Response to Referee #1 Comments on EGUSPHERE-2026-3629

In this manuscript, the authors compare the observational, quality-controlled data gathered by the ocean surface drifters from the well-known ‘Global Drifter Program’ with the output from various ocean current gridded products: AVISO altimetry-derived geostrophic currents, NeurOST machine-learning reconstructions, GlobCurrent and OSCAR multi-observational/mixed-layer total currents, and the GLORYS and TOPAZ reanalyses. The data comparison spans over a decade: years 2011-2021. In addition to the usual Eulerian evaluation of point-to-point similarity of current magnitude and direction, the authors use a Lagrangian method to show both discrepancy and coherence of the ocean surface’s circulation patterns. The identification of Lagrangian Coherent Structures (LCSs) as well as the use of the Lagrangian Uncertainty Quantification (LUQ) framework not only allows for trajectory-to-trajectory comparison between the observations and the chosen product but also permits transport assessment over a wider area (in the vicinity of the trajectory), organised by transport barriers and coherent structures.

The authors present and thoroughly analyse the statistical diagnostics, providing regional and seasonal differences between observations and the products as well. The presented case studies cover various dynamical regimes: from more energetic ones, where mesoscale processes have a major impact on the flow structure, to more coherent, large-scale, more stable patterns in the gateways to the Arctic Ocean. This leads the authors to two major conclusions: 1) Products incorporating observational constraints and wind effects outperform altimetry-only estimates; 2) good local velocity agreement does not guarantee realistic transport; and 3) no one product would provide the best results for all types of circulation patterns.

General assessment:

The manuscript's topic aligns well with the scope of Ocean Science and addresses the importance of validating publicly available gridded ocean current products. Despite the somewhat arbitrary choice of regional areas, the authors provide a logical explanation of the large number of metrics. Thanks to the application of two different methods, the authors successfully convinced the reader of their complementarity.

My bigger concern is the authors’ choice of the analysed case studies: why add these non-Arctic ones? How do they relate to the title of the article itself? I have a feeling that some of the cases were added not necessarily because they relate to the topic but rather because they are good to illustrate the differences between the products analysed. While the Lofoten Gyre and the Labrador Current are more closely connected to the Arctic (outflow and inflow areas), I would consider the Gulf Stream almost entirely out of scope of the paper. Perhaps adding ‘and sub-Arctic’ to the title would solve this issue?

Among others, the conclusion about the usefulness of the TOPAZ product for the ice-influenced cases, despite its weaker agreement in the global speed-based Eulerian statistics, is promising, especially due to the lack of observational data there, as illustrated by sparse GDP drifter coverage.

In my opinion, the manuscript is well written and illustrated and makes a proper research statement on the covered scientific topic. However, I also found it a bit too long and too repetitive in some parts. As mentioned above, some case choices seem slightly misleading. Therefore, I suggest that it can be published after some minor corrections.

We thank the reviewer for their thorough reading of our manuscript, the positive assessment of our work, and the constructive suggestions. We have addressed all the comments and suggestions made by the reviewer, as detailed below.

Minor comments and remarks:

Title: **'**Eulerian and Lagrangian Assessment of Arctic Surface Current Products Using Drifter Observations' – consider adding 'and sub-Arctic'.

We thank the reviewer for this suggestion. The main objective of including the non-Arctic case studies (such as the Gulf Stream extension) was indeed to test and illustrate how the products perform under different dynamical regimes, specifically contrasting highly energetic, mesoscale-dominated flows with the more stable or ice-influenced patterns characteristic of the Arctic and Arctic gateways. However, we agree that the Gulf Stream region lies outside the strict geographical definition of the Arctic and that this should be better reflected in the title and scope of the manuscript. Accordingly, we are adding it to the title, which is now: *"Eulerian and Lagrangian Assessment of Arctic and Sub-Arctic Surface Current Products Using Drifter Observations"*.

Introduction:

This section clearly describes the research background, motivation, data and methods and briefly introduces the next sections. The bibliography used is up to date and demonstrates a knowledge of the subject matter.

Data:

In this section, more information about the observations and gridded products is provided.

Line 182: 'This choice is made to ensure methodological consistency across all datasets'- Are all the other products similar? From their description, one may understand that they differ in the layers analysed.

We thank the reviewer for highlighting this point and allowing us to clarify our choice.

