We would like to thank the reviewer for your valuable comments that could further improve the quality of the paper. Please find below each reviewer comment in black and our response in red.

Some general / overall comments:

1. A basic trajectory model is described, and used to simulate the ctrajectories of both surface and drogued drifters with a geostrophic current field + permutations of ekman, Stokes and windage. The analysis is fairly lengthy, just to conclude with what is well known and should be expected: for the surface drifters, all components should be used, whereas for the drifters with drogue at 2m depth, direct wind drag should not be included. As this is all expected, one can question the relevance of all the different combinations used, e.g. using geostrophic current only. It still has some value to confirm what is known, but it could be made much more compact.

Thank you for this constructive comment. We agree that the original presentation was too lengthy and that some qualitative results were physically expected, particularly the importance of wind-related forcing for the undrogued surface drifters and the reduced direct wind response of the drogued drifters. Following the reviewer's suggestion, we substantially shortened and reorganized this part of the manuscript. Specifically, we combined the original Sections 3.5 ("Evaluation of particle trajectories") and 4.1 ("Optimal forcing combinations and ocean models") into a more compact Section 3.5, now titled "Data-driven optimal forcing combinations."

2. What is more of a question, is the calculation of the windage (wind drift factor) which is a critical component. This is calculated as 1.5% and 0.5% for surface and drogued drifters respectively, but without much details of the exact computation. In any case, this important coefficient will be associated with uncertainty, and it might be more relevant to vary (or even tune) this coefficient, than some of the above mentioned variants which are expected to be not good.

Thank you for this important comment. We agree that the original manuscript did not explain the windage coefficient clearly enough. In this study, the windage coefficient, r, was estimated from a first-order balance between aerodynamic drag on the exposed part of the drifter and hydrodynamic drag on the submerged part:

$$r = \left(\frac{\rho_a C_{d,a} A_a}{\rho_w C_{d,w} A_w} \right)^{1/2},$$

where ρ_a and ρ_w are the densities of air and seawater, $C_{d,a}$ and $C_{d,w}$ are the drag coefficients in air and water, and A_a and A_w are the projected areas above and below the sea surface, respectively. Because $C_{d,a}$ and $C_{d,w}$ were not measured directly for our custom drifters, we used $C_{d,a}/C_{d,w} \approx 1$ as a first-order approximation, consistent with previous studies (de Dominicis et al., 2016; van der Mheen et al., 2020).

This assumption ($C_{d,a}/C_{d,w} \approx 1$) is appropriate for the DX drifters because the exposed and submerged parts are both cylindrical, so the air-side and water-side drag coefficients are expected to be of comparable order. The cylindrical body has a diameter of 0.09 m and a height of 0.30 m, with approximately 0.03–0.05 m protruding above the water surface. Therefore, the projected area exposed to air is $A_a \approx 0.09 \times (0.03\text{–}0.05) = 0.0027\text{–}0.0045$ m^2, whereas the submerged projected area is $A_w \approx 0.09 \times (0.25\text{–}0.27) = 0.0225\text{–}0.0243$ m^2. Using $\rho_a/\rho_w \approx 1.2 \times 10^{-3}$ and $C_{d,a}/C_{d,w} \approx 1$, the estimated windage coefficient is $r_{DX} \approx 0.012\text{–}0.016$. Accordingly, $r = 0.015$ was adopted for the DX drifters.

For the DO drifters, the submerged flat drogue plate likely has a larger drag coefficient than the cylindrical float; therefore, assuming $C_{d,a}/C_{d,w} \approx 1$ gives a conservative upper-bound estimate of windage rather than an exact calibrated value. For the drogued DO drifters, the exposed area above the sea surface is similar to that of the DX drifters, but the submerged drag area is much larger because of the drogue positioned near 2 m depth. To avoid double-counting the two perpendicular plates, we used one effective projected drogue plate area: $A_{w,\mathrm{drogue}} = 0.30 \times 0.30 = 0.09$ m^2. The submerged cylindrical body contributes $A_{w,\mathrm{body}} \approx 0.09 \times 0.26 = 0.0234$ m^2. Thus, the total effective submerged projected area is approximately $A_w \approx 0.0234 + 0.09 = 0.1134$ m^2. Using $C_{d,a}/C_{d,w} \approx 1$, this gives $r_{DO} \approx 0.0053\text{–}0.0069$. The adopted value $r = 0.005$ is therefore a conservative geometry-based estimate for the DO drifters.

