Response to Review # 2 for OUTPACE LD Station Paper:

We thank Reviewer #2 for taking the time to read the submitted manuscript and providing their thoughts and comments. Below is the original response from Reviewer 2 (in italics), with our own responses interspersed within. Manuscript changes are shown with additions in blue, deletions in red strikethrough.

Reviewer:
The paper assesses the Lagrangian nature of drifting sediment traps using physical data collected on a drifter and from ship observations of the surrounding waters. While the motivation of the study is sound I found the presentation difficult to follow. If the intention is to both assess the Lagrangian nature of the OUTPACE deployments and provide a generic method for such an assessment then the presentation of the method needs to be improved to enable the reader to clearly follow the approach. I consider the paper requires major revisions to address this point and hence be acceptable for publication. In revising the paper, the authors need to address:

1. The way the method section is written it is not clear that you are assessing the Lagrangian nature of the SedTrap drifter and how you use the collected data to do this. I like the use of Spice in the analysis but how is the baseline data profile determine – first hour, day? Of the SedTrap measurements? Something else? Need to clearly define the data that went into the baseline definition before I can assess the validity of the method. Sounds like you used the CTD data but I would think you should use the initial profile of the SedTrap drifter and then look at changes in the water properties of the SedTrap Drifter to determine whether the float is Lagrangian.

Response:

Due to the large number of datasets used in our analysis, we recognize that it can be difficult to follow the methods section. You read correctly, the baseline data is derived from the entire ensemble of the CTD rosette data, not the data from the SedTrap drifter itself. We understand that logically, the most direct evaluation of the Lagrangian strategy's ability to stay in one water mass is to analyze the timeseries data from the SedTrap drifter, and set a baseline from the start over a prescribed time period. First, the decision to specifically use the CTD rosette data instead of drifter data was made for a few reasons:

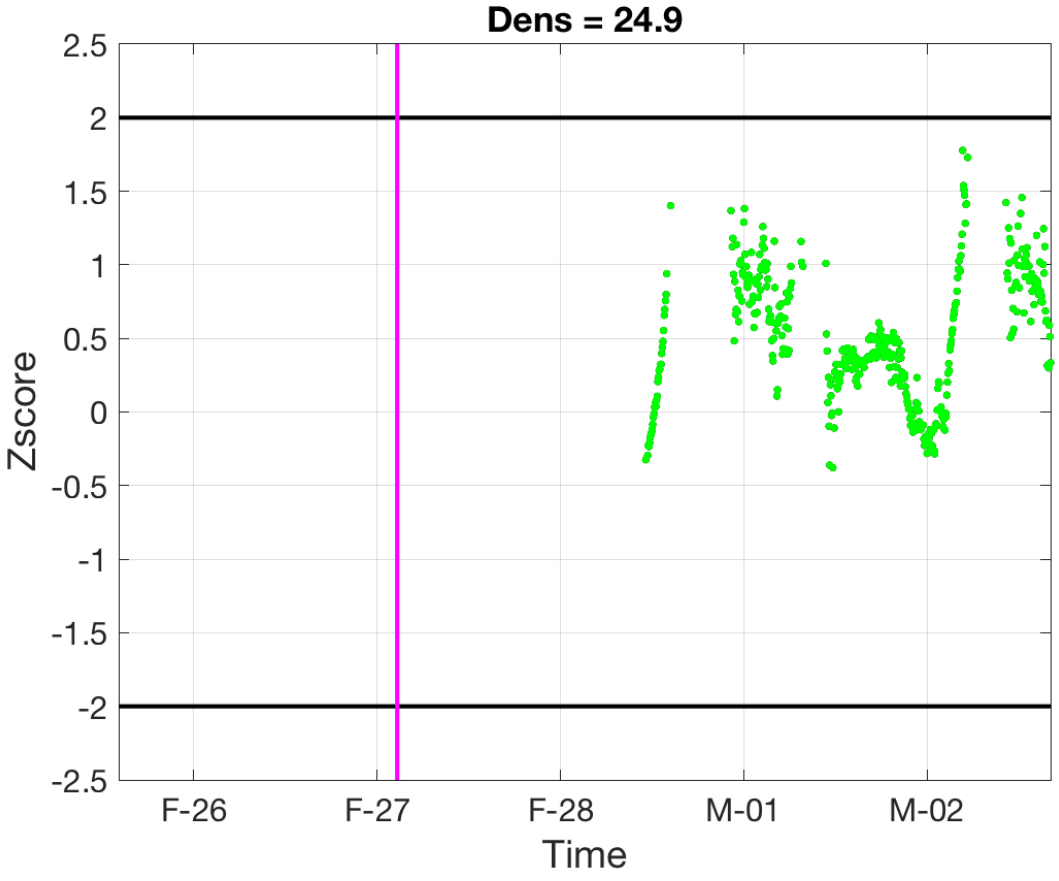
- CTD rosette data resolve a greater depth, and hence density, range and allow for comparison with the largest number of complementary datasets.
- While the station's sampling is defined by following the SedTrap drifter, the majority of the important measurements to achieve the goals of OUTPACE stem from water taken from the CTD rosette. Thus, while the drifter data are clearly important, the true ultimate need is to see whether the nearby CTD data also represent a single water mass.
- The sampling design for OUTPACE specifically included a large number of CTD casts, partly to provide the rich timeseries used in this analysis.

For these reasons, we think the CTD data is still central to our analysis, and should remain the benchmark. The suggestion by the reviewer to conduct timeseries analysis is a good one. By taking entire datasets and comparing their differences over space, our method removed time-dependence within the datasets. In constructing the methodology of this manuscript, a few underlying assumptions were applied and perhaps not adequately stated, as pointed out in another Reviewer's comments. These hypotheses include:

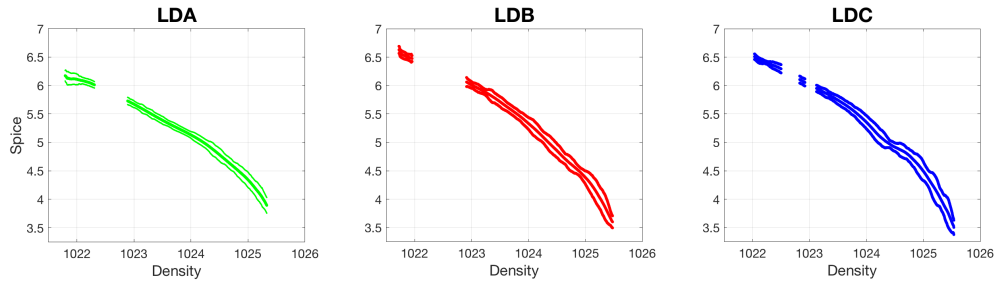
- The SedTrap drifter, with multiple objects providing drag at different depths, will clearly not be Lagrangian for a long time.
- Vertical shear will be the norm, not the exception, for currents in the Ocean. Thus, the drifter will experience water advecting past it at different rates since the drifter is not truly Lagrangian.
- Field campaigns for these drifters will not be always associated with an evident physical structure, such as mesoscale eddies or fronts/filaments, that indicates a physical and/or biogeochemical gradient (perhaps by design, as in our case). Physical variability will then be residual gradients resulting from complicated stirring not readily discernable.

As a result of these hypotheses, the method was constructed to see what gradients existed and at what scales; having determined the scale, available data on currents was used to see if water from beyond this scale could plausibly have advected past the drifter. Conducting timeseries analysis on the drifter and CTD data is a pre-requisite before moving on to our methodology. Therefore, we have conducted this analysis.

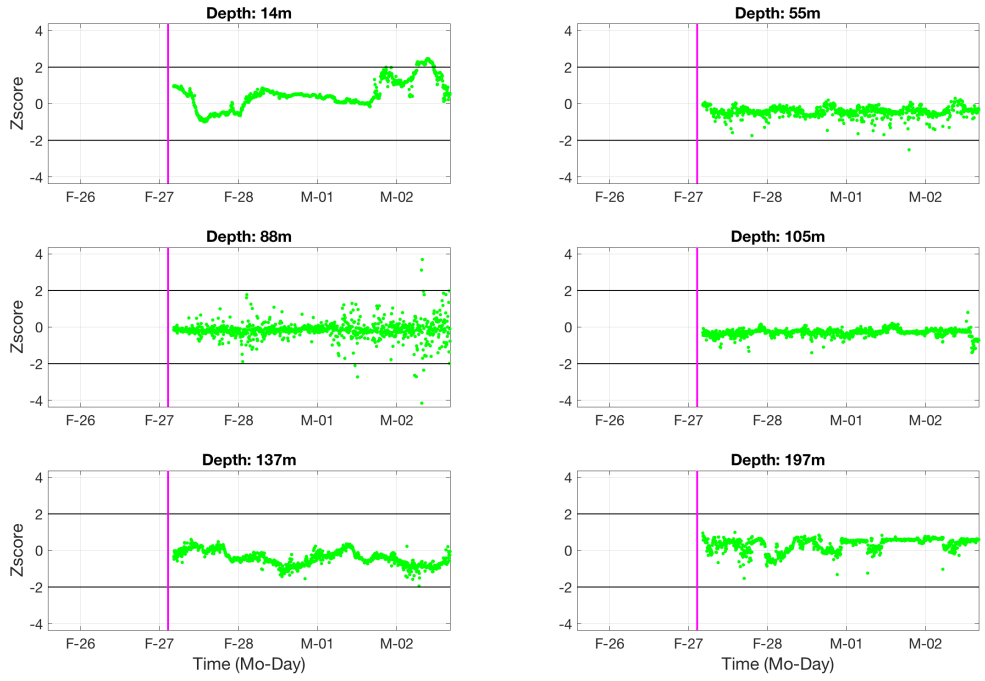
The initial time period chosen for establishing the statistical baseline is the local inertial period (36.5, 38.3, and 38.0 hours for LDA, LDB, and LDC, respectively), so that internal wave effects are taken into account. Therefore, we have the assumption that over an inertial period the water mass has not changed in the CTD data. We will start by comparing the baseline to the Drifter timeseries. Our initial plots of the Z-score timeseries show that the drifter sensors, fixed at a given depth, experienced different densities due to internal waves (see below for an LDA example; black lines are Z=2 thresholds, magenta line indicates end-time for CTD baseline determination).



