alpha D_t^{-1/2}W_tD_t^{-1/2} + (1 - \alpha)X_{t-1}X_{t-1}^T,
\]

where $X_{t-1}$ is obtained in the previous time step. The parameter $\alpha$ ($0 \leq \alpha \leq 1$) weights the significance of the current clustering in comparison to the preceding one, the latter of which is encoded in $X_{t-1}$. Thus, the matrix $\hat{W}_t$ is composed of $X_{t-1}X_{t-1}^T$ representing the previous clustering, weighted by $(1 - \alpha)$, and of $D_t^{-1/2}W_tD_t^{-1/2}$ representing the current connectivity of the nodes, weighted by $\alpha$. Note that while $W_t$ is sparse, $\hat{W}_t$ no longer is due to $X_{t-1}X_{t-1}^T$ being a full matrix. This problem can be overcome by using a sparse approximation of the eigenspace spanned by $X_{t-1}$ from the SEBA postprocessing\textsuperscript{37} (see Appendix A). Thereby, the computational cost of the analysis can be reduced extremely, making the study of large trajectory networks possible.
FIG. 4. Trajectory analysis in 2d case. (a) Temperature field at $t = 250T_f$ shown together with tracers with $S_{\text{max}}(X_i(t)) \geq 0.5$ for $t = 250T_f$. (b) Horizontal $x$-coordinates versus time of example trajectories that mostly belong to coherent sets. (c) Horizontal $x$-coordinates of trajectories that are mostly outside the coherent sets. (d) Probability density functions (pdfs) of the time-averaged local Nusselt number $\langle \text{Nu}_{\text{local}} \rangle_T$ averaged over $1500T_f$. The analysis is either conducted over all trajectories or two subgroups, either with respect to the cluster membership indicator $S_{\text{max}}$ obtained via SEBA or to the normalized node degree $d$, see legend. All pdfs are smoothed by a Gaussian kernel.

V. TWO-DIMENSIONAL RAYLEIGH-BÉNARD CONVECTION FLOW

We start with the 2d RBC case where 16,384 particle trajectories are analysed after an initial phase of $150T_f$ over a long time period of $1500T_f$, i.e., $T = \{150, 151, \ldots, 1650\}$, where time is given in free-fall times $T_f$. With a threshold of $\epsilon = 0.1$, we set up the weighted adjacency matrices $W_t$ for the observation time $\tau = 20$. Using the evolving network approach from section IV A, we consider the spectral properties of $D_t^{-1/2}W_tD_t^{-1/2}$ for $t = 160$ to $1640$ (in steps of 2). Figure 3(a) shows the time evolution of trajectory clusters as a result of the SEBA analysis of the leading $k$ eigenvectors, where $k \in \{6, 7, 8\}$ is chosen according to a spectral gap that shifts with progressing time (cf. Figure 3(c)). Light regions correspond to coherent features, whereas dark regions indicate the incoherent background possessing a low $S_{\text{max}}(X_i(t))$. These are the regions between the convection rolls where hot fluid rises and
cold fluid descents. Starting with eight clusters, the merger of the two right-most clusters between 450 and 550$T_f$ (more precisely, at about $t = 480T_f$, see Figure 3(a) (multimedia view)), as well as a further merger of the two left-most clusters between 1050 and 1150$T_f$, results in six final clusters that remain stable until the end of the simulation. Figure 3(b) visualizes the normalized node degree $d(X_i(t))$ for the same times as panel (a), offering a similar detection of the mergers of two coherent sets (see also Figure 3(b) (multimedia view)). The node degree gives – as expected due to its relation to FTLE$^{40}$ – a complementary picture with high values in the mixing and lower ones in the coherent regions. Figure 3(c) shows the temporal evolution of the leading eigenvalues. There is a prominent spectral gap between the 8th and 9th eigenvalue until about $t = 450T_f$, where the 8th eigenvalue (red curve) becomes significantly smaller and a new gap appears between the 7th and 8th eigenvalue. The former (green curve) also drops at about $t = 1150T_f$. Thus, the mergers of the coherent sets leave their clear footprints in the spectrum, confirming previous observations using transfer operator methods$^{43,44}$.

