Answers to the comments of Referee 2

Note: Line numbers refer to the track-changed version

Comment 1

A new and potentially useful method is introduced and discussed in relation to established methods, but there is no direct comparison, neither w.r.t. to the resulting clustering nor to computational run-times. While a detailed comparative study is certainly beyond the scope of this paper, the discussion should be extended in that respect. What are the advantages and what are the limitations of the method – in comparison to the established approaches?

Answer to comment 1

Thank you very much for that comment. A more detailed comparison was indeed lacking in the first version. In the revised version, we will add a subsection (3.4) in the methods section to point out the differences to previous methods, advantages and limitations.

Changes in text, line 236:

3.4 Comparison to existing methods

Our method aims to detect groups of particles, with trajectories of different groups having only little overlap. In this sense, our method detects groups of particles with little mixing between each other, which is close to detecting almost-invariant sets according to Froyland (2005). Yet our method is different from detecting almost-invariant sets with the transfer operator in several aspects. First, it is based on similarities between individual particles rather than spatial sets (bins), which allows us to cluster on the particle level rather than the bin level. As we will show in section 4.1, this can be used to resolve flow features down to scales much below the bin size. Secondly, our method employs the full trajectory information in terms of a particle's symbolic itinerary, rather than just the start and end points or symbols. In practical applications, this can be an advantage compared to the transfer operator framework, as there is no need in assuming Markovian behaviour of the flow given a state space partition, as done by e.g. Froyland et al. (2014).

There are also major differences between our method and other existing methods that cluster on the particle level (Froyland and Padberg-Gehle, 2015; Banisch and Koltai, 2017). First, these methods only compare particles at equal times, while we disregard the time information. This can be a significant advantage in situations where the major features of a flow are approximately stationary, i.e. can be seen as part of a (noisy) autonomous dynamical system. In this case, using the time information of drifter trajectories should not be necessary. Especially for the ocean drifter dataset, containing drifters of different starting times and lengths, it would be very difficult if not impossible to find sub-basin large scale structures when restricting to drifters that necessarily overlap temporally, although this is possible on the global scale to identify the basins themselves (Froyland and Padberg-Gehle, 2015; Banisch and Koltai, 2017). Note that simply placing all drifters at the same initial time and proceeding with one of the existing methods would lead to further problems, as there is ambiguity in which point of a trajectory should be taken for the initial time, probably requiring more data processing such as demanding an initially uniform particle distribution. Our method of simplifying the trajectories does not have these problems by construction, and can be readily applied to scarce drifter datasets.

From a computational perspective, setting up one of the sparse matrices $C(t)$ in eq. (1) is $O(N)$ such that computing the matrix G in eq. (2) is $O(N\tau)$. Computing $d[G^T]$ is $O(NM)$ as is the product $Gd[G^T]$). In total, computing R of eq. (10) is therefore of computational complexity $O(NM + N\tau)$. If we work with R directly, i.e. we use the simultaneous K-way clustering method described in algorithm 1 in section 3.3, computing this network is of lower computational complexity as the computation of the networks used by other studies (Padberg-Gehle and Schneide, 2017; Banisch and Koltai, 2017; Hadjighasem et al., 2016). These methods typically rely on comparing particle positions between all particles at all time instances, i.e. they scale with $O(N^2\tau)$ in the worst case, although a nearest-neighbour search as applicable to the studies of Padberg-Gehle and Schneide (2017) and Banisch and Koltai (2017) can reduce the N^2 term to something like $N \log N$. Further, the matrix R in eq. (10) is sparser

than $L_s[Q]$ or can have column dimension (= number of bins M) significantly lower than row dimension (= number of particles N), cf. section 4.1. In these cases, computing the SVD of R instead of the eigenvectors of L_s can lead to computational speed up. Finally, it is interesting to note that the computation of the network is faster for coarser partitions, i.e. when particles are connected in the network even when their trajectories are far apart, as the number of bins M decreases. This is opposite to the methods of Padberg-Gehle and Schneide (2017) and Banisch and Koltai (2017), where computing the network becomes more costly for larger spatial scale parameters (called ϵ in both studies).

The major drawback of our method is the dependence on a reference frame with respect to which the phase space partition and thus the symbolic itineraries are defined. This can be understood when imagining a time-independent flow from a rotating reference frame. The rotation of the reference frame contributes to a particle's itinerary, and, by averaging over different points in time, non-zero similarities between trajectories can result from the sole rotation of the reference frame. Due to this reason, our method can not be applied to strongly time-dependent systems such as the Bickley jet model flow where coherent vortices are transported in a periodic background flow. It is, however, still possible to detect transport boundaries in time-dependent flows such as the periodically driven double-gyre flow, as we show in section 4.1, where particle trajectories belonging to different invariant sets can still be distinguished with a fixed partition.

