Review of: *Applying dynamical systems techniques to real ocean drifters,* by Rypina et al.

This manuscript describes a variety of Lagrangian diagnostic fields computed from a dense drifter array deployed in the Gulf of Mexico. These types of fields have long been used to study flow properties and Lagrangian coherent structures (LCS) in simulations, but this is the first time they have been applied to this degree to an observed drifter dataset. As such, this is a welcome contribution to the field. The authors were able to identify a structure reminiscent of a hyperbolic region, as well as areas of strong cumulative convergence and divergence. However, these features were identified from the trajectories alone and only confirmed to a smaller or greater degree by the various diagnostic fields. It remains unclear to me whether any of the calculations added anything to the understanding of the flow. (I should emphasize that I find this a useful report, even if the conclusion is that even this dense drifter deployment is insufficient to extract any use from the diagnostic fields.) The authors' assessments that one field or another was "useful in identifying the dominant LCS" (or not) seemed a bit haphazard. Similarly, the comparisons between a dense and a sparse drifter simulation, presented in the supplemental materials, was rather subjective, and "good agreement" was hard to distinguish from "poor agreement". I believe the paper could benefit if the authors carefully defined what they mean by LCS (it seems to vary for different diagnostics), and clarify the specific contributions of each field, if any. I therefore recommend **major revisions** prior to acceptance of the paper. Detailed comments follow.

Main text

- Line 29: I suggest to name the two unsuccessful methods in the abstract.
- Lines 40 – 43: I find this statement full of jargon and hard to understand for anyone not familiar with Lagrangian ocean analysis.
- Lines 74 – 79: The dataset referenced on line 567 contains 144 drifters released that day in an approximately 12 x 12 grid by 3 boats in just under 3 hours. Maybe a subset of this grid is used here? If so, the authors should specify how and why they subsampled the data. (Regardless, it should definitely not be 7 boats on line 79.) The authors should also specify what date and time they chose as t_{start}.
- Line 96: Hadjighasem et al. (2017) also used an observed flow (wind velocity in Jupiter's atmosphere).
- Lines 107 – 109: Could the authors supply a reference for the statement about the spiraling FTLE structures?
- Line 116: Lekien and Ross (2010) is an odd choice for a reference here, since that paper specifically focused on using unstructured meshes. I believe Rypina et al (2021) also used the unstructured mesh method rather than dense regularly spaced orthogonal grids. On a related note, I am surprised the authors do not cite the latter paper as the first application of FTLE to an observational drifter dataset.

- Line 129: Singular values are by definition always positive.
- Line 132: Since the rest of the paper refers to 2D flows only, for consistency it might be clearer to stay in 2D here.
- Lines 145 – 146: Arclength was first proposed by Mendoza & Mancho (2010, doi: 10.1103/PhysRevLett.105.038501) as a Lagrangian descriptor. Mancho et al. (2013) explored several Lagrangian descriptors, including arclength.
- Lines 169 – 184: Can the encounter number be appropriately determined from such a limited sample? After just a short time period, as the drifters separate, the water masses encountered by each drifter are not captured by the other drifters. It seems that counting only parcels that were initially close to the target particle (i.e., within the deployment field) produces exactly the opposite effect of capturing mixing potential. Drifters that wandered off on their own are more likely to represent water masses mixing with new waters, no?
  In this context of limited sampling, the authors should discuss what this calculation represents.
- Line 186: Following the nomenclature of Huntley et al. (2015), this is the dilation *rate*, with units of inverse time.
- Lines 195 – 201: LAVD was introduced in the context of a flow field that contains coherent vortices as small subsets of the domain. It is not clear – as the authors note on lines 412 – 414 – how to interpret a vorticity deviation where the entire sample is within a vortex. Without such an interpretation, is the calculation meaningful?
- Line 203: This may be a bit nit-picky, but I would consider spectral clustering a data science technique rather than a dynamical systems technique.
- Line 237: What is the distance condition?
- Line 272: Could the authors explain why they are using t = 0 as the starting point for all three calculations, instead of considering the intervals [0, 0.5], [0.5, 1], and [1,3]? The latter approach would truly split the movement into 3 separate stages.
- Lines 275 – 277: This movie is not included in the supplementary materials (or I couldn't find it).
- Lines 278 – 283: My impression of the early FTLE field (top left of Fig. 3) is that it does not show any coherent structures, especially if one compares it to the top left field of Fig. S3, where even the SPLASH-like calculation shows clear patterns.
  Line 279 asserts generally smaller values in the south, but some of the highest values occur at the southern end, and larger values in the middle latitudes, but some of the lowest values are found there. Maybe if FTLE were plotted as a function of latitude only, such a pattern might arise, but I cannot discern it from the presented evidence.
  I am also not convinced that the drifters that are more tightly clustered at time 0.5 are associated with low FTLE (lines 281 – 283), since some of the highest FTLE values also occur in these tight clusters.

