

Answer to reviewer #1

Thank you for your detailed comments, which highlight a number of fundamental inaccuracies as well as more subtle issues which we have attempted to address in our revised manuscript. We have been particularly conscious on multiple comments made by the reviewers, which pertain to the manuscript as a whole, and made substantial changes throughout the manuscript. In particular, we have clarified the differences between FTLE and LCS, and clearly stated our analysis applies to the FTLE analysis and the reason for why this method was chosen. A special attention was also given to the integration length in the FTLE analysis. Detailed answers for each review comment are given below:

Referee comment:

1. The term "Lagrangian Coherent Structures" (henceforth LCSs) is used prominently in the title, and throughout the article. This term was coined by Haller in the early 2000s, and it is the continuing work of this group that the authors almost exclusively cite in this paper. However, the work of this group now very clearly defines LCSs in very specific ways, which are outlined clearly in Haller (Annu Rev Fluid Mech, 2015). These definitions, stated within the two-dimensional context in which the current paper is situated, relate to attempting to find curves towards/from there is extremal attraction/repulsion -- so-called "hyperbolic LCSs." It is these, and some allied entities, that are LCSs according to Haller and the group's definition, and so citing those papers and not using those definitions does not make sense. Moreover, Haller and collaborators in this and a range of other papers make clear that FTLEs are not necessarily these LCSs, and so prominently using the term LCSs in this paper is incongruous at best. Since the paper deals exclusively with FTLEs (and their averaging -- not even their ridges which some other authors may refer to as a type of LCS), the title and the paper should exclusively use the term FTLEs. Of course, one should position FTLEs in the context of LCSs -- which to a range of other authors represent an entire suite of techniques devised to extract coherent structures (based on differing definitions and intuition) from genuinely unsteady data. There are many recent reviews of this range of different methods for LCSs in this broader context in the literature: Balasuriya et al (Physica D, 2018), Hadjighasem et al (Chaos, 2017), Shadden (in: Transport and Mixing in Laminar Flows, Wiley, 2011, pp59). The authors are advised to consult these in positioning their work (as in line 85 when they say "Various methods have been proposed for LCS detection") and deciding whether the term LCS is actually appropriate.

Response:

We agree that in this manuscript, we are dealing with the variability of FTLE fields which are not directly equivalent to Lagrangian Coherent Structures. Furthermore, we agree that there are some fundamental issues with the FTLE as a tool for LCS detection, and that more complete methods exist. We have clarified that in the revised manuscript, discussed the weaknesses of FTLE with regards to LCS, compared it to other existing methods and adapted the manuscript's title accordingly.

Changes in manuscript:

We have made changes to the manuscript to highlight that we are dealing with FTLE, and not LCS. LCS and their properties are mentioned in the text (), as well as mentioning that FTLE ridges may be used for LCS detection under certain conditions and briefly discuss the weaknesses of this method as an LCS detection method (lines). The manuscript now focuses on FTLE, and we reason our choice to use FTLE analysis as method to investigate statistical properties of flow features pertinent for operational applications (line 44–48)

Referee comment:

2. *The main conclusion seems to be that taking FLTEs and averaging them gives a better diagnostic of "robust" coherent structures. In doing such an averaging, there are some scientific issues related to time-parametrization and that I will come back to in a later point. However, I have some comments with respect to the issue of averaging FTLE fields to smooth out what the authors seem to think of as non-robustness, and thereby extracting robust structures. First, the FTLE fields are definitely associated with a time-of-flow, and since the authors use 24 hours exclusively here, they will be identifying local exponential stretching rates over the past 24 hours. If the authors were *not* interested in stretching over that time-scale but rather over, say, a one-month period, then rather than taking a 24 time-window, they should take a one-month window. This will definitely smear over ephemeral stretching. (By the way, when the authors use the term "ephemeral" this depends on the context. It seems that this means structures which do not persist over a longer time-scale, say a week or two. Within this time-frame, perhaps 24 hours is ephemeral -- but the authors seem to have decided to use 24 hours in the FTLE calculation by choice, and so it's to be expected that features related to 24-hour time scales are what will be revealed in the FTLE field.) Second, the discussion seems to indicate that the authors want to find stationary structures which persist over a longer time. In other words, to think of the velocity field as predominantly steady, with unsteady variations, and the goal would be to find structures which are stable with respect to the unsteadiness (which is assumed smaller). Well, if so, there are better ways to approach this. Rather than calculating FTLE fields from the full unsteady Eulerian data, they can obtain a dominant steady part of the Eulerian velocity. One way to do this would be to average the Eulerian data over an explicit time period -- this has the added advantage of being able to specify a time-scale (the time of averaging) over which the Eulerian field is assumed "mostly" steady, and hence one can ascribe a time-scale to one's conclusions. Another alternative would be to use some smoothing technique -- and again, if using something like a spatial filter associated with an explicit length-scale, one can ascribe a scale (a length-scale in this instance) of "accuracy" of the processed data to enable any conclusions to be stated in relation to that. These methods work directly on the Eulerian velocity field, rather than applying techniques such as FLTEs on data which to all intents and purposes (based on the conclusions reached) seem to be viewed as "noisy." Remove the noise first to avoid amplification of inaccuracies when doing additional computations. Thirdly, if stationary objects are sought, it seems that the authors want to look at the dominant steady component, and if one has a steady velocity field, Lagrangian methods are irrelevant. Lagrangian issues only make sense if there is unsteadiness, and one needs to follow the flow. If steady, simple techniques (drawing streamlines, Okubo-Weiss criterion, etc) on the Eulerian velocity field give perfectly good information on what's going on.*

Response:

We realize that the choice of time (the integration time) must be discussed in the manuscript, and include an presentation of various choices for the integration time in the manuscript (Figure. 3). Furthermore, our objective is not to find structures which exist over longer time periods (e.g. longer than 24 hrs), but to identify predictable flow features. In our averaging in time over consecutive FTLE fields we sought to identify regions where high values in the FTLE appear *frequently*, while noting fact that individual structures will form, drift, and deform over various time scales. This has been clarified in the manuscript. We appreciate your suggestions on how we can approach this issue.