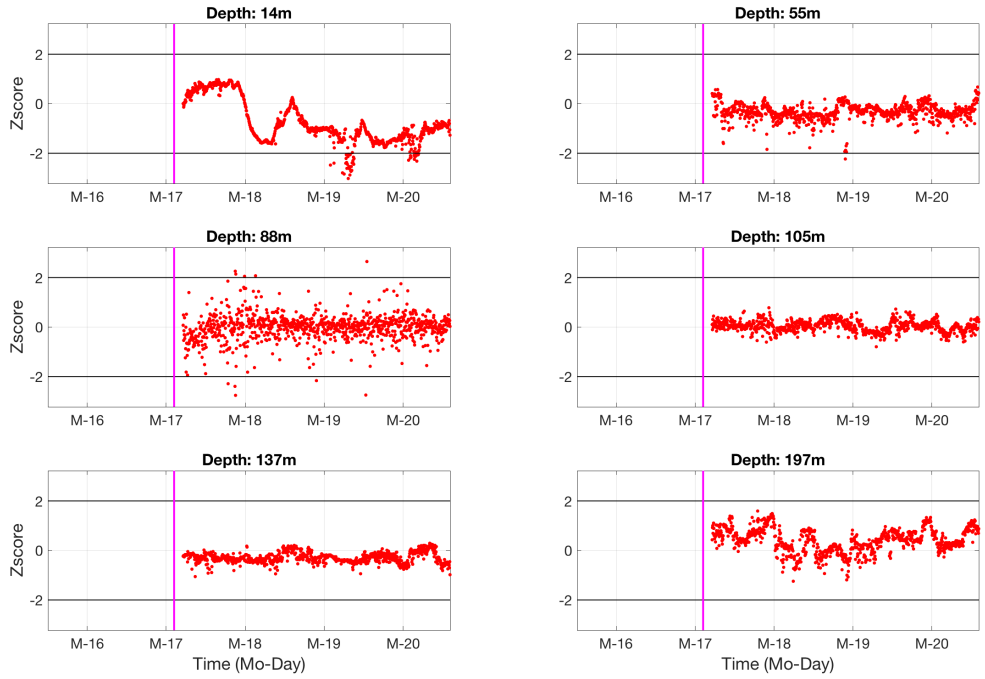
When these waves passed through, clear trends in Z-score were visible. This indicates a density-spice relationship present even at sub-bin scales. Therefore, the statistical baseline was redefined with a functional fit. This fit consisted of taking the ensemble of sorted density-spice pairs from the CTD baseline period, interpolating to a regular grid at 4 times the average density observation spacing, and smoothing with a moving average over the a bin width of +/- 0.1 kg m^{-3}. Standard deviations were calculated in a similar fashion, using observations within a moving window of the same bin-width. The new baseline distribution is shown below:



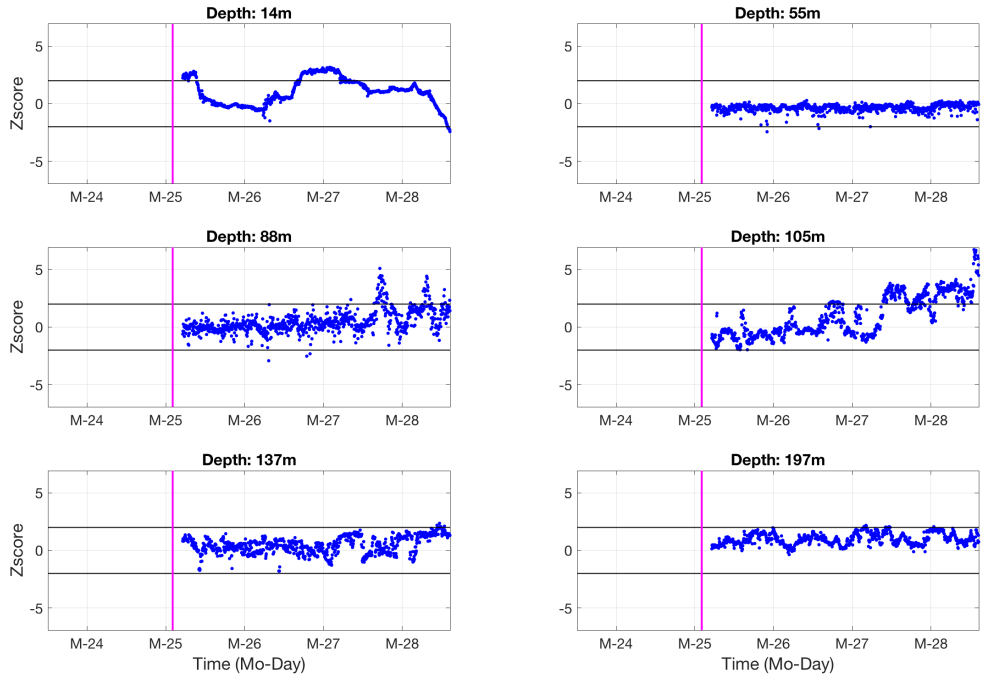
With this functional form, it is now possible to calculate Z-scores for the entire timeseries of the drifter data by looking up the corresponding functional spice value for observed density and comparing it to the observed spice. The resulting Z-score timeseries for the six drifting mooring sensors are shown for LDA below:



The sensor data generally show |Z|<2, though the sensor nearest the surface showed a departure with Z>2, though the Z-scores goes back down. For the third depth, (88 m), the data show some increasing variability with time though no trend is seen. The results for LDB are below:



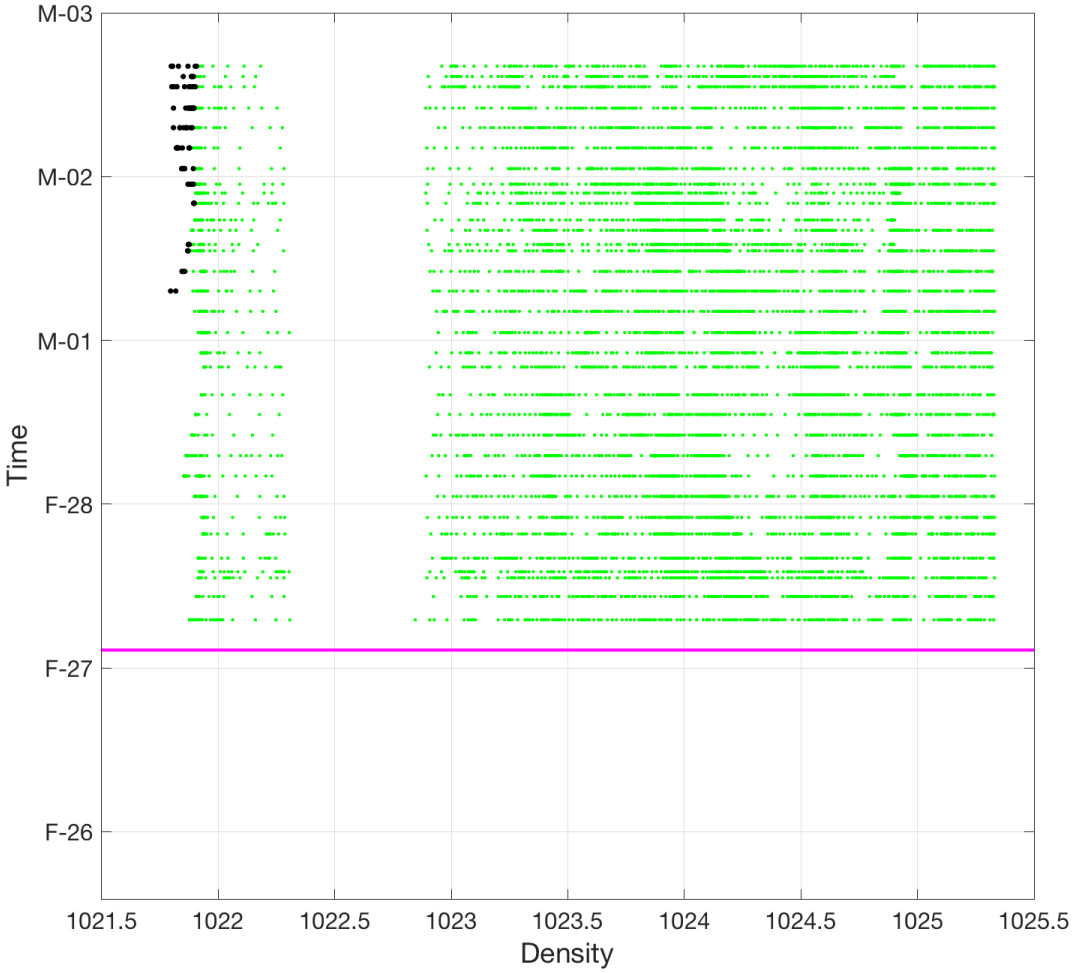
For LDB, the surface similarly shows some |Z|>2 departures, but this is not permanent; the rest of the depths show variability around Z=0, with some observations at |Z|>2, though this is likewise not an irreversible trend. Larger variability is again found at 86-m depth.