Figure 4(a) shows the temperature field at time $t = 250T_f$. Superimposed are Lagrangian particles with a cluster membership certainty $S_{\text{max}}(X_i(t))$ of at least 50% at this time. One observes that the extracted clusters lie in between the regions of ascending hot (light colors) and descending cold (dark colors) fluid and thus form the centers of convection rolls. In Fig. 4(b) we plot the horizontal $x$-coordinate of five selected Lagrangian tracers over time. These tracers exhibit a large average coherent set membership likelihood over the whole simulation time of 1500$T_f$ and tend to stay in the centers of convection rolls for long periods before switching to the center of a different roll. This is in contrast to particles with a cluster membership likelihood below average. These tend to switch frequently between different roles as illustrated in Fig. 4(c).

To study the turbulent heat transport, we compute the time-averaged local Nusselt number $\langle \text{Nu}_{\text{local}} \rangle_T$ for each particle trajectory (see Eq. 10). The additional average over all trajectories (L) results to $\langle \text{Nu}_{\text{local}} \rangle_{T,L} = 26.2 \pm 9.6$ which is very close to the Nusselt number $\text{Nu}_E$ obtained in the Eulerian frame of reference (see Table I). The probability density function (pdf) of the time-averaged local Nusselt numbers is shown in Fig. 4(d) (solid blue curve). To investigate the different contributions of coherent and complementary sets better, we compute both the time-averaged values $\langle S_{\text{max}}(X_i(t)) \rangle_T$ and $\langle d(X_i(t)) \rangle_T$ for every particle trajectory as well as the respective additional averages over all trajectories $\langle S_{\text{max}} \rangle_{T,L} = 0.39$.
FIG. 5. Trajectory analysis in 3d case. (a) Top view onto the layer at $t = 128, 130$ and $132 T_f$ indicating cluster membership certainty by color. Dashed circles mark a local splitting event. (b) 3d view of first figure of panel (a). (c) Leading eigenvalues plotted versus time. Marked in green and orange are the 70th and 71st eigenvalues, respectively. Inset: 65th to 90th eigenvalues are displayed for the first evolutionary step of $\tau/10$.

and $\langle d \rangle_{T,L} = 0.23$. Trajectories in coherent sets with $\langle S_{\text{max}}(X_i(t)) \rangle_T \geq \langle S_{\text{max}} \rangle_{T,L}$ exhibit a time-averaged local Nusselt number $\langle \text{Nu}_{local,c} \rangle_T = 21.7 \pm 8.0$ (solid yellow curve, 8179 tracers), whereas the remaining trajectories contribute significantly more to the turbulent heat transport and mixing with $\langle \text{Nu}_{local,m} \rangle_T = 30.7 \pm 9.0$ (solid red curve, 8205 tracers). The corresponding distinction by means of the normalized node degree $\langle d(X_i(t)) \rangle_T$ results in an even stronger separation of the heat transport contributions with $20.6 \pm 7.0$ (dashed yellow curve, 8699 tracers) and $32.6 \pm 8.0$ (dashed red curve, 7685 tracers), respectively. These results clearly indicate that coherent sets contribute significantly less to the turbulent heat transport with tracers trapped inside the convection rolls (cf. Figure 4(a)) for longer times, underlining the findings of Ref. [27] by the present framework.
VI. THREE-DIMENSIONAL RAYLEIGH-BÉNARD CONVECTION FLOW

The analysis of the 3d data turns out to be more complex, it is based on 65,536 trajectories over a time period $158T_f$, precisely $T = \{75.3, 77.2, \ldots, 212.9\}$. Here, we create a weighted network with parameters similar to the 2d case, $\tau = 19.65T_f$ and $\epsilon \approx 0.14$. This results again in sparse adjacency matrices with edge densities $\rho(\epsilon) < 10^{-4}$. The number of Lagrangian coherent sets is estimated based on physical assumptions to be approximately 80. In the first time step, centered at $t = 75.3T_f$, we detect a large difference between the 70th and 71st eigenvalue (cf. inset Figure 5(c)); therefore we take $k = 70$. In the subsequent analysis, this value will vary within $[0.9k_{t-1}, 1.1k_{t-1}]$. The large number of coherent sets in the 3d data set suggests the usage of the evolving cluster approach from section IV B to stabilize the investigated coherent sets. For our study, we advance in time in steps of $\tau/10$ and take $\alpha = 0.9$. As in the two-dimensional case, we investigate the temporal evolution of the leading eigenvalues, see Fig. 5(c). The variation of the 71st eigenvalue corresponds to splits and mergers of coherent sets which are clearly highlighted in plots of $S_{max}(X_i(t))$ in Fig. 5(a) at three consecutive times. A 3d image of $S_{max}(X_i(t))$ for the first time instance is given in Fig. 5(b). Multiple splits and mergers of coherent sets proceed at other instances as well. If the total number of coherent sets is unchanged these events are not traceable in the spectrum.