We also considered published drag-coefficient values for drifter drag-area calculations. McCaffrey and Koeberle (2024) used $C_D = 0.74$ for cylindrical float elements and $C_D = 1.98$ for flat-plate drogue elements. Applying these values to the DO geometry gives $r_{DO} = \left[\rho_a(0.74A_a)/\rho_w\left(0.74A_{w,\mathrm{body}} + 1.98A_{w,\mathrm{drogue}}\right)\right]^{1/2}$, which yields $r_{DO} \approx 0.0035\text{–}0.0045$. This value is slightly lower than, but close to, the adopted $r = 0.005$. Therefore, $r = 0.005$ does not underestimate the direct wind effect on the DO drifters; rather, it is a physically reasonable and slightly conservative estimate. We have added the discussion about r coefficient in Lines 148-176.

We agree that the windage coefficient is associated with uncertainty. However, we did not tune r to minimize trajectory error, because doing so would mix two different objectives: estimating a drifter-specific windage coefficient and evaluating the relative contributions of geostrophic current, Ekman current, Stokes drift, and windage. Instead, we used *a priori* geometry-based values of r and kept them fixed across the forcing-combination experiments. In response to the reviewer's comment, we added a sensitivity test for the undrogued DX drifter, which is more directly exposed to wind forcing. The geometry-based windage coefficient used for DX in the manuscript was $r = 0.015$. To examine the sensitivity of the results to this parameter, we repeated the DX simulations using a smaller value, $r = 0.005$, representing an underestimated windage effect, and a larger value, $r = 0.030$, representing an overestimated windage effect. The sensitivity test showed that the optimal forcing combination changed when r was substantially modified. For $r = 0.005$, the lowest average MAE was obtained for C8, $u_g + u_e + u_s + u_w$, with $\overline{\mathrm{MAE}} = 117.3$ km. For the geometry-based value, $r = 0.015$, the lowest

average MAE was obtained for C7, $u_g + u_s + u_w$, with $\overline{\text{MAE}} = 42.9$ km. For $r = 0.030$, the lowest average MAE was obtained for C4, $u_g + u_w$, with $\overline{\text{MAE}} = 88.2$ km. Although different values of r produced different optimal forcing combinations, the geometry-based value $r = 0.015$ produced the smallest overall error among the tested cases. Specifically, the optimal MAE values for $r = 0.005$ and $r = 0.030$ were 117.3 km and 88.2 km, respectively, both of which were larger than the optimal MAE of 42.9 km obtained using $r = 0.015$. These results indicate that model performance is sensitive to the windage coefficient, but they also support the use of the geometry-based value $r = 0.015$ for the DX drifter. We therefore used the geometry-based coefficient as the primary estimate and used the additional simulations only to evaluate sensitivity, rather than tuning r freely to obtain the best trajectory fit. This clarification and the sensitivity-test results have been added to the revised manuscript in Lines 473-479.

Table R1. MAE values for PTM-DX sensitivity test using $r = 0.005$. The maximum MAE value for each day is shown in bold.

Day	C1	C2	C3	C4	C5	C6	C7	C8
5	130.2	128.2	123.1	129.8	106.3	117.1	90.0	**29.2**
10	252.4	243.2	223.7	237.2	207.1	220.9	166.0	**138.2**
30	473.4	449.3	300.2	423.5	260.2	348.0	209.0	**161.1**
75	550.2	541.3	290.9	459.2	219.8	369.3	145.0	**140.7**
Avg.	351.6	340.5	234.5	249.9	198.4	211.1	122.0	**117.3**

Table R2. MAE values for PTM-DX sensitivity test using the geometry-based value $r = 0.015$. The maximum MAE value for each day is shown in bold.

Day	C1	C2	C3	C4	C5	C6	C7	C8
5	130.2	128.2	123.1	89.0	106.3	60.9	25.3	**20.4**
10	252.4	243.2	223.7	162.3	207.1	104.5	**29.5**	43.9
30	473.4	449.3	300.2	216.4	260.2	104.6	**55.6**	114.0
75	550.2	541.3	290.9	164.6	219.8	98.3	**61.2**	105.4
Avg.	351.6	340.5	234.5	158.1	198.4	92.1	**42.9**	70.9

Table R3. MAE values for PTM-DX sensitivity test using $r = 0.030$. The maximum MAE value for each day is shown in bold.