- Line 292: Should this be a more compact group in the <u>southern</u> part of the distribution?
- Line 296: It is not necessarily true that the drifters remained close together; they could have separated and then come back together. (Note that some of these data points are colored yellow in the top row.)
- Line 306: It is very hard to distinguish positive from negative FTLE in the plots. [See also my first comment below on the figures.]
- Line 315: While there is generally an increase from south to north, it is hardly monotonic at any of the analysis times across all longitudes.
- Line 330: What kind of "LCS" were the authors looking for that they did not find here? Of all the fields, Fig. 4 looks the most structured to my eye.
- Line 335 – 336: What is meant by a "slightly stronger variability"? To my eye, the variability looks comparable and, if anything, less for CD; but it depends completely on the chosen colors in the plot… How is the variability in two quantities that have different units like these compared? The standard deviation as a percentage of the mean?
- Lines 338 – 339: Why would one expect CD to be indicative of convergence? It seems like CD is not the right diagnostic for the LCS the authors were hoping to identify here, just based on its definition.
- Line 343: The variability at earlier times looks fairly small. But again, how should this be assessed? Is a range from 0.5 to 3 big or small in this context? Maybe it would help to determine the possible minimum and maximum values achievable over the different time intervals.
- Lines 353 – 354: It is counter-intuitive that filamenting water parcels would be encountering fewer neighbors than parcels remaining more compact. This is solely a function of undersampling of the encountered neighbors. (See comment above for lines 169 – 184.)
- Line 476: It would be helpful to explicitly identify/summarize the dominant LCS being referenced here, since most of the plots did not exhibit much obvious coherence.
- Line 491: Should this be the spectral clusters other than <u>green</u>?
- Lines 535 – 537: Another possibility is that LAVD is the wrong tool for identifying a coherent eddy core from a sample of significant mean vorticity.
- Line 550: I am not a big fan of calling a numerically derived model field the "true" model field, since it is also subject to errors, albeit smaller than in the "SPLASH-like" calculation. Maybe a better choice would be to call it a dense simulation.
- Lines 552 – 556: I think this is overstating the case a bit. The two model simulations showed good agreement of the coherent features in some cases, but not in others (e.g., late time FTLE, early and medium time V_{en}, early and late time clusters).
  I would also argue that, especially in the FTLE field, the model calculations show much greater coherence than those from the observations. I'm not sure what the similarities

in the geometries and types of features are that are being referenced on line 554. The patterns from the model seem quite different from those in the observations (e.g., mostly negative D in the early observations vs. mostly positive D in the early simulations; model fields are generally more coherent, especially at the late time). Lastly, I think the model experiment can show reliability but not robustness.

- It may be worth commenting in the Summary & Discussion on the differences between the frame-dependent and the frame-independent diagnostic fields and how they should be used differently.

References

- The references should be alphabetized.
- Filippi et al. (2021) should be given an 'a' and 'b' to differentiate the two publications.
- Froyland and Padberg-Gehle (2015) is not referenced in the manuscript.
- Essink et al. (2021) is missing from the references.
- It seems that the references for the supplemental materials are included here. If so, Beron-Vera et al. (2015) is also missing.

Figures

- Across all figures (main and supplemental), the choice of colormaps is not ideal. The authors try to compare fields between different rows when they have different scales on the colormaps. For some quantities (e.g. FTLE), the colormap is strangely designed to highlight gradients in very specific narrow bands only. Some of the quantities (FTLE and dilation rate), the distinction of positive and negative values is important, but hard to do with the given colormaps. Thus, I recommend modifying the colormaps such that they vary continuously, except that for FTLE and dilation rate there is a clear break at 0. I also recommend using uniform ranges on the colorbars for quantities that are scaled by or independent of the time interval (FTLE, correlation dimension, dilation rate, LAVD). For the clusters, the number of different colors in the colormap should equal the number of clusters. It would also help if they were chosen to be more easily distinguishable, as currently some of the shades of blue and some of the shades of red are hard to tell apart.
- Aside from providing the colorbars, the right column of the figures is not needed, since the drifter positions are placed in geographic context in Figs. 1 and S2.
- Figs. 7 and 8: It would help to have a reminder in the caption here what the different symbols mean and why some diamonds are left white.