Comment 2

The two case studies (double-gyre and drifters) are each treated with a different clustering approach (K-way clustering vs hierarchical clustering) and it remains open how these two choices influence the results, in particular as there is no obvious spectral gap in the double-gyre system indicating an appropriate choice of K. I also assume that the results depend very much on the trajectory length but this is only briefly mentioned for the drifter data (l. 286). These points could be addressed in a more detailed study of the double-gyre flow, taking into account different flow times and the two clustering approaches.

Answer to comment 2

Thank you very much for this suggestion. It, in fact, helped us to better understand our method. We now applied also the hierarchical clustering method to the double-gyre flow and found that there is not necessarily a global minimum in the NCut for the first split (along the transport boundary). We found that such a minimum appears only for larger bin sizes, see the figures below (B5-B7). The lack of a global minimum for this system was found before for the transfer operators by Froyland & Padberg (2009), and is partially also a consequence of the very idealized flow, where the objective function is symmetric around c=0. Nevertheless, because of this sensitivity on the bin size, we now also computed the clustering results for the North Atlantic drifters for different bin sizes, see the figure below (C2). We do not find any striking difference in the main features of the North Atlantic Ocean. The text will be changed accordingly.

Changes in text, line 328:

We also tested the algorithm for shorter trajectories, cf. fig. B4 in the appendix, showing an expected change of the boundary filaments between the left and right sides of the fluid, which mix less in the shorter period of time. To better understand

the differences between algorithms 1 and 2 introduced in section 3.3, we also applied the hierarchical NCut method to the non-autonomous double gyre flow. A problem arises here for the first split into two clusters, as there is no unique minimum in the objective function in eq. (6), i.e determining the cutoff c in algorithm 2, H5 (cf. section 3.3), is ambiguous for $\Delta x = \Delta y = 0.04$. The lack of a unique minimum for the NCut has been observed before for the same model flow by Froyland and Padberg (2009) in the transfer operator framework (see their fig. 15) corresponding to the lack of a unique maximum of the coherence ratio there (see proposition 1 in appendix A). Yet, for $\Delta x = \Delta y = 0.04$, in our case, the split between the left and right sides along the transport boundary (the $K = 2$ split in fig. 3) does not even have a local minimum for the NCut, cf. fig. B5, as opposed to the local maximum of Froyland and Padberg (2009). This however changes to a local minimum for $\Delta x = \Delta y = 0.1$ (fig. B6) and finally to a global minimum for $\Delta x = \Delta y = 0.2$ (fig. B7). A possible explanation of this dependence on bin size is that the addition of noise (i.e. larger bins) decreases the coherence of the gyre centres compared to the transport boundary, making the latter easier to be detected. Due to the sensitivity of the hierarchical clustering result to the bin size, we test different bin sizes for the clustering of the North Atlantic drifters in section 4.2 (cf. fig. C2).

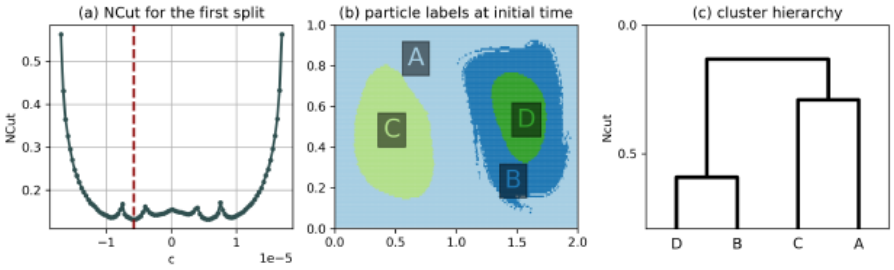

Figure B5. Hierarchical clustering result of the double gyre using algorithm 2 of section 3.3 and $\Delta x = \Delta y = 0.04$. a: NCut for the first split as a function of the cutoff c, where $c = 0$ corresponds to a cut along the transport boundary, the red line indicating the minimum found by our algorithm. b: result of the hierarchical clustering. c: dendrogram representing the hierarchy, where the horizontal lines correspond to the NCut value after each split. There is no global or local minimum at $c = 0$ in (a), which is why the transport boundary is not detected. Instead, two global minima are present in (a), but our implementation selects the left minimum for the first split due to the choice of sampling points.