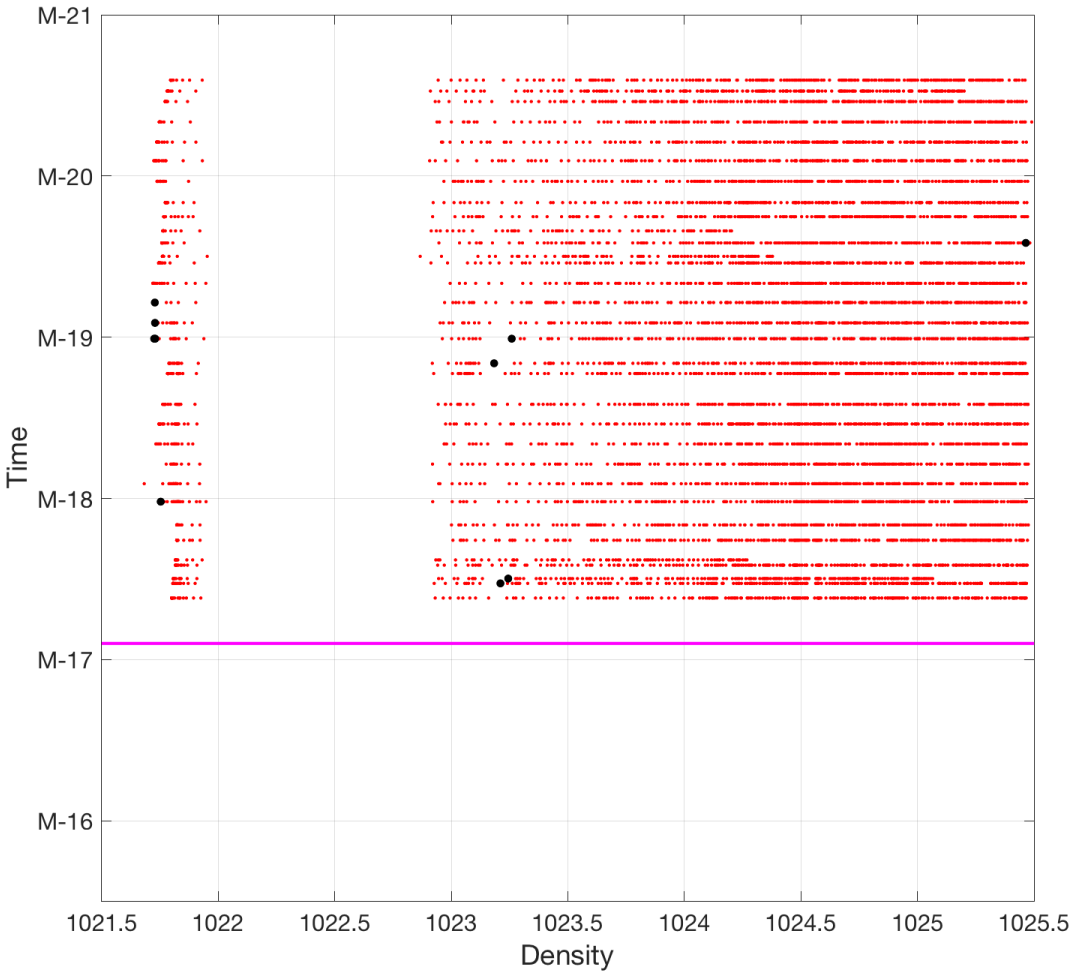


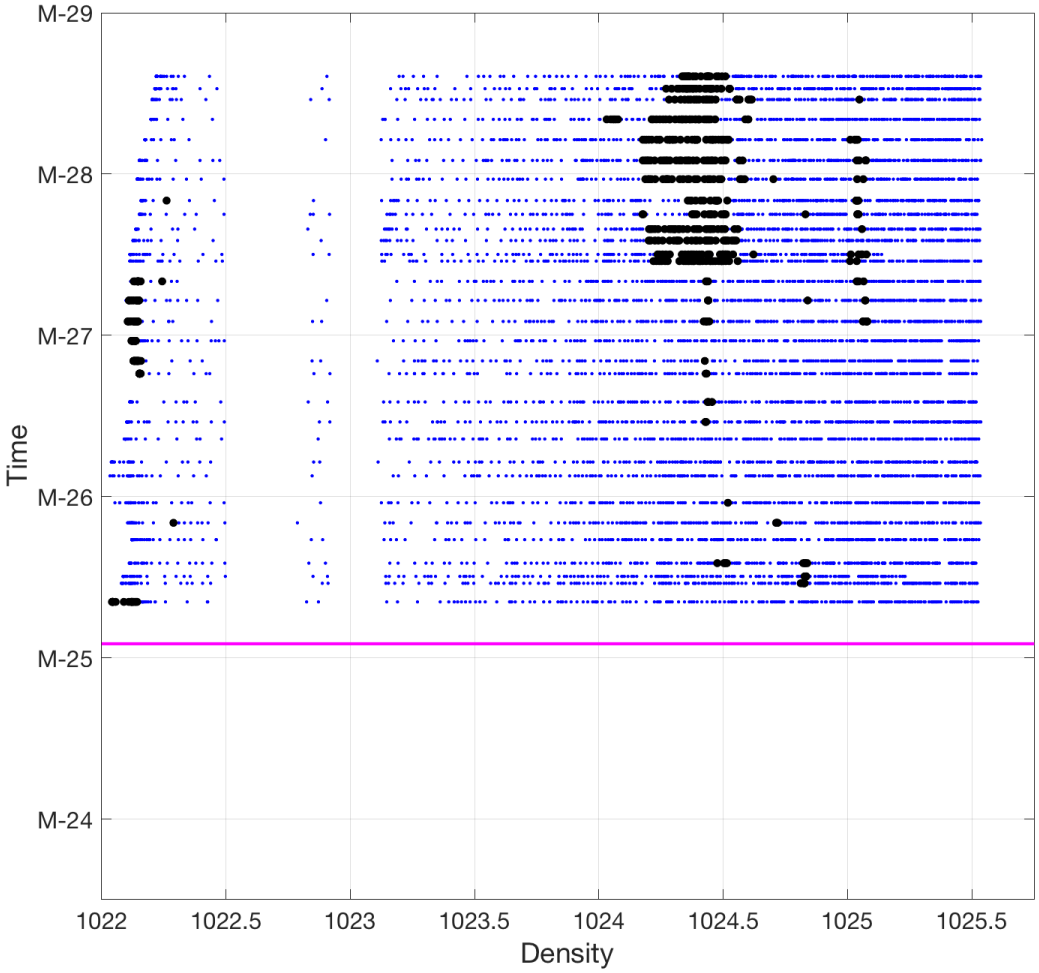
The surface for LDC is similar to the other stations. At depth, however, LDC shows a different story. At 83 m, there are large departures and a slight positive trend, and at 105 m there is a clear trend and oscillations. Therefore, it is possible that different water advected towards the drifting mooring at these depths. Fitting a smoothed version of the timeseries (2-

hour moving window), and finding where they cross |Z|=2, provides an initial timestamp of concern at 3/27/2015 16:05:01 and 3/26/2015 17:10:00 GMT for the two depths.

Performing a similar timeseries analysis for the CTD data, we find similar patterns, as shown below. All observations are plotted in green, and all |Z|>2 plotted in black.




M-21
M-20
M-19
M-18
M-17
M-16
Time
1021.5
1022
1022.5
1023
1023.5
1024
1024.5
1025
1025.5
Density



In these plots, we can see that LDA's Z-scores were close to zero, and that it is at the surface that we observe |Z|>2, similar to the SedTrap Drifter. During LDB, Z-scores were generally muted, with few observations surpassing |Z|>2. LDC shows a similar trend in increasing Z and variability as in the Drifters, and we can see that the variability is for densities between 1024 and 1024.5 kg m^{-3}, as well as a peak near 1025 kg m^{-3}.