Our choice to restrict the evaluation to the surface layer stems from the fact that the analyzed products represent slightly different vertical layers by definition (e.g., AVISO represents geostrophic surface currents, OSCAR represents the integrated upper 30 m flow, and 3D numerical reanalyses like GLORYS and TOPAZ resolve discrete vertical levels), as explained in each of the subsections. Our choice to use the uppermost surface currents from all products was to maintain a homogeneous evaluation baseline across purely surface/satellite-derived products (AVISO, NeurOST, GlobCurrent, OSCAR) and model-based reanalyses (GLORYS, TOPAZ). This allows us to assess the products strictly as they are distributed and most commonly utilized by end-users seeking surface current information.

We acknowledge that alternative choices, such as integrating over the upper few meters in reanalysis products, could introduce minor quantitative shifts in Eulerian metrics, particularly under strong wind-forcing events. However, our experience in the region and preliminary tests indicated that these differences do not alter the overall qualitative conclusions or the relative ranking of the products.

Methods:

In this section, the processing of the GDP drifters, Eulerian metrics, Langrangian validation framework and sensitivity analysis are described in detail. Furthermore, reasoning behind the chosen case studies is given, as well.

Line 272: 'In ocean transport problems, relatively small velocity differences can accumulate over time and generate large deviations in particle trajectories.' - This sentence is like the sentence in the previous section. Please consider rewriting/shortening it.

We thank the reviewer for pointing this out. We agree that the opening paragraph of Section 3.3 restated an argument already made at the end of Section 3.2. We have therefore removed it entirely rather than rewriting it, since the motivation for the Lagrangian analysis is already established in the previous section.

Line 334: 'This analysis supported the use of τ = 5 days and R = 10 km'. – Which other values were considered in your sensitivity analysis?

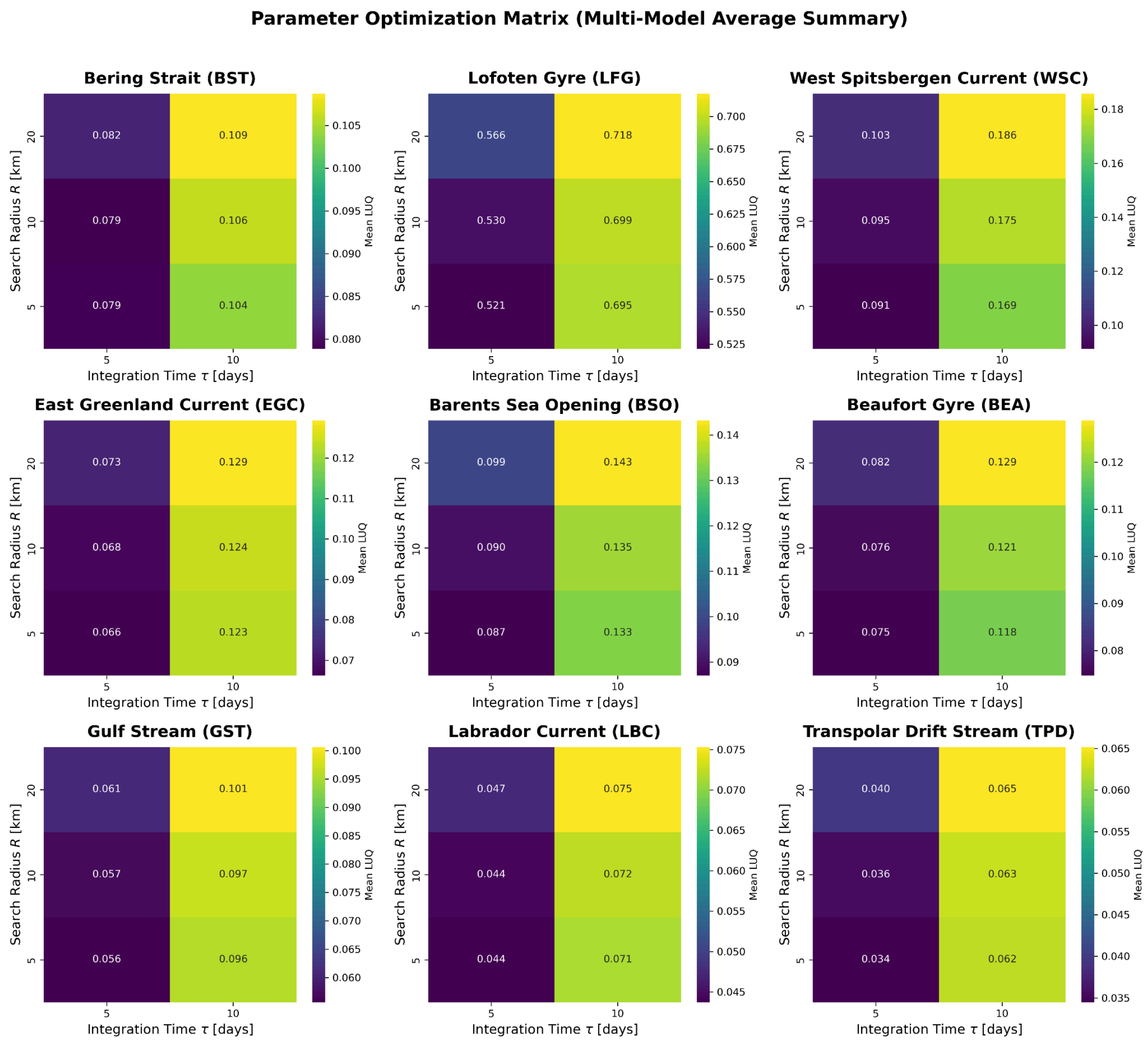
We thank the reviewer for asking for clarification regarding the parameter space explored during our sensitivity analysis.

