I would like to thank the referee for the time spent on the thorough review of this paper. In the following, referee comments are given in blue and responses in black.

The manuscript investigates Lagrangian metrics such as FTLE and absolute/relative dispersion using model results in the German Bight, showing the high variability of LCSs and enhanced space variability of trajectories close to ridges. The results are interesting but certainly not novel or unexpected, and the paper lacks in my opinion of clear focus and motivation.

I am not aware of any analysis of LCSs in the German Bight area. Results of such studies will always be site specific and I would not agree that the shapes of the LCSs obtained, partly constrained by the bathymetry in this specific region, are trivial or were foreseeable.

The author mentions several motivating applications, such as characterization or guidance for the observing system, but it is unclear how this would be carry out.

'Guidance for the observing system' suggests that the distribution of LCSs would be used for a characterization of certain locations. I revised the discussion to make clear that this is not the main objective of the study. The idea is rather that detailed simulations addressing a specific period of observations might support an analysis of these data, helping to better distinguish between real changes in a system and a simple shift in the range of influence. This point is now better addressed in the discussion.

I think the paper needs an extensive revision or even better a re-submission, where the motivations and the elements of novelty are clearly indicated and developed. Also, there are several specific points that need further clarification, as detailed in the following.

Main points

1) As mentioned above, there is an extensive literature showing the high sensitivity of particle trajectories to their seeding and the use of LCS to characterize it, so the results presented here are not new.

You are right, a large body of literature on LCSs already exists. However, results of practical studies are always site specific and I am not aware of any such FTLE study for the German Bight area. The present study describes LCSs that possibly affect the meaning of German Bight monitoring data. I wouldn't say that the shapes of these large LCSs, also constrained by the bathymetry in this specific region, are trivial or were foreseeable.

I think the paper needs a novel angle, and a more specific motivation to make the present results new and interesting. I thought that the angle mentioned by the author in the Introduction regarding the characterization of an observing system composed of fixed points is interesting. But it needs more focus and more practical applications.

With regard to the characterization of an observation system, please see my previous response. In this study, the focus is on hydrodynamic simulations being used in support of data analysis. I hope that this point is now better explained in both introduction and discussion. Optimizing the sites of an observation network would be another interesting task, but possibly more challenging. The analysis of a specific data set would go far beyond the scope of the present paper, needing a discussion of many specific problems related to the data of interest.

For instance, could the results be used to quantify uncertainty at the stations using as proxies the distance from ridges? Or could they be used to indicate areas of influence of the stations, in terms of LCS patterns of dispersion properties? Investigating this type of questions would be very useful from the application point of view and could lead to new results.

I see your point. However, in the light of high variability of the LCSs it seems difficult (if not impossible) to define such characteristic areas of influence for specific stations. Areas observations are representative for much depend on the recent history of winds. Therefore, in the revised manuscript now all figures provide a summary of wind conditions over the last 10 days . The revised discussion now refers to this additional information. Although pronounced LCSs are initiated by strong winds, they may survive and even sharpen under subsequent calm wind conditions. Instead of a static characterization of monitoring stations, a model based representation of wind driven, variable hydrodynamic conditions could be useful to support interpretation of monitoring data.

2) Since the results are based on the BSH model outputs, it is very important that the model set up and its validation are adequately described, significantly improving Section 2.2. This is especially relevant since the model based results are envisioned to be used in support of the observing system, possibly also in real time.

The description in Section 2.2 has been extended, it now provides more detailed information about the model.
BSHcmod has been in operational use for about two decades, with revisions being applied where needed and possible. Practical experiences were gained also in the context of search and rescue. The operational model was chosen exactly for the reason you mention: BSHcmod results are reliably available on an everyday basis. Of course, higher spatial resolution would be useful, but for the moment such more expensive simulations are not available operationally.

More specifically, has Lagrangian validation ever been performed using drifters?

