https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Ordering of trajectories reveals hierarchical finite-time coherent sets in Lagrangian particle data: detecting Agulhas rings in the South Atlantic Ocean

David Wichmann[1,2], Christian Kehl[1], Henk A. Dijkstra[1,2], and Erik van Sebille[1,2]

[1]Institute for Marine and Atmospheric Research Utrecht, Utrecht University
[2]Centre for Complex Systems Studies, Utrecht University

Correspondence: David Wichmann (d.wichmann@uu.nl)

Abstract. The detection of finite-time coherent particle sets in Lagrangian trajectory data using data clustering techniques is an active research field at the moment. Yet, the clustering methods mostly employed so far have been based on graph partitioning, which assigns each trajectory to a cluster, i.e. there is no concept of noisy, incoherent trajectories. This is problematic for applications to the ocean, where many small coherent eddies are present in a large fluid domain. In addition, to our knowledge none of the existing methods to detect finite-time coherent sets has an intrinsic notion of coherence hierarchy, i.e. the detection of finite-time coherent sets at different spatial scales. Such coherence hierarchies are present in the ocean, where basin scale coherence coexists with smaller coherent structures such as jets and mesoscale eddies. Here, for the first time in this context, we use the density-based clustering algorithm OPTICS (Ankerst et al., 1999) to detect finite-time coherent particle sets in Lagrangian trajectory data. Different from partition based clustering methods, OPTICS does not require to fix the number of clusters beforehand. Derived clustering results contain a concept of noise, such that not every trajectory needs to be part of a cluster. OPTICS also has a major advantage compared to the previously used DBSCAN method, as it can detect clusters of varying density. Further, clusters can also be detected based on density changes instead of absolute density. Finally, OPTICS based clusters have an intrinsically hierarchical structure, which allows to detect coherent trajectory sets at different spatial scales at once. We apply OPTICS directly to Lagrangian trajectory data in the Bickley jet model flow and successfully detect the expected vortices and the jet. The resulting clustering separates the vortices and the jet from background noise, with an imprint of the hierarchical clustering structure of coherent, small scale vortices in a coherent, large-scale, background flow. We then apply our method to a set of virtual trajectories released in the eastern South Atlantic Ocean in an eddying ocean model and successfully detect Agulhas rings. At larger scale, our method also separates the eastward and westward moving parts of the subtropical gyre. We illustrate the difference between our approach and partition based k-Means clustering using a 2-dimensional embedding of the trajectories derived from classical multidimensional scaling. We also show how OPTICS can be applied to the spectral embedding of a trajectory based network to overcome the problems of k-Means spectral clustering in detecting Agulhas rings.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

1 Introduction

Understanding the transport of tracers in the ocean is an important topic in oceanography. Despite large-scale transport features of the mean flow, on smaller scales, mesoscale eddies and jets play an important role for tracer transport (Van Sebille et al., 2020). Such eddies can capture large amounts of a tracer, and, while transported in a background flow, redistribute them in the ocean. Eddies have been shown to play an important role for the accumulation of plastic (Brach et al., 2018) and the transport of heat and salt (Dong et al., 2014). To quantify the effects of eddies for tracer transport in the ocean, it is necessary to develop methods that are able to detect and track them. Many methods exist to detect such *finite-time coherent sets* of fluid parcels based on different mathematical or heuristic principles (Hadjighasem et al., 2017). The term 'finite-time coherent set' is based on the work of Froyland et al. (2010), and is in our context defined as a set of particles that stay, in a sense to be made more specific, close to each other along their entire trajectories. In this article, we propose a new way to identify finite-time coherent sets in Lagrangian trajectory data. For the first time in this context, we make use of the density-based clustering algorithm OPTICS (Ankerst et al., 1999) which allows to detect coherent trajectories at different spatial scales at once, introducing a powerful computational tool to the geophysical fluid dynamics community.

The detection of coherent Lagrangian vortices using abstract embeddings of Lagrangian trajectories has received significant attention in the recent literature (Froyland and Padberg-Gehle, 2015; Hadjighasem et al., 2016; Padberg-Gehle and Schneide, 2017; Banisch and Koltai, 2017; Schneide et al., 2018). Examples include the direct embedding of trajectories in a high dimensional Euclidean space (Froyland and Padberg-Gehle, 2015), or more abstract embeddings based on related networks constructed from particle trajectories (Hadjighasem et al., 2016; Padberg-Gehle and Schneide, 2017; Banisch and Koltai, 2017). Using embedded trajectories for the detection of finite-time coherent sets is interesting as it allows to use scarce trajectory data, and it can in principle be applied to ocean drifter trajectories, as done by Froyland and Padberg-Gehle (2015) and Banisch and Koltai (2017). Yet, the methods proposed so far suffer from a major drawback: they cluster networks based on network partitioning, which does not incorporate the difference between coherent, clustered trajectories and noisy trajectories that should not belong to any cluster. Similar observations were made for the spectral clustering approaches of particle-based networks and transfer and dynamic Laplace operators by Froyland et al. (2019). Although some attempts have been made to accommodate such concepts in hard partitioning, e.g. by incorporating one additional cluster corresponding to noise (Hadjighasem et al., 2016), this approach is likely to fail for large ocean domains, as discussed by Froyland et al. (2019) and shown in section 4 of this paper. Froyland et al. (2019) have developed an algorithm based on sparse eigenbasis decomposition given the eigenvectors of transfer operators and dynamic Laplacians. By superposing different sparse eigenvectors, they successfully separate coherent vortices from unclustered background noise.

Here, we show how the density-based clustering method OPTICS (Ordering Points To Identify the Clustering Structure) developed by Ankerst et al. (1999) can be used to overcome the inherent problems of partition-based clustering. Density-based clustering aims to detect groups of data points that are close to each other, i.e. regions with high data *density*. Our data points correspond to entire trajectories, and groups of trajectories staying close to each other over a certain time interval are detected as such regions of high point density. Different from partition based methods such as k-Means or fuzzy-c-means, OPTICS does

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

not require to define the number of clusters beforehand. Further, density-based clustering has an intrinsic notion of a noisy data point: a point does not belong to any cluster (i.e. a finite-time coherent set) if it is not part of a dense region. The density-based clustering algorithm DBSCAN (Ester et al., 1996) has been applied to pseudo-trajectories in fluids to detect coherent sets (Schneide et al., 2018). Yet, DBSCAN is only able to detect clusters with a certain fixed minimum density, although clusters with varying densities might be present in a data set (Ankerst et al., 1999). Choosing a good value for the density parameter in DBSCAN is challenging if there is no underlying physical intuition for the density structure. In addition, OPTICS not only allows to detect clusters based on their absolute density, but also based on density changes. The main result of OPTICS, the *reachability plot*, can be used to derive any DBSCAN result (with similar parameter s_{min}, cf. section 3.3) without re-running the algorithm, as illustrated in section 4. Finally, clustering results from OPTICS are typically hierarchical, and the reachability plot provides this hierarchical information in a simple 1-dimensional graph. Indeed, finite-time coherent trajectories naturally come with a notion of hierarchy. For example, the surface flow in the North Atlantic Ocean can be seen as approximately coherent (Froyland et al., 2014), while mesoscale eddies and jets are also finite-time coherent sets of trajectories at smaller scales *within* the North Atlantic Ocean. This is also reflected in previous studies that apply methods to detect finite-time coherent sets to individual vortices and also to global drifter data, identifying the five major ocean basins (Froyland and Padberg-Gehle, 2015; Banisch and Koltai, 2017). The hierachical property of finite-time coherent sets has been studied in the transfer operator framework of Froyland et al. (2010) by Ma and Bollt (2013). Different from this approach, however, the clustering result derived from OPTICS is intrinsically hierarchical. This means that it shows in a smooth manner how the coherent structures change when zooming in or out, and it does not require to fix a certain partition to detect sub-partitions, e.g. as is typical for hierarchical applications of spectral clustering in the spirit of Shi and Malik (2000).

In section 4, we first show how OPTICS detects finite-time coherent sets at different scales for the Bickley jet model flow (also discussed e.g. by Hadjighasem et al. (2017)), successfully detecting the six coherent vortices and the jet as the steepest valleys in the reachability plot. The general structure of the reachability plot also reveals the large-scale finite-time coherent sets, i.e. the northern and southern parts of the model flow, separated by the jet. We then apply our method to Lagrangian particle trajectories released in the eastern South Atlantic Ocean, where large rings detach from the Agulhas Current (e.g. Schouten et al. (2000)). We detect several Agulhas rings, and on the larger scale also separate the eastward and westward moving branches of the South Atlantic Subtropical Gyre. While the traditional approach to study Agulhas rings is based on sea surface height analysis (see e.g. Dencausse et al. (2010)), several methods based on virtual Lagrangian trajectories have been applied to Agulhas ring detection before (Haller and Beron-Vera, 2013; Beron-Vera et al., 2013; Froyland et al., 2015; Hadjighasem et al., 2016; Tarshish et al., 2018). Our method is different from these approaches in that it is directly applicable to a trajectory data set. As the OPTICS algorithm is readily available in the sklearn package of SciPy, the detection of finite-time coherent sets can be done without much effort and with only a few lines of code. A further difference is the mentioned intrinsic notion of coherence hierarchy, which allows for simultaneous analysis of trajectory data at different scales. Finally, trajectory based approaches can in principle be applied to scarce trajectory data, i.e. to any Lagrangian particle simulation result without much care for the spatial coverage of the initial conditions. While we mainly focus on the direct embedding of trajectories in an abstract high-dimensional Euclidean space, we also show in section D in the appendix that OPTICS can be used to overcome the limits

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

of k-Means clustering in the context of spectral clustering of physically motivated trajectory based networks, such as the works presented by Hadjighasem et al. (2016), Padberg-Gehle and Schneide (2017) or Banisch and Koltai (2017).