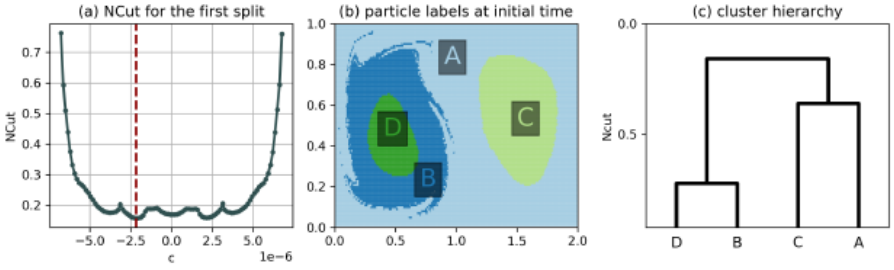

Figure B6. Hierarchical clustering result of the double gyre using algorithm 2 of section 3.3 and $\Delta x = \Delta y = 0.1$. a: NCut for the first split as a function of the cutoff c, where $c = 0$ corresponds to a cut along the transport boundary, the red line indicating the minimum found by our algorithm. b: result of the hierarchical clustering. c: dendrogram representing the hierarchy, where the horizontal lines correspond to the NCut value after each split. There is only a local minimum at $c = 0$, which is why the transport boundary is still not detected. This situation is similar to the one described by Froyland and Padberg (2009) for the transfer operator framework.

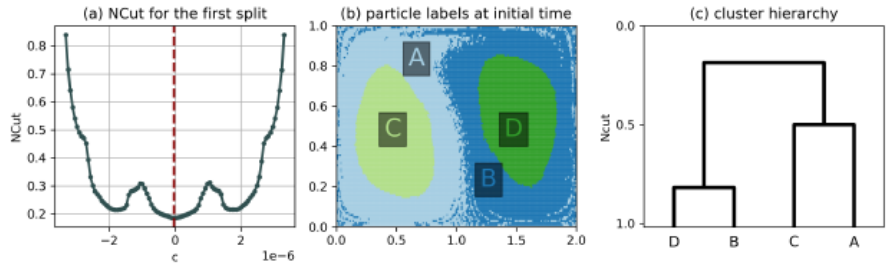

Figure B7. Hierarchical clustering result of the double gyre using algorithm 2 of section 3.3 and $\Delta x = \Delta y = 0.2$. a: NCut for the first split as a function of the cutoff c, where $c = 0$ corresponds to a cut along the transport boundary. b: result of the hierarchical clustering. c: dendrogram representing the hierarchy, where the horizontal lines correspond to the NCut value after each split. There is a global minimum at $c = 0$, so that the transport boundary is detected as the first split.

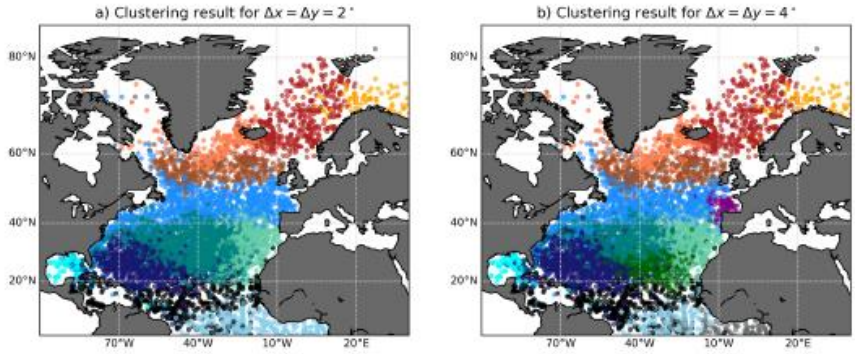

Figure C2. Clustering of Q for the drifter data set and different bin sizes, with trajectory lengths restricted to one year, 20 clusters as in fig. 6, plotted at initial time and the drifter release location. a: $\Delta x = \Delta y = 2°$. b: $\Delta x = \Delta y = 4°$. The main features of fig. 6 are still visible, although some structures change: the Bay of Biscay is not detected in (a), and the structure of the clusters in the centre of the Subtropical Gyre change compared to fig. 6. There are also new clusters in black (a, b) and dark green (b). Note that the main features of the North Atlantic Ocean such as the Subtropical-Subpolar Gyre boundary, the Western Boundary Current region, the Caribbean Sea and the Nordic Seas are still detected, see the result in fig. 6.

Comment 3

I don't understand how figure 7 relates to the hierarchical clustering that is carried out for the drifter data and where the indicated separations between the different geographical regions come from. Does fig 7 show the results for the K-way clustering? Some more explanations are required in my view.

