Overall comment:
Note that Fig. 10 and some values in sections 5 and 6 have slightly changed due to using an older version of a data file. Now all time scales, diffusivities and R(0) are consistent throughout the paper. Note also that for better readability we have divided Section 3 (Theory and Methods) into subsections. For further clarity we have also reorganized Section 5 (The effects of the mean flow on eddy diffusivities) with respect to subsections.

<u>Editor Comments</u>

We thank the editor for their helpful comments and suggestions to improve our manuscript. Please find our point-by-point responses to your comments below. Editor comments are shown in black and our responses are shown in blue. Edited text in the manuscript is shown in purple. Unchanged text that has been copied in for completeness is shown in gray.

I'm only doing writing comments:

* "x" and "y" are as far as I can tell not explicitly defined in the text, although there are implicit references suggesting these are "zonal" and "meridional" respectively. I raise this because since the work talks about major and minor axis and mean flow suppression, "x" could plausibly mean "direction along the mean flow". There is also "i,j=1" to mean "zonal" which doesn't help.

Much safer to explicitly define them. Would suggest doing "i,j=x,y", then defining "x" and "y" here. Otherwise, define them around Equation 12
We are not sure where exactly the editor refers to with 'here'. According to the editor's suggestion we defined the terms x and y just after equation 12. L206: … where x is the zonal direction and y is the meridional direction

* abstract: Consider rewriting abstract such that references to "xx" "xy" etc. are dropped completely, unless the authors decide the directions are going to be defined explicitly here. (The investment in words to make the details make sense self-consistently within the abstract is, to me, not entirely worth those details being in the abstract at all.)
We agree with the editor and adjusted the abstract accordingly. Please see the changes made to the abstract in the tracked changes document.

* general: there are various places where the sentences are long, and could really do with adding some commas to break it up a bit (e.g. line 54 after "T", line 168 before "which", some others in the results sections I didn't note down when I went through the article)
We have corrected the suggested lines accordingly. We further checked the results section and added commas where helpful to ensure a smooth read through. We hope that unrecognised comma issues will be resolved during typesetting.

* line 99: formatting, "Grisel et al (2010), Grisel et al (2014) AND Chen et al (2014, 2015)" or similar
Changed accordingly.

* line 110: add an "and" after the comma (like the second bullet point in line 112-114)
Changed accordingly.

* line 142: footnote unnecessary and inconsistent, since acronyms are defined in text for all other cases. Change to "Coordinates Ocean-ice Reference Experiments (CORE)" probably
We apologize for the inconvenience and changed the footnote according to editors suggestion.

* line 147: formatting and brackets, "...Runge-Kutta scheme (Grisel et al, 2010)."
Changed accordingly.

* line 150-151: remove ", hereinafter refereed to as OSCAR,", since the use of acronym makes that part redundant
Removed.

* captions of Fig 1: "for the 28/11/2016" and "is shown 28/11/2016" doesn't make sense as is, edit accordingly. ("for the DAY OF" and "is shown FOR THE DAY OF" maybe)
Changed accordingly.

* line 155: "in-situ" with a hyphen?
Changed accordingly.

* Eq 3 and 4: consider using

\left\langle symbols \right\rangle to have full height brackets.
Changed accordingly.

* line 201: "...Rypina et al (2012) AND Grisel et al (2010)..."
Changed accordingly.

* Eq 17: full stop after the equation
Changed accordingly.

* Eq 19: comma after equation
Changed accordingly.

* line 228: no indent (probably one too many empty lines in LaTeX)
Indent removed.

* line 484: formatting of the degrees symbol inconsistent with the one at line 485
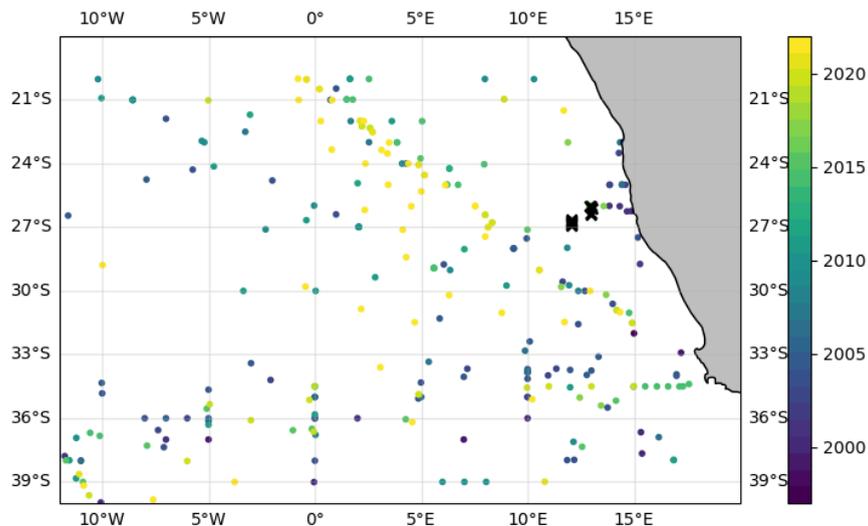Changed accordingly.

Reviewer 1

We thank reviewer 1 for their very helpful comments and suggestions to increase the overall clarity and readability of our manuscript. Please find our point-by-point responses to your comments below. Reviewer comments are shown in black and our responses are shown in blue. Edited text in the manuscript is shown in purple. Unchanged text that has been copied in for completeness is shown in gray.

This is my first review of a manuscript by Oelerich et al. titled "An estimate of the eddy diffusivity tensor from observed and simulated Lagrangian trajectories in the Benguela Upwelling System". The authors use observed and simulated Lagrangian particle trajectories to estimate the diffusion tensor in the Benguela upwelling region. The authors find that the background mean flow significantly impacts the trajectories and reliable diffusion tensor can only be estimated once the mean flow impact is removed. Likewise, the authors find that motions below 0.3 deg (unresolved by the OSCAR product) contribute significantly to the estimated diffusivities. Although some questions remain as the results are based on relatively short timeseries, I think the manuscript is well written, methods are well established, and the results support the conclusions drawn by the authors. Therefore, I only have a few editorial comments and consider this to be a minor revision.

