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 are prepared to address in a revised manuscript. Initial responses to your various points are given below:

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.

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 realize that there are some fundamental issues with the FTLE as a tool for LCS detection, and that more modern methods exist. We intend to clarify that in the manuscript, discuss the weaknesses of FTLE with regards to LCS, compare it to other existing methods and adapt the manuscript’s title accordingly.

*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 smoothening 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 at 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.*

We realize that the choice of time (the integration time) is an issue, and we addressed this in comment #4. However, our intention was not to find structures which exist over longer time periods (e.g. longer than 24 hrs). In our averaging in time over consecutive FTLE fields we sought to identify regions where high values in the FTLE appear *frequently*, fully aware of the fact that individual structures will form, drift, and deform over various time scales. We will clarify this in a revised manuscript. We appreciate your suggestions on how we can approach this issue, and intend to investigate the proposed simple techniques to see how these compare to FTLE.

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 FLTE 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_0s 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 FLTE field, which gives a field at time $ t_0 $). Of course, one might argue that calculationally 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.*

In a revised manuscript we will tone down the interpretation of FTLE maxima as transport barriers, especially in the core of the paper. We will 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 will also be more careful about how we define FTLE and how it describes the flow field. Finally, we also plan to elaborate on the methods used to conduct the time averaging, so that it becomes clear to the reader. Several of the reviewer's ways of formulating these definitions and interpretations are actually right to the point and will form the basis for revised text.

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).

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 will now provide an analysis figure with examples of various integration lengths for the FTLE (time-of-flow). The figure (in its current form) is added here, and it 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 (and will discuss) 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. Wheres 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 like to 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. In our initial analysis, we did indeed experiment with different periods from one hour up to a few days, and some of these examples are presented below (see Fig. A).

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 many 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!*

As outlined in relation to an earlier reviewer comment, we intend to rework the manuscript to clarify that FTLEs and LCSs are not the same thing, and focus on the properties of the FTLE only. We thank the reviewer for providing the paper by Balasuriya et al (J Fluid Mech, 2016), which we intend to use when discussing FTLE. We use the term “FTLE ridges” loosely to discuss “curves” 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.
We also agree that it is unjustified to identify eddies through the FTLE, which was never intended as the main focus in this manuscript. The original purpose of identifying the eddies was that these are easily identifiable flow structures appearing in the velocity data, such that they could be used to infer something about FTLE ensemble variability. 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 will focus less on the eddies and clarify any misconceptions written about FTLE regarding eddies.

6. The impression given throughout is that the authors are examining robustness of LCSs. I've already talked about why it's actually FLTEs 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 no calculated any uncertainty (i.e., have not evaluated anything to do with uncertainty in the input data).

We would like to take this opportunity to clarify that we suggest 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 will discuss them in our revised manuscript. Our goal was 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.
We plan to clarify this distinction and put less emphasis on time variability as this subject has been more extensively been addressed in other studies. Furthermore, the reviewer suggested including an analysis of current velocities, which we would like to adopt in a revised manuscript to show how variability in the flow field translates into variability in FTLE fields.

7. There are several statements around Equation (2) and (3) which are incorrect. The statements in line 101 are all incorrect: $ \delta x $ 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 $$ x(t) = x_0 + \int_{t_0}^t u \left(x(\tau), t(\tau) \right) \mathrm{d} \tau $$, 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 \left(x(t), y(t) \right) $, $ \dot{y} = v \left(x(t), y(t) \right) $ and the initial condition $ \left(x(t_0), y(t_0) \right) = \left(x_0, y_0 \right) $ connects better with the discussion, perhaps.

We will look through our equations in section 2.3 again and fix mistakes.

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.

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 intend to conduct a major revision of the manuscript, and downplay or question the term "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). So, 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 that structures like these do not appear in nature. 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 will be reworked.

9. The fact that a large standard derivation 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.

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

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 smoothening process will of course get rid of the smaller wavenumbers. Similarly one expects more smoothness when averaged over more and more days.

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.