Yes, an identical model setup was successfully used for an evaluation of drifter experiments. References to two recent studies (Callies et al. 2017b, 2019) have been added at the end of Section 2.2.

It is also important to be up front regarding model limitations. For instance, given the 1 km resolution we can expect that coastal submesoscale is only partially resolved at best.

Thanks, this is indeed an important aspect. It is now addressed in the discussion (starting at line 304). A clear classification with regard to spatial scales seems hardly possible, as even large scale currents may generate LCS on a scale smaller than resolution of the grid used for underlying hydrodynamic simulations. This would need, however, the FTLE field being defined with higher resolution, which was not been explored in the present study.

Also, if the model is hydrostatic, we cannot expect that near surface divergence processes are correctly described.

The model is hydrostatic, but in the absence of significant bathymetric variations implying high vertical velocities, this shouldn't be a major restriction. The present analysis focusses on open sea conditions, where currents should be reliable (see former studies by Callies et al. (2017b) and Callies et al. (2019)).

3) The description of the used techniques in Section 2.4- 2.5 should be improved, indicating also possible limitations and clarifying definitions. For instance, is the definition of FTLE in eq (3) valid in the case of 2-dimensional flows (as the text at line 120 seems to imply)?

Section 2.4: It is not fully clear to me which kind of limitations and clarifying definitions this comment refers to. I think that all equations needed are already there. The formalism deals with a kinematic description of flow motions (according to Haller (2015): "…*mimicking experimental flow visualization by tracers"*) so that there are no limitations with regard to the underlying physical dynamics. The whole study deals with two-dimensional surface currents, which is now stated more explicitly: "Considering two-dimensional surface currents, this tensor is also two-dimensional."

Section 2.5: Contains Eqs. (4) and (5), please see my response to your next point.

Also, what is the difference between eq (4) and (5) for dilation? From the text (line 144-145), they seem to indicate the same thing, but it is unclear. Indeed the results in Fig.1b,c are quite different.

The dilation rate, defined in the former Eq. (4) has been removed from the paper in order to not overload the presentation.

4) In general, I think that the text commenting the results should be more realistic throughout the paper. For instance the comparison between LCS ridges, and the 2 forms of dilation in Fig.1 (lines 188-195 and lines 312-13) is very positive, while I fail to see a good comparison between the figures. I do not see a "striking similarity" between Fig.1a and 1b, where the main North-South ridge is absent.

I do not understand why you do not see the North-South ridge in Fig. 1b. Actually, the FTLE ridge from Fig. 1a is very clearly reproduced (a more or less perfect copy) in terms of a (green) line of convergence. The good agreement is really striking.

The author acknowledges the clear difference between Fig.1b and 1c, but I do not understand the point of the comparison, given that the model itself is not well suited for this diagnostics.

Fig. 1c showing the dilation rate has been removed and is now replaced by the temperature distribution (formerly Fig. 4a, see above).

Also the comments on Fig.4 do not seem very grounded to me. In a case with very little gradients, except for the obvious coastal ones, as in Fig.4a and at some extent 4c, it is impossible to draw any meaningful conclusion.

Indeed the display of temperatures in Fig. 4 is problematic, the challenge being resolution of small temperature differences in the presence of large gradients towards the coast. Therefore, Fig. 4 was removed in the revised manuscript keeping, however, Fig. 4a which now occurs as Fig. 1c. The colour scale was changed in such a way that now features corresponding with the FTLE field in Fig. 1a are clearly recognizable.

5) Finally, and very importantly in my opinion, new diagnostics and metrics should be investigated, related to the observing system as mentioned in point 1). How can LCS be used to evaluate the observing system? How do LCSs vary on time? At which scales? Which proxies can we use to quantify these changes?