General

The authors use a drifter dataset that they have contributed in collecting. However, these drifters were all deployed around the same time, and I wonder if the Global Drifter Program has data in the region from other years? I think the manuscript is okay even without additional data, but I would imagine that downloading a few additional GDP trajectories and repeating the computation would be quite straightforward and would make the manuscript more robust and likely more influential.

We appreciate the reviewer's suggestion to expand our analysis to cover additional years. While we would have gladly included more years in our study, the available drifter data in or near the Benguela Upwelling System is unfortunately very limited. The GDP dataset contains 300 to 350 drifters released between 1995 and 2022 (see figure below) in the Benguela Region. However, only a small fraction of these drifters were deployed near our specific study location or in the vicinity of the upwelling region. Moreover, the number of drifters for each deployment is limited and the number is not enough to provide statistically significant results for our analysis.

Map of deployment locations of all drifters in the GDP data set colour coded by year of deployment. Black crosses indicate the location of the drifter used in the manuscript.

Also, the POP simulation is for a different year than the trajectories/OSCAR data. It is likely that there is a fair bit of interannual variability (see e.g. https://www.science.org/doi/10.1126/sciadv.aav5014), and, therefore, it doesn't make sense to make any direct comparison between the model results and the observational results. It would increase the robustness of the results, if the authors would, for example, show timeseries of (E)KE or some other meaningful parameter/index in the region, such that the reader gets an idea how 'normal' the years 1996/2016 might have been (altimetry derived eddy trajectories might be a source for interesting data – see also the comment below). The OSCAR product is available for both 1996 and 2016, so one could, for example, show the (E)KE in the upwelling region throughout the OSCAR timeseries.

We agree with the reviewer regarding the interannual variability present in this region, which is largely driven by the influence of Agulhas rings. Consequently, there are discrepancies between the POP simulation and the altimetry product, primarily due to differences in spatial and temporal resolution. Moreover, it is important to note that the altimetry era only began in 1993, which poses additional challenges when attempting to reproduce conditions for 1996 and 1997. Specifically, as we go further back in time, the spatial resolution of the altimetry data decreases, making direct comparisons increasingly difficult and less robust. However, we would like to point out that the kinetic energy (KE), both in terms of the temporal mean and snapshot values from the POP simulations for 1996/1997, shows strong similarity to the observations from the altimetry product in 2016 (Fig. 2). These similarities provide a reasonable justification for making this comparison, as it suggests that the key dynamic features of the region are being captured consistently across these datasets, despite the temporal gap.

In addition, as suggested by the reviewer we have included a new Fig. 3, which compares the KE in the OSCAR product and the POP simulation. For further clarification we added the following text to the manuscript:

L286-297: The POP simulation with numerical trajectories is only available for 1996/97, while the observational dataset was obtained in 2016/17. Regardless, as Fig.2 illustrates, the horizontal distributions of kinetic energies are similar. There is visible interannual variability for the region (Fig.

3a) that is largely driven by the variability of the Agulhas rings. While there are discrepancies between OSCAR and POP particularly around 2004, the area-averaged energy levels in 1996 in POP (red solid and dashed lines in Fig. 3a) and in 2016 for the OSCAR product (black solid and dashed lines in Fig. 3a) are similar, and so are the total kinetic energies of POP and OSCAR averaged along the drifter locations (red and cyan diamonds in Fig. 3a respectively). However, the total kinetic energy of the surface drifter dataset at hourly resolution averaged over the surface drifter locations in the observations is substantially higher (black diamond in Fig. 3a). This is largely due to inertial oscillations, visible as semidiurnal oscillations in the time series of the observational drifter velocities (black lines in Fig. 3b,c) as compared to the velocities obtained from POP and OSCAR that are similar in magnitude (red and cyan lines in Fig. 3b,c) but do not capture motions on these time scales.
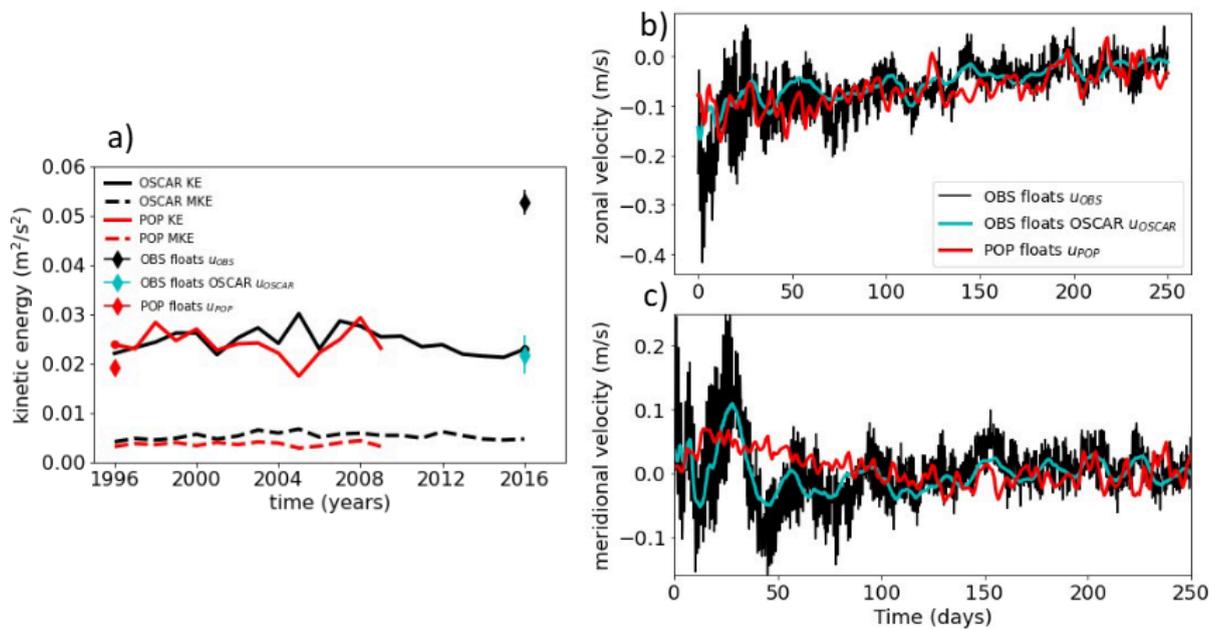

Fig. 3: The time series of annual mean kinetic energies for POP (red) compared with the OSCAR product (black) averaged over the Benguela upwelling region (as shown in Fig.1) are shown in panel a), where the total kinetic energies (solid lines) and the annual mean background flows (dashed lines) are displayed. Additionally, the total kinetic energies averaged over all float locations for POP, the observations and the OSCAR velocities interpolated to drifter locations (red, black and cyan diamonds) are shown. The time series of zonal and meridional velocities averaged over all drifters are displayed in panels b) and c) for POP (red), the observations (black) and OSCAR velocities interpolated to drifter locations (cyan).