Changes in manuscript:

New analysis of the effect of the choice of integration time, and discussion on how this parameter should be chosen to fit the application in section 3.1.

Referee comment:

3. Based on the conclusions that the authors seem to be reaching, there seems to be an incomplete understanding of what the FTLEs represent. The FTLE field $\sigma_{t_0}^t$ represents a field at time t_0 , associated with the stretching rate experienced by infinitesimal fluid parcels beginning at time t_0 and flowing until time t . This rate is converted to an exponential one via the logarithm, and the time-of-flow $T = |t-t_0|$ is used to time-average this quantity so that it's the average exponential rate over the time period considered. Note that there is absolutely no mention here of anything being a flow barrier, or a "coherent structure." Thus, the FTLE identifies regions in the time t_0 based on stretching rates over the time-of-flow (in this case, 24 hours in the past). When one time-averages the FTLE field, what exactly does that mean? Presumably (and this is not clearly stated), FTLE fields are generated for differing t_0 s but the same time of flow (this is sometimes called time-windowing in the coherent structure community), but then are the authors averaging over t_0 ? (Give an explicit formula for the averaging, so that this is clear.) If so, they are taking scalar fields which are defined over different times t_0 , each of which is associated with flow over a different time-window (t_0-T, t_0) , and averaging them. If this is what is done, it needs to be explained clearly. But then, the interpretation of this needs to be carefully stated, since one is averaging over different initial times, and what one gets cannot be associated with a particular time instance (unlike one calculation of an FTLE field, which gives a field at time t_0). Of course, one might argue that computationally the averaging tells us of how 24-hour motion calculated over several initial times (say 1 January to 31 January 2023) is averaged to give an "average exponential rate of motion in January 2023 when a time-scale of 24 hours is considered for the exponential rate". This would need to be explained, because it's quite awkward and hard to interpret.

Response:

Motivated by this comment, we revised our interpretation of FTLE maxima as transport barriers. We instead expand the introductory section with more mention of previous works that investigate this relationship—and then return to the issue in the Discussion/Conclusions section. We are also more careful about how we define FTLE and how it describes the flow field. We now provide more details on our objective by applying time averaging, We

adopted several of the reviewer's ways of formulating these definitions and interpretations, and like to thank the reviewer for providing clarifications.

Changes in manuscript:

We avoid interpretation of high values in the FTLE field as strict transport barriers, and investigate whether these act as accumulation regions for particle transport. We have added equations showing how the FTLE averages are conducted to make the method description more concise (Eqs. 6 and 7).

Referee comment:

4. The above point related to one aspect of time-parametrization: the t_0 in the field $\sigma_{t_0}^t$. Another time-parametrization issue is the t , and thus the time-of-flow. Everything in this paper has used a time-of-flow of 24 hours. Hence, everything is slaved to this time-scale -- the exponential rate is computed based on time-of-flow for this time-scale. This issue is buried in the paper, with the multitude of plots not mentioning this explicitly. If a time-of-flow of 48 hours were chosen instead, how do things change? Basically, the calculation of an FTLE field explicitly picks out a time-scale, and this is not something which the authors have clarified. The results are explicitly associated with this time-scale, and no other. If the results are to be used in forecasting, why is this the correct time-scale? Or is this method robust to changing the time-scale? How does the time-scale interact with the time associated with the time-averaging as discussed in my previous point? Basically, the issue of TIME (initial plus time-of-flow in calculating the FTLE and the appropriate interpretation of the FTLE field, the times of computation chosen for averaging) is crucial, and needs to be carefully explained, interpreted, and robustness evaluated (if appropriate).

Response:

The choice of 24 hours as integration period is not completely arbitrary but motivated by typical uses of ocean forecasting models. These are decision support tools for search-and-rescue operations, oil-spill modeling, ice-berg trajectory forecasts, and similar trajectory analysis which often require forecasts of a few hours up to a few days. But the comment is definitely warranted, and we now provide an analysis figure with examples of various integration lengths for the FTLE (time-of-flow). The figure indicates that our results are not overly sensitive to the integration period within integration length of 12 hours to 72 hours. But we do see some interesting differences. For example, for much longer integration periods of 15 to 30 days, we see that the FTLE analysis yields distinct linear features in low current velocity regions, especially in the deep basins off the continental slope. In contrast, in the energetic flow regions over the shelf and slope, longer integrations tend to smear out features. Our interpretation is that the ability of the FTLE field to pick up LCSs (in a broad sense) depends on the integration period matching the time scale of the dynamics—which varies depending on the environmental conditions.

We also compare FTLE analysis based on long integration periods with the time-average of several short integration periods and see distinct differences in the two approaches.

Whereas the time-average provides a description of regions that are typically abundant with FTLE ridges, the long-time integrations do not distinguish areas in the high-velocity region

but instead allow to better characterize low-velocity regions, which are less pertinent to time-critical contingency modeling.

We do not take an opinion on right or wrong time integration, but in the revised manuscript we discuss its impact on the analysis, and provide a view on how an appropriate time period may be selected depending on the application in focus.

Changes in manuscript:

As our study is motivated by typical ocean forecasting models, we have not changed the integration time in our results. We have been more clear about this motivation in the text (), and also provided a brief analysis of how the FTLE field changes based on the integration time and discussed the impact of this analysis on our results, included in section 3.1.