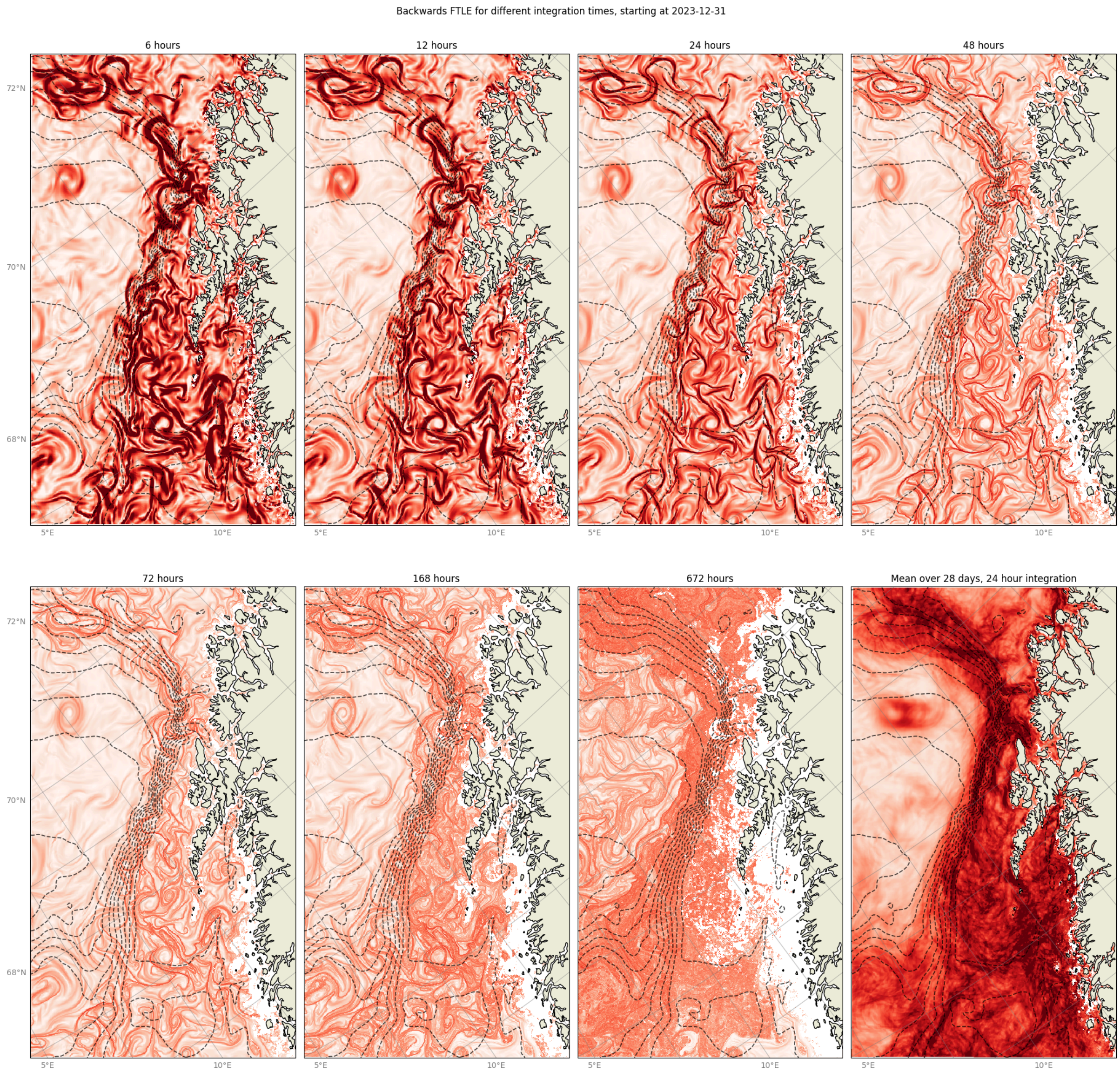

Fig. A: Backwards FTLE for different integration times (6, 12, 24, 48, 72, 168, and 672 hours). The final panel shows the monthly average of 24-hours long FTLE computations.