In our preliminary assessment, we tested combinations of integration times $\tau \in \{5,10\}$ days and spatial radii $R \in \{5,10,20\}$ km. The selection of τ=5 days and R=10 km represents a trade-off grounded in the dynamical scales of the studied features.

A radius of R=20 km was found to be excessively large, often spanning across distinct mesoscale dynamic regimes or exceeding the characteristic scale of the local oceanographic structures under study. Conversely, R=5 km was overly restrictive, limiting the spatial domain around the initial trajectories and capturing too few synthetic particles to provide robust statistical Lagrangian metrics. Performance metrics across R=5 km and R=10 km showed high consistency, making R=10 km the optimal intermediate scale that ensures adequate particle sampling while remaining strictly smaller than the sub-mesoscale/mesoscale features being evaluated.

Regarding the temporal horizon τ, the integration window to τ=10 days frequently resulted in particle trajectories advecting far away from the localized circulation features or coherent structures under investigation, introducing shear dispersion effects unrelated to the local feature evaluation. A 5-day horizon (τ=5 days) retained high particle coherence within the regional structures while providing sufficient time for diagnostic transport metrics to develop clear dispersion signatures.

We provide a figure below showing the comparative diagnostic performance across the tested combinations of (τ,R).



We have revised the text in Section 3.3 (Line 334) to explicitly state the parameter space tested. The revised text now reads as follows (changes in **bold**):

L328-330: "*To assess the robustness of the LUQ diagnostics to the choice of methodological parameters, we performed a sensitivity analysis* ***testing integration time scales of τ∈{5,10} days and neighborhood sizes of R∈{5,10,20} km****. This analysis supported the use of τ=5 days and R=10 km.*"

Results:

This section provides a number of good illustrations of the differences in the seasonal, temporal and regional variabilities emerging from all products after the methodology has been used. However, the time series of the normalised LUQ values do not cover the same periods for all cases but are scattered among several years (in subplots of one figure). Is there a reason for this?

We thank the reviewer for this question. The two features have different origins.

Regarding the different periods across panels: each time series is computed from one real drifter selected as representative of a given dynamical regime (Table 1). The evaluation period is therefore dictated by the availability of that particular drifter record and not chosen by us. Because no single period provides simultaneous drifter coverage across all nine regimes, the case studies necessarily span different years. This does not affect comparability, since the LUQ metric is normalised by the characteristic trajectory length L of each drifter.

Regarding the gaps within individual curves: these reflect differences in product coverage along the drifter track rather than interruptions in the drifter record. They occur mainly where the trajectory approaches the coast or narrow passages, and their extent depends on how each product is masked. AVISO, for example, uses a coarser land mask that leaves nearshore grid points undefined, so no LUQ value can be computed when the velocity field is undefined. Similar limitations affect other products in ice-affected and high-latitude nearshore areas.

Line 439: ‘narrow Arctic gateways’ – I think you may add a sentence or two about the worst results that were obtained for S9 (the Canadian Archipelago); it really stands out in Figure 9.