Day	C1	C2	C3	C4	C5	C6	C7	C8
5	130.2	128.2	123.1	**25.6**	106.3	31.8	48.4	58.6
10	252.4	243.2	223.7	**69.8**	207.1	87.2	117.7	109.7
30	473.4	449.3	300.2	**131.9**	260.2	165.9	207.7	198.8
75	550.2	541.3	290.9	**125.5**	219.8	173.7	207.7	198.8
Avg.	351.6	340.5	234.5	**88.2**	198.4	91.7	116.3	113.2

3. The Stokes drift is calculated directly from a wave model, and apparently given by Eq 1, but this equation is missing the left side, presumably being the Stokes drift itself. Despite only about 10% of the drifter sticking out of water (~3 of 30 cm), Stokes drift is evaluated at z=0 (surface) and 1.5m ("weighted mean" of surface and drogue at 2m). For the undrogued drifter, a depth of ~15 cm should be more relevant. Difference might not be huge, but (along with uncertain/postulated wind drift factor) might be enough to give wrong conclusions about the optimal configuration, e.g. to support the claim/finding that the above model is better than using CMEMS and HYCOM models.

Thank you for this comment. We agree that Eq. (1) was incomplete in the original manuscript. We have corrected the equation by adding the missing left-hand side and explicitly defining it as the Stokes drift velocity.

We also agree that the depth at which Stokes drift is applied should be consistent with the submerged geometry of each drifter type. In the original manuscript, the DX simulation was described as using surface forcing at z=0 m. However, the Stokes drift used in the actual calculation was taken from the shallowest subsurface level of the wave-model output, z=−0.1 m (Line 223), rather than exactly at the air–sea interface. The DX drifter is 0.30 m high, with only about 0.03–0.05 m protruding above the water surface. Therefore, most of the drifter body is submerged within the upper 0.25–0.27 m of the water column, giving an effective submerged center of approximately 0.12–0.14 m below the surface. Thus, the use of z=−0.1 m provides a physically representative estimate of the Stokes drift acting on the DX drifter and is close to the reviewer's suggested depth of approximately 0.15 m.

4. The end of the paper performs simulations also with "CMEMS and HYCOM" models, as an alternative to the above "data driven" approach (according to title). However, this "CMEMS model" is the GLORYS reanalysis, and "HYCOM" is here the GOFS analysis. These more descriptive names should be given in the manuscript, and also show that these "models" are in fact analyses which should also be based on geostrophic currents and Ekman currents, just as the "data" that is here used as an alternative to the "models". In other words, the distinction between "data" and "model" is unclear and partly un-real. The discussion would then be why the "analysis" constructed in this paper (geostrophic+ekman+stokes+wind) is better than these other analyses. It should also be noted that the used geostrophic current has hourly time step, whereas the GLORYS CMEMS has only daily data (as far as I can tell) - which can also partly explain the poorer performance.

Thank you for this helpful comment. We agree that the terms "data-driven" and "model" were not sufficiently clear in the original manuscript. We have revised the manuscript to avoid this ambiguity. Specifically, the simulation previously referred to as "CMEMS" is now described as the GLORYS12V1 global ocean reanalysis distributed by Copernicus Marine, and the simulation previously referred to as "HYCOM" is now described as the GOFS analysis based on HYCOM and NCODA. GLORYS12V1 is a global eddy-resolving ocean reanalysis product, while GOFS/HYCOM is an operational analysis/forecast system based on HYCOM with data assimilation. Therefore, we replaced the general expression "global ocean models" with more

accurate terms such as “ocean analysis/reanalysis products” and “model-derived velocity products” throughout the revised manuscript.

We also agree that the distinction between our “data-driven” approach and these analysis/reanalysis products should be stated more carefully. GLORYS and GOFS/HYCOM are not free-running models; they assimilate observations and provide dynamically consistent ocean circulation fields. Therefore, the revised manuscript no longer presents the comparison as a simple contrast between “data” and “models.” Instead, we describe our approach as an observation-constrained surface-forcing reconstruction, in which geostrophic current, Ekman current, Stokes drift, and windage are explicitly combined and evaluated against observed drifter trajectories.