2 Trajectory datasets

2.1 Bickley jet

We apply our method to a model system that has been used frequently in studies to detect finite-time coherent sets (Hadjighasem et al., 2017; Padberg-Gehle and Schneide, 2017; Hadjighasem et al., 2016; Banisch and Koltai, 2017). The velocity field of the Bickley jet is defined by a stream function $\psi(x,y,t)$, i.e. $\dot{x} = -\frac{\partial \psi}{\partial y}$ and $\dot{y} = \frac{\partial \psi}{\partial x}$, with $\psi(x,y,t) = \psi_0(y) + \psi_1(x,y,t)$ consisting of a stationary eastward background flow

$$\psi_0(y) = -UL\tanh(y/L), \tag{1}$$

and a time-dependent perturbation

$$\psi_1(x,y,t) = UL\,\mathrm{sech}^2(y/L)\,\mathrm{Re}\left[\sum_{n=1}^{3} f_n(t)\exp(ik_n x)\right], \tag{2}$$

where $\mathrm{Re}(z)$ denotes the real part of the complex number z. We use the same parameter values as Hadjighasem et al. (2017), with $U = 62.66$ m/s the characteristic velocity of the zonal background flow, and $L = 1770$ km. The parameters in eq. (2) are given by $k_n = 2n/r_0$, $f_n(t) = \epsilon_n \exp(-ik_n c_n t)$ with $\epsilon_1 = 0.075$, $\epsilon_2 = 0.4$, $\epsilon_3 = 0.3$, $c_3 = 0.461U$, $c_2 = 0.205U$, $c_1 = 0.1446U$. The domain of interest is $\Omega = [0, \pi r_0] \times [-3000\,\mathrm{km}, 3000\,\mathrm{km}]$, where $r_0 = 6371$ km is the radius of the Earth, and left and right edges of Ω are identified, i.e. the flow is periodic in x-direction with period πr_0. Similar to Banisch and Koltai (2017), we seed the domain with an initial number of 12,000 particles on a uniform 200×60 grid. For this choice, the initial particle spacing is slightly above 100 km in both directions. We compute the trajectories for 40 days with a time step of one second using the SciPy integrate package. We output the trajectories every day, i.e. we have $T = 41$ data points in time for each trajectory.

2.2 Agulhas rings in the South Atlantic

To test our method with a more realistic ocean flow, we simulate surface particle trajectories in a strongly eddying ocean model. Surface velocities are derived from a NEMO ORCA-N006 run (Madec, 2008), which has a horizontal resolution of $1/12°$ and velocity output for every five days. The model is forced by reanalysis and observed data of wind, heat and fresh water fluxes (Dussin et al., 2016), i.e. the currents do not only contain the geostrophic component, as is the case in altimetry-derived currents (Beron-Vera et al., 2013; Froyland et al., 2019). For the advection of virtual particles, we use version 1.11 of the open source Parcels framework (Lange and van Sebille, 2017), see oceanparcels.org. The 2-dimensional surface current

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics Discussions
Open Access
EGU

velocity is interpolated in space and time with the C-grid interpolation scheme of Delandmeter and van Sebille (2019), using a 4th order Runge-Kutta method with a time step of 10 minutes. We initially distribute particles uniformly in the ocean on the vertices of a $0.2° \times 0.2°$ grid in the domain $[30°W, 20°E] \times [40°S, 20°S]$, which corresponds to a total number of 23,821 particles. At 30°S, a spacing of $0.2°$ corresponds to roughly 20 km. The particles start at January 5, 2000 and are advected for two years. We output the trajectories with a time interval of five days. We only use the first 100 days as data to detect the finite-time coherent sets, i.e. we have $T = 21$ data points for each trajectory, but also look at later times to see how long the rings need to disperse. The data is a subset of the trajectory output of our previous paper (Wichmann et al., 2019), and we refer to that paper and the code references in there for the details of the particle simulation. We provide the used trajectory data for the Agulhas flow as numpy file on zenodo (Wichmann, 2020).

3 Methods

3.1 Detecting coherent structures in trajectory data: an overview

For N trajectories of dimension D and length T, the trajectory information can be stored in a *data matrix* $X \in \mathbb{R}^{N \times DT}$, where each row results from a particle trajectory by concatenating the different spatial dimensions. The analysis of trajectory data to detect finite-time coherent sets of trajectories (Froyland and Padberg-Gehle, 2015; Banisch and Koltai, 2017; Hadjighasem et al., 2016; Padberg-Gehle and Schneide, 2017; Schneide et al., 2018; Wichmann et al., 2020) can be split into two essential steps:

Step 1 **Embedding** of the trajectories in an abstract (metric) space, i.e. $X \to \bar{X} \in \mathbb{R}^{N \times M}$ where $M \leq DT$. If one uses a dimensionality reduction method, $M < DT$.

Step 2 **Clustering** of the embedded data with a clustering algorithm.

Figure 1 shows a few possible options for these two steps that have partially been explored before (see the footnotes in the figure for the combinations used in related studies). For a given trajectory dataset, one can in principle apply an arbitrary combination of embedding and clustering method. Only a few of the different combinations have been explored so far, and many more options for embedding and clustering as those shown in fig. 1 exist. It is important to note that a good choice of embedding and clustering might well depend on the specific problem at hand, and there might be no combination that performs well for all possible situations.

Here, we explore the OPTICS clustering algorithm for the first time in the context of finite-time coherent sets. We test it for three different kinds of embeddings:

E1 A direct embedding of the trajectory data in a high dimensional Euclidean space, i.e. $M = DT$ (cf. section 3.2.1).

E2 A reduction of the trajectory data to a 2-dimensional embedding space using classical multidimensional scaling (MDS, cf. section 3.2.2). This is mainly to visualize the difference to partition based k-Means clustering.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

E3 An embedding of the network proposed by Padberg-Gehle and Schneide (2017), which is in section D in the appendix for the sake of brevity.

In the following sections, we explain in detail the embeddings E1 and E2 and the OPTICS algorithm. We introduce the network embedding E3 together with the corresponding results in section D in the appendix.

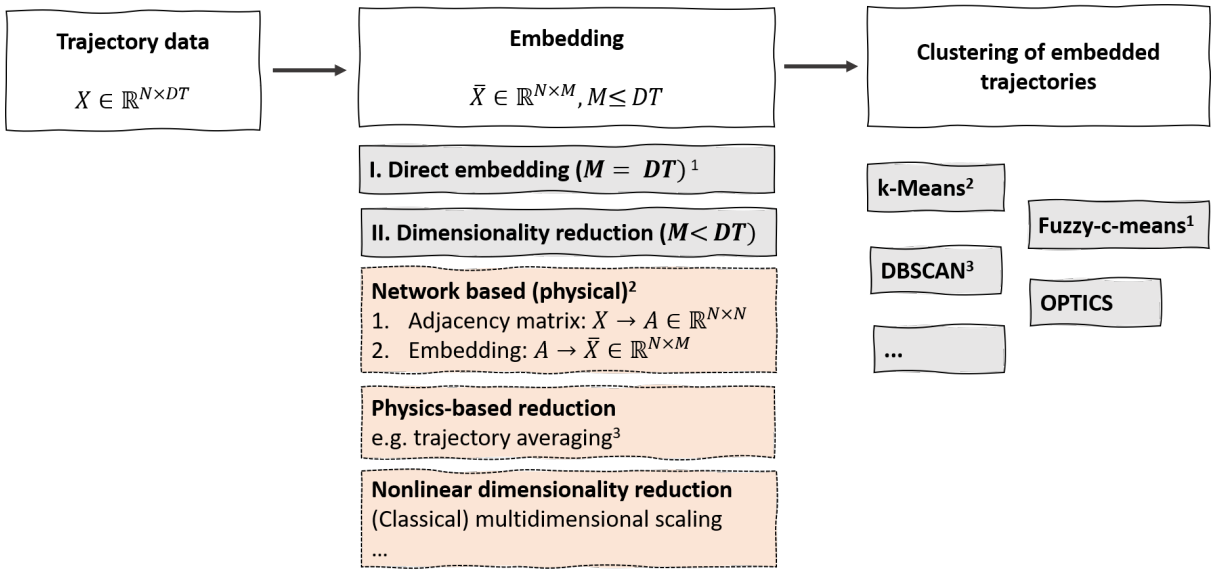


Figure 1. Different steps to detect coherent trajectories in Lagrangian data with trajectory clustering. The figure is non-exhaustive, and many more options for embedding and clustering exist. Footnotes: [1] Froyland and Padberg-Gehle (2015). [2] Hadjighasem et al. (2016), Padberg-Gehle and Schneide (2017) and Banisch and Koltai (2017) all define networks with spectral embedding and subsequent k-Means clustering. [3] Schneide et al. (2018).

3.2 Trajectory embedding

3.2.1 Direct embedding

The direct embedding of each trajectory in $\mathbb{R}^{DT}$ is the most straight forward embedding as it requires no further pre-processing of the trajectory data. For simplicity, assume we are given a set of N trajectories in a 3-dimensional space, i.e. $(x_i(t), y_i(t), z_i(t))$ where $i = 1, \ldots, N$ and $t = t_1, \ldots, t_T$. We then simply define the embedding of trajectory i in the abstract $3T$-dimensional space as

$$u_i = (x_i(t_0), x_i(t_1), \ldots, x_i(t_T), y_i(t_0), y_i(t_1), \ldots, y_i(t_T), z_i(t_0), z_i(t_1), \ldots, z_i(t_T)) \in \mathbb{R}^{3T} \tag{3}$$

and impose an Euclidean metric in $\mathbb{R}^{3T}$ to measure distances between different embedded trajectories. The resulting embedded data matrix $\bar{X}$ is then simply given by the vertical concatenation of the different embedding vectors. This kind of embedding

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics Discussions
Open Access
EGU

was also explored by Froyland and Padberg-Gehle (2015), together with a fuzzy-c-means clustering.