Referee comment:

5. *Returning to the definition of the FTLE mentioned earlier. It represents a field at time t_0 , associated with the average exponential rate over flow from time t_0 to t . Note that there is absolutely no mention here of anything being a flow barrier, or a "coherent structure." Yes, there are papers in the literature which seem to indicate such a connection, but the reality is that it is unjustified. The early papers in this area used STEADY toy models which had saddle points with one-dimensional stable and unstable manifolds emanating from them, and since these manifolds are associated with exponential decay rates, came up with the idea that FTLE ridges had something to do with stable and unstable manifolds. And these manifolds were flow barriers in some way. However, this argument does not hold water, since there are examples such as in Haller (Physica D, 2011) which show that the stable/unstable manifold interpretation sometimes fails even in infinite-time flow. (And, getting back to a previous point related to hyperbolic LCSs, the fact that repelling/attracting do not necessarily occur as expected are also shown.) Real data is much worse: it is unsteady, and finite-time. Finite-time aspects are awkward for inferring exponential rates of growth, since any function over a finite-time can be bounded by an exponential. Unsteadiness is yet another problem, because (even in the infinite-time context) saddle points generalize to hyperbolic trajectories, and their stable/unstable manifolds move around (another reason why time-averaging is questionable). Furthermore, it is not clear what an "FTLE ridge" is -- one never gets a genuinely one-dimensional curve which is well-defined, but rather gets regions of larger FTLE values. Balasuriya et al (J Fluid Mech, 2016) provide an assessment and interpretation of what the FTLE means, with an emphasis on fluid motion, which helps understand these issues. In particular, for finite-time, unsteady data sets, using FTLEs and their ridges cannot easily reach conclusions regarding flow barriers and coherence. FTLEs explicitly look at exponential growth rates with respect to the time-of-flow considered, and that's about it. So when one takes FTLE fields, as done in this paper, and tries to reach conclusions regarding coherent structures such as eddies (as the authors do towards the end of the paper), this needs to be treated with suspicion, because it is not on any firm scientific grounds. The idea that eddies can be demarcated by FTLE ridges -- notwithstanding my earlier comments on finite-time and unsteadiness -- may go back to work in the 1990s (such as del-Castillo-Negrete, Knobloch, Pierrehumbert) who analyzed perturbed toy models. In these cases, the unperturbed models were explicit and steady, and had saddle points with stable and unstable manifolds. In some cases, the geometry was such that these manifolds encircled an eddy. Since the manifolds may be*

discoverable using FTLE ridges (again subject to various caveats), in such cases ONLY, one might think of an eddy as being found using FTLE fields. However the interior of the eddy does NOT typically contain exponential stretching, and hence FTLE fields by themselves cannot be used to reliably identify eddies as the authors here are doing in their later figures. Actually, in the standard fluid-mechanical dichotomy between stretching and rotation, the eddies have the opposite of stretching, and thus the FTLE is exactly the wrong thing to use (Okubo-Weiss and related criteria may help; however as noted by many authors unsteadiness is a problem in using such Eulerian characteristics). Basically, statements such as "two additional regions are identified and considered as robust in Fig.7 " (line 211 in the paper) are, in this vein, questionable. Indeed, high FTLE values indicate greater separation, and hence LESS certainty!

Response:

As outlined in relation to comment #1 and #2, we revised the manuscript to clarify that FTLEs and LCSs are not the same thing, and focus on the properties of the FTLE only as this is what we have investigated. We thank the reviewer for providing the paper by Balasuriya et al (J Fluid Mech, 2016), which we refer to when interpreting FTLEs. We use the term "FTLE ridges" loosely after a brief definition as "elongated areas of high FTLE values" (line 34-35) to discuss features formed by locally high FTLE values which can be seen in our figures of the FTLE field. However, we agree that we don't get a genuine and well-defined 1D curve in the FTLE field, and will make changes in the manuscript to reflect this, as our analysis is centered around a statistical analysis of FTLE fields to characterise their robustness and persistence.

We also agree that it is unjustified to identify eddies through the FTLE, which was not intended as the main focus in this manuscript, hence we removed this interpretation. We see that this point is not clear, and instead it seems that we are attempting to use the FTLE to identify eddies. In a revised manuscript, we do not see a need to focus on analysis of eddies.

Changes in manuscript:

Changes have been made to address the distinction between FTLE and LCS in the introduction and throughout the manuscript. Since we don't conduct any ridge detection in our manuscript, and the term "FTLE ridge" appears to be a common and well defined term in the scientific community, we have defined "FTLE ridge" as "elongated areas of high FTLE values" (line 35). Furthermore, we heed the advice from the reviewer about eddies. Eddies are important structures of the flow field, and their mention has not been completely removed from the manuscript, but we have made changes to the manuscript so that it does not seem like we are trying to identify eddies through FTLE

Referee comment:

6. The impression given throughout is that the authors are examining robustness of LCSs. I've already talked about why it's actually FTLEs and not LCSs, but in this point I want to question whether robustness is the right thing. There are many recent papers which examine robustness of FTLE fields (Balasuriya, J Comp Dyn, 2020; Guo et al, IEEE Trans Visual Comp Graphics, 2016; Raben, Exp Fluids, 2014), but this paper is not one of them.

(There are a few other papers, from oceanographic situations, which the authors speak to in lines 339-344.) By "robustness," the authors seem to mean smearing over small time scale motion, in other words looking for entities which persist over longer times. This needs to be made clear throughout the manuscript. (I've talked about the time-parametrization issue previously; to smear over smaller time scales, one needs to simply choose appropriate times for the FTLE which are relevant to what one is looking for.) In Section 4.3, also, the word "uncertainty" is questionable because of this same reason -- the authors have not calculated any uncertainty (i.e., have not evaluated anything to do with uncertainty in the input data).

Response:

We would like stress that our recommendation is a distinction between 'robustness' and 'persistence' of FTLE fields as calculated from ocean model forecasts. Much of the discussion in literature regards what we identify as persistence, i.e. time variability. In our discussion, robustness refers to similarities of FTLE features across an ensemble of model flow realizations. We are aware of the studies which address uncertainties in FTLE computations, and have included them in our manuscript. Our goal is to investigate this issue from an operational perspective and investigate how an FTLE field manifests itself in an ensemble prediction system, directly addressing uncertainty in forecasts.

Clarify this distinction and put less emphasis on time variability as this subject has been more extensively been addressed in other studies and needs no repetition. Furthermore, the reviewer suggested including an analysis of current velocities. We have chosen to not go into details in this analysis, but included figures showing velocity fields and their averages, as well as particle trajectories in different ensemble members.