Sample trajectories viewed from the top are shown in Fig. 6. We indicate their membership in a coherent set by solid colored lines and their membership in the spatial complement of the coherent sets by dashed lines. One can clearly see how the individual tracers switch between different coherent sets in the course of their dynamical evolution. The correlation of clusters and temperature contour lines is displayed in Fig. 7 which provides a view from the top onto the 3d convection flow. All Lagrangian coherent sets are illustrated together with isocontours of $T = 0.5$ in the mid-plane at $z = 0.5$. These figures illustrate the increased complexity of the 3d case.

Finally, we repeat the analysis of the conditional heat transport along the trajectories. The Nusselt number is now $\langle Nu_{local}\rangle_{T,L} = 4.3 \pm 1.0$. The corresponding pdfs are plotted in Fig. 8(a). Also shown are again the contributions by trajectories assigned to coherent and complementary sets. Based on the cluster membership certainty $S_{max}$, we obtain $\langle Nu_{local,c}\rangle_T = 4.0 \pm 1.0$ and $\langle Nu_{local,m}\rangle_T = 4.5 \pm 1.0$. Based on the normalized node degree
FIG. 6. Top-on view of particle trajectories with either large or small \( \langle S_{\text{max}}(X_i(t)) \rangle_T \). Different colors represent different coherent sets. Gray dotted sections of trajectories represent times spent in the mixing region. The trajectories were selected such that there is no overlap of different trajectories in this two-dimensional image.

FIG. 7. Lagrangian clusters at time \( t \), determined by \( S_{\text{max}}(X_i(t)) \geq 0.7 \), with temperature contour lines \( T = 0.5 \) at mid-plane averaged over \( \mathbb{T}_{t,\tau} = \{ t - \frac{\tau}{2}, \ldots, t + \frac{\tau}{2} \} \). The temperature contours indicate the centers of the convection rolls in the Eulerian frame of reference.

d it follows \( \langle \text{Nu}_{\text{local},x} \rangle_T = 4.0 \pm 1.0 \) and \( \langle \text{Nu}_{\text{local},m} \rangle_T = 4.6 \pm 1.0 \). These numbers support the conclusions drawn from the 2d analysis, see also Ref. 27. They also demonstrate the applicability of evolutionary clustering to turbulent 3d flows. In addition, we study the heat transport for smaller time intervals. Figure 8(b) shows therefore the pdfs of a single interval \( \tau \). The separation of the heat transport contributions are less pronounced in comparison to
FIG. 8. Heat transport statistics in 3d. (a) pdf of $\langle \Nu_{\text{local}} \rangle_T$. Line styles are analogous to those in Fig. 4(d). (b) The same analysis as in panel (a), but now for $\langle \Nu_{\text{local}} \rangle_\tau$. All pdfs are smoothed by a Gaussian kernel.

the full time period $T$. Nevertheless, the distinction of contributions from coherent sets and their complements is still observable with $\langle \Nu_{\text{local},c} \rangle_\tau = 3.3 \pm 3.3$ based on $S_{\text{max}}$ compared to $\langle \Nu_{\text{local}} \rangle_{\tau,L} = 4.2 \pm 3.5$ for all particles. By choosing a very high threshold value on $S_{\text{max}}$ ($S_{\text{max}} \geq 0.9$) one can improve considerably the separation into regions with different contributions to heat transport. However, the coherent sets in this case are very diffuse and monitoring the evolution of these sets is difficult. The effect of the evolving cluster approach in contrast to the evolving network approach for the analysis of heat transport contributions is illustrated in Fig. 9 that shows for one time span $\tau$ the pdf of the instantaneous local Nusselt number. The contributions of the respective distributions for the evolving cluster approach are the red and yellow solid lines. Dashed lines correspond to the equivalent dis-
FIG. 9. Comparison of the heat transport statistics using the evolving cluster approach (solid lines) and the evolving network approach (dashed lines). Red lines correspond to the trajectories with $S_{\text{max}}(X_i(t)) < 0.7$, yellow lines to the trajectories with $S_{\text{max}}(X_i(t)) \geq 0.7$. All pdfs are smoothed by a Gaussian kernel.

tributions using the evolving network approach. One can see that for this time span the separation into sets with different contributions to heat transport works better using the evolving cluster approach. This is the case for almost all time intervals and also for the analysis of the time averaged local Nusselt number $\langle \text{Nu}_{\text{local}} \rangle_{T,L}$ (not shown).