To take into account the πr_0-periodicity in x-direction of the Bickley jet flow, we first put the individual 2-dimensional data points on the surface of a cylinder with radius $r_0/2$ in $\mathbb{R}^3$, and interpret the resulting ($D = 3$) trajectories in a 3-dimensional Euclidean space. The resulting data matrix is $\bar{X} \in \mathbb{R}^{N\times 3T}$, with $N = 12,000$ and $T = 41$. For the Agulhas particles, we put the single data points on the earth surface in a 3-dimensional Euclidean embedding space by the standard coordinate transformation of spherical to Euclidean coordinates. The resulting data matrix is thus $\bar{X} \in \mathbb{R}^{N\times 3T}$ with $N = 23,821$ and $T = 21$.

3.2.2 Classical multidimensional scaling

To get an intuition for the clustering results of the OPTICS algorithm, we visualize the density structure of the trajectories in the 2-dimensional plane by a common method of nonlinear dimensionality reduction, called classical Multidimensional scaling (MDS), see e.g. chapter 10.3 of Fouss et al. (2016). Classical MDS tries to find an embedding of the high-dimensional data points in a low dimensional space such that the pairwise distances are approximately preserved. Classical MDS makes use of the eigenvectors corresponding to the largest eigenvalues of the kernel matrix

$$B = -\frac{1}{2} H \Delta^2 H, \tag{4}$$

where $\Delta^2 \in \mathbb{R}^{N\times N}$ is a matrix containing all squared distances between the points, and H is the centring matrix with $H_{ij} = \delta_{ij} - 1/N$, where δ_{ij} denotes the Kronecker delta. We compute Δ^2 with the Euclidean embeddings described in section 3.2.1 and restrict ourselves to the first two dimensions, i.e. the embedding is defined by

$$u_i = (w_{0,i}, w_{1,i}),\ i = 1, \ldots, N, \tag{5}$$

where $Kw_j = \lambda_j w_j$, and $\lambda_0 \geq \lambda_1 \geq \lambda_k$ for all $k = 2, \ldots N-1$. This choice of embedding ensures to capture the main variance of the data points, and we therefore also expect to capture the main structure in terms of data density. For large particle sets however, computing the spectrum of H in eq. (4) is computationally not feasible, as the matrix B is in general dense and computing the spectrum scales with $O(N^3)$. We apply classical MDS to the 12,000 particles of the Bickley jet model flow, and a random selection of the equal number of particles for the Agulhas flow. In our context, the method is most useful for visualization purposes, as it provides a good 2-dimensional approximation of the point distances, i.e. also the density structure of the embedded trajectories.

3.3 Clustering with OPTICS

The detection of dense accumulations of points that are separated from each other by non-dense regions (noise) is the main goal of density-based clustering. We use the OPTICS (Ordering Points To Identify the Clustering Structure) algorithm by Ankerst et al. (1999) to detect these regions. The OPTICS algorithm can be seen as an extension of DBSCAN (Ester et al., 1996), with important advantages for the detection of finite-time coherent sets, as discussed in the introduction and will become clear in

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

section 4.

As we have no prior information on the density structure of the embedded nodes, we set the 'generating distance' of OPTICS to infinity and our presentation here is limited to this case. The general OPTICS algorithm with finite generating distance is slightly more complicated, and we refer to Ankerst et al. (1999) for more details. For $\epsilon \in \mathbb{R}$, the ϵ-neighbourhood of a point $p \in \mathbb{R}^M$ is defined as the M-dimensional ball around p. Define $M_\epsilon(p)$ as the number of points that is in the ϵ-neighbourhood of p, including p itself. OPTICS requires one parameter, an integer s_{min} (called MinPts by Ester et al. (1996)), that defines the *core-distance* of a point p as

$$c(p) = \{\min(\epsilon) \mid M_\epsilon(p) \geq s_{min}\}. \tag{6}$$

The ordering of the points is based on the *reachability distance* of a point p w.r.t. another point q, defined as

$$r(p|q) = \max(c(q), ||p-q||), \tag{7}$$

where $||p-q||$ in our case denotes the Euclidean distance between p and q. The ordering of points is then constructed with the following scheme

Step 1 Pick a point p_1. This is the first point in the order, and is arbitrary.

Step 2 Compute the core-distance $c(p_1)$ of p_1.

Step 3 Define an ordered seed list containing all other points, p_l, $k = 2, \ldots, N$. For each point p_l, define the reachability value $r(p_l)$ as the reachability distance (eq. (7)) w.r.t. p_1, $r(p_l) = r(p_l|p_1)$. Order the list in ascending order of the $r(p_l)$.

Step 4 Pick the first point on the ordered seed list as p_2 and compute the core-distance $c(p_2)$. For all remaining points p_l, $l = 3, \ldots, N$, update the reachability value $r(p_l) \to \min(r(p_l), r(p_l|p_2))$.

Step 5 Update the ordered seed list according to the new reachability.

Step 5 Repeat steps 4-5 to obtain p_3. Continue until all points are processed.

The main result of the OPTICS algorithm is a *reachability plot*. This plot is the graph defined by $(i, r(p_i))$. The reachability plot is a powerful presentation of the global and local distribution of a set of points at once. The valleys in this plot correspond to dense regions, which we will relate to finite-time coherent sets. We will show examples of reachability plots in section 4. Given the reachability plot $(i, r(p_i))$, we use two common ways to derive a clustering result:

1. **DBSCAN clustering**: Choose a cut-off parameter ϵ and define all points p_i with $c(p_i) \leq \epsilon$ as core points. All points that are not in the ϵ-neighbourhood of a core point are defined as noise. This is equal to all points p that are not core points and have a reachability value $r(p)$ with $r(p) > \epsilon$. A cluster of size L is defined as a consecutive set (in the sense of the

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics Discussions
Open Access
EGU

ordering) of non-noise points $(p_j, p_{j+1}, \ldots, p_{j+L-1})$, with adjacent points p_{j-1} and p_{j+L} being noise. This is similar to the clustering result of a DBSCAN run with equal values for s_{min} and ϵ. All possible realizations of DBSCAN clusters (with the same s_{min}) can therefore be derived from the reachability values, core distances and the ordering determined by OPTICS. Up to boundary points, a DBSCAN clustering result can be obtained by drawing horizontal lines in the reachability plot, cf. section 4.

2. **ξ-clustering**: While the DBSCAN clustering method looks for deep valleys in the reachability plot, this method looks for valleys with steep boundaries. In short, the larger a parameter ξ with $0 < \xi < 1$, the steeper the boundary of a valley has to be to be classified as a cluster. In more detail, a ξ-cluster is defined as a consecutive set of points $(p_j, p_{j+1}, \ldots, p_{j+L-1})$ that has steep boundaries in the sense that for a parameter ξ, $0 < \xi < 1$:

   (a) The start of the cluster p_j is in a ξ-steep downward area. A ξ-steep downward area is a maximal set of consecutive points $(p_l, p_{l+1}, \ldots, p_{l+k})$ where: 1. p_l and p_{l+k} are ξ-steep downward points, i.e. $r(p_l) \leq (1-\xi)r(p_{l-1})$ and $r(p_{l+k}) \leq (1-\xi)r(p_{l+k-1})$, 2. $p_{l+i} \leq p_l$ for all $i = 1, \ldots, k$ and 3. not more than s_{min} consecutive points in the set are no ξ-steep downward points.

   (b) The end of the cluster p_{j+L-1} is a ξ-steep upward area. The definitions are the reverse of the ξ-steep downward area, with the definition of a ξ-steep upward point is changed to $r(p_j) \leq (1-\xi)r(p_{j+1})$.

   (c) The cluster contains at least s_{min} points.

   (d) Every point in the inside of the cluster is at least a factor of $(1-\xi)$ smaller than the boundary points p_j and p_{j+L-1}. All points that do not belong to a cluster are classified as noise.

We refer to Ankerst et al. (1999) for a more detailed discussion of the ξ-clustering method with illustrations for example data. Note that the full ξ-clustering method presented by Ankerst et al. (1999) does contain some more details related to the choice of the start and end points, which we did not mention here.

Functions to derive both clustering results from an OPTICS output are also available in the SciPy sklearn package. Note that the implementation in sklearn allows for a minimum cluster size different from s_{min} (item (c) for the ξ-clustering), but we will not make use of this additional freedom to reduce the number of parameters. Note that, different from k-Means, both clustering methods do not require an a priory determination of the number of clusters. For the ξ-clustering method, a larger ξ requires steeper boundaries to form a cluster, i.e. will typically lead to a reduction of the number of resulting clusters. For DBSCAN clustering with very large ϵ, one will detect one large global cluster. Making ϵ smaller leads then to consecutive splits of this cluster, forming (up to noise) a cluster hierarchy. We will demonstrate the properties for both clustering methods in section 4 for different situations. In the following applications, we use an estimation of the minimum number of particles per finite-time coherent set for the parameter s_{min}.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

4 Results

4.1 Bickley jet flow

We start with the direct embedding of the trajectories. As explained in section 2, the data matrix has dimension $X \in \mathbb{R}^{12,000\times 143}$. We apply the OPTICS algorithm to the resulting points, together with DBSCAN clustering, choosing $s_{min} = 80$ as a minimum size of the finite-time coherent sets. In the following, all axis units are in 1000 km. Figure 2 shows the reachability plot, together with the DBSCAN clustering result of three different choices of ϵ. The six vortices and the jet are clearly visible as the major valleys in the reachability plot. The hierachical structure of the DBSCAN clustering with decreasing ϵ is visible in the figures from top (larg scale coherence) to bottom (small scale coherence). Being able to study this hierarchical structure with one run of OPTICS is a major advantage compared to DBSCAN and other methods to detect finite-time coherent sets. Note again that one run of OPTICS provides the DBSCAN clustering result of any parameter ϵ (with the same s_{min}).