The authors maybe interested to read a recent paper by Zhang and Wolfe https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS004004 in terms of timescales. In the future, it could also be interesting to estimate diffusivity from eddy tracks in the Benguela upwelling region.
We thank the reviewer for this reference, which is indeed interesting. We are now citing this study in the discussion and conclusions section.

L608-610: A promising approach was recently suggested by Zhang and Wolfe (2024) who have shown that eddy diffusivity can be inferred from the eddy tracks without the need for drifter deployments or altimetry-derived geostrophic velocity measurements.

L225 This might be a naive question, but is K really symmetric by construction? I have worked before only with diffusivity tensor inversion, and then one needs to still separate the symmetric and antisymmetric part of the tensor. So, I am wondering that if the authors actually rotate K by the angle

alpha (equation 20), are the off-diagonal components 0? It is not obvious to me that the equation 18 would automatically produce a symmetric tensor.

It is true that diffusivity tensors usually used in the parameterization of eddy effects in ocean models can be divided in symmetric (eddy-driven diffusion) and antisymmetric (eddy-driven advection) parts as outlined in in the introduction (L51-60). Our diffusivity tensor here is indeed symmetric and parameterizes only the horizontal diffusion. We have added the following line to improve clarity.

L253: The diffusivity tensor κ, which is symmetric by construction, can be diagonalized…

L475 The last two sentences of this paragraph are a bit puzzling. I would probably rather use 'entrain' than 'pull' (I guess it is still the wind driven upwelling that brings the cold water to the surface, and these waters are then entrained into westward propagating eddies). It is also somewhat well known that the eddies organize into bands (see e.g. citations below) and I wonder if these results also reflect this fact. i.e. there are mixing barriers (fronts) linked to preferred eddy paths.

We agree with the reviewer and have replaced all 'pull' with 'entrain' as suggested (L305,527,644). It is true that the filaments in the Benguela Upwelling region are mainly zonally oriented due to the action of often counter-rotating eddies that entrain the cold water, and associated with the filaments are also zonally oriented jets at the eddy rims. As far as the so-called striations (alternating zonal jets) are concerned, they are features that usually appear after time-averaging (and additional low-pass filtering). When we average the velocities in time we arrive at Fig. 2a,b). One could argue that there are zonal bands of jets that might be associated with preferred eddy paths (Anticyclonic Agulhas eddies on the one hand and cyclonic eddies that originate from the upwelling region on the other). However, these jets are not alternating in sign and it is unclear whether they would be called striations. We feel that the discussion of striations goes beyond the scope of this paper.

https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2016JC012348 (recent paper in upwelling region of Chile)
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2005GL022728 (original paper)
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014JC010088 (bands in SST)
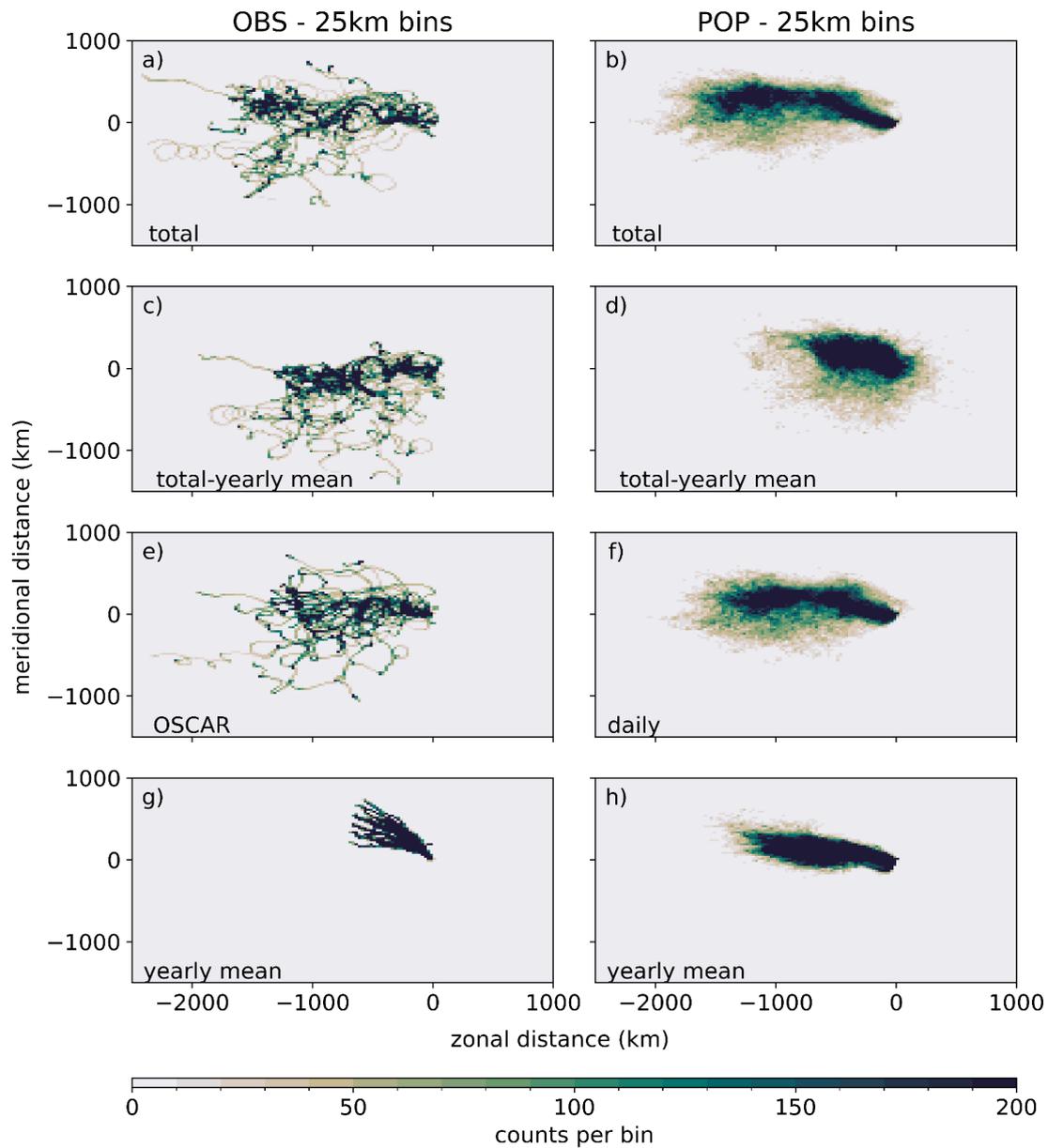
Figures:

Figs. 4-5.: The labels are confusing, I would suggest 'U included' and 'U subtracted'

We agree with the reviewer and have changed the labels to a concise and consistent format. We have included a new Table 1 that clarifies the different flow components used in this study and have consistently labeled all figures throughout.

I wonder if instead of using the individual trajectories in Fig. 3, it would be more informative to show heat maps/contours of trajectory concentration (bin the count of the trajectories). Especially for POP, there are so many trajectories that it is hard to say if there is an underlying pattern in the saturated parts of the figure.

We see the difficulty in visibility of underlying patterns the reviewer addressed. According to the reviewers suggestion we produced a figure showing the trajectory concentration in 25 km bins.

Number of trajectories in each bin over all timesteps and trajectories with a used bin size of 25 km for observations (left column) and 10km for POP (right column) for each type of trajectory: full (a,b), total-yearly mean (c,d), OSCAR (e), POP daily (f) and yearly mean (g,h).

However, we have decided to produce a new version of Fig. 3 (now Fig. 4) showing a subset of the drifter trajectories from the POP Simulation (every 5th trajectory) to improve visibility on underlying trajectory patterns. Note that the coloring of trajectories was chosen such that every trajectory has a slightly different color in the range of blue or red so that patterns of individual trajectories can be shown more clearly. We also produced a new version of Fig. 7 (now Fig. 8) showing the pseudo-trajectories from the integration with the residual velocity u'' to improve the visibility of underlying patterns.
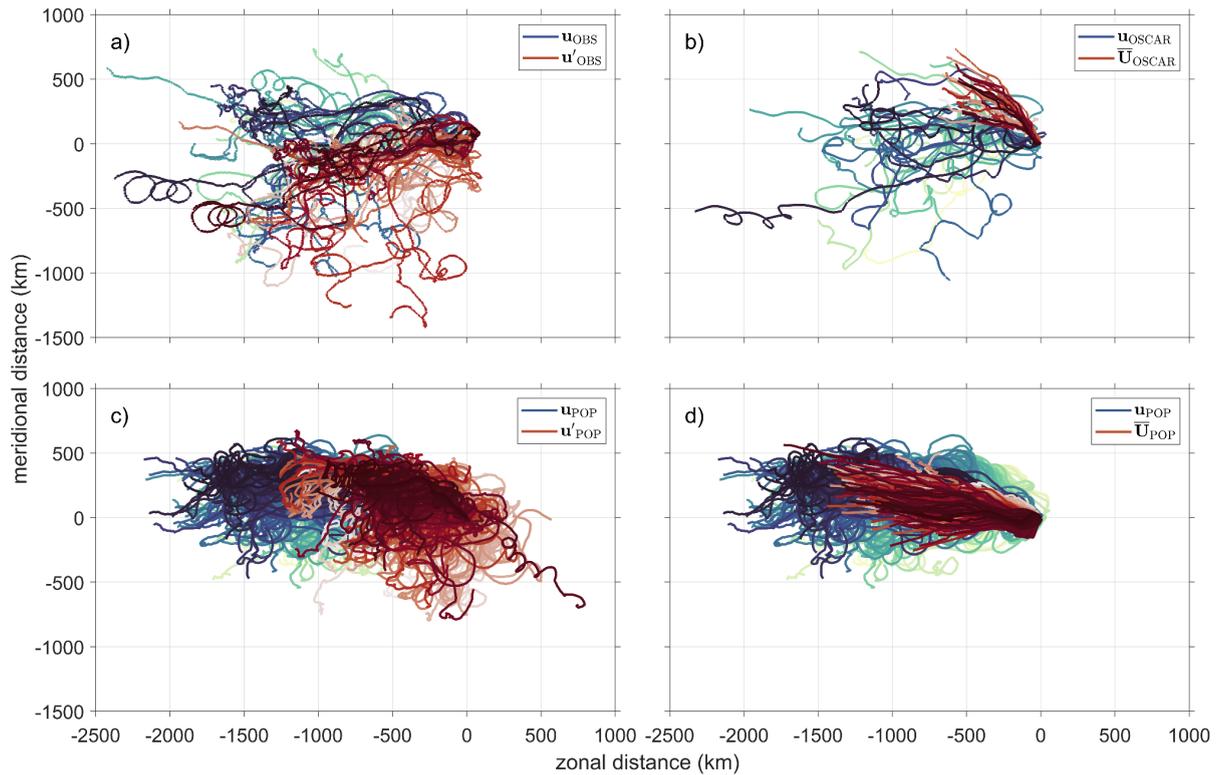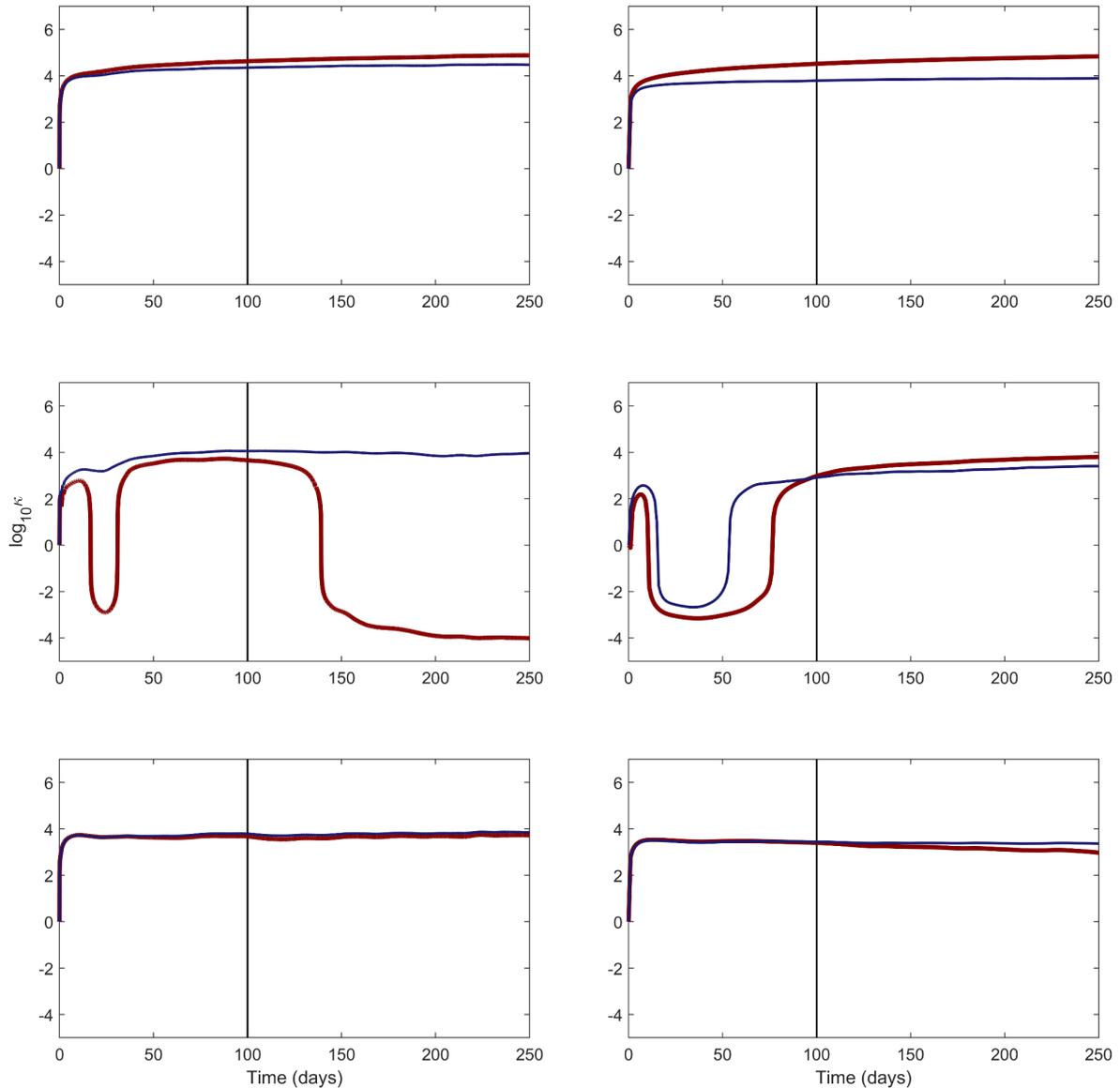
Fig. 4: The pseudo-trajectories (compare Equation 13) for the observations (a,b) and the POP simulation (c,d). Note that all trajectory components are shown in the cartesian coordinate system. For the POP every 5th trajectory is shown to improve visibility of underlying patterns. The original trajectories are marked in shades of blue (a,c) and the pseudo-trajectories from the integration with u' (Equation 12) are marked in shades of red (a,c). The pseudo-trajectories from the integration with the OSCAR surface currents u_{OSCAR} and the POP simulation daily mean velocities u_{POP} are highlighted in shades of blue (b,d) and the ones with the annual mean velocities interpolated to the drifter locations $\overline{U}|r(t)$ are shown in shades of red (b,d). All pseudo-trajectories have been constructed such that (0,0) is the origin. Note that the coloring of trajectories was chosen such that every trajectory has a slightly different color in the range of blue or red so that patterns of individual trajectories are highlighted.