The reason why the reconstructed forcing combination can outperform GLORYS and GOFS/HYCOM in this specific case is not that GLORYS and GOFS/HYCOM lack observational information. Rather, the difference arises because the reanalysis/analysis velocity fields are dynamically adjusted by the model physics, data-assimilation procedure, vertical mixing, spatial resolution, and temporal averaging. Even when sea-surface height is assimilated, the resulting near-surface velocity gradients and finite-time Lagrangian transport structures can differ from those obtained from the explicitly reconstructed forcing field. We therefore added a direct FTLE-based comparison of the data-driven velocity combinations and the GLORYS/GOFS velocity products. This comparison shows that the data-driven fields better reproduce the hyperbolic-type Lagrangian structure associated with the early drifter bifurcation, whereas the corresponding FTLE ridges in GLORYS and GOFS/HYCOM are weaker, displaced, or less coherent in the bifurcation region (Lines ??).

We also clarified the temporal resolution of the products. The GLORYS12V1 reanalysis used for PTM-CMS provides daily fields, whereas the data-driven reconstruction includes wind-related components derived from hourly ERA5 wind forcing. Thus, differences in temporal resolution and in the representation of high-frequency near-surface wind-driven transport may partly contribute to the lower trajectory skill of the GLORYS-based simulation (Lines 517-520).

5. To evaluate the performance of the various simulations, mean absolute separation and a skill score is used. However, these are apparently calculated only as a single number from the beginning of each the trajectory, albeit also given at some discrete “time steps” (rather “end times”). As a bifurcation point is seen very early (apparently better in geostrophic current field than in “models”, the simulated and observed drifters will quickly be far apart. And when these are separated, continued calculation of skillscore (or MAE) is less relevant, as you are comparing a drifter in one place, to a drift calculation using currents from a very different place - i.e. data becomes uncorrelated. Much more robust to evaluate the optimal forcing, would be to chunk the trajectories in segments of a couple of days. For this reason - and especially due to the early bifurcation “incident” - and also considering the above points about windage and Stokes drift - the conclusion that given “data-driven analysis” is better than the “models” becomes questionable.

Thank you for this important comment. We agree that cumulative trajectory metrics such as MAE and skill score should be interpreted with caution when an early bifurcation occurs. Once the observed and simulated drifters enter different dynamical branches, the subsequent separation may partly reflect pathway divergence rather than only local velocity error. Therefore, we agree that a segment-based evaluation provides a more robust diagnostic of short-term trajectory skill.

Following the reviewer's suggestion, we performed an additional segment-based MAE analysis for the DX drifters. The observed and simulated trajectories were divided into 3-day segments, and the MAE was calculated independently for each segment. This test was added as a sensitivity analysis to reduce the dominance of accumulated downstream errors after the early bifurcation.

The new results are summarized in Table R4. The segment-based MAE confirms that the simulations including geostrophic current, Stokes drift, windage, and Ekman current performed best overall. The GSWE case produced the lowest average segment-based MAE, 28.7 km, and also the smallest standard deviation, 3.7 km. The GSW case was the second-best case, with an average MAE of 31.3 km and a standard deviation of 8.3 km. In contrast, the G-only and GE cases showed much larger average MAE values, 53.5 km and 51.9 km, respectively. The improvement is especially clear during the first 0–3 day segment, when the early bifurcation occurs: the MAE decreased from 120.0 km for G and 113.1 km for GE to 46.4 km for GSW and 33.6 km for GSWE. This indicates that the improved performance is not only a result of long-term accumulated separation, but is already evident during the early bifurcation period.

This additional analysis slightly refines our original interpretation. In the cumulative MAE analysis, GSW showed the best DX performance, whereas the segment-based MAE analysis identifies GSWE as the best and most stable case. This difference indicates that the two metrics emphasize different aspects of trajectory prediction: cumulative MAE evaluates the integrated pathway prediction, including whether the simulation follows the correct transport branch after bifurcation, whereas segment-based MAE evaluates short-term local trajectory skill.

We have therefore revised the manuscript to clarify that the cumulative metrics are not interpreted as purely local velocity-error measures. We also softened the conclusion by describing the comparison with model-derived velocity products as a benchmark comparison rather than an absolute demonstration that the data-driven analysis is always superior. The revised interpretation is that the data-driven forcing combinations, particularly GSW and GSWE, better reproduce the early bifurcation and short-term DX trajectory evolution, while the cumulative metrics remain useful for evaluating the integrated transport pathway.