Answer to Comment 3

Thanks a lot for this comment. You were absolutely right that the figure had little meaning in the context of hierarchical clustering. We will remove it in the revised manuscript.

Comment 4

The chosen bin size of 0.4 for the sparse data case (l. 246) means that some bins have to cropped in order to fit into the domain and as a result the bins are not equally sized. How is that done and how does that influence the computation?

Answer to comment 4
Thanks for that comment. Indeed the bins are not equally sized then, or we artificially extend the domain to y = 1.2. The text will be adapted accordingly.

Changes in text, line 316:

the bin size such that the network becomes connected again. Therefore, we choose $\Delta x = \Delta y = 0.4$~~for this case.~~. In doing so, we effectively extend the domain in the y-direction to $y = 1.2$ and disregard the fact that the top row of bins is not completely covered with initial conditions. Figure 5 shows the result for the clustering of Q for the incomplete data set, plotted on top of

Comment 5
Probably not all readers are familiar with the concept of almost-invariant sets, so this should be briefly motivated in the introduction.

Answer to comment 5
We will include a brief explanation in the introduction.

Changes in text, line 57:

method is naturally extendable to incomplete trajectory data and thus readily applicable to ocean drifters. Conceptually, our method is close to detecting minimally mixing fluid regions, so called almost-invariant sets (Froyland, 2005), although there are also important differences to this method, cf. section 3.4. In our case, almost-invariant sets are represented by the initial conditions of groups of particles, with trajectories of different groups having only little overlap.

Comment 6
The other anonymous referee already pointed out that S is not defined in the denominator of equation (6). In that context it would be helpful for the reader if the authors briefly explain the cost function.

Answer to comment 6
Thank you, this will be fixed in the revised version.

Changes in text, line 155

Assume we are given an undirected network ~~with~~ defined on a discrete set S containing N vertices~~and~~, with edges given by the symmetric adjacency matrix ~~$Q_{ij} \in \mathbb{R}^{N\times N}$~~ $Q \in \mathbb{R}^{N\times N}$. We assume that Q is connected. If it is not connected, we focus on ~~the largest~~ each connected component separately~~(see section 4.2)~~. According to Shi and Malik (2000), the normalized cut of a partition of the nodes into K sets $S_1, \ldots, S_K$, $S = \cup_{i=1}^{K} S_i$, is defined as

$$\mathrm{NCut}(S_1, \ldots, S_K) := \sum_{i}^{K} \frac{Q(S_i, S_i^C)}{Q(S_i, S)}. \tag{6}$$

Here, $Q(S_i, S_j)$ is the sum of all weights connecting S_i and S_j, i.e. $Q(S_i, S)$ is the sum of all weights connected to S_i. S_i^C denotes the complement of S_i. The term $\frac{Q(S_i, S_i^C)}{Q(S_i, S)}$ appearing in the definition of the NCut in eq. (6) is simply the total weight of all the edges connecting a set S_i to its complement relative to the total weight of the set S_i. Clustering a graph according to the NCut refers to finding a partition $\{S_k\}$ such the objective function in eq. (6) is minimized. Note that for an increasing

Comment 7

I also realized that the concept of “similarity” is used in a rather sloppy sense.

Answer to comment 7
Thank you for noting, indeed we were not very inconsistent in the usage of ‘similar’ and imprecise. We will change this in the revised version.

Changes in text, line 185:
... it follows that these eigenvectors are equal to the left singular vectors corresponding to...

Changes in text, line 438:
Minimizing the generalized normalized cut of A_P defined in eq. (6) with spectral relaxation is equal to maximizing the generalized coherence ratio

Changes in text, line 447:
We now show that the right singular vectors of R defined in eq. (10) are equal to the right eigenvectors of $\hat{P}$ if the particle measure is invariant and uniform.

Comment 8
The color figures are not appropriate for gray-scale printouts

Answer to comment 8
Thanks for the comment. As we need to use quite a few colours to illustrate the clusters, and as papers in NPG are available online for free, we do not think this is a major problem and would like to leave this issue to the Editor.

Comment 9
I suggest some critical proof-reading (e.g. capitalization in reference list).

Answer to comment 9
Thank you for also checking the references so carefully. We have gone through them again and will make appropriate changes, also regarding capitalization and doi’s in the references.

References

Froyland, G., & Padberg, K. (2009). Almost-invariant sets and invariant manifolds - Connecting probabilistic and geometric descriptions of coherent structures in flows. *Physica D: Nonlinear Phenomena*, *238*(16), 1507–1523. https://doi.org/10.1016/j.physd.2009.03.002