VII. CONCLUDING REMARKS

Studies of turbulent convection in the Eulerian frame of reference identified the formation of slowly evolving patterns termed turbulent superstructures. They are formed by rising and falling thermal plumes which subsequently create large-scale convection rolls. The detection of superstructures and the analysis of their long-term evolution in the Lagrangian frame of reference, however, is limited when conventional spectral clustering is applied. The reason is the immanent dispersion in turbulent flows that destroys coherence of material transport for longer time intervals. In the present work, we applied two evolutionary trajectory-based network frameworks to overcome this constraint and to follow detected coherent sets in 2d and 3d turbulent convection flows in their gradual time evolution. This approach eases the mathematical rigor of material transport on the one hand. It allows, on the other hand,
to monitor a long-term dynamics of coherent sets including their splits or mergers. These events were traced back to the changes of spectral properties of the corresponding network matrices. Thus, we were able to detect long living sets with lifetimes of more than $10^3 \, T_f$ in the 2d case. We showed that Lagrangian coherent sets have a different dynamical relevance compared to their spatial complements, the strong mixing regions. They effectively suppress the turbulent heat transport across the domain. Their contribution to the turbulent heat transfer is reduced by about 30% in the two-dimensional and 13% in the three-dimensional case, respectively.

Particularly in the three-dimensional turbulent flow case, the application of evolutionary spectral clustering as introduced in section IV B stabilizes the coherent sets and leads to their identification as sets with minimal contribution to the heat transfer. This is different to the successive spectral clustering. Furthermore, it allows us to study the dynamical behavior of these macroscopic flow structures, which has not been possible by simple time-extension of the spectral clustering approach due to turbulent dispersion. These persistent evolving sets are thus good candidates to serve as the Lagrangian counterpart of the Eulerian turbulent superstructures in turbulent convection flows. The continued investigation of the robustness of this approach as a function of the Rayleigh number for two- and three-dimensional cases as well as with respect to the number of tracer particles is part of our future work. The latter point would be particularly relevant for investigations of the robustness of the method for sparse data records. The presented trajectory-based dynamical framework might also be well-suited when several transport processes with different material parameters have to be considered, such as temperature and salinity in the ocean which differ in their diffusion constants by a factor of $\sim 10^2$. These measurements are often carried out by drifters and the proposed clustering approach can be adapted to deal with such sparse and possibly incomplete multivariate data sets.

**ACKNOWLEDGMENTS**

CS and PPV are supported by the Priority Programme DFG-SPP 1881 “Turbulent Superstructures” of the Deutsche Forschungsgemeinschaft. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC).
on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). KPG thanks Alexandra von Kameke for fruitful discussions on the weighted network construction.

**DATA AVAILABILITY STATEMENT**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Appendix A: Sparse Eigenbasis Approximation and evolutionary clustering**

The main idea of the sparse eigenbasis approximation (SEBA) is to transform the set of orthonormal eigenvectors $u_1, \ldots, u_k$, which span a subspace $U \subset \mathbb{R}^N$ to a basis of sparse vectors $s_1, \ldots, s_k$ for a subspace $S \subset \mathbb{R}^N$ which approximates $U$, i.e. $U \approx S$. To be more precise, given the matrix $U \in \mathbb{R}^{k \times N}$, the goal is to find a rotation matrix $R \in \mathbb{R}^{k \times k}$ and a sparse matrix $S \in \mathbb{R}^{N \times k}$ such that $U \approx SR$, see Ref. [37] for the exact setting. It immediately follows that

$$UU^T \approx SR(SR)^T = SRRTS^T = SS^T.$$  \hspace{1cm} (A1)