I would suggest log scaling for all the plots with diffusivity tensor terms (in python one can use symlog to have log scaling for both positive and negative values).

Thank you for suggesting the use of log-scaling to represent the eddy diffusivities. We tested the representation of eddy diffusivities using a log scale, as recommended. However, we found that the log-scaling did not improve the visibility or interpretability of the eddy diffusivity behavior. In fact, due to the nature and distribution of our data, the log-scale representation either compressed critical features or amplified minor variations, making it more difficult to extract meaningful insights. Therefore, we have opted to retain the original linear scaling, as it provides a clearer and more accurate depiction of the eddy diffusivity patterns and their uncertainties.

Example of single-particle eddy diffusivities with log-scaling.

Style:

There are a few occasions with citations in brackets within brackets, like L107. I would suggest removing the inner brackets around the citations in these cases. We agree with the reviewer, inner brackets of citations appearing in brackets have been removed.

L505 This is a rather long sentence. I would suggest breaking it into two. We agree with the reviewer. Due to a comment from another reviewer this sentence has now changed and is also shorter. Please see lines below:

L593-595: This relates to the research question, which addresses the role of smaller scale motions that are detected by drifters but not by current altimeter products in influencing diffusivities and anisotropy. The study highlights the significant contribution of these unresolved motions.

Note that Fig. 10 and some values in sections 5 and 6 have slightly changed due to using an older version of a data file. Now all time scales, diffusivities and R(0) are consistent throughout the paper. Note also that for better readability we have divided Section 3 (Theory and Methods) into subsections. For further clarity we have also reorganized Section 5 (The effects of the mean flow on eddy diffusivities) with respect to subsections.

Reviewer 2

We thank reviewer 2 for their very helpful comments, suggestions and raised questions. Please find our point-by-point responses to your comments below. Reviewer comments are shown in black and our responses are shown in blue. Edited text in the manuscript is shown in purple. Unchanged text that has been copied in for completeness is shown in gray.

Analyzing the drifter and numerical Lagrangian data, this manuscript studies the horizontal 2D diffusivity tensor with a particular focus on the scale effect.

Also, it is interesting to know the departure of the single-and pair-diffusivity relation from the theory.