We can also see that besides the time-dependence of the large Z-scores, that more and more density layers begin to be affected, starting halfway through March 27, 2015. Likewise, the peak near 1025 kg m^{-3} begins around this time. Previously in our manuscript, we accepted this variability as part of the baseline's definition, and noticed that since the drifters had lower Z, that this would be acceptable. We thank the reviewer for their question that allows us to better understand the time-dependent nature of the variability.

These new findings require alterations to the Methods, Results, and Discussion. Since our responses to the reviewer's other points also require changes to these sections, the detailed amendments will be appended at the end. Here we provide a quick preview summary: Methods will be re-organized to describe the time-series as a first step in the analysis. Results will include the discussion above, with accompanying figures. The Discussion will include the implications of the gradients found during the drifter deployments.

Reviewer:

2. The second step to the method is also not clear "evaluation of whether that scale was surpassed by the SedTrap Drifter …". What does this mean, why is it useful, how is it determined, and why does it differ from looking at the changes in the water properties of the SedTrap Drifter? Please expand this discussion so I can follow how the current data collected on the different platforms are used to assess the Lagrangian nature of the SedTrap drifters. The trajectories from the different velocity datasets appear significantly different and not consistent with the drift of SedTraps, what does this mean?

Response:

As mentioned in our response to item 1, our assumption is that the drifter will not be Lagrangian for long periods. As a recourse, the original method's first step investigates at what spatial scale the water mass changes, and the second step uses complementary velocity data to see if water from these areas could have come into contact with the drifter. Logically, after conducting the timeseries analysis it may seem that if we do not observe changes in the timeseries, the analysis is finished. However, the physical changes determined by the baseline are relative and specific to the OUTPACE data. Therefore, in absolute terms, not observing a change in the timeseries by this metric does not mean that a change has not occurred.

Since we are constrained to using the cruise data (i.e. baseline), one approach to better contextualize it is to analyze data from around and outside the deployment. If a strong physical change (e.g. Z-score) is seen outside of the sampling region, but is not observed in the CTD or drifter timeseries, then we are more assured that the water mass has not changed. If there is no change over large distances where one would expect at least some change, then this calls into question the timeseries results. Ultimately, this is why extending the Z-score analysis to greater spatial scales is used to create a spatial scale: since our baseline is relative, more data from different environments is needed to shore up the timeseries conclusions.

Due to the failure of the drifter to be Lagrangian for long periods, and the presence of shear, it is plausible that certain density layers have advected from beyond the actual trajectory of the SedTrap drifter. This is why the velocity reconstruction is useful: if perceived changes in the timeseries are weak or marginal (so $|Z|<2$ or $|Z|\sim2$), then an additional check is to see if it is possible that sufficiently different water (as roughly determined by the spatial scale) could have reached the drifter. Therefore, this tests not so much if the drifter is Lagrangian, but rather that surrounding water masses did not impinge on the station's sampling in a region of apparently weak gradients.

The fact that the trajectories are different from each other and the drifters reflects the general trend that a) velocities are intensified near the surface, and b) vertical shear at different layers is leading to different trajectories. The sum of the vertical shear acting upon the drifter might lead to less net displacement, though the water passing by indeed comes from farther afield.

To clarify and better motivate also this second step, we have added the above discussion to the new version of the manuscript, in the Methods section. As before, these changes are reflected in the appended pages at the end.

Reviewer:

3. The method section also provides much additional information that does not help clarify what you are doing. What is the value in comparing to climatology for the Lagrangian assessment? Why discuss water mass breakdown if it is not used? Why even show the remotely sensed maps and evolution if they are not used to assess the Lagrangian nature of the SedTrap Drifters?

Response:

The climatology was added to show inherent variability in the region to inform whether the statistical baseline, which we have noted is relative, was reasonable. Having a second climatology, however, is perhaps not necessary, and so the second supplemental figure with the CARS data has been removed.

The water mass breakdown was mentioned to highlight the need for developing a new method in identifying the water mass seen during OUTPACE and how to establish whether it had changed during the SedTrap drifter's deployment. Since this is not necessary for the development of the method, we will remove mention of it in the Methods section.

The remotely sensed maps and evolution were included because a) as part of the OUTPACE special issue the spatio-temporal context of the three stations would be useful for readers (and authors) of the accompanying articles, and b) to demonstrate that no apparent gradients or changes (in SST or Chl-a) were visible around the SedTrap drifter. As another reviewer has pointed out, biogeochemical gradients do not necessarily coincide with physical ones. By including these data, we can show that no obvious variability was present for our stations, and suggest that the use of our method is valid.

Reviewer:

4. Is the Lagrangian nature satisfied for all depths in the SedTrap drifter deployment? One may expected a surface drifter may not represent the flow at depth. Please discuss.

Response:

The analysis conducted in response to item 1 shows that water mass changes could occur at some depths but not others (particularly at LDC!). Yes, a surface drifter may not represent flow at deeper depths, which is partly reflected by the trajectories of the SVP (bottom row of Fig. 8) against the SADCP trajectories (middle row of Fig. 8). As shown above, we have broken down our analysis to resolve depth-dependence, which will be reflected in the amended Results, Discussion, and Figures appended at the end of this document.

Reviewer:

5. The SedTrap drifters all drift less than 5.6 km – hard to think it is not Lagrangian since it barely moved. Why do analysis out to 1000 km when the drifter barely moves? Need to provide some context as to why you extend the analysis to much larger scale.

Response:

The SedTrap drifters did not drift very far, this is true. However, the trajectories calculated for different depths show that water at different depths probably did move much farther, and in different directions. The resultant drag force of these motions potentially cancel out, and so the fact the drifter did not move much does not preclude the drifter timeseries from observing water sourced from far away.

Some specific comments.

Abstract. L1-5 – not clear what a quasi-Lagrangian drifter approach is – deploy a drifter and follow it. Add a sentence to explain what this is.

Response:

We have added the following to the Abstract:

Pg1, Abstract, Line 2:

A popular experimental design is the quasi-Lagrangian drifter, often mounted with in situ incubations that follow the flow of water over time. ~~The ship then~~ After initial drifter deployment, the ship tracks the drifter for continuing measurements that are supposed to ~~remain in~~ represent the same water environment.

Pg3 line 23 – state how close the production line was to the SedTrap drifter?

Response:

Since the production line was recovered on a daily basis, it was followed closely by the vessel, and re-deployed close to the SedTrap drifter. While we do not have telemetry from the production line, the CTD positioning is the closest proxy. These distances range from 300 m to 5.7 km, averaging 1.3 km for the entire cruise, so we have added this to the manuscript:

Pg. 3, Sect. 2.1, Line 22:

The Production Line was redployed on a daily basis close to the SedTrap Drifter. While no telemetry exists for the Production Line, the CTD casts from which incubation water was drawn ranged from 300 m to 5.7 km from the SedTrap drifter. After 5 days, …

Pg4. How is the remote sensing data used to assess the Lagrangian nature of the SedTrap drifters?

Response:

As mentioned in our response to item 3, the remote sensing data were used to both justify the use of our method, and to provide context as part of the manuscript's role within the special issue of OUTPACE.

- Where do you use Mixed Layer depth in your analysis?

Response:

The Mixed Layer depth was mentioned on Pg. 11, Sect. 3.3, Lines 10-11 to discuss how it was not resolved by the ADCP trajectories.

Pg5 line8 – to make it easier to follow say you are assessing the SedTrap Drifters since it is confusing when you use the quasi Lagrangian drifting mooring.

Response:

This specific line has been removed as part of the larger changes to the Methods section, as shown above.

Line 14 – how is comparing to climatology useful for your analysis? Line 18-25 – again interesting information but not relevant here. Could go in the introduction as a way of testing Lagrangian nature of the sampled water. It is not part of your method.

Response:

As previously discussed in the response to item 3, the climatology was used to motivate the statistical baseline. The section on the water mass breakdown will be removed (see above changes).

Pg 6 line 14-25 – Z scores are functions of density do they vary significantly with density
Line 34 – does the SedTrap Drifter have greater variability than TSG

Response:

The method has now been altered to retain differences in density, and are shown in the new figures presented in response to Item 1. For density-dependence in the baseline, we refer to Fig. 5, which has been updated and included in our response to Item 1. TSG variability, being near the surface, is generally greater than that of the SedTrap Drifter at depth.

Pg 8 why is the satellite data needed?

Response:

As mentioned in our response to item 3, the remote sensing data were used to both justify the use of our method, and to provide context as part of the manuscript's role within the special issue of OUTPACE.

Table 2 variation in Dist(z=2) is huge what does this mean? It implies the calculations are very dependent on the data source? Or you have not used the data appropriately

Response:

The variation in Dist(z=2) is due to our applying multiple functional fits. We did this because we did not assume a specific functional form for the Z-score vs. Distance relationship, and some variability is to expected. Also, as can be seen in Fig. 6, the TSG data is more variable, and the Z-scores increase much faster. Since surface water (in the mixed layer) is subject to stronger forcing than at depth, this makes the two datasets complementary but not necessarily comparable. However, since we are now considering all the density layers, fitting functions to each sub-set of data is somewhat intractable to present, and with the re-analysis conducted during the response to this review, it was more practical to choose a smallest scale were |Z|>2 between the TSG and CTD data. As a result, Table 2 has been removed, and the Methods and Results have been updated (See amendments above).