Changes in manuscript:

The revised manuscript provides an early definition of robustness and persistence (lines 66-70), and what motivates us to investigate robustness. New figures 2 and 8 show particle trajectories. Figures 4, 5 and 6 now show time and ensemble averaged velocity fields. We point the reader to a detailed analysis of particle trajectories in line (110).

Referee comment:

7. There are several statements around Equation (2) and (3) which are incorrect. The statements in line 101 are all incorrect: Δx is not the final location, but x is, and the (1,1) term in Equation (2) then represents the partial derivatives of x with respect to x_0 , for example. The integral bounds in (3) are unclear and inconsistent with the previous equation. Presumably the authors mean something like $x(t) = x_0 + \int_{t_0}^t u(x(\tau), y(\tau)) d\tau$, and then one needs to also clarify that $x(\tau), y(\tau)$ are the evolving trajectory locations. However, this integral formulation may not be the most natural. The differential form with $\dot{x} = u(x(t), y(t))$, $\dot{y} = v(x(t), y(t))$ and the initial condition $(x(t_0), y(t_0)) = (x_0, y_0)$ connects better with the discussion, perhaps.

Response and changes in the manuscript

Thank you for spotting the errors around Eqs. 2 and 3. We have corrected the equations in section 2.3 according to the reviewers comments.

Referee comment:

8. The discussion at the end of Section 2.3 is fraught. What is a "2D curve" [line 126]? The statement that "averaging smooths the ridges into fields of attraction/repulsion" [line 128] is incorrect because, as mentioned previously, claiming that FTLEs have anything to do with attraction/repulsion is questionable. "Making these ridges more certain" [line 130] relates to an earlier comment that what is being done here has nothing to do with robustness or certainty. The authors comment that the time- and ensemble-averaged FTLE fields are not transport barriers (which is correct -- but neither is the FTLE field -- and again it's not clear to me how a field can be a barrier), but then talk about things being barriers over larger regions. Figure 2 is strange. One does not usually get FTLE ridges (even in idealized steady toy models in 2D) which intersect and pile on each other like this. Intersections in such cases occur at saddle points. If unsteady, intersections that one gets, analogous to intersections between stable and unstable manifolds, can relate to chaotic motion -- but these are between forward-time FTLE ridges and backwards-time FTLE ridges, rather than self-intersections within one of these.

Response:

We acknowledge that we have been too quick on equating high values in the FTLE field to attraction or repulsion, and that we used the term "transport barriers" too loosely. We have in fact only looked at the FTLE field, and not conducted any ridge detection or investigated further criteria from e.g. Farazmand and Haller (2012) to distinguish any potential LCS. Furthermore, as the second reviewer pointed out, strong values in the FTLE field may be produced by horizontal velocity shear, which does not result in attraction/repulsion. We have undertaken a major revision of the manuscript with regard to interpretation of FTLEs, and avoid interpretation as "transport barriers". We do, however, believe there is such a thing as a dynamical transport barrier, specifically related to the strong potential vorticity (PV) gradient associated with a steep continental slope (as we have in our domain). A PV barrier is not impenetrable, but it does inhibit tracers and particles - as documented throughout the dynamical literature (we will add references). And we do believe that the high FTLE values over the continental slope, also when averaging over time or over the ensemble, do in fact arise from the strong PV gradient (or 'PV barrier' if one allows a more loose use of the term). In the revised manuscript, we will be more careful in explaining how we use the term, aware that different parts of the research community might be using various degrees of rigor in their definitions.

We agree that Fig. 2 is misleading and have removed it from the manuscript. Instead, we present examples of FTLE ensemble realisations from the ocean model along with the ensemble average and a time average around this situation in Fig. 8.

Changes in manuscript:

We have made it more clear that FTLE is not the same as LCS, and focused on discussing FTLEs. We also added a discussion about potential vorticity and FTLEs produced by horizontal velocity shear. Fig. 2 has been removed from the manuscript. New Figure 8 shows example of ensemble realisations of FTLE with particle positions along with ensemble average and time average.

Referee comment:

9. *The fact that a large standard deviation of the time-averaged FTLE usually is close to where the FTLE is large [line 169-171] is no surprise. Large FTLE relates to larger uncertainty in the results, because any errors (based on interpolation to a grid, say) increase exponentially. Hence, the values one assigns to the FTLE tend to be less certain. This issue is well-known, and described in some of the papers I've mentioned previously on robustness of FTLEs.*

Response:

Thank you for this remark. This will be considered when reworking the manuscript.

Changes in manuscript:

After closer consideration, we decided that the standard deviation did not add much to our analysis, and have removed all figures of standard deviation.

Referee comment:

10. *For the discussion on Section 3.4, I can't quite understand what the k in the figures is. Since in 2D, one has two wavenumbers -- say l and m , associated with the eastward and northwards coordinates. I don't understand the discussion [lines 230 onwards] about averaging over rows and columns (over l and m). Is $k = \sqrt{l^2 + m^2}$? The spectral plots in Figure 9 are used to infer robustness in some way, based on the fact that one gets decay with k in the bottom figures, say. Any smoothing process will of course get rid of the smaller wavenumbers. Similarly one expects more smoothness when averaged over more and more days.*

Response:

The procedure for producing Fig. 9 has been poorly described in the manuscript. The transformation of the FTLE field to Fourier space has been done using the 2D discrete cosine transformation method. As pointed out, this results in two wavenumbers, m and l . These are combined to create a radial wavenumber $k = \sqrt{m^2 + l^2}$. We intend to revise this section, include equations and references, and clarify it in the discussion.

Changes in manuscript:

Equations and references are included in section 3.4. Text in the section has been revised to explain how what we do and how we obtain the Fig. 9 (now Fig. 7) more clearly.