In the trajectory clustering framework, different flow components, i.e. regions in the flow with different dynamical behavior, originally encoded in the eigenvectors of the network matrix $D^{-1/2}WD^{-1/2}$, are approximated by individual sparse vectors. The algorithm in Ref. [37] returns the sparse vector $s_i$ as column $i$ of a matrix $S$ where the columns are ordered by decreasing reliability of the identified flow component. Due to a rescaling used in Ref. [37], the entry $S_{ij}$ can be interpreted as the likelihood of the data point connected to index $i$ to be an element of component $j$. Therefore, $S_{\text{max},i} = \max_j S_{ij}$ represents the maximum likelihood of data point $i$ to belong to any flow component. By choosing a threshold on $S_{\text{max}}$, data points can be assigned to coherent regions or to the complement, the incoherent mixing region.

In the framework of preserving cluster memberships, the optimal solution to the normalized cut problem for time $t$ is the matrix $X_t$ whose columns are the eigenvectors of matrix

$$\hat{W}_t = \alpha D_t^{-1/2}W_tD_t^{-1/2} + (1 - \alpha)X_{t-1}X_{t-1}^T.$$  

20
where $X_{t-1}$ is obtained in the previous time step. In order to speed up the computation, we calculate a sparse approximation of $\hat{W}_t$. First, we approximate the subspace spanned by $X_{t-1}$ in terms of the corresponding sparse subspace spanned by the columns of the matrix $\tilde{S}_{t-1}$. $\tilde{S}_{t-1}$ is obtained via the SEBA output matrix $S_{t-1}$ from $X_{t-1}$, setting possibly occurring negative entries in $S_{t-1}$ to zero, and a subsequent column normalization of $S_{t-1}$. Using (A1) this results in

$$X_{t-1}X_{t-1}^T \approx \tilde{S}_{t-1}\tilde{S}_{t-1}^T.$$  

(A2)

Second, we insert this approximation into the derivation of $\hat{W}_t$

$$\hat{W}_t \approx \alpha D_{t}^{-1/2}W_t D_{t}^{-1/2} + (1 - \alpha)S_{t-1}S_{t-1}^T =: \bar{W}_t.$$  

(A3)

An approximation of the optimal solution for time $t$ is now given by the matrix whose columns are the eigenvectors of the normalized matrix $\tilde{D}_t^{-1/2}\bar{W}_t\tilde{D}_t^{-1/2}$, where $\tilde{D}_t$ is the degree matrix of $\bar{W}_t$. By this normalization $\tilde{D}_t^{-1/2}\bar{W}_t\tilde{D}_t^{-1/2}$ is a stochastic matrix.

Appendix B: Number of relevant eigenvectors and temporal evolution of eigenvalues

Spectral clustering makes use of the eigenvectors corresponding to the $k$ leading eigenvalues of the network matrix, where $k$ is heuristically chosen according to a spectral gap criterion. To this end, we compute the $K$ largest eigenvalues $\lambda_{t,1} = 1 > \lambda_{t,2} \geq \ldots \geq \lambda_{t,K}$ for the network with respect to the time window centered at $t \in \mathbb{T}$. The number $K$ is significantly larger than the expected $k_t$, where $k_t$ denotes the number of relevant eigenvectors for this time step. In this work, we use $K = 10$ for the 2d case and $K = 100$ for the 3d case. For each time step $t$, we calculate the distances between successive eigenvalues via

$$g_t(i) = \lambda_{t,i} - \lambda_{t,i+1}, \ i = 1, \ldots, K - 1.$$  

(B1)

and determine the arguments of the (local) maxima of $g_t$. For the 2d case, we initially choose $k_t = 8$ as this maximizes $g_t$ for the initial time window centered at $t = 160T_f$. Under the time evolving network $k_t$ is chosen at each time step to be the maximizer of $g_t$ when $k_t \in \{k_{t-1} - 1, k_{t-1}, k_{t-1} + 1\}$. In the 3d case, we start with $k_t = 70$ as a local maximizer of $g_t$ for the initial time window centered at $t = 75.3T_f$, taking into account prior knowledge of the approximate number of clusters from Ref. [27]. As the evolving cluster approach stabilizes the
number of clusters, we choose $k_t \in [0.9k_{t-1}, 1.1k_{t-1}]$ again as a local maximizer of $g_t$, allowing us to monitor significant changes in the time-dependent clusterings. Further heuristics on the spectral-based choice of the number of clusters are discussed in Ref. 37.

