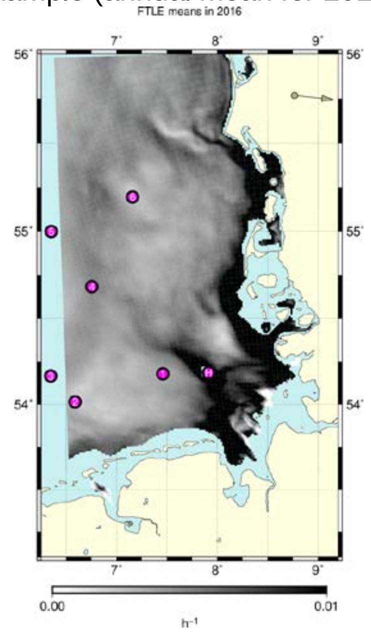
Dear Jens Meyerjürgens,

I appreciate your comments on this study. Please find below my answers to the main points you raised.

Why not using temporal averaged fields?
Due to the strong variability of spatial structures, average fields are very smooth, the following distribution gives an example (annual mean for 2016):



A LCS at a given time depends on a specific recent history of wind forcing. Winds are not constant over longer times, therefore I am not sure that FTLE fields specified for different (constant) wind directions would be very informative. But it might nevertheless be an interesting experiment.

The citation of earlier studies of the German Bight with focus on Lagrangian methods:
You are right, the recent paper by Ricker et Stanev (2020) is a useful reference because the authors also discuss the relevance of wind forcing.

I'm missing representative studies with a more practical and oceanographic background; an example is I. Rypina. Such studies would probably enable direct comparisons of the results between different regions of interest.
Do you mean articles like Serra et al. (2020), *Search and rescue at sea aided by hidden flow structures*, co-authored by I. Rypina? That could indeed be an interesting. A comparison of different regions, however, would definitely be beyond the scope of this paper.

Why is not the density shown instead of T and S?
This study was based on archived data, the hydrodynamic model wasn't re-run to generate additional information.

I'm missing an explanation of the underlying dynamics favouring the LCS structures. There is no discussion taking into account the physical oceanography of the German Bight. What about frontal dynamics?
For analysing frontal dynamics, for instance, one should go back to the original model rather than just archived model output. This study suggests using data that are available operationally and could be used even for real time analyses. FTLE fields were produced for a couple of years with one field every 7 hours. From this data base, then few example fields were discussed in the paper. A subsequent dynamical study might focus on an analysis of different LCSs with possibly different dynamic backgrounds. Also using different hydrodynamic models would be interesting.

The hydrodynamic model: The number of vertical layers is missing. How can the use of a 5 m surface layer be reasoned if the region of interest has depth range of 20-40 m? Is the setup able to realistically resolve vertical processes which, probably, lead to pronounced LCS patterns?
In the German Bight area the model has a maximum of 24 vertical layers. Only for archiving these data were interpolated to another grid in which the top layer now combines results from different finer layers. The model dynamics, however, were implemented on a grid sufficiently fine to resolve all relevant processes.

Why are the three quantities (FTLE, FDLD and dilation rate) chosen if they are so similar?
FTLE and dilation rate are indeed very similar and dilation rate could possibly discarded. However, FTLE and FDLD are calculated quite differently and it is nice to see that results are so similar.

Are there more practical conclusions that can be drawn from the study? What about oil and floating marine litter?
It is hard to draw general conclusions when spatial FTLE patterns are so variable. The conclusion with regard to monitoring is that interpretation of actual measurements from multiple stations should take into account any substantial LCS that possibly separates one in situ station from another close by. I am not sure whether such considerations could affect the analysis of marine litter distributions, where high resolution monitoring data are missing.

Page 3, line 66: However, distinct long term current patterns do exist in the German Bight. They are important for long term tracer dynamics. There are several examples in the literature also in case of the North Sea.
As you say, this applies to long term behaviour. On a short term scale, however, currents and residual currents are very variable.

Is there a reason for the1 km seeding grid?
The resolution seems reasonable in the light of the spatial resolution of the underlying hydrodynamic model.

What is the link between observations and modelling? Which role do the MARNET stations play?
Here, the MARNET stations were shown to facilitate the description of LCS locations and a comparison of different situations. From an applied perspective one may be interested to see which stations have probably higher chances to be close to certain LCS structures. For this purpose, however, one has to look into longer times series of FTLE distributions (see the video in the supplement, for instance).

Is there a proof of different water masses separated by LCS in the station data?
No, the present analysis refers just to surface currents.

It is mentioned that tides have a strong influence on the LCS and dispersion patterns. What does this influence mean for the analysis of FTLE and dispersion?
Tidal movements are obviously important in combination with stirring effects of topographic structures, for instance near the island of Helgoland.

The discussion also lacks of Lagrangian studies (GPS drifters as well as simulations) of the German Bight and how the presented results could be related to them.
I am afraid that so far the data available in the German Bight area do not enable the discussion you are asking for. Callies et al. (2017b, 2019) successfully employed BSHcmod surface currents (also used here) to simulate observations from drifter experiments. These

studies are mentioned in the manuscript. However, the very limited number of drifters is definitely not sufficient to identify LCSs. In any case, it can never be expected that model results provide a perfect surrogate reality. Crosschecking LCSs obtained from different models and probably also from radar observations would be very useful.

Kind regards
Ulrich Callies