This discussion has been added in Lines 487-495.

Table R4. Segment-based mean absolute error (MAE) for the DX drifter simulations. The observed and simulated DX trajectories were divided into 3-day segments, and the MAE was calculated independently for each segment to evaluate short-term local trajectory skill. G, E, S, and W denote geostrophic current, Ekman current, Stokes drift, and windage, respectively. AVG and STD indicate the mean and standard deviation of the segment-based MAE across the

listed time intervals. Lower MAE values indicate better agreement between the simulated particles and observed DX drifters.

duration(day)	0-3	3-6	6-9	27-30	72-75	AVG	STD
G	120.0	37.0	31.5	41.8	37.4	53.5	33.4
GE	113.1	38.1	29.9	37.1	41.3	51.9	30.8
GS	103.0	32.2	30.6	27.5	27.3	44.1	29.5
GW	83.4	27.4	31.6	31.9	53.4	45.5	21.0
GSE	91.6	33.1	29.4	25.8	28.9	41.8	25.0
GEW	67.6	27.9	30.1	28.6	57.2	42.3	16.8
GSW	46.4	22.4	33.3	27.4	27.1	31.3	8.3
GSWE	33.6	23.7	31.9	25.4	29.0	**28.7**	**3.7**

6. When using CMEMS and HYCOM models, the Stokes drift is varied as 1.5 and 3.0% Why not use the same variations as for the "data-driven" approach with geostrophic current?

Thank you for pointing this out. We apologize for the confusion caused by the wording in the original manuscript. The values of 1.5% and 3.0% refer to windage coefficients, not to variations of Stokes drift.

7. Mean current and wind fields are shown in both Figure 1 and Figure 2, apparently redundant. Figure 2 shows average wind and current for November along with surface drifters and December along with drogued drifters. This is confusing, as both sets of drifters were active the whole/same period. Also the value of showing monthly average to support discussion of specific events such as a bifurcation point and sudden wind change is questionable.

Thank you for this constructive comment. We agree that the original presentation of the mean current and wind fields was partly redundant and potentially confusing. In the revised manuscript, we reorganized Figures 1 and 2 to separate their purposes more clearly. Figure 1 now focuses on the deployment region and the environmental conditions during the initial release and early bifurcation period, using one-week-averaged geostrophic current, wind, SST, and SSH fields after drifter release. Figure 2 was revised to focus only on the observed drifter trajectories and the Lagrangian structure associated with the early bifurcation, without repeating monthly mean current and wind fields.

Some specific comments:

1. Line 48: Unclear sentence: "Stokes drift influences the direction and speed of surface currents". Stokes drift is a lagrangian current adding to the Eulerian current, not to be confused with Coriolis-Stokes which can modify the Eulerian current.

Thank you for your comment. The sentence has been revised to 'The wave-induced current known as Stokes drift adds to the Eulerian surface current, thereby contributing to the transport of floating objects' (Line 52).

2. Line 90: Drogued drifters are throughout paper referred to as “DO” and undrogued as “DX”. However, the naming logic is not clear, and thus the reader can easily forget which is which. More descriptive terms should be used, or instead the user should be reminded regularly (e.g. in figure texts) which is which.

Thank you for your constructive comment. We agree and have clarified the terminology. We now consistently refer to them as undrouged (DX) and ‘drogued (DO)’ throught the manuscript.

3. Line 137: What exactly is the SST used for?

Thank you for your question. The SST was not used as an input forcing in the particle tracking model. It was only used as a background field in Figure. 1 to describe the surface thermal structure near the deployment region. In particular, the SST helps show the location of thermal fronts around the EKWC, which are useful for interpreting the local current structure and the early drifter bifurcation.

4. Line 145: diffusion term is added - but is this used for the simulation with just a single “particle” per trajectory? If so, the random numbers will degrade the simulation, which is then not deterministic. Diffusion (or random numbers in general) only makes sense when using many particles.

Thank you for this important comment. We agree that using a stochastic diffusion term with only one numerical particle per drifter initial location can make the trajectory comparison sensitive to a single random realization. In the main simulations, one numerical particle was released from each observed drifter initial position so that the simulated particle set corresponded directly to the observed drifter set. However, because the random-walk diffusion term includes Gaussian random numbers, the exact trajectory and MAE values may be affected by stochastic variability.