Answer to reviewer #2

Thank you for your detailed comments, which highlight a number of fundamental inaccuracies as well as more subtle issues which we have attempted to address in our revised manuscript. We have been particularly conscious on multiple comments made by the reviewers, which pertain to the manuscript as a whole, and made substantial changes throughout the manuscript. In particular, we have clarified the differences between FTLE and LCS, and clearly stated our analysis applies to the FTLE analysis and the reason for why this method was chosen. However, we realize that the quasi-steady LCS method might be more proper, and while we did not compute them here, we have given the method more attention in the discussion. A special attention was also given to the integration length in the FTLE analysis. Detailed answers for each review comment are given below:

Referee comment:

20 I would suggest you emphasize the sensitivity of trajectories and not the sensitivity of the velocity because even with a hypothetically perfect velocity, a small error in a trajectory's initial position can grow exponentially into large errors. The idea of your paper is to introduce perturbations in the velocity and measure how trajectories respond. Note the interest is in the trajectory response as visualized through LCS given a velocity ensemble spread. The main concern is trajectory uncertainty, even if explored in terms of velocity uncertainty.

Response:

This is a good point and we will adapt this change in wording.

Changes in manuscript:

Wording has been adapted (lines 50-53).

Referee comment:

In 35 you mention:

Previous studies often discuss the LCS methodology and their practical applications, but rarely touch upon the topic of LCS estimates being inherently affected by uncertainties in the velocity fields they aim to describe. Furthermore, short-lived flow features constantly develop, drift, and dissipate in real oceanic flow (Chen and Han, 2019). Given their time-dependency, LCSs might appear and disappear just as quickly. This brings up two important questions: (1) Given the velocity field uncertainty, how robust, i.e. predictable, are LCSs derived from ocean models at a particular time?; (2) Given their time-dependency, how persistent are LCS in ephemeral flows?

I don't think robust can be equated with predictable. Robust in your study means that different realizations of a simulation (i.e. similar simulations) result in the same LCS. There is no predictive capability (i.e. estimates of future information based on past information) in this analysis

Response:

We agree that predictability does not directly follow from robustness. This is only be the case if the underlying circulation model is proven to have both predictive skill on both the time and the spatial scales of interest, and if the ensemble spread exhibits reliability. Both criteria are valid to a limited extent in terms of surface currents, which has been investigated by Idzanovic et al. (2023) for the model system that is used here. Our study is interested in whether the EPS yields similar FTLE fields at a specific time, and does not focus on their predictability. However, when using the method in concrete applications one must confirm that the underlying EPS has reliability in the ensemble spread for surface currents. This point is clarified in the manuscript.

Changes in manuscript:

Wording has been changed in the manuscript and this point has been clarified. A statement on the requirement of reliability has been added to section 4.5 §3.

Referee comment:

The short-lived structures you mention are not a problem, or even interesting, as it is straightforward to filter them and find the prominent deformation patterns, without the need to average see e.g. Olascoaga & Haller (2012; <https://doi.org/10.1073/pnas.111857410>) or Kunz et al (<https://doi.org/10.5194/egusphere-2024-1215>) that discusses the importance of persistence when it comes to attracting hyperbolic patterns and the lack of meaningful influence with short-lived structures (in particular while hyperbolic structures are forming or decaying). These are results without ensembles or any type of averaging. In particular, Olascoaga & Haller (2012) get rid of short-lived LCS by increasing the integration time T to 15 days. Thus, the choice of $T=1$ day in your study raises the question of how do your results depend on your choice of T ? Are the transient FTLE features that you filter through averaging unnecessarily increased by this choice? Wouldn't it be better to use a longer integration time to filter those features instead of time averaging? Also, as I will mention below, there are several papers showing 1) how to find persistent (or quasi-steady) LCS and 2) that persistent LCS are ubiquitous and meaningful. It is true however that we do want to be able to discern which are short-lived and not meaningful, and there is more than one way to get there.

Response:

Reviewer #1 raised a similar concern about the integration period. The choice of 24 hours as integration period is not completely arbitrary but motivated by typical uses of ocean forecasting models. These are decision support tools for search-and-rescue operations, oil-spill modeling, ice-berg trajectory forecasts, and similar trajectory analysis which often require forecasts of a few hours up to a few days. But the comment is definitely warranted, and now provide an analysis figure with examples of various integration lengths for the FTLE (time-of-flow). The figure indicates that our results are not overly sensitive to the integration period within integration length of 12 hours to 72 hours. But we do see some interesting differences. For example, for much longer integration periods of 15 to 30 days, we see that the FTLE analysis yields distinct linear features in low current velocity regions, especially in the deep basins off the continental slope. In contrast, in the energetic flow regions over the shelf and slope, longer integrations tend to smear out features. Our interpretation is that the

ability of the FTLE field to pick up LCSs (in a broad sense) depends on the integration period matching the time scale of the dynamics—which varies depending on the environmental conditions.

We also compare FTLE analysis based on long integration periods with the time-average of several short integration periods and see distinct differences in the two approaches.

Whereas the time-average provides a description of regions that are typically abundant with FTLE ridges, the long-time integrations do not distinguish areas in the high-velocity region but instead allow to better characterize low-velocity regions, which are less pertinent to time-critical contingency modeling.

We do not take an opinion on right or wrong time integration, discuss the impact of this analysis in the revised manuscript, and provide a view on how an appropriate time period may be selected depending on the application in focus.

Changes in manuscript:

As our study is motivated by typical applications in operational oceanography, we have not changed the integration time in our results. We have been more clear about this motivation in the text (lines 143-147), and also provided a brief analysis of how the FTLE field changes based on the integration time and discussed the impact of this analysis on our results. This analysis added into section 3.1.

Referee comment:

Schematic 2 is not correct, the average of any number of zeros is still zero, i.e. regions where FTLE is zero in the left side of the schematic should also be zero on the right. Notice that in the caption of Figure 2 you introduce a concept that is not mentioned, or used, in any other part of the paper and that is “the average region covered by them”. There must be a better way to convey the idea you want to convey.

Response:

We agree that Fig. 2 is misleading and that structures like these do not appear in nature. This is also a concern that was raised by reviewer #1. The idea behind Fig. 2 was to show what FTLEs might look like over the ensemble dimension, where each drawn line represents FTLE ridges from different ensemble members. This is not shown clearly enough in neither the figure nor the text, and has been removed.