Supplementary materials

- Lines 21 – 24: Doesn't the topography also play a role here, in addition to the density gradients? And wind?

- Line 51: Why were different time intervals chosen in the model than for the observations? Line 47 suggests possibly using 1.5 days instead of 1 day, but why 2 days, and then 4 days?
- Lines 63 – 64: I don't see the similarities in the patterns of high and low values. At early times: observations – scattering of a few high values near the corners of the deployment; model – coherent swath of high values in the center of the domain. At late times: observations – mostly high values throughout the domain, with a region of lower values in the lower half away from the edges; model – mostly low values, with a swath of very high values along the SE edge.
  The magnitude of the values is also not comparable, if one can use the ranges on the colorbars as a guide, especially for negative FTLE values.
- Line 83: It is not clear in what ways the agreement was favorable. E.g., one could not identify the coherent structures at the middle time from the SPLASH-like simulation alone, even if they are weakly reflected in it.
- Line 96: I am not clear what standards of agreement are being used here. The structures in the dense simulation (S-shaped and longitudinal ridges) are not identifiable in the SPLASH-like simulation. Magnitudes are impossible to compare, since different colorbar ranges are used. Maybe a scatter plot of values at the positions of the SPLASH-like data from the two simulations would make that clearer. (Such scatter plots could replace the right columns of Figs. S3 – S8 to directly assess the reliability of the sparse calculations for all the Lagrangian diagnostics.)
- Line 109: The SPLASH-like fields shows patches of increased V_{en} similar to those deemed meaningful in the observations. Why are they considered negligible here?
- Line 125: Are they confined to a smaller area, or are they hidden behind the large markers?
- Line 156: I would not say that the SPLASH-like simulation successfully mapped out the dominant features at 4 days, although it did seem to show a hint of the region of lower LAVD values in the middle.
- Line 168: Fig. S9 shows only 4 clusters for the second version at intermediate times.
- Line 191: Is it meaningful to compare the number of clusters found in the two simulations over different domains? Maybe it would be better to apply the spectral clustering in the dense simulation only to the subdomain also sampled by the SPLASH-like simulation.
- Line 203: I disagree that the clusters look similar in the figures. The only similarity I find is the split around 29N.
- Line 213: I suggest citing the filtering method developed for these kinds of trajectories by Yaremchuk & Coelho (2015, doi: 10.1109/JOE.2014.2353472).

- To draw comparisons with the observations (e.g. "similar magnitudes"), it would be helpful to use the same colormaps and ranges in the supplemental figures as are used in the main manuscript for the corresponding diagnostics.
- Fig. S4: I don't understand why the values shown along the longitudinal filament in the south at the late time don't agree between the two simulations. Coarse sampling should have no impact on pathlength. Is this a plotting artifact?

Language

Below are a few suggestions to fix grammar and spelling issues. Overall, this is one of the most readable manuscripts I have reviewed lately; so these are just minor things.

- Line 68: "and then <u>use</u> observation<u>s</u> of their trajectories"
- Line 91: "parallels were <u>drawn</u> between…"
- Line 123: "coordi<u>nate</u> system
- Line 150: "frame-independen<u>t</u>"
- Line 210: "distance between <u>the</u> i^{th} and j^{th} trajector<u>ies</u>"
- Line 216: "the <u>cluster</u> sizes"
- Line 225: "<u>The</u> LLS methods"
- Lines 226 and 229: The vectors for U and A should be transposed. (They need to be n x 1 and 3 x 1, respectively.)
- Line 240, 247: "LL<u>S</u>"
- Line 262: "Some clustering temporarily occurs" (no "of")
- Lines 338 – 339: "neither" and "nor" should be replaced with "either" and "or"
- Line 463: Strike "the" from "the NOAA's Global Drifter Program"
- Line 474: Strike "respectively" (respective to what?)
- Line 526: "<u>The</u> SPLASH experiment was…"
- Line 542: "In <u>the</u> SPLASH experiment"
- Line 710: "experiment <u>sight</u>"
- Line 731: "(bottom) <u>3</u> days"
- Supp, Line 8: "the site of <u>the</u> SPLASH experiment"
- Supp, Line 29: "as <u>is</u> evident"
- Supp, Line 91: "Fig. <u>S3</u>"
- Supp, Line 117: "blue value<u>s</u>"
- Supp, Line 121: "a bit <u>of</u> red"
- Supp, Line 206: I suggest spelling out STD.
- Supp, Line 208: "V_{en}" instead of "N_{en}"