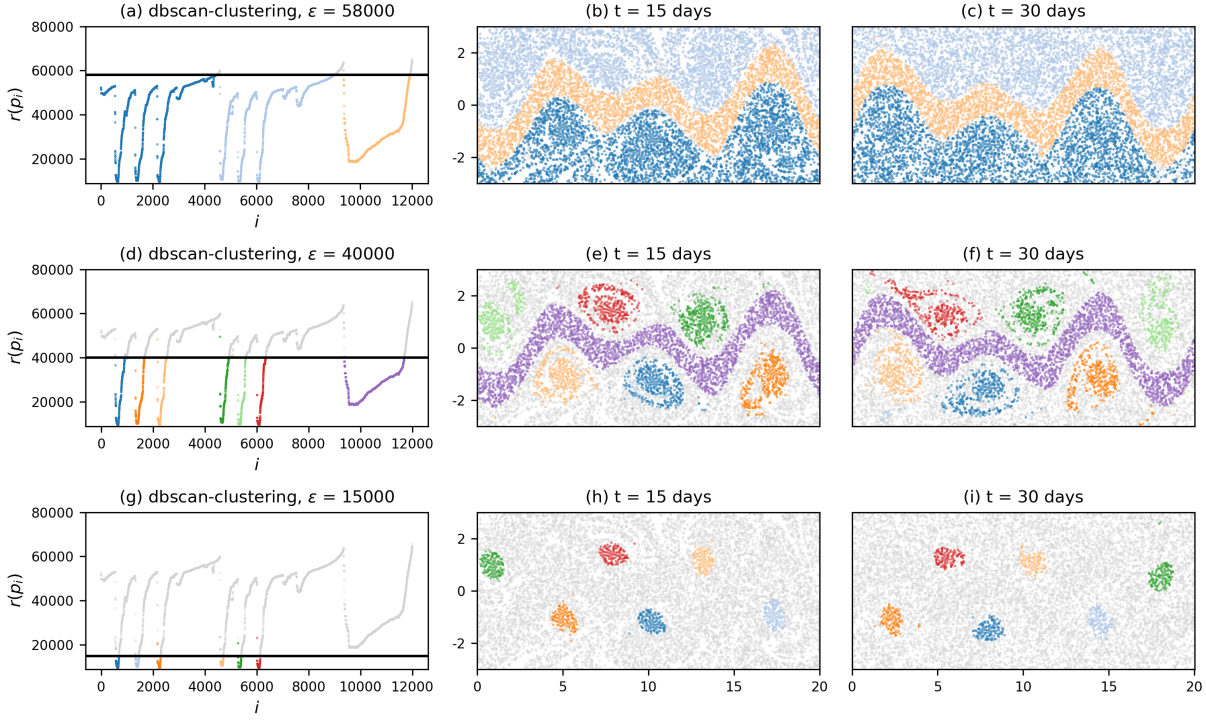


Figure 2. Result of the OPTICS algorithm applied to the direct embedding of the trajectories. (a), (d) and (g) show the reachability plot with different DBSCAN clustering results, indicated by the black horizontal line. The corresponding clustering results of each choice of DBSCAN parameter ϵ is shown on the right of the reachability plots for different times. Grey particles correspond to noise. Axis units in the centre and right column are in 1000 km.

Next, we use the embedded trajectories and apply classical MDS to obtain a 2-dimensional embedding. As mentioned in section 3.2.2, this assures to capture the major variance along the embedding axes. The spectrum of B in eq. (4) is shown in fig.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

C1 in the appendix, with two clearly dominant eigenvalues. Figure 3 shows the result of OPTICS for this case of embedding. Most notably, applying MDS has lead to the vortices and the jet having comparable depth in the reachability plot, such that a single DBSCAN clustering result detects all six vortex centres and the jet in the middle. Figure 4a shows the corresponding embedding in the 2-dimensional embedding space, and fig. 4b the cluster labels for OPTICS with DBSCAN clustering at $\epsilon = 1000$, as shown in fig. 3. The jet and the six vortices are clearly recognizable as dense accumulations of points in this 2-dimensional space. Figure 4c shows the result for a k-Means clustering with $K = 8$ clusters, which corresponds to the six vortices, the jet, and one noise cluster as suggested by Hadjighasem et al. (2016). The corresponding clustering result is shown in fig. B1 in the appendix, showing that the clusters corresponding to the vortices are much less focussed. In addition, each of the eight clusters in fig. 4c contains some of the noisy points of fig. 4b, which shows that using one additional cluster for noise does not really address the issue of not detecting the vortices properly for this case of embedding. We emphasize again that we use classical MDS here mostly for visualization purposes, as the computation of the classical MDS embedding is difficult for large particle sets. In our case, a dense $12,000 \times 12,000$ symmetric matrix has to be diagonalized, which already takes a significant amount of computation time.

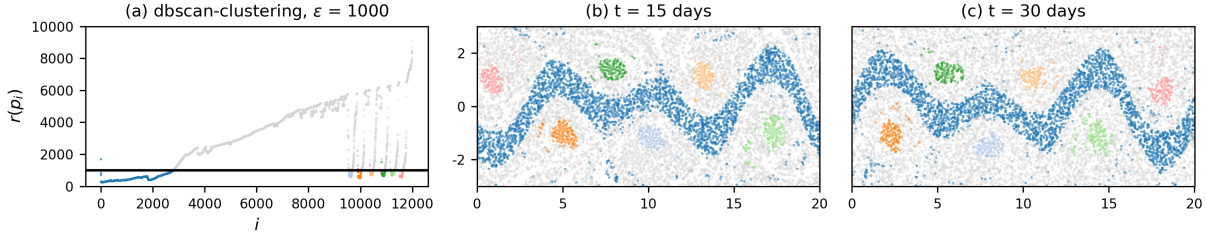


Figure 3. Result of DBSCAN clustering of the 2-dimensional embedding of the classical MDS method. a: reachability plot with black line representing the DBSCAN parameter ϵ. b-c: corresponding clustering results at different times. Grey particles represent noise. Axis units are in 1000 km.

4.2 Agulhas rings

We next apply OPTICS to the Agulhas trajectories. As described in section 2, we have $\bar{X} \in \mathbb{R}^{N \times 63}$ with $N = 23,821$. We choose $s_{min} = 100$ in the following, which corresponds to a square cell of $2° \times 2°$, i.e. a reasonable size of an Agulhas ring. Figure 5 shows the result of the direct embedding. The reachability plot in 5a is much more jagged than for the Bickley jet model flow (cf. fig. 2a). The narrow deep valleys and the wider valleys in the reachability plot indicate the presence of large and small scale coherence patterns. Figure 5a-c shows the DBSCAN clustering result for a relatively large value of ϵ. The main separation of fluid domains is between the red and the blue particles, with a few vortices at their boundary. These two water masses are the northern and southern parts of the subtropical gyre in the South Atlantic, the red particles moving to the west, the blue particles to the east. The second and third rows of fig. 5 show the results for the ξ-clustering method with different values of ξ. The valleys with steep boundaries correspond to eddy-like structures, separated by background noise. Our method also performs well when almost 80% of the trajectories is removed. Figure A1 shows a similar result for a reduced data set

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

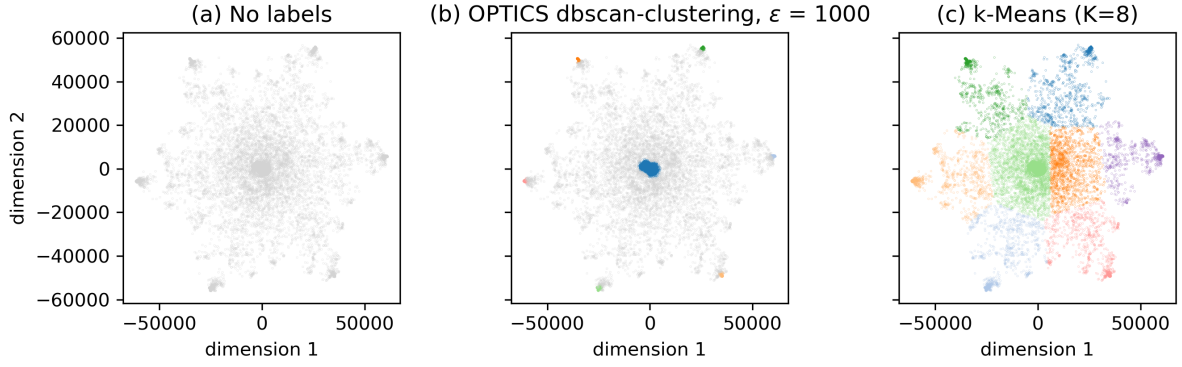


Figure 4. a: 2-dimensional embedding of the classical MDS method (cf. section 3.2.2) of the trajectories. b: with labels according to the DBSCAN result of fig. 3. The six vortices and the jet are clearly visible as dense regions. Grey particles correspond to noise. c: k-Means clustering result for K=8, see fig. B1 for the spatial clustering result of k-Means.

where we keep 5,000 randomly chosen particles and set $s_{min} = 20$ to account for the reduction in the number of particles. The large-scale structure as well as many of the eddies shown in fig. 5 are still visible.