Figure 1. state it is a weighting of the 42 days of data based on inverse distance squared from the ship track. What remotely sensed data is used and what is it resolution in time and space?

Response:

We have re-phrased the caption to highlight the weighting of the data corresponding to the cruise, and have added the data source:

Figure 1. Weighted average s~~S~~atellite surface (a) chl-*a* and (b) SST for the 42 days of OUTPACE ~~cruise~~. ~~Pixel~~ D~~d~~ata are weighted ~~by the~~ based on normalized inverse distance squared ~~between each pixel and~~ from the RV *L'Atalante*'s daily position ~~over the 42 days of OUTPACE~~. Shiptrack shown in white. LD station locations shown with black +'s. Domains used in Fig. 2 are shown by color-coded rectangles, with green for LDA, red for LDB, and blue for LDC. Chl-*a* and SST provided by CLS with support CNES.

Figure 2. Is there weighting of the remotely sensed data? If you need to show remotely sensed data I would use figure 2 only.

Response:

Figure 2's data is the same as Figure 1, and this will be added to the caption. The other reviewer wanted additional information that has been placed in Figure 1, so we will keep it.

Figure 3. why show this plot? It is not a Lagrangian view of the data. It simply shows the variability in the region around the time of drifter deployment. How do you use it in your assessment?

Response:

As mentioned above, the satellite data were used to justify the use of our method and provide context for the cruise. For example, the heating at the end of LDA was reflected by the Z-scores near the surface, so having an independent timeseries showing this informs the Results and Discussion.

Figure 4. Spice is not orthogonal to density – why not?

Response:

Spice is not orthogonal to density due to the aspect ratio of the plot. Properly adjusted, this becomes more apparent. Also, since density is a non-linear function of temperature and salinity, some of the inherent curvature makes it difficult to see the orthogonality. Please see the figure below showing a zoom-in and changing of the aspect ratio that demonstrates the orthogonality.

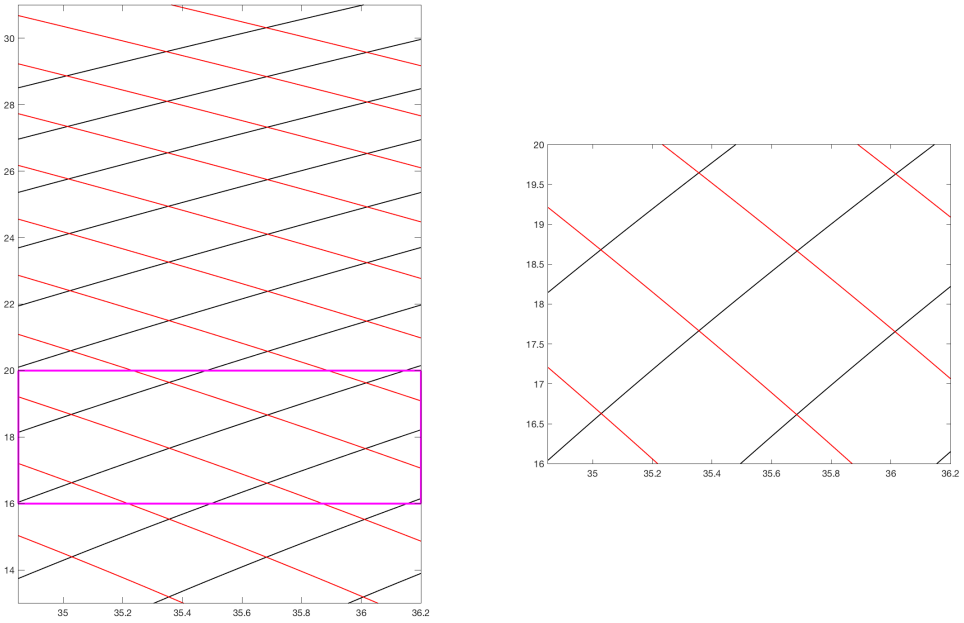


Figure 6. How do you measure distance for SedTrap drifter? How is distance defined for the other data? Why go out to 1000 kms when the acceptable distance is less than 100 km?

Response:

Distance for the SedTrap drifter is determined from the closest CTD cast in time. The distance to the TSG, SD/LD stations are from the initial CTD cast location for the given LD station. However, since our analysis now just uses the SedTrap drifter for its timeseries, this

information for the SedTrap drifter is no longer necessary. Including data out to 1000 km is needed to help constrain the determination of R_Z. As visible in the LDA Z-score vs. distance plot of the CTD data, density layers at depth can show low Z-scores for 1000s of kms.

Figure 7. Why are there such large differences between velocities in h, I, k, l panels?

Reponse:

The differences between the velocities are likely due to small-scale unresolved motions, as mentioned in Pg. 11, Sect. 3.3, Lines 18-19.

Figure 8. rewrite caption to clarify what you are showing. What are the 3 rows showing? The trajectories appear quite different between the rows, why? And what implications does this have for the Lagrangian assessment?

Response:

The trajectories appear quite different between the rows because they represent either satellite-derived geostrophic flows (top row), ADCP currents throughout the water column (middle row), or SVP drifters representing the surface. The top row, showing only the geostrophic surface current, does not include waves. These waves, with depth-dependent structure, lead to the various trajectories seen in the ADCP-derived middle row. The bottom row's SVP trajectories represent the surface, but in the mixed layer not resolved by the ADCP. As a result, we feel it is not entirely surprising that these trajectories are so different, and so each contribute in their own way to the analysis. The implication is that satellite data are not necessarily sufficient for a Lagrangian assessment, and that different depths will travel different distances, so figuring out where the physical gradients are for each depth is important for the Lagrangian assessment.

The caption will be rewritten as follows:

Figure 15 (former 8). Observed and calculated trajectories for currents analysis. Data from LDA, LDB, and LDC, shown in left, center, and right columns, respectively. Top row (a-c): Observed trajectory of SedTrap drifter plotted in magenta. Time-averaged altimetry-derived surface currents shown with black arrows. Rossby radius R_D traced as a color-coded circle, and R_Z, the calculated spatial scale, in black. Starting position of SedTrap drifter shown with a star. Middle row (d-f): Calculated SADCP 38 kHz trajectories of water at each depth down to 600 m. Bottom row (g-i): Observed SVP drifter trajectories, with mean trajectory plotted in black.

of the world Ocean, and so many biogeochemical measurements will be taken elsewhere. For the campaigns where sites are far (possibly by design) from obvious, organized mesoscale structure, there is still a need to conduct an independent, post-cruise validation of the ~~advent of satellite oceanography, structures harboring enhanced gradients such as eddies and fronts were difficult to detect before sampling . As a result, the risk always existed that sites could be chosen close to these structures, putting the~~ drifter's ~~mission into jeopardy. In recent years, the incorporation of near-real-time satellite data having become routine minimizes this (in the list above, since the BOUM 2008 campaign; Moutin et al., 2012). The second step, after deployment of the drifting mooring, consists of an independent , post-cruise validation of the data, which is the~~ success, which is the focus of the ~~present study~~present study.

Before proceeding into this study's description of our methodology, a few remarks are needed regarding its applicability. We already mentioned that we will consider regions away from strong, organized mesoscale structure. Additionally, the method relies upon independent physical, not biogeochemical, measurements to indicate a change of water mass due to the drifter not being Lagrangian. This approach does not detect the existence of biogeochemical gradients in water that might exist on smaller scales, so application of our method requires the user to apply contextual knowledge of their sampling region and keep this possibility in mind. For this study, a regional biogeochemical gradient was expected (Moutin et al., 2017) and rationales for this method's application will be provided.

The Oligotrophy to UlTra-oligotrophy PACific Experiment (OUTPACE) cruise provided an opportunity to assess the success of the quasi-Lagrangian sampling strategy. Conducted from 18 February to 3 April 2015 in the western tropical South Pacific (WTSP), one of the goals of OUTPACE was to assess the regional contribution of nitrogen fixation as a biogeochemical process to the biological carbon pump (Moutin et al., 2017). During the cruise, three long duration (LD) stations employed the quasi-Lagrangian strategy. In the subsequent discourse regarding these stations, we proceed as follows. Sect. 2 describes how the drifting mooring was deployed, our methodological strategy, how concurrent data were collected, and the analyses undertaken to answer our central question of whether we sampled a single environment. We then present the data and results in Sect. 3, followed by a discussion in Sect. 4. The paper finishes in Sect. 5 with a summary of our recommendations regarding future implementations of this sampling strategy.

2 Materials and Methods

In this section, we begin by describing the general manner in which the three LD stations were conducted during the OUTPACE cruise. ~~We then outline the different sources of data presented in this study and how they were processed~~Following an outline of the methodological strategy, we present the different data sources and their processing. Additionally, we describe in detail the analyses needed to answer our central question regarding sampling in a coherent environment.