The first question (*How do LCSs vary on time?*) can hardly be answered based on a limited number of example situations shown in static figures of the manuscript. This is why a video was provided in the supplement, showing the evolution of FTLE fields over the full year 2016. This video also answers your second question (*At which scales?*). Features at very different scales can be distinguished, arising from an interaction between changing wind conditions and topographic constraints. In the revised manuscript, a summary of wind conditions during the last ten days has been added in each example and also in the supplementary video.

Your third question (*Which proxies can we use to quantify these changes?*) is the most difficult one. It is addressed in the last paragraph of the conclusions. It would be extremely useful if LCS could be predicted via a simple dependence on atmospheric forcing. However, as the LCSs may depend on wind histories over extended periods of time, such a relationship cannot easily be established. The display of recent wind history in each example illustrates the difficulty of establishing a clear relationship.

More specific points

The Introduction (and possibly also the title) should be re-written with more focus to-ward point 1) above.

The introduction has been revised. Parts of the discussion were moved to the introduction and shortened. A special paragraph was added (starting at line 56), focussing on the use of drift simulations for an improved interpretation of observational data. It is now clearly stated that the study does not aim at the optimization of a monitoring network.

Title: Slightly changed, now explicitly mentioning 'backward simulations'.

More in details, many phrases are unclear. Some examples are listed below - below line 20: "deficiencies of the underlying hydrodynamics...". Is this phrase indicating subgrid uncertainties or what? Deficiencies is certainly not the right word - around line 35.

'Deficiencies' has been replaced by 'inaccuracies'. Indeed these also arise from subgrid uncertainties, as explained by the following half sentence "… including the effects of unresolved sub-grid scale hydrodynamic structures".

Discussion on local-versus nonlocal is not very precise. Indeed, local relative dispersion has been shown to be much faster at small scales and initial times than non local (Poje et al., 2014).

I agree that definition of local relative dispersion was missing. This has been added.

It should also be clarified throughout the text whether the emphasis is on mesoscale or submesoscale dynamics

In the discussion it is mentioned (starting at line 304) that the scale of ridges in the FTLE field can be smaller than grid resolution in the underlying field, due to the fact that separation of tracer particles released vary close together may evolve along drift paths that much exceed numerical grid resolution. This fact makes it very difficult to classify FTLE ridges in terms of either mesoscale or submesoscale dynamics.

Section 2.1. It would be useful to mention from the beginning (lines 80-85) the ge-ographical extension of the German Bight (lat/long are now mentioned at line 128 in Section 2.4), and clarify that the area is depicted in all the figures.

This information has been added in Section 2.1.

Section 3, on results. The author shows 3 examples of LCS (Fig1, 2) for three different flow realizations and dates, 1 example of particle stats (Fig.3) for another realization, and finally SST (Fig.4) for a mix of realizations. It would be better to focus on 3 cases only, and compare LCS with particle stats, as well as SST.

Thanks for your suggestion. In the revised manuscript, the temperature field in the former Fig. 4a has been moved to Fig. 1 to be combined with the corresponding FTLE and FDLD fields shown in panels (a) and (b). The other panels in former Fig. 4 have been removed. In Fig. 3, a third panel has been added that now complements the analysis by showing the corresponding FTLE field. This FTLE field was previously provided only in the supplement. The FTLE field in Fig. 2a is new, it nicely illustrates the origin of FTLE ridges in Fig. 2b. All changes were done to achieve a better description of the relevance of changing wind conditions.

Section 4, provides a broad discussion on FTLEs and their applications, but there is no clear connection with the present results. Indeed, most of the information are more suitable for the introduction, and in any case should be trimmed and focused on the paper's goals.

Section 4 (Discussion) has been thoroughly revised. A key change the discussion now refers to is that now all examples (figures) include kind of wind rose that summarizes the recent history of winds that lead to the FTLE field shown. The examples suggest that even though strong winds generate the pronounced FTLE ridges, these patterns continue to exist for longer times even under subsequent calm meteorological conditions.

Following the reviewer's advice, some paragraphs were moved from the discussion to the introduction.