To examine whether this stochastic variability influenced the main conclusions, we performed an additional ensemble-sensitivity test using 10 numerical particles for each drifter initial location. In this test, the deterministic velocity fields were kept identical to those used in the original simulations, while only the Gaussian random sequences in the diffusion term were varied. The trajectory error was then recalculated using the ensemble-mean MAE.

The comparison between the original single-particle results and the 10-particle ensemble-sensitivity results showed that the absolute MAE values changed slightly, but the main optimal forcing combinations did not change. For PTM-DX, C8 produced the lowest MAE at the early time of 5 days, whereas C7 produced the lowest MAE at later times and also had the lowest average MAE. Thus, except for the very early stage, C7 remained the optimal forcing combination for DX. For PTM-DO, C8 consistently produced the lowest MAE in both the original single-particle simulation and the 10-particle ensemble-sensitivity test.

These results indicate that the stochastic diffusion term affects the exact MAE values but does not control the main conclusion regarding the optimal forcing combinations. We have clarified this point in the revised manuscript by adding the ensemble-sensitivity test in Section 3.5 (Lines 508-515).

Table R5. Original MAE table using one numerical particle per drifter initial location.

	Day	C1	C2	C3	C4	C5	C6	C7	C8
PTM-DX	5	134.5	107.3	122.7	90	106.1	57	22	**19.3**
	10	251.4	242.6	221.1	166	209.6	92.6	**24.1**	35.8
	30	475.8	460.7	308.6	209	265.6	92	**54**	91
	75	562.5	606.1	347	145	243.9	90.8	**73.8**	83.2
PTM-DO	5	75.7	64.1	79.3	68.7	68.8	59.2	68.7	**55.3**
	10	146.5	135.6	140.2	123.1	129.4	113.8	120.1	**107**
	30	318	286.5	304.9	235.6	216.9	175.5	201.4	**110.7**
	75	537	487.9	487.9	377.1	459.6	322.6	410.3	**185.2**

Table R6. Revised MAE table using a 10-particle ensemble and ensemble-mean MAE.

	Day	C1	C2	C3	C4	C5	C6	C7	C8
PTM-DX	5	130.2	128.2	123.1	89	106.3	60.9	25.3	**20.4**
	10	252.4	243.2	223.7	162.3	207.1	104.5	**29.5**	43.9
	30	473.4	449.3	300.2	216.4	260.2	104.6	**55.6**	114
	75	550.2	541.3	290.9	164.6	219.8	98.3	**61.2**	105.4
PTM-DO	5	70.5	64.8	77.4	67	67.5	57.8	67.5	**53**
	10	142.9	133.5	143.1	125.7	129.4	114.3	120.1	**106**
	30	315.6	273.7	296.2	276	211	166.6	194.9	**117.7**
	75	528.1	488.7	477	472.9	359.2	285	330.9	**182.6**

5. Line 147: “each step ormal distribution”

Thank you for pointing this out. We have corrected it.

6. Line 148-150: Windage is here said to be part of “total current velocity” - this should probably instead be “total drift velocity”.

Thank you for pointing this out. We have corrected it.

7. Line 153-158: Some more details about calculation of r would be welcome, e.g. what values are used for Cd,a and Cd,w.

Please see the answer to Q2.

8. Line 162: Reference to Table 1 - but both this and Table 2 are missing!

Thank you for pointing this out. We apologize for the missing information about Tables 1 and 2.

9. Line 174: the distance of a degree of longitude is different from a degree of latitude. Thus, the given MAE will “punish” longitudinal errors more than the same latitudinal errors.
Thank you for this comment. We agree that one degree of longitude and one degree of latitude do not correspond to the same physical distance. Although the MAE equation is written in terms of longitude and latitude differences, the calculation was not performed directly in

angular units. Before computing the MAE, both ΔLon and ΔLat were converted to physical distances in kilometers, accounting for the latitude-dependent length of a degree of longitude. The MAE therefore represents the trajectory separation in physical distance space and is not biased by the different spatial scales of longitude and latitude. We have clarified this (Lines 240-241).