Changes in manuscript:

Fig. 2 has been removed. New Figure 8 shows example of ensemble realisations of FTLE with particle positions along with ensemble average and time average.

Referee comment:

40 should be Gulf Stream

Response:

This will be fixed.

Changes in manuscript:

Fixed.

Referee comment:

40 The paper by Badza et al 2023 does not present reliable results (this is a paper that should have been rejected in my view) because:

They use a stochastic differential equation for the velocity to measure the robustness of LCS methods to noise. This is a big mistake. The mathematical theory of variational LCS explicitly states the results are only valid for a deterministic velocity, i.e. the results only hold for the typical ordinary differential equation $dx/dt = v(x(t),t)$. This is a very basic, yet fundamental mistake that renders their results meaningless.

Even if the theory of hyperbolic LCS were to hold for a stochastic vector field, a stochastic component is not representative of the uncertainty typically encountered in a geophysical velocity field, neither simulated nor observed. An ensemble of simulations is a much better choice for uncertainty.

In their Gulf Stream case, they allow for a very long integration time (three months) while computing LCS within a limited spatial region, Thus the results are plagued by fictitious boundary effects, as is evident from their figures. A simple computation shows the inadequacy of their choice: Their domain is 30 degrees wide (note the flow is mainly west to east in their domain). That means their domain is less than $30 \times 111 = 3330$ km wide. Yet their integration time is 90 days, which means you would only need a velocity of 37 km/day (0.43 m/s) to traverse the whole domain, from west to east. The Gulf Stream commonly reaches a velocity of over 150 km/day (above 1.5 m/s). Indeed, boundary effects in their results are apparent, and they mention it themselves: "Most of these streaks appear to look like diagonal lines, which is likely attributable again to the exodus of particles over the large period of flow considered." They also mention in their discussion that: "As with most of the previously discussed methods, this can be attributed to the exodus of particles from the domain over our 90 day flow period." Even without the two mistakes mentioned above, there is not much that can be learned from results plagued by unphysical boundary effects.

In your study, Badza et al. 2023 are cited to say that hyperbolic LCS detection is not reliable. However, as explained above their conclusion is meaningless. I therefore, as a reviewer, make the extraordinary suggestion that you delete, or at least adequately discuss, any sentence citing the Badza et al. paper, to avoid amplifying misleading results. The other papers you cite such as Harrison & Glatzmaier are better and are adequate for the point you wish to make regarding FTLE. In particular, the representation of uncertainty they choose is realistic and does not involve a fundamental dynamical-systems mistake.

Response:

Thank you for this comment. We did not realize this issue with Badza et al. (2023), and removed the citation from our manuscript.

Changes in manuscript:

The Badza et al. (2023) citation and discussion around it has been removed from our manuscript. We instead conducted more literature research and found other articles to reference to.

Referee comment:

55 what do you mean by dynamically active shelf region? Is there such a thing as a dynamically inactive sea? Either explain clearly the idea you want to convey or delete statements that don't add useful information, yet leave the reader wondering.

Response:

What we meant here is that the shelf and slope region has more variability due to the presence of a complex coastline and the juxtaposition of different water masses (setting up frontal zones) that generate various forms of flow instability. We have adapted the wording in the manuscript.

Changes in manuscript:

This has been fixed (line 71).

Referee comment:

In the caption of Figure 1 you mention Moskstraumen has been indicated by an arrow, consider mentioning in the text what is this region. why is it important?

Response:

The velocity field through Moskstraumen is highly tidally dependent, and we investigated whether this will yield periodic and thus predictable FTLE fields. Figures of this are not included in the manuscript, but are mentioned in the text (line 207-211).

Changes in manuscript:

We now mention Moskstraumen in lines 89-90.

Referee comment:

100 Although it is true that fluid parcels need to be advected, Equation 3 is not an accurate description of how Equation 2 is computed. It needs to be clear that you are time-integrating the velocity along a path which is not the same as integrating the velocity with respect to time at a fixed location, as your equation suggests. Importantly, you need to integrate two trajectories to be able to compute the distance ∂x , and it is not enough to just integrate the velocity as your equation reads.

Response:

Thank you for pointing this out. We have updated the equation and text.

Changes in manuscript:

The Eq. 3 (now Eq. 1) and corresponding text has been updated and moved to the beginning of section 2.3. Lines (116-121)

Referee comment:

105 Although deformation is indeed given by the singular values of the Jacobian of a Flow map, there is no such information as a “speed of deformation” embedded in the Flow map (note speed has units distance/time).

Response:

This will be fixed.

Changes in manuscript:

Fixed in line 135.

Referee comment:

110 there is nothing to show, that is the definition of FTLE.

Response:

This will be fixed.

Changes in manuscript:

Fixed in line 137.

Referee comment:

115 “Largest FTLE = LCS” is not true. This needs to be explained in detail throughout the paper so that your conclusions are not misleading. See my comments about strong FTLE produced by large horizontal shear in coastal regions in what follows.

Response:

The other reviewer also commented on this. We realize that the largest FTLE does not necessarily imply an LCS and that there are additional criteria which an FTLE ridge must satisfy to qualify as an LCS approximation. We have clarified this difference between FTLE and LCS in the introductory section, and made remarks about the weaknesses of FTLE.

Changes in manuscript:

We have been more mindful about equating FTLE to LCS, and made the distinction between these two more clear. We have made changes throughout the manuscript reflecting that we are working with FTLE and not LCS throughout the manuscript. Weaknesses of the FTLE are mentioned in lines 43-45 and discussed section 4.1.

Referee comment:

125 “infinitesimally thin” is an unusual description. Although the width of a line within a plane indeed has measure zero, just like the width of a point within a line has measure zero, it is better to just say co-dimension 1 and leave it at that. In addition, co-dimension 1 is true for proper LCS, yet FTLE ridges tend to be coarse (as you suggest in your Figure 2 and other parts of the paper) and therefore your “infinitesimal” description is confusing. Best to omit this part.