We note that for the sake of a simple and clear visualization of the temporal behavior of the spectrum we have just connected the corresponding (ordered) eigenvalues $\lambda_{t,i}$ and $\lambda_{t+1,i}$ from successive time windows in Figures 3(c) and 5(c). This is correct as long as these curves are well separated. In order to accurately follow how the eigenvalues evolve in time in the case of crossings (i.e. changes in the ordering of the eigenvalues), one would have to take into account the evolution of the corresponding eigenvectors as proposed in Ref. 43. However, as we do not consider the temporal behavior of individual eigenvectors in our study, crossings of eigenvalues are difficult to interpret, and moreover, such an investigation is unfeasible for the 3d case where a very large number of eigenvalues and corresponding eigenvectors is studied.

REFERENCES

1. G. Ahlers, S. Grossmann, and D. Lohse, “Heat transfer and large scale dynamics in turbulent Rayleigh-Bénard convection,” Rev. Mod. Phys. 81, 503–537 (2009).
2. F. Chillà and J. Schumacher, “New perspectives in turbulent Rayleigh-Bénard convection,” Eur. Phys. J. E 35, 58 (2012).
3. B. Stevens, “Atmospheric moist convection,” Annu. Rev. Earth Planet. Sci. 33, 605–643 (2005).
4. J. Schumacher and K. R. Sreenivasan, “Colloquium: Unusual dynamics of convection in the sun,” Rev. Mod. Phys. 92, 041001 (2020).
5. D. H. Kelley and T. Weier, “Fluid mechanics of liquid metal batteries,” Appl. Mech. Rev. 70, 020801 (2018).
6. M. R. Allshouse and T. Peacock, “Lagrangian based methods for coherent structure detection,” Chaos 25, 097617 (2015).
7. G. Haller, “Lagrangian coherent structures,” Annu. Rev. Fluid Mech. 47, 137–162 (2015).
8. 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).
9E. Tew Kai, V. Rossi, J. Sudre, H. Weimerskirch, C. Lopez, E. Hernandez-Garcia, F. Marsac, and V. Garçon, “Top marine predators track Lagrangian coherent structures,” Proc. Natl. Acad. Sci. USA 106, 8245–8250 (2009).

10S. Shadden and C. Taylor, “Characterization of coherent structures in the cardiovascular system,” Ann. Biomed. Eng. 36, 1152–1162 (2008).

11E. S. Gawlik, J. E. Marsden, P. C. D. Toit, and S. Campagnola, “Lagrangian coherent structures in the planar elliptic restricted three-body problem,” Celest. Mech. Dyn. Astron. 103, 227–249 (2009).

12M. J. Olascoaga and G. Haller, “Forecasting sudden changes in environmental pollution patterns,” Proc. Natl. Acad. Sci. USA 109, 4738–4743 (2012).

13E. Zambianchi, M. Trani, and P. Falco, “Lagrangian transport of marine litter in the Mediterranean sea,” Front. Environ. Sci. 5, 5 (2017).

14G. Froyland, S. Lloyd, and N. Santitissadeekorn, “Coherent sets for nonautonomous dynamical systems,” Physica D 239, 1527–1541 (2010).

15G. Froyland, “An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems,” Physica D 250, 1–19 (2013).

16D. Karrasch and J. Keller, “A geometric heat-flow theory of Lagrangian coherent structures,” J. Nonlinear Sci. 30, 1849–1888 (2020).

17G. Haller, D. Karrasch, and F. Kogelbauer, “Material barriers to diffusive and stochastic transport,” Proc. Natl. Acad. Sci. USA 115, 9074–9079 (2018).

18S. Balasuriya, N. T. Ouellette, and I. I. Rypina, “Generalized lagrangian coherent structures,” Physica D: Nonlinear Phenomena 372, 31–51 (2018).

19G. Froyland, “Dynamic isoperimetry and the geometry of Lagrangian coherent structures,” Nonlinearity 28, 3587 (2015).

20G. Froyland and P. Koltai, “Detecting the birth and death of finite-time coherent sets,” Preprint (2021), arXiv:2103.16286 [math.DS].

21G. Froyland and K. Padberg-Gehle, “A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data,” Chaos 25, 087406 (2015).

22A. Hadjighasem, D. Karrasch, H. Teramoto, and G. Haller, “Spectral-clustering approach to Lagrangian vortex detection,” Phys. Rev. E 93, 063107 (2016).