10. Line 195: Please explain what is the "radially symmetrical variance"

Thank you for pointing this out. The term "radially symmetrical variance" was unclear, and we have revised the manuscript to define it more explicitly. Following Okubo (1971) and LaCasce (2008), σ_r^2 represents the horizontal radial variance of a particle or drifter distribution, defined as the mean squared distance of particles from the center of mass of the distribution. In other words, an irregular particle cloud is represented by an equivalent radially symmetric spreading measure, and σ_r^2 quantifies the radial spread of that cloud. It was used to estimate the horizontal dispersion coefficient from the temporal growth rate of the particle-cloud variance (Lines 261-264).

11. Line 202: For FSLE-figures below, I assume L/delta equals delta_0, but what is the value of delta_f ?

Thank you for pointing this out. We agree that the definition of the final separation distance was not sufficiently clear. In the FSLE calculation, δ_0 denotes the initial separation scale, and δ_f is the threshold separation distance used to define the growth time. In this study, δ_f was not prescribed as a fixed absolute distance; instead, it was defined relative to each initial separation scale as $\delta_f = 2\delta_0$. Thus, the FSLE was calculated as $\lambda(\delta_0) = \ln(2)/\langle\tau(\delta_0, 2\delta_0)\rangle$, where $\tau(\delta_0, 2\delta_0)$ is the time required for a particle pair separation to grow from δ_0 to $2\delta_0$. We have revised the manuscript to explicitly state this definition and to avoid ambiguity in the FSLE formulation (Lines 269-273).

12. Line 208: Heading "Results, or a descriptive heading about the results"

Thank you for your comment. We changed the section heading to "3. Results".

Section 3.1: The bifurcation is discussed, but as mentioned above, it is not easy to assess this without seeing the actual currents (including CMEMS and HYCOM) zoomed at the time and location of the passing of this point.

Thank you for this insightful comment. We agree that, because GLORYS and GOFS/HYCOM assimilate satellite altimetry, their sea-surface-height fields may contain information similar to that used to derive the altimetric geostrophic velocity. However, assimilation of sea-surface height does not necessarily ensure that the resulting near-surface velocity gradients, finite-time transport barriers, and hyperbolic-type Lagrangian structures are reproduced in the same way.

To examine this point, we added a direct comparison of FTLE structures computed from the data-driven velocity fields and the model-derived velocity products (Figure 10). In the data-driven cases, particularly for G, GES, and GESW (Figure 10a1–a3), the FTLE fields show a clear hyperbolic-type structure near the region where the observed drifters bifurcated. The

attracting and repelling FTLE ridges form a coherent transport-separating structure, consistent with the observed early separation of the drifter pathways.

In contrast, the corresponding FTLE structures obtained from CMS/GLORYS and HYC/GOFS are weaker, displaced, or less coherent in the bifurcation region (Figure 10b1–c3). Adding windage and Stokes drift to the model-derived velocities modifies the FTLE ridge patterns, but it does not fully recover the same transport-separating structure shown by the data-driven velocity fields. This indicates that the poorer performance of the model-derived simulations is related not only to differences in velocity magnitude, but also to differences in the finite-time Lagrangian structure controlling the early trajectory separation.

Therefore, although GLORYS and GOFS/HYCOM assimilate sea-surface height, the model-derived surface velocities do not necessarily reproduce the same geostrophic velocity gradients and hyperbolic-type transport structure as the altimetry-derived geostrophic product. We have clarified this point in the revised manuscript and added Figure 10 to support the interpretation (Lines 565-591).