We again emphasize that the main result of OPTICS is the reachability plot itself. Fig. 6 shows a colour map at initial time of the reachability values. We clearly see Agulhas rings as the dark regions corresponding to lowest values of reachability. The regions of large reachability correspond to trajectories that are relatively noisy compared to all the other trajectories.

In order to illustrate again the difference between OPTICS and k-Means for this example, we choose 12,000 random trajectories and again embed the trajectories in a 2-dimensional space with classical MDS (cf. section 3.2.2). The reduction of the particle set is necessary to simplify the eigendecomposition of the matrix B in eq. (4), and we therefore choose $s_{min} = 30$. The corresponding spectrum of B is shown in fig. C2 in the appendix, showing that there are again two dominant eigenvectors. Figure 7 shows the embedded trajectories together with OPTICS / DBSCAN clustering (fig. 7b) and k-Means (fig. 7c) for K=40. Figs. B2 and B3 show the corresponding clustering results in the fluid domains. It is visible that k-Means does not detect a single vortex, but splits the fluid domain into regions of similar size. OPTICS easily detects multiple Agulhas rings by finding the steepest valleys in the reachability plot.

Spectral embeddings derived from networks together with partition based clustering have a similar problem as the one illustrated in fig. 7c (Froyland et al., 2019). Similar to the case discussed here, OPTICS can be used to overcome the problems of k-Means. We show this in appendix D for the network proposed by Padberg-Gehle and Schneide (2017) for the Agulhas region, together with a brief introduction of the network and how to construct spectral embeddings. In summary, k-Means again fails to detect any of the vortices, while OPTICS detects many of the coherent vortices in the spectrally embedded network. Yet, other flow features are also present that result from the physical motivation of the network definition, see the results in appendix D.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

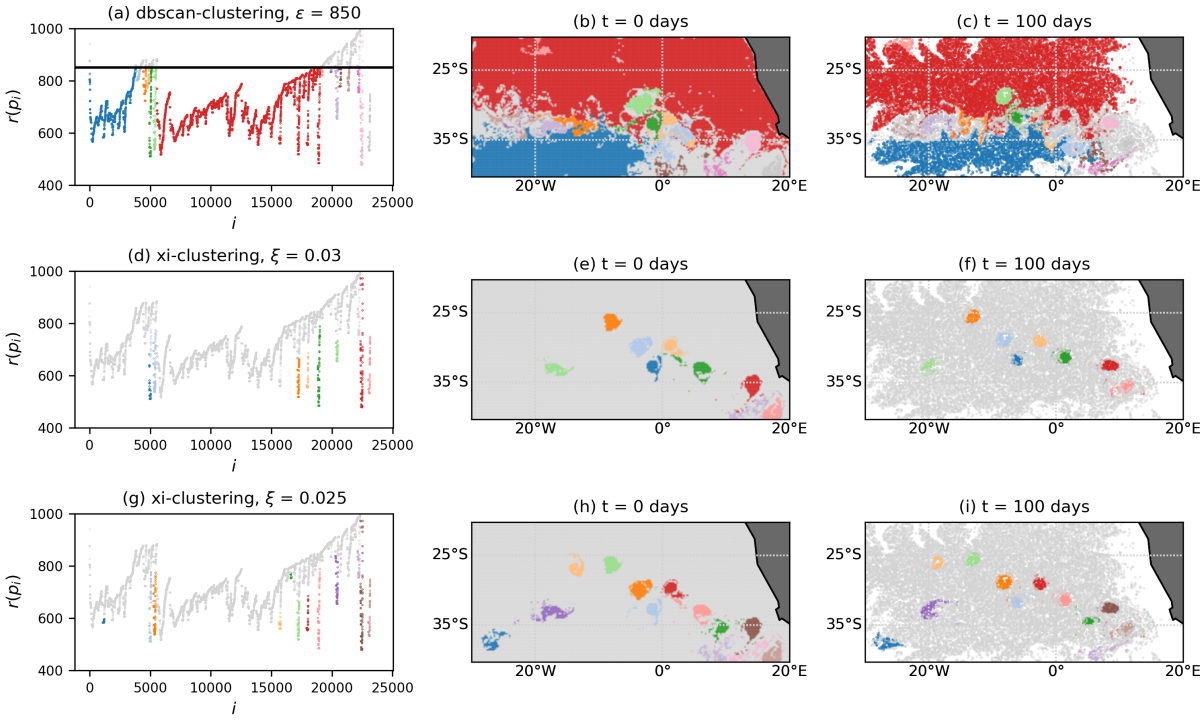


Figure 5. Result of the OPTICS algorithm applied to the direct embedding of the trajectories, with different clustering methods. Grey particles correspond to noise.

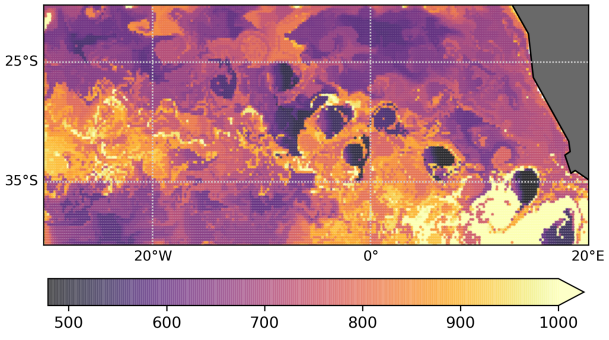


Figure 6. Reachability values at initial time, resulting from the OPTICS algorithm applied to the direct embedding of the trajectories. The regions with lowest values clearly correspond to Agulhas rings. The colour bar is cut off at a reachability of 1000 to show the relevant structure of variations.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

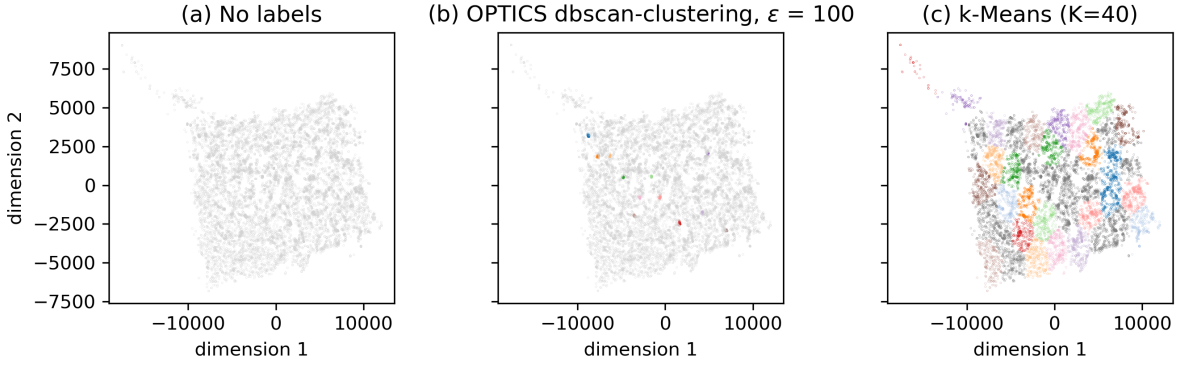


Figure 7. Embedding of the Agulhas trajectories in the 2-dimensional space defined by the leading eigenvectors of the MDS Kernel matrix B. a: no labels. b: clustering labels of OPTICS / DBSCAN, see fig. 7 for the corresponding plot in the Agulhas region. Grey particles represent noise. c: k-Means with $K = 40$, see fig. B3 for the corresponding plot in the Agulhas domain.

5 Conclusions

The abstract embedding of particle trajectories in a metric space with subsequent clustering is a promising field of research for the detection of finite-time coherent sets in oceanography, as it can be potentially applied to sparse sets of trajectories e.g. from drifter release experiments. Yet, most of the existing methods lack the ability to separate finite-time coherent structures from noisy trajectories that do not belong to any such structure, which hampers the application to large ocean domains. This is because the clustering methods proposed so far have been based on graph partitioning, which treats noisy, unclustered data points insufficiently. In this article, we presented a simple way to overcome this problem by using the density-based clustering algorithm OPTICS (Ankerst et al., 1999). Different from partition based clustering methods such as k-Means, OPTICS detects the clustering structure of the embedded trajectories by looking for dense accumulations of points, i.e. groups of trajectories that are close to each other in embedding space. Coherent groups of particle trajectories can be identified as valleys in the reachability plot computed by the OPTICS algorithm. This plot also has a natural interpretation in terms of cluster hierarchies, i.e. finite-time coherent sets that are by themselves part of a larger scale finite-time coherent set. Such hierarchies are present in the surface ocean flow, where the subtropical basins are approximately coherent and at the same time comprise other finite-time coherent structures such as eddies and jets. This hierarchical property is a clear advantage compared to DBSCAN, which has been used before to detect coherent sets (Schneide et al., 2018). One run of OPTICS can in principle produce all possible results of DBSCAN clustering for the same parameter s_{min}. In addition, different from DBSCAN, OPTICS can detect clusters of varying density, and detect them by locating the valleys with the steepest boundaries in the reachability plot with the ξ-clustering method of Ankerst et al. (1999).

We apply OPTICS to Lagrangian particle trajectories directly, in the spirit of Froyland and Padberg-Gehle (2015). OPTICS successfully detects the expected coherent structures in the Bickley jet model flow, separating the six vortices and the jet from

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics Discussions
Open Access
EGU

background noise. We also apply our method to the eastern South Atlantic and successfully identify Agulhas rings, separated by noise. We visualize the difference of OPTICS to k-Means with a 2-dimensional embedding of the trajectories based on classical multidimensional scaling. We also show how OPTICS can be applied to the spectral embedding of the particle based network proposed by Padberg-Gehle and Schneide (2017), providing a necessary amendment to this method to detect coherent vortices in a large ocean domain, i.e. when k-Means fails.

