This manuscript has inaccuracies, and shortcomings and lacks proper scientific context and some depth. However, the idea of ensemble-averaging and time-averaging to characterize LCS is a good one, and potentially, a good contribution. I will err on the side of the authors and recommend a major revision. My comments follow:

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.

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

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.

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.

40 should be Gulf Stream

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*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, Badwa 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.

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.

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?

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 \partial x, and it is not enough to just integrate the velocity as your equation reads.

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).

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

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.

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.

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.

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?

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?

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.

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.

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?

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

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.

Line 243 and 252, could it be that FTLE seems to be more robust than persistent due to your choice of T=1day? 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.

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

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

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.

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.

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.

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.

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?