However, I am confused about issues in the manuscript, such as the definition of average.
We agree with the reviewer that the definition of average should be clarified in this study. To define the background mean flow we use a time mean, where the time interval is one year, which is larger than the typical eddy timescales in the study region. A certain amount of spatial coarse-graining is also implied since in the observations the time mean is based on 1/3° resolution (OSCAR product) and in the simulation the time mean is based on 1/10° resolution (POP simulation).

We have added a more specific description to the definition of the background mean flow in L52-53 in the introduction and also added further clarifications on the background mean flow in the methods sections (L207-211). A table (Table 1) summarizing the description of the different flow components used in this study was added.

L50-51: ... where the overline denotes an average over intervals in time and space that are larger than typical eddy scales, and ' the deviation from that average, while Q describes the sources and sinks of the tracer T.

L210-214: In the observational drifter data set $\overline{U}$ is the time mean velocity of the 1/3° OSCAR surface currents interpolated to the drifter locations, while in the POP simulation $\overline{U}$ is the time mean of the 1/10° Eulerian currents interpolated to the numerical drifter locations (Table 1). Note the current altimeter-derived products, such as the OSCAR product, could have effective resolutions in space and time more close to about 100 km and 30 days (Ballarotta et al. 2019) and thereby also imply a spatial coarse-graining.

I will recommend this manuscript's publication after addressing the following questions and comments.

1) Definition of mean.
Please see response above to the definition of averages used in this study.

Line 113: " ... since it should be independent from the mean flow"
This statement depends on what "mean" means. If a time average to decompose roughly the internal waves and mesoscale eddies then the mesoscale mean flow also introduces pair diffusion.
We agree that clarification for the averaging and mean definition is needed. Please see responses

above to the definition of the background mean flow. We have edited the research question (L116-118) to clarify that the background mean definition does not resolve the motions induced by the eddies.

L116-118: How do the diffusivities depend on eddy-mean flow decompositions considering both single and pair particle statistics, and can the pair particle diffusivity emerge as an alternative to the single particle diffusivity since it should be independent from the mean flow that does not resolve the motions induced by the eddies?

Equation (12): I am confused about the definition of U. A five-day average is used?
For the OSCAR product yes, for the POP simulation the resolution is daily. We have clarified this information by adding a table (Table 1) that states the spatial and temporal resolution of products used.

In parameterizations, Eulerian spatial average should be considered, this manuscript uses time mean of Lagrangian data. They may not have a direct link.
We use the time mean of Eulerian data together with the Lagrangian trajectories that provide space-time information. The theoretical underpinning to estimate Lagrangian diffusivities in a spatially inhomogeneous mean flow and the connection to the Reynolds-averaged tracer transport equation (1) was developed by Davis (1987, 1990). We have already pointed this out in the introduction (L95-99).

L96-100: Davis (1987) and Davis (1991) devised the underlying theory how to compute the diffusivity tensor in principle in the presence of an inhomogeneous background mean flow where, instead of diagnosing the statistics from the absolute Lagrangian velocities and displacements, diffusivities are computed from residual velocities and displacements after the Eulerian mean has been subtracted.

For further clarification we have added the following lines in the methods section:
L200-201: In that sense, the diffusivity introduced by Davis (1987) is a mixed Eulerian-Lagrangian quantity that can directly be related to the eddy tracer fluxes in a diffusive parameterization, as derived in Davis (1987).

With respect to the Eulerian spatial averaging, please see our response on the definition of the background mean flow above.

Line 506: only time average is used, right? Then how can we know the spatial information of the deformation radius? This looks like a pure conjecture.
We use a time average, but spatial information is also included (please see our response on the definition of the background mean flow above). To avoid confusion we have changed the sentence which now reads as:
L591-594: This relates to the research question, which addresses the role of smaller scale motions, which are detected by drifters but not by current altimeter products in influencing diffusivities and anisotropy, the study highlights the significant contribution of these unresolved motions.

2) Inhomogeneity influences the diffusivity a lot, especially for the current Lagrangian data, which captures information of different locations at different times. Is there any way to check the homogeneity of the residual velocity field?
We agree that inhomogeneity of the residual velocity field could influence our diffusivity calculations. To address this, we applied a bootstrapping method to ensure the robustness of our results. This approach allows us to capture any variations of the residual velocities by including them in the error calculations. Therefore, even if there are inhomogeneities in the residual velocities, they are reflected in the resulting error estimates, providing a reliable measure of robustness despite any underlying

variability. In principle, one could test the homogeneity assumption by dividing the data into spatial bins and recalculating the diffusivity for these bins to obtain the spatial distribution in diffusivities as in Griesel et al. (2010, 2014). However, there is not enough observational data for binning.

3) Figure 3: It is strange to observe that the total-mean does not equal eddy?
It is not strange because in the observations the difference total-mean are the small scale processes not resolved by the OSCAR product. To improve clarity we have now changed the figure labels. The label 'eddy only' for the observations was changed to u_{OSCAR} . In addition, the different velocity components are explained in Table 1 for consistency and clarity.

4) Line 417: What about the parameterization effect, such as the unavoidable numerical dissipation, that enters the difference between measurement and numerics?
We believe that numerical dissipation and diffusion do not impact our results. As we show in Fig. 2 the velocity fields from the POP simulation compare well with the velocity fields from the OSCAR product (L289). In addition we added a new Fig. 3 that compares the kinetic energy levels, which also match(L290-293). The numerical drifters in the POP simulation are advected and not diffused, so the overall impact of numerical diffusion on eddy diffusivity is negligible.

L288: Regardless, as Fig. 2 illustrates, the horizontal distributions of kinetic energies are similar.

L288-293: There is visible interannual variability for the region (Fig. 3a) that is largely driven by the variability of the Agulhas rings. While there are discrepancies between OSCAR and POP particularly around 2004, the area-averaged energy levels in 1996 in POP (red solid and dashed lines in Fig. 3a) and in 2016 for the OSCAR product (black solid and dashed lines in Fig. 3a) are similar, and so are the total kinetic energies of POP and OSCAR averaged along the drifter locations (red and cyan diamonds in Fig. 3a respectively).