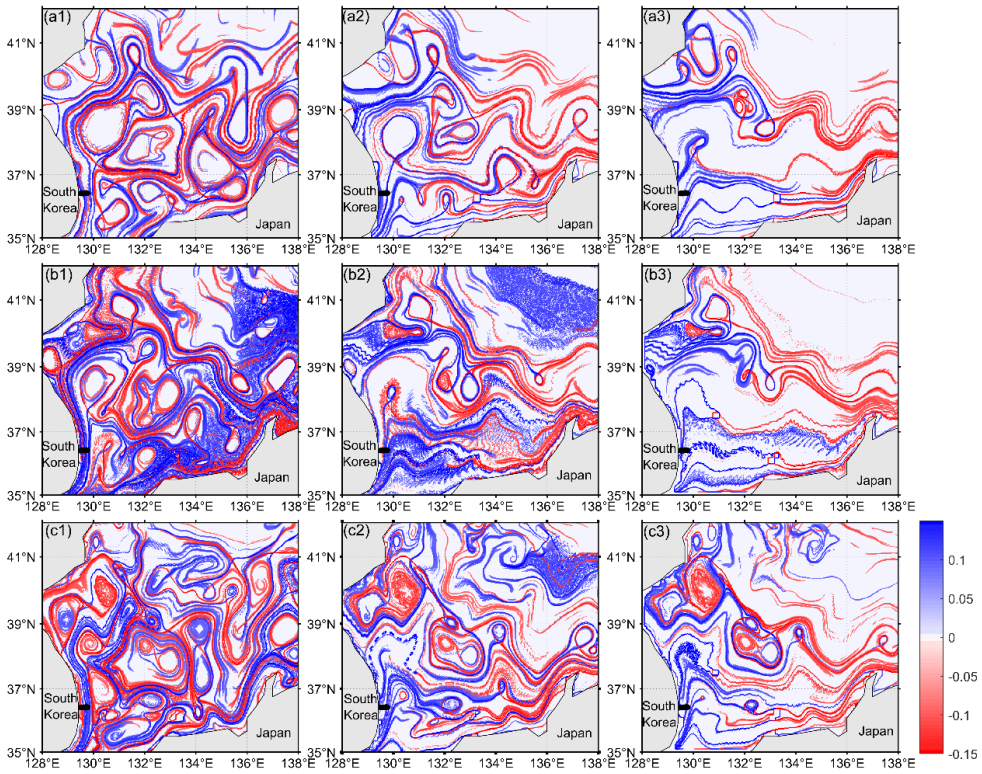

Figure 10. Finite-time Lyapunov exponent (FTLE) structures computed from one-week-averaged velocity fields. The panels show FTLE fields for (**a1**) u_g, (**a2**) $u_g + u_e + u_s$, (**a3**) $u_g + u_e + u_s + u_w$, (**b1**) CMS, (**b2**) CMS-w1, (**b3**) CMS-w1-st, (**c1**) HYC, (**c2**) HYC-w1, and (**c3**) HYC-w1-st. Here, u_g, u_e, u_s, and u_w denote geostrophic current, Ekman current, Stokes drift, and windage, respectively. CMS denotes the GLORYS reanalysis distributed by Copernicus Marine, and HYC denotes the GOFS analysis based on HYCOM.

13. Line 224: current are white vectors, not black.

Thank you for your comment. We have corrected the Figure 2.

14. Line 318: A model grid resolution of 25 km is mentioned. Which model does this refer to? The geostrophic current used has a grid spacing of 0.125 degrees.

Thank you for pointing this out. The “25 km grid resolution” mentioned in the manuscript refers to the nominal grid spacing of the gridded geostrophic velocity field used in the PTM calculations, not to the effective physical resolution of the altimetry product. At the time the trajectory calculations were performed, the geostrophic velocity field used in the PTM was on a 0.25° grid, which corresponds to approximately 25 km in the East/Japan Sea. We agree that this wording was unclear, particularly because the current Copernicus Marine SEALEVEL_GLO_PHY_L4_MY_008_047 product is now distributed on a 0.125° grid. We have therefore revised the manuscript to distinguish the nominal grid spacing used in the PTM calculations from the current product grid description (Lines 141-143, Lines 418-419, Lines 425-429).

15. ‘Figure 11 text: “The raletive”

Thank you for pointing this out. We have corrected the typo.

References

de Dominicis, M., Bruciaferri, D., Gerin, R., Pinardi, N., Poulain, P.-M., Garreau, P., Zodiatis, G., Perivoliotis, L., Fazioli, L., Sorgente, R., and Manganiello, C.: A multi-model assessment of the impact of currents, waves and wind in modelling surface drifters and oil spill, Deep Sea Research Part II, 133, 21–38, 2016.

van der Mheen, M., Pattiaratchi, C., Cosoli, S., and Wandres, M.: Depth-dependent correction for wind-driven drift current in particle tracking applications, Frontiers in Marine Science, 7, 305, 2020.

McCaffrey, L. and Koeberle, A. L.: An economical open-source Lagrangian drifter design to measure deep currents in lakes, Water Resources Research, 60, e2024WR037429, 2024.