We thank the reviewer for pointing this out. We agree that the Canadian Archipelago (S9) stands out clearly in Fig. 9 and deserves further discussion. This result may partly reflect the challenging conditions in this region, including complex dynamics, ice–ocean interactions, and persistent sea-ice cover during much of the year, which limits the availability of drifter observations. As shown in Fig. 2, the observational sampling in this region is relatively sparse. We have therefore added the following text between L438–442 to clarify and discuss this result:

“*The Canadian Archipelago (S9) stands out as a region of particularly low skill in Fig. 9, with larger errors and lower correlations than in most other regions. This result should, however, be interpreted with caution, as the Canadian Arctic Archipelago is characterized by persistent and seasonally varying sea-ice cover, which strongly limits the spatial and seasonal availability of drogued-drifter observations in this region (Fig. 2). The reduced sampling may therefore contribute to the apparent degradation in the regional validation statistics, in addition to the complex bathymetry and ice–ocean interactions characteristic of the area.*”

I would also advise the authors to consider removing some repetitions that protract the text in this section as well: for example, these two sentences below are very similar:

Line 477: 'Rather than evolving as a random diffusive cloud around the ground-truth trajectory, transport deviations are organized by the transport geometry of the underlying velocity field and by the Lagrangian coherent structures governing particle evolution.'

Line 506: 'Accordingly, transport deviations are not uniformly distributed around the observed drifter position, but are organized by the underlying recirculation and its associated transport barriers.'

We agree. The sentence at line 506 restated the general property already introduced at line 477 and added no new information in the context of the Lofoten Gyre case, which illustrates it directly. We have removed it and merged the surrounding text.

Line 574: 'that products with with similar local velocity' – remove repetition

Thank you for noticing this. We have deleted this repetition.

Discussion:

In this part of the manuscript, the results are thoroughly discussed, together with the added value and limitations of the performed analysis. There are some repetitions from the previous parts, but this is probably justified and aims to highlight the importance of proposals.

This section lacks a broader context and a comparison of the results obtained with similar studies and some implications for future work. This might be helpful for the readers conducting similar data experiments or using similar methods.

We thank the reviewer for this suggestion. We have expanded the Discussion to situate our results within the broader validation literature and to make the implications for future work more explicit.

To address the comparison with similar studies, we added the following text at the end of the first paragraph of Section 5.2 (L667-671):

"This behaviour is consistent with independent saildrone-based validation of OSCAR in the western Arctic marginal seas, which found reduced skill over shallow topography and in freshwater-influenced regimes, attributing it to unresolved salinity/buoyancy effects (Chi et al., 2023). Similarly, Lagrangian-based comparisons on the south Greenland shelf found that incorporating the Ekman component substantially improved trajectory skill relative to geostrophic-only estimates (Coquereau and Foukal, 2023), reinforcing the role of ageostrophic contributions identified here."

To address the implications for future work, we added the following text in Section 5.5 (L733-738"):

"A broader intercomparison of global ocean current products against drifter observations found that including Stokes drift substantially improves skill scores, while tidal currents contribute little at the global scale (Aijaz et al., 2023), suggesting an additional unresolved source of variability not addressed here. More generally, the LUQ metric (García-Sánchez et al., 2025) remains a recent development that has not yet been widely applied across different ocean regions and current products; extending its use to other study areas would allow systematic intercomparison of transport-based skill across contrasting dynamical regimes."

Line 668: Section title - make it a verbless sentence to be consistent with previous subsection titles

Following this suggestion, we have changed the section title to *"Eulerian agreement and transport realism"* to ensure consistency with the titles of the other sections and subsections.

Conclusions

This part is a summary of the obtained results; here the repetitions are unavoidable.

Figures and Tables:

The manuscript contains 17 figures and 2 tables, and they are correctly prepared for publication. The fonts are large enough, the lines are not too thin, and the colours might probably be better in some of them (Figures 6, 7, 14-16). Please check the best colour palettes for scientific figures and data; this may help in your future submissions.

We thank the reviewer for the positive evaluation of our figures and tables, as well as for the thoughtful advice regarding scientific color palettes.

Regarding Figures 6, 7, and 14-16, these panels require the simultaneous display of distinct oceanographic products alongside observational data. While testing alternative colorblind-friendly palettes with 6 distinct hues, we found that reducing color contrast reduced the visual separation between individual product lines, ultimately compromising readability. To ensure maximum legibility for this multi-product comparison, we have maintained the current palette, while we appreciate the reviewer's recommendation for future submissions.