23R. Banisch and P. Koltai, “Understanding the geometry of transport: Diffusion maps for
Lagrangian trajectory data unravel coherent sets,” Chaos 27, 035804 (2017).

24K. L. Schlueter-Kuck and J. O. Dabiri, “Coherent structure colouring: identification of coherent structures from sparse data using graph theory,” J. Fluid Mech. 811, 468–486 (2017).

25K. Padberg-Gehle and C. Schneide, “Network-based study of Lagrangian transport and mixing,” Nonlin. Processes Geophys. 24, 661–671 (2017).

26C. Schneide, A. Pandey, K. Padberg-Gehle, and J. Schumacher, “Probing turbulent superstructures in Rayleigh-Bénard convection by Lagrangian trajectory clusters,” Phys. Rev. Fluids 3, 113501 (2018).

27P. P. Vieweg, C. Schneide, K. Padberg-Gehle, and J. Schumacher, “Lagrangian heat transport in turbulent three-dimensional convection,” Phys. Rev. Fluids 6, L041501 (2021).

28S. L. Brunton, B. R. Noack, and P. Koumoutsakos, “Machine learning for fluid mechanics,” Annu. Rev. Fluid Mech. 52, 477–508 (2020).

29R. J. A. M. Stevens, A. Blass, X. Zhu, R. Verzicco, and D. Lohse, “Turbulent thermal superstructures in Rayleigh-Bénard convection,” Phys. Rev. Fluids 3, 041501 (2018).

30A. Pandey, J. D. Scheel, and J. Schumacher, “Turbulent superstructures in Rayleigh-Bénard convection,” Nat. Commun. 9, 2118 (2018).

31G. Green, D. G. Vlaykov, J. P. Mellado, and M. Wilczek, “Resolved energy budget of superstructures in Rayleigh-Bénard convection,” J. Fluid Mech. 887, A21 (2020).

32D. Krug, D. Lohse, and R. J. A. M. Stevens, “Coherence of temperature and velocity superstructures in turbulent Rayleigh-Bénard flow,” J. Fluid Mech. 887, A2 (2020).

33P. P. Vieweg, J. D. Scheel, and J. Schumacher, “Supergranule aggregation for constant heat flux-driven turbulent convection,” Phys. Rev. Res. 3, 013231 (2021).

34E. Fonda, A. Pandey, J. Schumacher, and K. R. Sreenivasan, “Deep learning in turbulent convection networks,” Proc. Natl. Acad. Sci. USA 116, 8667–8672 (2019).

35T. 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).

36Y. Chi, X. Song, D. Zhou, K. Hino, and B. L. Tseng, “Evolutionary spectral clustering by incorporating temporal smoothness,” in Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY, USA, 2007) pp. 153–162.
G. Froyland, C. P. Rock, and K. Sakellariou, “Sparse eigenbasis approximation: Multiple feature extraction across spatiotemporal scales with application to coherent set identification,” Commun. Nonlinear Sci. Numer. Simul. 77, 81–107 (2019).

J. D. Scheel, M. S. Emran, and J. Schumacher, “Resolving the fine-scale structure in turbulent Rayleigh-Bénard convection,” New J. Phys. 15, 113063 (2013).

G. Iacobello, L. Ridolfi, and S. Scarsoglio, “A review on turbulent and vortical flow analyses via complex networks,” Physica A 563, 125476 (2021).

R. Banisch, P. Kolta, and K. Padberg-Gehle, “Network measures of mixing,” Chaos 29, 063125 (2019).

J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell. 22, 888–905 (2000).

F. R. Bach and M. I. Jordan, “Learning spectral clustering, with application to speech separation,” J. Mach. Learn. Res. 7, 1963–2001 (2006).

C. Blachut and C. González-Tokman, “A tale of two vortices: How numerical ergodic theory and transfer operators reveal fundamental changes to coherent structures in non-autonomous dynamical systems,” J. Comput. Dyn. 7, 369–399 (2020).

M. Ndour, K. Padberg-Gehle, and M. Rasmussen, “Spectral early-warning signals for sudden changes in time-dependent flow patterns,” Fluids 6, 49 (2021).

R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, “Ambiguities in recurrence-based complex network representations of time series,” Phys. Rev. E 81, 015101 (2010).