2.1 Sampling strategy

The OUTPACE cruise occurred aboard the RV *L'Atalante* from 18 February to 3 April in late austral summer, starting in New Caledonia and finishing in Tahiti, traveling over 4000 km. Stations were conducted in a mostly zonal transect traveling west to

east, with the ship track averaging near 19° S. The three LD stations, entitled LDA, LDB, and LDC, and lasting 5 days each, were designed to resolve a regional zonal gradient in oligotrophy, the existence of which is reflected in the surface chlorophyll-*a* (chl-*a*) data (Fig. 1a). As described in the introductory article of this special issue (Moutin et al., 2017), site selection for the LD stations involved identifying physical structures by use of the SPASSO software package (http://www.mio.univ-amu.fr/SPASSO/) using near-real-time satellite imagery, altimetry, and Lagrangian diagnostics (Doglioli et al., 2013; d'Ovidio et al., 2015; Petrenko et al., 2017).

Before starting each LD station, surface velocity program (SVP; Lumpkin and Pazos, 2007) drifters were deployed adjacent to the site. ~~Two~~ The number of drifters deployed are summarized in Table 1, and their mean initial positions were 1.1, 1.6, and 0.9 km away from the first CTD of station LDA, LDB, and LDC, respectively. At the start of each station, two quasi-Lagrangian drifting moorings were deployed during the OUTPACE LD stations with surface floats. The first drifting mooring, hereafter referred to as the SedTrap Drifter, had a 'holey sock' attached at 15 m depth. It was followed actively by the ship and is the emphasis of this study. It had three sediment traps (Technicap PPS5/4) fixed at 150, 250, and 500 m depth, along with onboard conductivity-temperature-depth (CTD) sensors and current meters, described below in Sects. ~~2.3.1 and~~ 2.4.1 and 2.6, respectively. ~~It~~ The SedTrap Drifter was deployed at the beginning of each station and was left in the water until the station's completion. The second ~~drifter~~drifting mooring, referred to as the Production Line, housed in situ incubation platforms for measuring primary production, nitrogen fixation, oxygen, and other biogeochemical measurements (see Moutin and Bonnet, 2015, for more documentation). The Production Line was redeployed on a daily basis close to the SedTrap Drifter. While no telemetry exists for the Production Line, the CTD casts from which incubation water was drawn ranged from 300 m to 5.7 km from the SedTrap drifter. After 5 days, the SedTrap Drifter was recovered, and the LD station completed. Occasions when the exact implementation of this general strategy was not realized will be mentioned in following sections for the relevant measurements. A summary of time duration for each data source can be found in Table 1.

Between the LD stations, 15 short duration (SD) stations lasting approximately 8 h each were interspersed along the ship's trajectory in roughly equidistant sections ~~.~~ (Fig. 1b). Among the measurements made, CTD casts from SD stations will figure into the validation process in this study. Most casts (both LD and SD) were at least 200 m, with at least one 2000 m cast for all stations. These casts were conducted with the same CTD rosette platform described more fully below in Sect. 2.3.1.

Throughout the cruise, surface conductivity-temperature measurements from the themosalinograph (TSG) and currents from shipboard acoustic Doppler current profilers (SADCP) were collected. Their processing is described in Sects. ~~2.3~~2.4.1 and ~~2.4~~2.6, respectively.

2.2 ~~Remote Sensing Data~~Post-validation method

~~Satellite~~ The goal of this study is to evaluate whether the three LD stations during OUTPACE sampled in a homogeneous body of water. In order to achieve this goal, a number of steps were undertaken:

- Validity of application and environmental context. As mentioned in the Introduction, if a physical structure such as an eddy or front is present, its dynamics will dominate and must be taken into account. Additionally, since we used physical

water properties in this study, we must determine whether biogeochemical gradients existed at smaller scales. For this purpose, we looked at remote sensing data.

- Establishment of statistical baseline. To evaluate whether station sampling remained in one water mass, the water mass itself needed to characterized. This was achieved by initializing a baseline within the timeseries of hydrographic properties. The subsequent time evolution of these properties within the defining dataset served as a first test for whether sampling stayed in one environment.

- Spatial scale determination and baseline context. If timeseries analysis showed no change in water properties, complementary data from farther away were compiled to evaluate the spatial scale at which the water mass did change. These data were also used to contextualize whether the statistical baseline over-estimated or under-estimated water mass variability.

- Currents analysis and Lagrangian risk. The spatial scale of the water mass determined, water trajectories were used to evaluate at what point the observed flow regime might have brought another water mass into contact with LD station sampling near the SedTrap drifter.

The following sections in the Methods are organized around these steps, detailing the data and analyses involved for each step.

2.3 Validity of method application and environmental context

Detection of physical structures and biogeochemical gradients used satellite measurements of sea surface temperature (SST), surface chl-*a*, and sea surface height ~~and~~ with its associated geostrophic currents~~and diagnostics,~~. These data were also used in the LD site selection phase (Sect. 2.1)~~, provided temporal context for the LD stations~~. All processed satellite data were provided by CLS with support from CNES. SST was derived from a combination of AQUA/MODIS, TERRA/MODIS, METOP-A/AVHRR, METOP-B/AVHRR sensors, with the daily product produced being a weighted mean spanning 5 days (inclusive) previous to the date in question. Weighting was greater for more recent data. Similar to SST, chl-*a* was a 5 day weighted mean produced by the Suomi/NPP/VIIRS sensor. The SST and chl-*a* products had a 0.02° resolution, equivalent to $\sim 2\mathrm{km}$. These satellite products spanned from 1 December 2014 to 15 May 2015. In order to compress the daily satellite products, weighted temporal means were calculated. For each grid location, the weight for a given day was inversely proportional to the distance from the grid point to the ship's daily position.

~~The temporal~~ Temporal fluctuations of SST and surface chl-*a* were determined by producing timeseries of both variables within a given spatial range surrounding the starting position of the three LD stations. ~~Square regions of sides~~ The spatial range consisted of a 120 x 120 km ~~long,~~ box centered at each LD station~~, were chosen as the boundaries~~. Satellite pixels falling within this region were used to create a probability distribution function. The 120 km square size was chosen because 60 km is a typical size of the Rossby radius of deformation for the region (Chelton et al., 1998). Sudden changes in SST and

chl-*a* distributions indicated strong surface forcing or the passage of gradients, which could invalidate the applicability of the method.

Local surface currents derived from altimetry were also provided by CLS with support from CNES. These data come from the Jason-2, Saral-AltiKa, Cryosat-2, and HY-2A missions, cover a domain from 140° E to 220° E, and 30° S to the equator, covering the yearlong period of June 2014 to May 2015. The velocity grid had a $\frac{1}{8}^{\circ}$ resolution, applying the FES2014 tidal model and CNES_CLS_2015 mean sea surface. Ekman effects due to wind were also added using ECMWF ERA INTERIM model output.

2.4 ~~Hydrographic data and methodology~~Establishment of statistical baseline

2.4.1 ~~CTD data sources~~

~~The shipboard CTD~~ Water mass characterization depended upon observations of conservative temperature (C_T) and absolute salinity (S_A), or T-S measurements. The statistical baseline, which served as the reference for each LD station, needed to reflect the initial state of the water near the SedTrap drifter. While the SedTrap drifter itself had CTD sensors onboard, these were fixed in depth and did not resolve the full variability of the water column. Additionally, although the SedTrap drifter served as the moving station's location, water derived from the shipboard CTD-rosette ultimately served as the source material for the biogeochemical measurements taken during the cruise. The shipboard casts were always positioned near the SedTrap drifter, averaging 1.2 km over the entire cruise. For these reasons the CTD cast data were chosen for the baseline calculation, while both SedTrap drifter and CTD cast data were included in the timeseries analysis.

2.4.1 CTD data for timeseries analysis

The shipboard CTD employed during OUTPACE was a Seabird SBE 9+ CTD-rosette, with two CTDs installed. Data from each cast were calibrated and processed post-cruise using Sea-Bird Electronics software into 1 m bins. All CTD data from other instruments mentioned later were likewise processed using Sea-Bird Electronics software. ~~Final absolute salinity (S_A), conservative temperature (C_T)~~, and potential density (σ_θ) were calculated using the TEOS-10 standard (McDougall and Barker, 2011). In total, over 200 CTD casts were performed during OUTPACE. Most SD stations had three or four casts, except for SD13, which had time for only one cast owing to a medical emergency. The LDA, LDB, and LDC stations had 46, 47, and 46 casts, respectively, each approximately 3 h apart. During LDA, the two drifting moorings accidentally collided and, due to the time necessary to disentangle them, there is a gap of 9 h in the timeseries. The majority of CTD casts were to 200 m depth, with at least one 2000 m cast per station. Mixed layer depth was determined using de Boyer Montégut et al. (2004)'s method, by finding the depth where density has changed more than 0.3 kg m^{-3} from a reference value, which was chosen to be the value at 10 m depth. The 10 m reference was chosen because post-processed CTD casts did not always include the surface.