<u>Reviewer 3</u>

We thank reviewer 3 for their thoughtful comments and raised questions to improve our manuscript. Please find our point-by-point responses to your comments below. Reviewer comments are shown in black and our responses are shown in blue. Edited text in the manuscript is shown in purple. Unchanged text that has been copied in for completeness is shown in gray.

Oelerich et al. examine a longstanding problem, attempting to estimate a subgrid eddy tracer diffusivity tensor using Lagrangian data, both from observations and an eddy-rich model experiment. They explore Lagrangian particle statistics and their relation to tensor anisotropy in the presence of a background mean flow, including the presence of flow structures and implied links between flow kinematics, transport and mixing. Overall, this endeavour is worth pursuing and the basic message - to carefully consider the form of the eddy tracer diffusivity tensor in regions of significant mean flow - is worth continued communication to the modelling community. However, I have some concerns over the clarity and substantive novelty of this message. I wonder whether the manuscript could be revised to address my major concerns, as I would like to see these results published, if possible.

Major comments:

1. Novelty: I would like to see the aspects which are novel, compared to, e.g. Klocker and Abernathey (2014; https://doi.org/10.1175/JPO-D-13-0159.1), emphasized more in this study.
We agree with the reviewer to emphasize the new results of our study more. We would like to highlight that key differences between our study and Klocker and Abernathy (2014) do exist. First, Klocker and Abernathy (2014) focus on the quantification of the meridional diffusivity and mixing suppression by zonal flow, whereas we estimate the full tensor (L108-109) and discuss the effect of the mean flow on all tensor components. As far as we know, our study is the first to quantify these diffusivities for the Benguela upwelling region. Additionally, our study compares the effect of the mean flow on both single and pair particle diffusivities for the first time simultaneously. Furthermore, we specifically quantify the contribution not resolved by the altimeter product to the diffusivity, whereas Klocker and Abernathey (2014) only used altimetry-derived velocity fields. We discussed the relation to Klocker and Abernathey (2014) in the discussion and also added a citation in the introduction with regards to mixing suppression by mean flows. We have added the following lines in our manuscript to emphasize novel aspects:

L109-111:Additionally, our study compares the effect of the mean flow on both single and pair particle diffusivities for the first time simultaneously. Furthermore, we specifically quantify the contribution not resolved by the altimeter product to the diffusivity.

In addition, we added two new tables (Table 2 and 3) that summarize the results for the different flow components in terms of mixing length theory, the first one for the major-minor axis decomposition and the second one for all the tensor components.

We discussed the relation to Klocker and Abernathy (2014) in the discussion and conclusion section. Based on the results highlighted by Table 2 and 3. The following lines have been added:

L559-581: The diffusivity can be written as the product of kinetic energy, i.e. the velocity autocorrelation at zero time lag, and the integral time scale and we summarize the results for all flow components in Table 3. The results show that after mean flow subtraction, the kinetic energies for the zonal and meridional directions become more similar to each other with ratios of 1.1-1.2 for the observations, the POP simulation and the OSCAR product. However, the zonal integral time scales are 3.6 (observations), 2.6 (POP) and 1.5 (OSCAR) times larger than the meridional ones. The reduction in time scale for the meridional direction can be related to mixing suppression by a zonal mean flow as discussed in (Klocker 2012b, Klocker and Abernathy 2014 and Griesel et al. 2015).

The degree of mixing suppression can be based on mixing-length theory but taking eddy propagation relative to the mean flow into account. In the presence of a zonal background flow, the meridional diffusivity can then be written as:

$$\kappa_{yy} = R(0)\frac{\gamma}{\gamma^2 + k^2 \overline{(U-c})^2}$$

where γ is a typical Lagrangian decorrelation time scale and is equal to the growth rate of unstable waves in linear instability theory (Griesel,2015), k is related to a typical eddy size and can be regarded as the wavenumber of maximum growth in linear instability theory (although both might differ in the presence of an inverse energy cascade), U is the zonal background mean flow, c is a typical translational speed of the eddies and α=0.35 as diagnosed by Klocker and Abernathey (2014).. It is the difference of U and c that leads to the mixing suppression effect, otherwise the diffusivity is equal to the velocity autocorrelation at zero lag times the Lagrangian decay scale γ. For the POP simulation we use the eddy sizes and translational speeds as diagnosed from an eddy tracking algorithm (Griesel, 2015), which amount to 90 km for the average eddy radius and -0.033 m/s for the zonal translational speed. With γ as the decay scale (time to first zero crossing) for the meridional direction of about 1/(12 days) and using our mean flow values averaged over the drifter trajectories, the time scale is about 3 days, which is close to the Lagrangian integral time scale obtained for POP for yy (Table 3). On the other hand, if we apply equation 21 to the suppression by the mean meridional flow, using the region-averaged meridional translational eddy speeds of 0.003 m/s and γ =1/(182 days), we arrive at a time scale of 14 days, which is indeed close to Tyy from Table 2. These time scales are also consistent with the ones obtained by Klocker and Abnernathey (204) and Ruehs et al. (2018).

To emphasize the new aspects provided by the Tables 2 and 3 we have edited and added the following lines to the manuscript:

L499-509:The diffusivities can be written as the product of major and minor axis kinetic energies and integral times scales, as in equation (7) (Table 2). It is found that after mean flow subtraction, the ratio of major to minor axis kinetic energies is 1.24 and 1.25 for the observations and the POP simulation, respectively. It is mainly the integral times scales that explain the large anisotropy in diffusivities. After mean flow subtraction, the major axis time scales are about 14 days for the observations and 12 days for the POP simulation, while the minor axis time scales are -0.21 and 3 days for the observations and POP simulation respectively. It should be noted that while the small anisotropy in the kinetic energy components is similar in the POP simulation and the observations and the major integral time scales are the same similar, the kinetic energies are larger in the

observations (Table 2) explaining the larger major axis diffusivities compared to the POP simulation. The $R(0)$ are significantly larger in the observations than in the POP simulation, even after subtraction of the OSCAR velocities (third row in Table 2). This implies that the kinetic energy associated with inertial oscillations might contribute to the diffusivities in section 5.2, even though the integral time scales for u'' are small.