Our method is different from previous approaches used to detect finite-time coherent sets in ocean models from Lagrangian trajectory data as it has a clear interpretability in terms of clustering hierarchy, where large-scale and small scale structures are visible in the reachability plot produced by OPTICS. Our method can also be applied to scarce trajectory data sets, i.e. without too much concern about the spatial coverage of a fluid domain with initial conditions. Finally, our method is very simple to implement in Python, as OPTICS is available in the SciPy sklearn package. While we here present the results of OPTICS with three different kinds of embedding, it is likely that OPTICS also works for other trajectory embeddings, or even other methods using clustering such as transfer operator based finite-time coherent sets (Froyland et al., 2010) or dynamic Laplacians (Froyland et al., 2019).

Extending our method to datasets with more trajectories can be made more efficient by choosing a finite generating distance for OPTICS (Ankerst et al., 1999). While this is better from a computational point of view, it requires some knowledge or intuition about the spatial distribution of the embedded trajectories. A major challenge for the method proposed here is the embedding dimension. For very long trajectories, it is important to reduce the dimensionality of the trajectories before applying OPTICS. A complication here is the desired property of an embedding to preserve both local and global distances in order to make full use of the hierarchical properties of OPTICS. This means, for example, that the popular method of a locally linear embedding (Roweis and Saul, 2000) is not suitable, unless only the small scale (densest finite-time coherent sets) are to be detected. Using classical multidimensional scaling (MDS), as we did here to visualize the clustering results, in principle preserves local and global distances, but is not an option for very large data sets as it requires the diagonalization of a dense symmetric square matrix of size equal to the particle number. Spectral embeddings of derived networks such as the ones of Hadjighasem et al. (2016), Padberg-Gehle and Schneide (2017) and Banisch and Koltai (2017) are useful to achieve lower-dimensional embeddings, but they come with the introduction of additional parameters for the network construction. Further research into other non-linear dimensionality reduction techniques that have not been explored in the context of finite-time coherent sets can lead to more efficient methods.

Code and data availability. All code is available at https://github.com/OceanParcels/coherent_vortices_OPTICS, including the code to generate the Bickley jet trajectories. The data for the virtual particles in the South Atlantic is available on zenodo (Wichmann, 2020). Details on the Parcels simulation for the virtual trajectories in the ocean can be found at the github repository of our previous paper, https://github.com/OceanParcels/near_surface_microplastic. The data from the NEMO ORCA-006 run are available at http://opendap4gws.jasmin.ac.uk/thredds/nemo/root/catalog.html

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Appendix A: Agulhas rings with smaller particle set

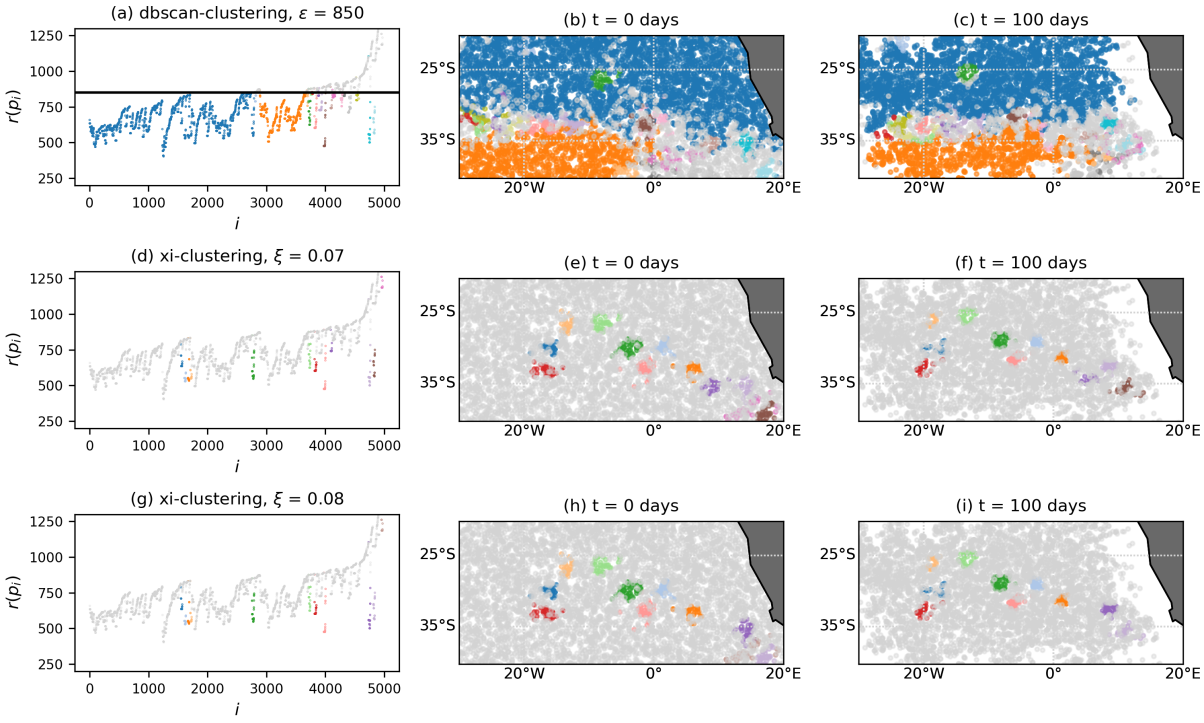


Figure A1. Result of the OPTICS algorithm for the Agulhas particles, applied to 5,000 randomly selected trajectories and $s_{min} = 20$. The large-scale structure as well as many of the eddies are very similar to the full dataset case, see fig. 5. Grey particles are classified as noise.

Appendix B: Additional figures for the classical MDS embedding

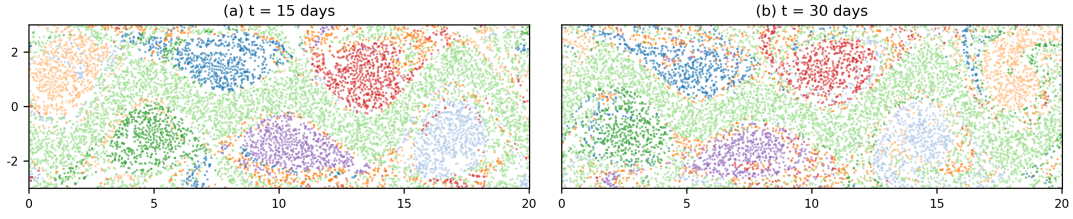


Figure B1. Result of $K = 8$ k-Means clustering of the 2-dimensional embedding from classical MDS, cf. fig. 3. Axis units are in 1000 km.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

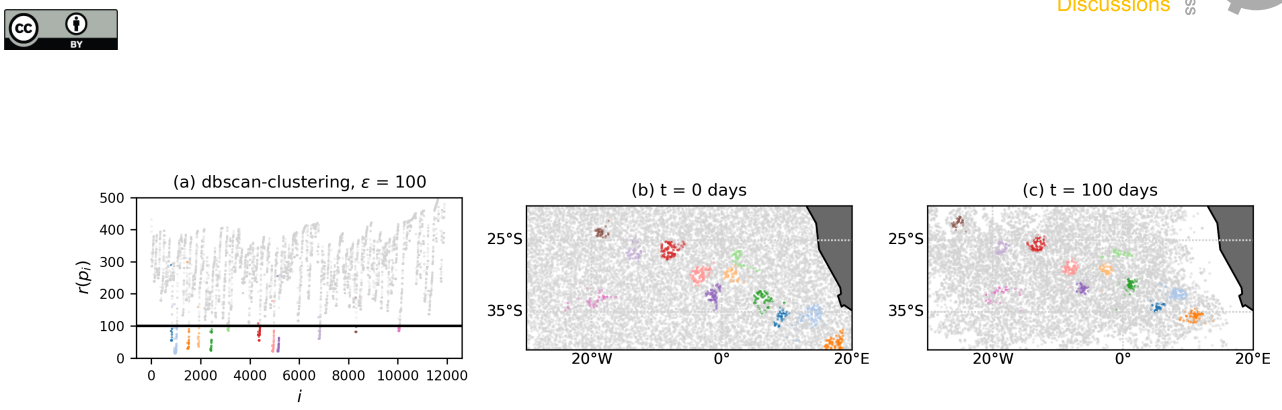


Figure B2. Result of OPTICS applied to the 2-dimensional embedding of 12,000 randomly selected particles with the classical MDS method, cf. fig. 7b, and $s_{min} = 30$. The corresponding spectrum is shown in fig. C2 in the appendix, showing that there are two dominant eigenvectors. Grey particles are classified as noise.

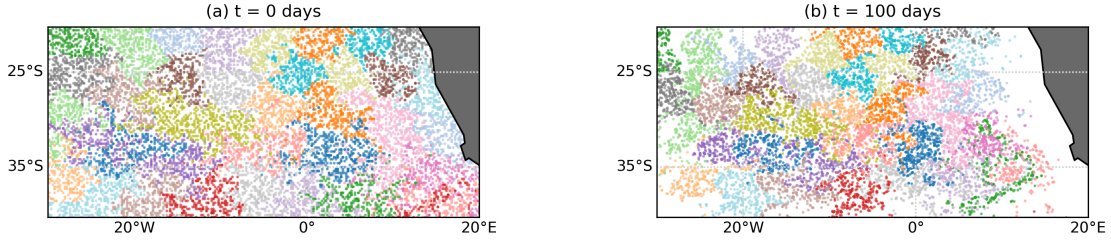


Figure B3. Result of the k-Means clustering with $K = 40$ applied to the 2-dimensional embedding with classical MDS, cf. fig. 7c.