~~Surface C_T, S_A, and σ_θ for the entire cruise were provided from a Seabird SBE 21 SeaCAT Thermosalinograph (TSG), with SBE 38 thermometer using the ship's continuous surface water intake. Subsequent to post-cruise processing of TSG data as detailed in Alory et al. (2015), the timeseries was available in 2 intervals. Additionally, the~~ The SedTrap Drifter had on board

six SBE 37 Microcat CTDs. Their depths, as determined by mean in situ pressure, were ~ 14, 55, 88, 105, 137, and 197 m. These instruments yielded data every 5 min during their deployments. As mentioned in the previous paragraph, during LDA, the SedTrap Drifter tangled with the Production Line, and so the data presented here from LDA came from its re-deployment until the end of LDA. No gap in the data occurred for LDB or LDC.

2.4.2 Tracer analysis

~~**The central question in this article is whether for each LD deployment the quasi-Lagrangian drifting mooring stayed within a similar physical environment. For our purposes, the physical environment is defined by the water's specific C_T and S_A (T-S) values. Generally, over the upper 200 of the water column, the depth range of most of our CTD casts, a given profile of T-S values will vary along a curve (Stommel, 1962). This reflects how each profile is made up of increasingly denser layers over depth, each with distinct histories. In some sense, these layers could be considered their own physical micro-environment; however, since the biogeochemical measurements of OUTPACE spanned the entire euphotic zone, the ensemble of these layers, the entire T-S profile, had to be combined to represent the physical environment. One way to contextualize T-S profiles is to compare them with previous measurements or climatologies of the region. What occurred for OUTPACE, however, was that the profiles of T-S values did not necessarily coincide with the~~ **baseline definition**

The need for a baseline within the OUTPACE dataset can be shown by comparison of the CTD data with climatologies such as the World Ocean Atlas ~~or CSIRO Atlas of Regional Seas (supplementary material, Figs. S1-S2~~(Fig. S1; Boyer et al., 2013~~; Ridgway et al.,2002). When measures of climatological~~). While OUTPACE observations were consistent with these previous observations, when metrics of variability were available ~~, the envelope of~~ they produced envelopes of max/min ~~/max~~T-S values ~~were~~ large enough to preclude distinguishing between different stations~~during OUTPACE. Another, more standard, method for breaking down~~ . Since no other sufficiently fine data were available to compare T-S measurements~~involves identifying linear combinations of previously defined water masses associated with specific~~ , data from within each station were used to create a reference baseline of T-S ~~values (Mackas et al., 1987; Poole and Tomczak, 1999). While this method provides a least-squares best estimate of constituent water types, it requires a priori knowledge and sufficiently distinct, conservative parameters to define these water masses. The method also does not provide a clear basis by which similar profiles will be considered different (e.g., if two profiles have 5% and 10%, respectively, of a given water mass, does this signify a sufficiently different physical environment?). Moreover, the associated water mass eigenvector basis precludes capturing additional, previously unobserved water masses. Given these concerns,~~ variability. Given the lack of fine variability data and the need to work within the dataset of a single cruise, another approach ~~is~~ was needed to condense T-S variability so that physical environments can be distinguished. ~~As mentioned earlier, each density layer in~~

Generally, over the upper 200 m of the water column, the depth range of most of our CTD casts, a given profile ~~has its distinctive~~ of T-S ~~value, and together these layers constitute~~ values will vary along a curve (Stommel, 1962). This reflects how each profile is made up of increasingly denser layers over depth, each with distinct histories. In some sense, these layers could be considered their own physical micro-environment, and their ensemble constitutes the physical environment. Assuming that ~~these~~ the density layers were not subject to strong forcing such as diapycnal mixing events or atmospheric effects, their

values should have remained constant until isopycnal exchange or diffusion could occur over longer timescales. Treating these density layers as separate entities, variations of T-S along isopycnal surfaces can provide an approach to distinguish physical environments, the goal of our analysis.

Using density as one variable, another is needed to fully describe a water parcel's characteristics, ideally one which is independent of density. Spice, a variable constructed from T-S, is well suited for this purpose. Spice is defined such that hot and salty water is 'spicy', a convention dating to Munk (1981). In the formulation proposed by Flament (2002), its isopleths are everywhere perpendicular to isopycnals, and it effectively both encapsulates and accentuates T-S variability at a given density into a single value. Therefore, in our analysis, spice variability in a given density layer was used to ~~to~~ define the statistical baseline, and determine whether a physical environment changed during OUTPACE.

The statistical ~~determination of whether the physical environment changed during the OUTPACE LD stations relied on a two-step protocol: first, the determination of a spatial scale, R_Z, over which the LD station 's physical environment was the same, and second, evaluation of whether that scale was surpassed by the SedTrap Drifter, in situ currents, or the SVP drifters. In order to find the R_Z of~~ baseline was defined as a functional fit between σ_θ and spice measurements at the beginning of each LD station in the upper 200 m of the ~~physical environment, a baseline profile of spice-density values was defined for each LD stationusing the most abundant data source, namely the timeseries of CTD casts spaced every ~3 . Independent spice measurements were compared to the baseline using the traditional Z-score for a normally distributed variable. The relationship between Z-score and the horizontal distance between observations was used to determine R_Z, the spatial scale beyond which the physical environment changed. The second step, determining whether R_Z was crossed during each LD station, is detailed in Sect. 2.4. The statistical baseline at each LD station was established by collecting spice observations into density bins of width~~ water column spanning the euphotic zone. The period of time used for defining the baseline was chosen to be the local inertial period, so that internal wave effects would be minimized. For each station, this meant that the first 13, 15, and 15 casts were used for LDA, LDB, and LDC, respectively. A regular grid of density values was created, with one-fourth the number of values as the total number of observations. The fit of baseline spice, or $S_{base}(\rho)$, was calculated inside a moving window of $\pm$ 0.1 kg m^{-3}, ~~ranging over 1021.5-1026.3 (bin centers). For each density bin, a mean S_{LD} and standard error $SErr_{LD}$ was calculated, retaining only values in the bins where~~ along with the standard error in spice, $SErr_{base}(\rho)$. Only values corresponding to windows with at least 50 observations were ~~available. The means and standard errors in the baseline were used to compare independent spice observations as summarized by~~ kept.

Comparisons between the baseline and new σ_θ-spice measurements were made using a Z-Score, following the general formula

$$Z(\rho_{obs}) = \frac{S_{obs} - S_{base}}{SErr_{base}} \tag{1}$$

where, for a ~~specific density layer ρ, S is the independent spiceobservation. Variations of this formula were used depending on the dataset in question. Typically, if multiple observations were available in a given density bin, the mean value was used for S, similar to the process used in defining the baseline S_{LD}. When a dataset spanned multiple isopycnals, the distribution of Z-scores was summarized again by the mean of their absolute magnitude. This two-step averaging was done, as opposed~~

~~to averaging all the Z-scores in a profile at once, so that individual density layers were on equal footing.~~ density observation ρ_{obs}, S_{obs} is the observed spice, with S_{base} and $SErr_{base}$ being the linearly interpolated functional baseline spice value and standard error. The assumption applied in this analysis is that ~~isopycnal layers are~~ while a continuous curve in T-S, or σ_θ-S, is to be expected and can be fit to a function, the isopycnal layers were independent of each other, and ~~represent new information from different populations. If one layer had significantly more observations than another, a straightforward mean would bias the~~ represented different physical sub-populations. Keeping track of variability through Z-score ~~toward the value of the layer with more values. To calculate the~~ tied to a functional σ_θ-S relationship produced a flexible metric. For sensors fixed at a certain depth, such as for the SedTrap drifter, a Z-score ~~of other LD CTD rosette timeseries , the reference baseline values were used to calculate~~ could be computed irrespective of whether internal waves were shoaling or deepening isopycnals.

2.5 Spatial scale determination and baseline context

The Z-scores ~~for each density layer, and the mean of this distribution of Z-scores reported. Since there are two possible Z-scores for each LD pair (depending on which is chosen as the reference for the $SErr_{LD}$) , the largest of the two was retained. At an SD station, with usually 3-4 CTD profiles, for each density bin the mean Z-score was used across CTD casts, and the mean of the~~ derived from the CTD and SedTrap timeseries provided a first-order evaluation of physical variability in the immediate surroundings of the SedTrap drifter as it moved through time. If large variability (|Z|>2, in the traditional α = 0.05) was observed, then the physical environment likely had changed. If |Z|<2, however, this was not a guarantee that the physical environment had not changed. Since the functional fit of σ_θ and spice was based only upon the data from OUTPACE, Z-score ~~profile reported.~~ is a relative measure of variability. In order to test whether the σ_θ-spice relationship was robust, it was necessary to extend the Z-score analysis farther in space to include complementary density and spice measurements. If Z-scores ~~from the SedTrap Drifter's six onboard Microcat CTDs are calculated with a 15 window before and after each CTD cast used in the baseline. Since the SedTrap Drifter data represents the smallest distance scales (ranging 0.3 to 5.6~~ remained low for large distances, then the $SErr_{base}$ was too large. By compiling independent Z-scores over larger distances, we can test whether there is a relationship between Z-score and distance.