Response:

This is a good point, and we agree that saying “co-dimension 1” is better phrasing than “infinitesimally thin”. Furthermore, we agree that this only applies to LCS, whereas the FTLE is just a field of values, and as per the previous comment, we intend to make the distinction between FTLE and LCS more clear here as well.

Changes in manuscript:

Section 2.3 now focuses directly on FTLE instead of FTLE as a tool for LCS detection.

Referee comment:

Line 152, you mention larger FTLE at initial time is due to a large velocity gradient, this suggests high FTLE is produced by the velocity horizontal shear in which case it is not an LCS. Also, if it is an LCS then attraction rather than accumulation would be better.

Response:

Thank you for this remark. The note on a horizontal velocity shear is a good point, which we did not consider in our manuscript. While it is true that a horizontal velocity shear (shear dispersion) would yield large FTLE, this does not cause attraction/repulsion, and is therefore not an LCS. This brings us again back to the point that FTLE does not necessarily equate to LCS, which is a distinction that was lacking in our manuscript. We now discuss this in the revised manuscript.

Changes in manuscript:

Discussions about FTLE produced by horizontal velocity shear are now included in lines 43-45, 206, 287-299 and 252.

Referee comment:

Line 160 you claim longer averaging periods effectively decrease variability. Although this makes sense intuitively, it is hard for me to see this by just looking at the figures. For example, there does not seem to be a large difference between the 7-day standard deviation and the 28-day one. Can you quantify this further?

Response:

The idea here is that both the average and standard deviation appear to be concentrated with large values at particular locations (e.g. along the continental slope) for shorter time periods, but spreads out more throughout the domain for longer periods of time. Fig. 4

shows an example of this, whereas Fig. 9 intends to quantify this decrease in variability. We agree that this can be hard to see from Fig. 4, so this figure will be remade and we will consider presenting it alongside Fig. 9.

The idea here was the both the average and standard deviation appeared to be concentrated with large values at particular locations (e.g. along the continental slope) for shorter time periods, but spread out more throughout the domain for longer periods of time. Fig. 4 showed an example of this, whereas Fig. 9 intended to quantify this decrease in variability. However, we did not feel that the short term average or the standard deviation added much value to our analysis, and have been removed.

Changes in manuscript:

Fig. 4 has been replaced by the new Fig. 4. After consideration, we decided that the standard deviation did not add much to the analysis, and figures of standard deviation have been removed. Furthermore, we no longer show time averages for shorter time periods, but kept Fig. 9 (now Fig. 7) to quantify the variability.

Referee comment:

Lines 167-168 can you discuss further the relation between a persistent current and persistent FTLE? Why is not surprising that they co-locate? Is it due to velocity shear?

Response:

Thanks for the reminder. We are convinced that the underlying cause of both persistence and robustness is the strong potential vorticity (PV) gradient set up by the steep continental slope. A steady current and recurring FTLEs are two of many results of this PV gradient. But, in relation to your cautionary point on shear dispersion, the velocity shear along the continental slope is probably a large contributor to FTLE formation here.

Changes in manuscript:

We added a discussion about the persistent PV gradient set up by the steep continental slope in section 4.1, where we argue that this will have an effect on particle transport. Here, we discuss that this PV gradient likely is one of the reasons that persistent FTLEs are seen along the continental slope (lines 294-300). Velocity shear is also mentioned in lines 43-45, 206, 287-299 and 252.

Referee comment:

Figure 6, some colors are saturated (especially e and f) so we can't get a sense of how large the values are, also the mean and the std deviation are not too far off, consider plotting them with the same scale (say 0.01 to 0.07 or whatever is needed so colors are not saturated over large regions). This will aid comparisons between mean and standard deviation.

Response:

We decided that the standard deviation did not add much value to our discussion.

Changes in manuscript:

This Fig. 6 has been replaced by the new Fig. 6.

Referee comment:

172-173 It should be easy enough to test whether it is truly a transport barrier. How about releasing synthetic drifters on both sides of the barrier candidate and testing this directly? It would be nice to see results from individual members and some trajectory ensemble averages.

Response:

This is a very nice suggestion which would aid our discussion. We started Lagrangian particle calculations for this purpose, but ended up not including results indicating transport barriers. Instead we include references where this experiment has been conducted. We did however add a Figs. 2 and 8, showing an example of particle trajectories from individual ensemble members.

Changes in manuscript:

We added Fig. 2, which shows a simple example of how particles are advected differently in different ensemble members, and point the reader to de Aguiar et al (2023) (line 110), where the authors conduct a detailed analysis of particle advection in different Barents 2.5 EPS ensemble members. Also added Fig. 8 which shows a case where particles from different ensemble realizations are attracted to an ensemble averaged FTLE feature.

Referee comment:

Figure 9a, the legends for daily winter and summer seem the same color. Also, why do the spectra for ensemble members (c and d) seem to decrease monotonically from member 1 at the top to the last member on the bottom?

Response:

The spectral graphs are computed after first averaging over a set number of ensemble members. That is, the graph for “ensemble members = 1” is just showing the spectral variance from one ensemble member, but the graph for “ensemble members = 24” shows the spectral variance taken after first averaging all 24 ensemble members. The averaging effectively decreases variability over the domain.

Changes in manuscript:

The legends have been fixed. We expanded upon the method used to create Fig. 9 (now Fig. 7), and the text in section 3.4 is now more clear.

Referee comment:

Line 241 can you describe the dependence on the averaged members when only a few members are averaged?

Response:

Here we pick the members at random, and there is a possibility that any member might deviate largely from the other members. When only picking a few members, e.g. four, there is a possibility that one of those might be largely different from the other three members. This can have a large impact on the calculated variability between these four members, and the results can differ if all four members are more similar. The impact on the results from one deviating member will be larger when few members are considered as opposed to when many members are considered.

Changes in manuscript:

We added a note about the dependency of the ensemble average on the specific members included for few members in lines 258-260.

Referee comment:

Line 242, it would be nice to see the three regions used for the spectra, as you mention, some circulation patterns are highly predictable for example circulation along a slope tends to be quite predictable along large portions of the slope.