Appendix C: Spectra of the MDS kernel matrices

Appendix D: Detecting Agulhas rings with a particle based network

To demonstrate that OPTICS can also be applied to the spectral embedding of a particle based network, we use the network proposed by Padberg-Gehle and Schneide (2017). If we have a set of particle trajectories $x_i(t)$, where $i = 1,\dots,N$, $t = t_1, t_2, \dots, t_T$ with N the number of particles and T the number of time steps, the network $A \in \mathbb{R}^{N\times N}$ is defined as:

$$A_{ij} = \begin{cases} 1, & \text{if } \exists t \in \{t_1, t_2, \dots, t_T\} \ \ s.t. \ \ ||x_i(t) - x_j(t)|| < d, \\ 0, & \text{otherwise.} \end{cases} \tag{D1}$$

Here, $||.||$ denotes the Euclidean norm and $d \in \mathbb{R}$ is a fixed pre-determined cut-off parameter, see Padberg-Gehle and Schneide (2017) for a discussion on the choice of d (called ϵ in Padberg-Gehle and Schneide (2017)). Similar to Padberg-Gehle and Schneide (2017), we embed the nodes in a lower dimensional space $\mathbb{R}^K$ by means of the eigenvectors of its random walk Laplacian, (see e.g. Von Luxburg (2007))

$$L_r = D^{-1}A, \tag{D2}$$

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

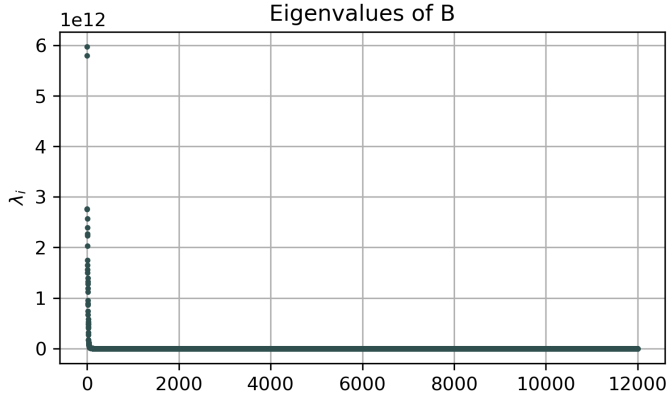


Figure C1. Spectrum of the classical Multidimensional Scaling Kernel matrix K for the Bickley Jet example. It is visible that there are three dominant eigenvalues. In the manuscript, we choose the vectors corresponding to he first two for visualization purposes.

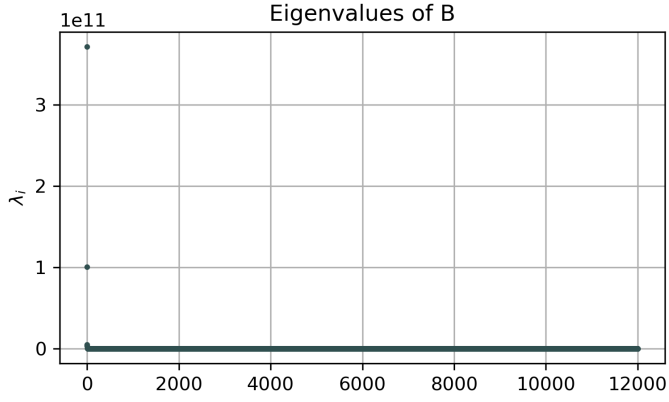


Figure C2. Spectrum of the classical Multidimensional Scaling Kernel matrix K for the Agulhas flow, where we first constrain the particle data to 12,000 randomly selected trajectories. There are again two dominant eigenvalues, for which we choose the corresponding vectors for the embedding.

where D is a diagonal matrix with $D_{ii} = \sum_j A_{ij}$. The embedding of node i is defined by

$$y_i = (v_{1,i}, v_{1,i}, \ldots, v_{K,i}) \in \mathbb{R}^K, \tag{D3}$$

where, the v_i, $i = 0, \ldots, N-1$ are the right eigenvectors corresponding to the largest eigenvalues λ_i of L_r. The eigenvalues are assumed to be ordered in descending order, i.e. $1 = \lambda_0 > \lambda_1 \geq \ldots, \geq \lambda_N$. This is the most common network embedding for the detection of finite-time coherent sets so far (Padberg-Gehle and Schneide, 2017; Banisch and Koltai, 2017; Hadjighasem et al., 2016). The classical simultaneous K-way normalized cut proceeds with applying the k-Means algorithm to the embedding defined in eq. (D3) to detect K clusters (Von Luxburg, 2007), resulting in an approximate solution to the normalized cut

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

problem (Shi and Malik, 2000).

Figure D1 shows the spectrum of the resulting random walk Laplacian with $d = 200$ km. No obvious spectral gap is visible that would suggest a truncation of the embedding space. Figure D2 shows the clustering result if we apply a k-Means algorithm as suggested by Padberg-Gehle and Schneide (2017) to detect $K = 40$ clusters. It is visible that the partition based k-Means clustering method does not detect any individual Agulhas rings, but partitions the state space into regions of approximately equal size.

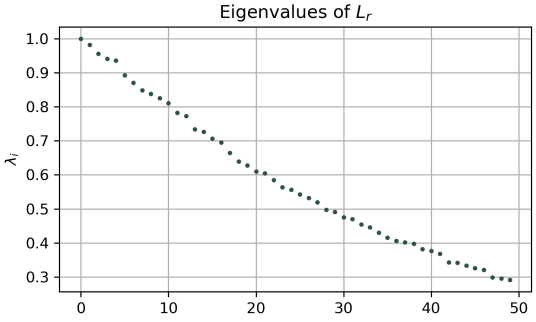


Figure D1. Spectrum of the random walk Laplacian, cf. eq. (D2) of the network proposed by Padberg-Gehle and Schneide (2017) applied to the Agulhas trajectory data. No clear gap exists that suggest a truncation of the embedding.

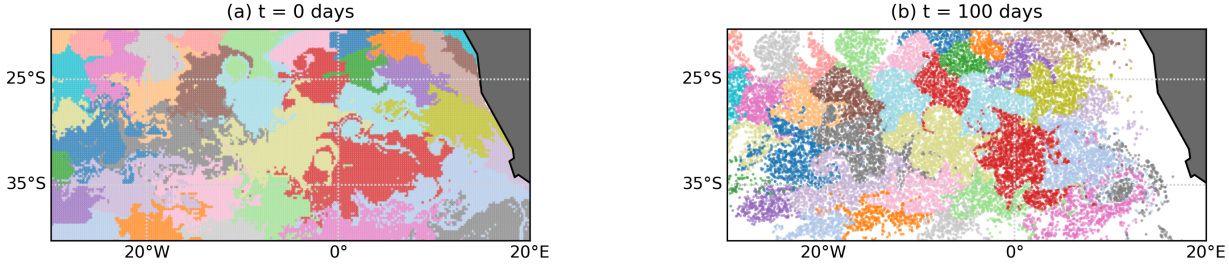
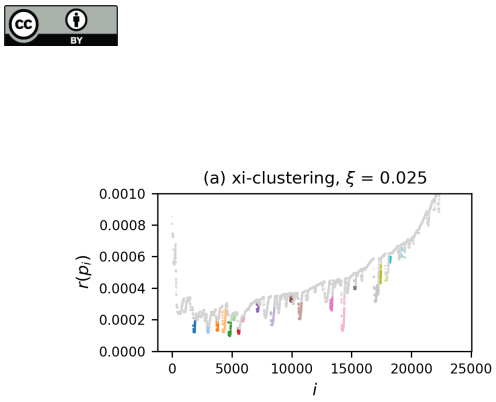


Figure D2. Result of k-Means clustering applied to the 40 leading eigenvectors of the random walk Laplacian, cf. eq. (D2), looking for 40 clusters. No individual vortices are detected.

Applying OPTICS instead ot k-Means with a subsequent ξ-clustering detects some of the Agulhas rings, see fig. D3, where we choose $s_{min} = 100$. Note that also other structures than typical circular eddies are detected. While this depends on the clustering parameter ξ (or ϵ for DBSCAN), this is also a consequence of the *physically motivated* network defined by eq. (D3), where particles are connected equally if they are close to each other at least once in time. This is different from the direct embedding, where we require particles to stay close along the entire trajectory.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020



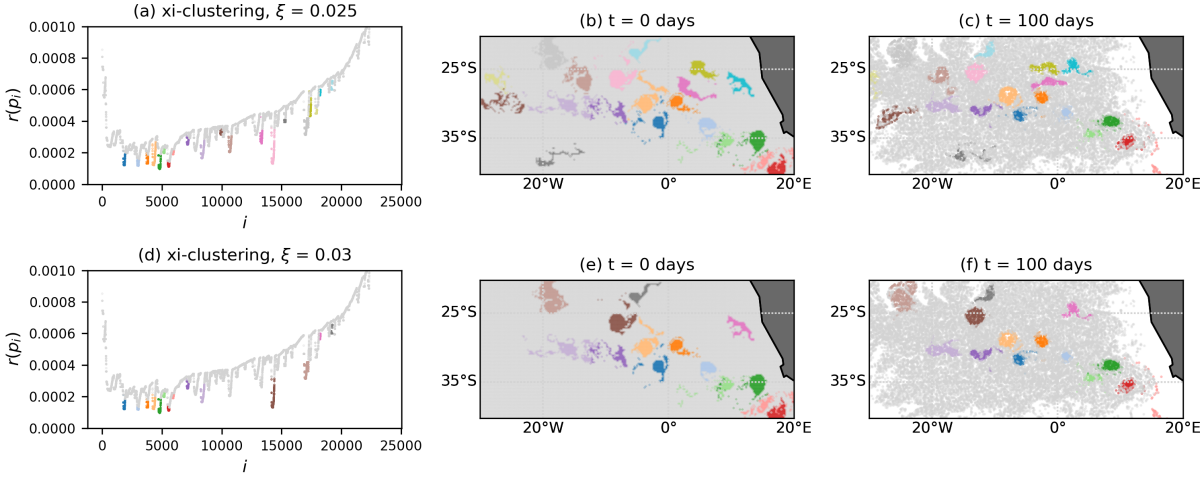


Figure D3. Result of optics applied to the $K = 40$ spectral embedding of the network defined in eq. (D1) with $d = 200$ km and $s_{min} = 100$. Grey particles are classified as noise.