L583-592:We find that the diffusivities in the observed drifter data set are larger than the ones inferred from the POP trajectories, particularly after mean flow subtraction. After mean flow subtraction, the integral times scales are slightly larger in the POP simulation compared with the observations, but overall similar (second and fifth row of Table 3). However, the velocity autocorrelation at zero lag after mean flow subtraction, is 2-3 times larger in the observations than in the POP simulation and is thus explaining the larger eddy diffusivities in the observations. This is partly due to the fact that the background flow in the POP simulation is larger and more highly resolved in the POP simulation than in the OSCAR surface currents product, hence leaving a smaller EKE residual, and illustrates the importance of using high resolution background flow components. However, also the total kinetic energy ($R(0)$ for u_{OBS} in Table 3) is larger in the observations than in the POP simulation, while the $R(0)$ when considering the contribution from the OSCAR velocities alone (last two rows in Table 3) are more comparable to the POP simulation.

L595-602: We found that these motions contribute 8% to the xx component and 42% to the yy component of diffusivities after mean flow subtraction. These small scale motions include inertial oscillations, which are hypothesized to not contribute much to the net diffusion, since they may only lead to oscillations in the velocity autocorrelation, which average out in the integral over time lag. Indeed, as Table 3 shows, the integral times scales for u''_{OBS} are only about 1 day for the xx and yy components. However, the velocity autocorrelation at zero lag is significant and is largely due to the kinetic energy in the inertial motions. Further studies with more trajectories are needed to clarify the role of inertial motions for mixing.

2. Clarification of the similarities and differences between Eulerian and Lagrangian representations of the eddy tracer diffusivity(*) tensor. Typically, climate models use an Eulerian or quasi-Lagrangian (e.g. if moving vertical coordinate) form, but much of the manuscript refers to either a Lagrangian tensor or is not clear whether Eulerian or Lagrangian is referred to. Similar goes for references to diffusivity, diffusion and similar in the Introduction, e.g. lines 70-95.
We use the time mean of Eulerian data together with the Lagrangian trajectories that provide space-time information. The theoretical underpinning to estimate Lagrangian diffusivities in a spatially inhomogeneous mean flow and the connection to the Reynolds-averaged tracer transport Eq. (1) was developed by Davis (1987, 1990). We have already pointed this out in the introduction (L95-99).

L96-200: Davis (1987) and Davis (1991) devised the underlying theory how to compute the diffusivity tensor in principle in the presence of an inhomogeneous background mean flow where, instead of diagnosing the statistics from the absolute Lagrangian velocities and displacements, diffusivities are computed from residual velocities and displacements after the Eulerian mean has been subtracted.

For further clarification we have added the following lines in the methods section:
L200-201: In that sense, the diffusivity introduced by Davis (1987) is a mixed Eulerian-Lagrangian quantity that can directly be related to the eddy tracer fluxes in a diffusive parameterization, as derived in Davis (1987).

We further discussed the different kinds of diffusivity estimation methods in the introduction.
L61-65: Diffusivities can be quantified using both Eulerian and Lagrangian methods (Griesel et al. 2014, Abernathey et al., Griesel et al. 2019). The Lagrangian approach is based on the spreading of floats or tracers as they follow the flow (Taylor 1921,1953; Nakamura 1996), whereas Eulerian diffusivities can be quantified e.g. from Eulerian eddy tracer fluxes (e.g. Eden 2006, Griesel 2014).

3. As has been stated in the other reviews, I am also concerned about the definition of the "mean" flow and associated implications.   This needs further exploration and justification. Also, the time analysis window mismatch needs addressing, as suggested in the other review, for example.
We agree with the reviewer that the definition of average should be clarified in this study. To define the background mean flow we use a time mean, where the time interval is one year, which is larger than the typical eddy timescales in the study region. A certain amount of spatial coarse-graining is also implied since in the observations the time mean is based on 1/3° resolution (OSCAR product) and in the simulation the time mean is based on 1/10° resolution (POP simulation).

We have added a more specific description to the definition of the background mean flow in L50-51 in the introduction and also added further clarifications on the background mean flow in the methods sections (L207-211). A table (Table 1) summarizing the description of the different flow components used in this study was added.

With respect to the time analysis window mismatch we would like to clarify that for the OSCAR product the resolution is 5-daily and for the POP simulation the resolution is daily. We have clarified this information by adding a table (Table 1) that states the spatial and temporal resolution of products used.

L50-51: ... where the overline denotes an average over intervals in time and space that are larger than typical eddy scales, and ' the deviation from that average, while Q describes the sources and sinks of the tracer T.

L210-214: In the observational drifter data set $\overline{U}$ is the time mean velocity of the 1/3° OSCAR surface currents interpolated to the drifter locations, while in the POP simulation $\overline{U}$ is the time mean of the 1/10° Eulerian currents interpolated to the numerical drifter locations (Table 1). Note the current altimeter-derived products, such as the OSCAR product, have effective resolutions in space and time in this region of about 100 km and 30 days (Ballarotta et al. 2019) and thereby also imply a spatial coarse-graining.

4. In several places, e.g. line 199, "submesoscale" motions are mentioned.   These usually have a relatively strong vertical velocity component, compared to mesoscale eddies.   Although a 3x3 tensor is mentioned early in the manuscript, only a 2x2 tensor appears to be diagnosed.   How should this be reconciled?   Also, are there any implications for using 2-particle pair statistics with a 3x3 tensor, e.g. that there might be insufficient statistical information to capture flow deformation gradients in 3D or even in 2D? Surely 2-particle stats only sample deformation rate gradient in one direction, don't they?

We agree that it would be useful to also estimate the diffusivities involving the vertical velocities. However, unfortunately we do not have advection by the 3D velocity field available in the observations since the drifters stay on a fixed depth of 15 m. Hence, we are restricted to observing only the horizontal mixing. In the POP simulation, the floats are advected by the 3D model fields but they mostly stay within the 10 - 20 m depth range since vertical velocities in the POP simulation are very small (submesoscale higher Rossby number processes are not resolved). With both single and

pair particle statistics we consider the dispersion in different directions. In principle, this would also be possible in 3D, except that one might need more particles than in 2D. Unless we misunderstand the question, we believe that 2-particle statistics can in principle sample the deformation rate gradient in different directions, however it depends on the number of particles.

I would be happy to review a revised version of this manuscript.
I broadly agree with the comments already made by the other two reviewers.