Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2020-18-RC2, 2020

Nonlinear Processes in Geophysics
Discussions

Interactive comment on "Detecting flow features in scarce trajectory data using networks derived from symbolic itineraries: an application to surface drifters in the North Atlantic" *by* David Wichmann et al.

Anonymous Referee #2

Received and published: 5 July 2020

The authors introduce a novel method for the identification of almost-invariant sets in fluid flows based on sparse and incomplete trajectory. By binning the particle positions and ignoring the temporal information a bipartite network is obtained that connects particles and bins. Links are weighted with the probability that the particle can be found in that bin. By solving either a standard or hierarchical NCut clustering problem, dominant flow features are detected from groups of trajectories that behave in a similar manner. The approach is successfully applied to the double-gyre benchmark flow and to surface drifter data in the North Atlantic and shown to perform well for the respective

setting, even in the case of scarce trajectory data. The paper is very well and clearly written and the method is certainly of great interest to the readers of NPG. However, a moderate revision addressing the points outlined below is required before the paper can be accepted for publication. My major criticism is that a more exhaustive study of the method is required.

Major points:

- 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?

- 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.

- 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.

- 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?

Minor points:

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

- 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.

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

- The color figures are not appropriate for gray-scale printouts.

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

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2020-18, 2020.