Author contributions. DW performed the analysis, with support from CK, EvS and HD. DW wrote the manuscript and all authors jointly edited and revised it

Competing interests. The authors declare no competing interests

Acknowledgements. David Wichmann, Christian Kehl and Erik van Sebille are supported through funding from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No 715386). This work was partially carried out on the Dutch national e-infrastructure with the support of SURF Cooperative (project no. 16371). We thank Andrew Coward for providing the ORCA-N006 simulation data.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

References

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J.: OPTICS: ordering points to identify the clustering structure, ACM Sigmod record, 28, 49–60, 1999.

Banisch, R. and Koltai, P.: Understanding the geometry of transport: Diffusion maps for lagrangian trajectory data unravel coherent sets, Chaos, 27, https://doi.org/10.1063/1.4971788, 2017.

Beron-Vera, F. J., Wang, Y., Olascoaga, M. J., Goni, G. J., and Haller, G.: Objective detection of oceanic eddies and the agulhas leakage, Journal of Physical Oceanography, 43, 1426–1438, https://doi.org/10.1175/JPO-D-12-0171.1, 2013.

Brach, L., Deixonne, P., Bernard, M. F., Durand, E., Desjean, M. C., Perez, E., van Sebille, E., and ter Halle, A.: Anticyclonic eddies increase accumulation of microplastic in the North Atlantic subtropical gyre, Marine Pollution Bulletin, 126, 191–196, https://doi.org/10.1016/j.marpolbul.2017.10.077, http://dx.doi.org/10.1016/j.marpolbul.2017.10.077, 2018.

Delandmeter, P. and van Sebille, E.: The Parcels v2.0 Lagrangian framework: new field interpolation schemes, Geoscientific Model Development Discussions, pp. 1–24, 2019.

Dencausse, G., Arhan, M., and Speich, S.: Routes of Agulhas rings in the southeastern Cape Basin, Deep-Sea Research Part I: Oceanographic Research Papers, 57, 1406–1421, https://doi.org/10.1016/j.dsr.2010.07.008, http://dx.doi.org/10.1016/j.dsr.2010.07.008, 2010.

Dong, C., McWilliams, J. C., Liu, Y., and Chen, D.: Global heat and salt transports by eddy movement, Nature Communications, 5, 1–6, https://doi.org/10.1038/ncomms4294, http://dx.doi.org/10.1038/ncomms4294, 2014.

Dussin, R., Barnier, B., and Brodeau, L.: The making of Drakkar forcing set DFS5, Tech. rep., LGGE, Grenoble, France., 2016.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., and Others: A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd, vol. 96, pp. 226–231, 1996.

Fouss, F., Saerens, M., and Shimbo, M.: Algorithms and models for network data and link analysis, Cambridge University Press, 2016.

Froyland, G. and Padberg-Gehle, K.: A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data, Chaos, 25, https://doi.org/10.1063/1.4926372, 2015.

Froyland, G., Santitissadeekorn, N., and Monahan, A.: Transport in time-dependent dynamical systems: Finite-time coherent sets, Chaos, 20, 1–10, https://doi.org/10.1063/1.3502450, 2010.

Froyland, G., Stuart, R. M., and van Sebille, E.: How well-connected is the surface of the global ocean?, Chaos, 24, 033 126, https://doi.org/10.1063/1.4892530, 2014.

Froyland, G., Horenkamp, C., Rossi, V., and van Sebille, E.: Studying an Agulhas ring's long-term pathway and decay with finite-time coherent sets, Chaos, 25, 083 119, https://doi.org/10.1063/1.4927830, 2015.

Froyland, G., Rock, C. P., and Sakellariou, K.: Sparse eigenbasis approximation: Multiple feature extraction across spatiotemporal scales with application to coherent set identification, Communications in Nonlinear Science and Numerical Simulation, 77, 81–107, https://doi.org/10.1016/j.cnsns.2019.04.012, 2019.

Hadjighasem, A., Karrasch, D., Teramoto, H., and Haller, G.: Spectral-clustering approach to Lagrangian vortex detection, Physical Review E, 93, 1–17, https://doi.org/10.1103/PhysRevE.93.063107, 2016.

Hadjighasem, A., Farazmand, M., Blazevski, D., Froyland, G., and Haller, G.: A critical comparison of Lagrangian methods for coherent structure detection, Chaos, 27, https://doi.org/10.1063/1.4982720, 2017.

Haller, G. and Beron-Vera, F. J.: Coherent Lagrangian vortices: The black holes of turbulence, Journal of Fluid Mechanics, 731, 1–10, https://doi.org/10.1017/jfm.2013.391, 2013.

https://doi.org/10.5194/npg-2020-28
Preprint. Discussion started: 29 June 2020


Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Lange, M. and van Sebille, E.: Parcels v0.9: Prototyping a Lagrangian ocean analysis framework for the petascale age, Geoscientific Model Development, 10, 4175–4186, 2017.

Ma, T. and Bollt, E. M.: Relatively coherent sets as a hierarchical partition method, International Journal of Bifurcation and Chaos, 23, 1–18, https://doi.org/10.1142/S0218127413300267, 2013.

Madec, G.: NEMO ocean engine, Note du Pôle de modélisation, No 27, 2008.

Padberg-Gehle, K. and Schneide, C.: Network-based study of Lagrangian transport and mixing, Nonlinear Processes in Geophysics, 24, 661–671, https://doi.org/10.5194/npg-24-661-2017, 2017.

Roweis, S. T. and Saul, L. K.: Nonlinear dimensionality reduction by locally linear embedding, science, 290, 2323–2326, 2000.

Schneide, C., Pandey, A., Padberg-Gehle, K., and Schumacher, J.: Probing turbulent superstructures in Rayleigh-Bénard convection by Lagrangian trajectory clusters, Physical Review Fluids, 3, 1–12, https://doi.org/10.1103/PhysRevFluids.3.113501, 2018.

Schouten, M. W., de Ruijter, W. P. M., van Leeuwen, P. J., and Lutjeharms, J. R. E.: Translation, decay and splitting of Agulhas rings in the southeastern Atlantic Ocean, Journal of Geophysical Research: Oceans, 105, 21 913–21 925, https://doi.org/10.1029/1999jc000046, 2000.

Shi, J. and Malik, J.: Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 888–905, https://doi.org/10.1109/34.868688, 2000.

Tarshish, N., Abernathey, R., Zhang, C., Dufour, C. O., Frenger, I., and Griffies, S. M.: Identifying Lagrangian coherent vortices in a mesoscale ocean model, Ocean Modelling, 130, 15–28, https://doi.org/10.1016/j.ocemod.2018.07.001, 2018.

Van Sebille, E., Aliani, S., Law, K. L., Maximenko, N., Alsina, J. M., Bagaev, A., Bergmann, M., Chapron, B., Chubarenko, I., Cózar, A., Delandmeter, P., Egger, M., Fox-Kemper, B., Garaba, S. P., Goddijn-Murphy, L., Hardesty, B. D., Hoffman, M. J., Isobe, A., Jongedijk, C. E., Kaandorp, M. L., Khatmullina, L., Koelmans, A. A., Kukulka, T., Laufkötter, C., Lebreton, L., Lobelle, D., Maes, C., Martinez-Vicente, V., Morales Maqueda, M. A., Poulain-Zarcos, M., Rodríguez, E., Ryan, P. G., Shanks, A. L., Shim, W. J., Suaria, G., Thiel, M., Van Den Bremer, T. S., and Wichmann, D.: The physical oceanography of the transport of floating marine debris, Environmental Research Letters, 15, https://doi.org/10.1088/1748-9326/ab6d7d, 2020.

Von Luxburg, U.: A tutorial on spectral clustering, Statistics and Computing, 17, 395–416, https://doi.org/10.1007/s11222-007-9033-z, 2007.

Wichmann, D.: Lagrangian particle dataset (2 years) for Agulhas region surface flow, https://doi.org/10.5281/zenodo.3899942, https://doi.org/10.5281/zenodo.3899942, 2020.

Wichmann, D., Delandmeter, P., and van Sebille, E.: Influence of near-surface currents on the global dispersal of marine microplastic, J. Geophys. Res. Oceans, 124, 6086–6096, 2019.

Wichmann, D., Kehl, C., Dijkstra, H. A., and van Sebille, E.: Detecting flow features in scarce trajectory data using networks derived from symbolic itineraries: an application to surface drifters in the North Atlantic, Nonlinear Processes in Geophysics Discussions, 2020, 1–20, https://doi.org/10.5194/npg-2020-18, https://www.nonlin-processes-geophys-discuss.net/npg-2020-18/, 2020.