Response:

Good point. This has been included.

Changes in manuscript:

A figure showing the three regions used for the computation has been added to Fig. 7.

Referee comment:

Line 243 and 252, could it be that FTLE seems to be more robust than persistent due to your choice of $T=1\text{day}$? Short T should be expected to result in more transient features relative to longer integration times. See for example the papers cited in the comments above for lines starting at 35.

Response:

This could definitely be the case, and could be investigated more. As stated previously, we are motivated by typical applications which are on these time scales. However, we have now added a brief analysis of the integration time.

Changes in manuscript:

We ended up not pursuing this topic into more detail, but have made it clear that we are motivated by short-term applications which motivates our choice of T . An analysis of the integration time has been included in section 3.1.

Referee comment:

Line 276, a repelling and attracting LCS cannot be parallel at the same location, if you go back to Dong et al you will see they describe an attracting LCS

Response:

Thank you, this was an error on our side.

Changes in manuscript:

This has been fixed.

Referee comment:

277-278 and again we have the issue of T being relatively short at 1 day.

Response:

This has been discussed (as outlined above).

Changes in manuscript:

A discussion around the choice of T has been added in section 3.1.

Referee comment:

283 You mention other methods for detecting persistent LCS could yield more nuanced results. You also cite (line 343) a paper by Gouveia et al to say that large-scale features give rise to quasi-steady LCS. If you read the Gouveia et al paper carefully you will see they use a method to find quasi-steady LCS that was published in 2018 (<https://doi.org/10.1038/s41598-018-23121-y>). This later paper has over 40 citations according to Google Scholar, suggesting that many other studies are using that method to extract quasi-steady LCS. This topic is directly relevant to your study, so it seems your literature review is lacking. As you will see throughout my comments, the use of FTLE is a problematic issue that keeps coming up. The methodology published in 2018, and used by Gouveia et al, does not rely on FTLE, although there is some averaging. The difference in approach suggests a worthwhile discussion regarding the differences between that approach and your approach, including the strengths and weaknesses of each method.

Response:

While other methods provide better means to identify LCS as linear features, our reason for using FTLE fields is to obtain a gridded spatial description of Lagrangian transport characteristics that can be analyzed using simple statistical methods, e.g. to calculate a mean and standard deviation of FTLE fields. In addition, we see the FTLE analysis as a practical tool for use in operational applications (e.g. contingency modeling), and we want to highlight the potential use of FTLE analysis in operational oceanography. Nevertheless, we agree that there are problems with the FTLE, and are familiar with better and more modern approaches for detecting LCSs. We have been more careful about the connection between LCS and FTLE, and provide a more detailed discussion on shortcomings of FTLE analysis as opposed to methods that identify LCS as linear features, such as quasi-steady LCS.

Changes in manuscript:

We now argue why we use FTLE in the introduction (lines 45-48), but try to highlight the problems of FTLE as an LCS detection tool (lines 43-45). The manuscript now focuses on FTLE themselves instead of LCS detected from FTLE.

Referee comment:

288 it is also possible that a very strong FTLE shows up in the average, even if it does not persist much in time, especially if it recurs.

Response:

This is a good point..

Changes in manuscript:

We added a note about this in lines 330-331.

Referee comment:

358-359 Strong, persistent FTLE can be caused by persistent horizontal shear in which case it would not be indicative of LCS. Strong persistent FTLE can be expected in many coastal regions to be caused by horizontal shear. High FTLE next to the coastline, as in Fig. 5 during the summer or around 69.4N in the winter, for example, should be particularly suspect. You need to clarify throughout your paper that strong FTLE may be caused by horizontal shear, in which case FTLE is NOT indicative of an LCS, and mention that horizontal shear can be persistently high at certain locations such as a coastline. These locations, according to your suggestion, would have persistent LCS due to the persistent high FTLE, yet FTLE is not indicative of LCS if it is solely due to shear.

Response:

Clearly a good point. As outlined above, we have been more careful about the distinction between FTLE and LCS, and specifically mention the case of shear dispersion.

Changes in manuscript:

A discussion about horizontal velocity shear has been added as outlined above.

Referee comment:

It is not enough to mention FTLE ridges only approximate LCS and to reference where to find the distinction between the two (line 144), because this distinction directly impacts the interpretation of our results, as has been explained above.

Response and changes in manuscript:

We have made changes in the manuscript to clearly distinguish between FTLE and LCS. Furthermore, we have highlighted issues with the FTLE as a tool for LCS detection and discuss other LCS detection methods from literature.

Referee comment:

382 *“...by combining LCS analysis with ensemble prediction methods.” The 2018 method to find quasi-steady LCS mentioned above should be discussed in this context as well. Can that method be used to find robust features in operational forecasting? Or is it complementary information to the ensemble methods you propose? Or are these two methods for differing purposes? Can you expand on how to use these methods to detect robust or persistent LCS in terms of operational oceanography? How concretely can these methods be applied? Can you suggest step-by-step instructions on how to implement these methods in an operational application? Can you give an example of how they have been used or can be used in operational oceanography? How feasible, useful, and accessible are these methods in operational oceanography?*

Response:

We have expand on the possible operational uses/applications of FTLEs in an ensemble prediction system. As for the 2018 method to find quasi-steady LCSs, we do not understand it well enough yet to be able to ascertain exactly how this method would respond to an ensemble average. However, as the 2018 paper handles time averages, we see the potential from an operational perspective of combining ensemble averaging with quasi-steady LCS to find robust features. We have not attempted to compute these in this study, but have discussed the two approaches in a revised manuscript.

Changes in manuscript:

We have not computed quasi-steady LCS, but seeing as this is a more modern method for LCS detection, we added a discussion about quasi-steady LCS and their ability to detect persistent features. We also expand on the idea that because an averaging is conducted when finding persistent quasi-steady LCS in Gouveia et al (2020), a similar average over ensemble members can be conducted to find robust quasi-steady LCS. This is included in section 4.1.