During OUTPACE, complementary σ_θ-spice observations stemmed from neighboring stations, the SD stations (Fig. 1). Additionally, surface measurements for the entire cruise ~~), instead of calculating a Z-score with S_{LD} being the station mean, the matching individual cast data was used. This choice was made to better test whether sampling close to the SedTrap Drifter results in greater variability or not. TSG Z-scores were calculated with data within a 500 radius of the LD station position, both before and after sampling. Due to ship transiting, Z-scores were not calculated for a given time period, but rather split into distance bins~~ were provided from a Seabird SBE 21 SeaCAT Thermosalinograph (TSG), with SBE 38 thermometer using the ship's continuous surface water intake. Subsequent to post-cruise processing of TSG data as detailed in Alory et al. (2015) , the timeseries was available in 2 min intervals.

The relationship between Z-score vs. distance was used to evaluate the baseline. Distances were calculated from the ship position of observation and the initial CTD cast position for the LD station. ~~For the first 100 from~~ For the SD stations, the ~~LD station position, the bin width was~~ Z-scores found by the functional fit of spice for each cast were binned by density, in a

regular grid with bins of 0.25 $\mathrm{kg\ m^{-3}}$ width. TSG data from during the LD stations were excluded, and Z-scores were binned by distance from the station, in 10 km increments for the first 100 km, and then ~~became~~ 20 km ~~. Since the TSG represents near-surface values and spans few density bins, the mean of each layer's Z-scores was retained as opposed to taking a second mean to combine layers.~~ afterwards until 500 km. The spatial scale R_Z ~~for each LD station was found through fitting both linear and exponential models to the Z-score vs distance relationship. Due to the differences between TSG and SD/LD Z-score distributions, various fits were applied to the entire dataset as well as subsets. A conservative version of R_Z was ultimately estimated using the SedTrap Drifter and TSG datasets. The distances at which Z= 2 (α=0.5 rejection level) for both linear and exponential models were averaged together, weighted by their coefficient of determination, or r^2, value. The definition of r^2 used~~ was

$$r^2 = 1 - \frac{\sum_i^N (Z_i - \hat{Z}_i(x))^2}{\sum_i^N (Z_i - \bar{Z})^2}$$

~~where Z_i is the Z-score~~determined where Z≥2, ~~$\hat{Z}_i$ is the modeled Z-score, and $\bar{Z}$ is the Z-score mean.~~ and used to evaluate the ability of the statistical baseline to discern gradients in physical properties. A natural spatial scale to serve as a useful reference to the ~~fitted~~ empirical distance is the first Rossby radius of deformation, approximated via Wentzel-Kramers-Brillouin (WKB) method by Chelton et al. (1998) as

$$R_D = \frac{1}{\pi f} \int_{-H}^{0} N(z) dz \tag{2}$$

where f is the local coriolis parameter and $N(z)$ is the depth-dependent Brunt-Väisälä frequency. R_D was calculated for each LD station using the deepest cast available: 2000 m casts for LDA and B, and a deep 5000 m cast for LDC. N was calculated with centered differences of the 1-m binned density profiles.

2.6 ~~Velocities~~ Currents analysis and ~~drifter positions~~lagrangian risk

The previous step analyzed the relationship between Z-score and distance, providing an estimated distance over which the physical environment changed. In order to evaluate the risk that the SedTrap drifter encountered different water masses, an analysis of the local currents was needed. Since it is clear that the SedTrap drifter was not perfectly Lagrangian, and that vertical shear could transport layers at different rates, it was necessary to see if water at specific depths could have advected the distance over which different water masses appear.

The in situ velocities for each LD station were derived from the shipboard acoustic Doppler current profilers (SADCP), two Ocean Surveyors at 150 kHz and 38 kHz. Timeseries data for the SADCPs were post-processed using the CASCADE software package (Le Bot et al., 2011; Lherminier et al., 2007) and binned into 2 min intervals. The 150 kHz SADCP provided a depth resolution of 8 m, with bins starting from 20 m, and reliable data coverage down to 200 m depth. Since the SedTrap Drifter had sediment traps extending down to 500 m depth, the 38 kHz data was also used, albeit with reduced depth resolution of 24 m bins, extending from 52 m down to 1000 m. Additional in situ velocities came from six Nortek AQUADOPP current meters,

positioned at 11, 55, 88, 105, 135, and 198 m on the SedTrap Drifter. The post-processed AQUADOPP timeseries provided observations every 5 min.

Velocities were integrated using a first-order Euler method to calculate the theoretical trajectories of water subsequent to the beginning of each LD station. Since the object of these calculations was to see whether water could have ~~surpassed~~ traveled a critical spatial scale, for each dataset the maximal amount of time was given for the time integration. SADCP timeseries spanned between the first and last CTD of the LD station, using the ship position as the initial position. The AQUADOPP integrations spanned the entirety of valid data and used the corresponding SedTrap Drifter satellite fix for an initial position.

To compare the integrated velocity positions with the realized positions of the SedTrap drifter and SVP drifters, GPS positioning was achieved by use of Iridium telemetry. Positions were successfully found for LDA before and after the SedTrap Drifter's re-deployment, along with all of LDB. During LDC, the battery of the positioning antenna ran out and so the timeseries for LDC positions of the SedTrap Drifter was shortened. Since only the initial position is needed for the velocity integration, the AQUADOPP integration was continued beyond this positioning failure until the SedTrap Drifter was recovered. Positions of the SVP drifters deployed at each station were successfully retrieved for all three LD stations. Satellite fixes were available spaced about 1 h apart for both datasets. Both SedTrap Drifter and SVP positions were interpolated to hourly timeseries. SVP positions were used to compute relative dispersion (supplementary Fig. ~~S3~~S2) using the definition for N particles (LaCasce, 2008),

$$D(t) = \frac{1}{2N(N-1)} \sum_{i \neq j}^{N} [x_i(t) - x_j(t)]^2 \tag{3}$$

where N here is the total number of SVP drifters, and x the timeseries of the drifter i's x, y position.

3 Results

3.1 Satellite data~~and~~, temporal context, and method applicability

The regional distributions of SST and surface chl-a as seen during the OUTPACE cruise are shown in Fig. 1. The data in Fig. 1 are weighted means, with the weight being the inverse square of the ship's daily distance to each pixel. A north-south meridional gradient was found in SST, with warmer water near the equator ($\sim 30°\mathrm{C}$) and cooler water poleward ($\sim 25°\mathrm{C}$). This gradient was uninterrupted for the duration of the OUTPACE cruise. Due to the zonal shiptrack the surface thermal conditions observed by the ship during OUTPACE were relatively homogeneous. Furthermore, no strong temperature gradients, indicative of frontal or eddy structures, were visible in the vicinity of the stations. While a north-south regional gradient was found in SST, the opposite was found in chl-a. Chl-a values were around 0.3 $\mathrm{mg\ m^{-3}}$ in the western portion of the domain, west of 190° E. Stations LDA and LDB were in this region, with LDB positioned inside a bloom with values near 1 $\mathrm{mg\ m^{-3}}$. More details concerning the LDB bloom can be found in de Verneil et al. (2017). Chl-a values dropped precipitously, over an order of magnitude to 0.03 $\mathrm{mg\ m^{-3}}$, just east of LDB near LDC. The low value of chl-a was indicative of the South Pacific Gyre (SPG; Claustre et al., 2008).

Since SST was relatively unchanging during OUTPACE, Fig. 2 provides zoomed-in views of the chl-*a* data for the three LD stations, with domains chosen to include the nearest SD stations. The spacing of the SD stations was relatively regular along the OUTPACE transect ~~.~~ (Fig. 1b). In Fig. 2a the enhanced chl-*a* was distributed evenly inside the domain, so no clear surface gradients are present. In Fig. 2b, the chl-*a* was concentrated in the aforementioned bloom, with values higher than those seen in Fig. 2a near LDA. The size of the bloom was large enough to cover most of the 120 km x 120 km region shown in Fig. 2b, so the bloom edges were far away from station LDB's initial position. By contrast, waters outside the bloom had chl-*a* values lower than in Fig. 2a. The low chl-*a* values near LDC in Fig. 2c were typical of the SPG, and no visible patches of chl-*a* indicated sharp gradients.

The timeseries of chl-*a* and SST for the three stations are shown in Fig. 3. Comparing the three LD stations, a few patterns emerge. SST showed similar trends across the three LD stations. All stations experienced warming trends from December 2014 to mid-March 2015, consistent with summer heating. The lack of data from cloud cover sometimes led to abrupt drops in the distribution of daily SST shown. The timing of maximum temperature, and the magnitude of that warming, however, did differ between LD stations. A rapid heating in December 2014 occurred around LDA's position, which then slowly continued until early March 2015, at which point temperatures began to drop. Towards the end of sampling at station LDA the SST rises, indicating possibly another warming event occurred ~~.~~ or the arrival of a warm patch of water. Depth-resolved application of our method in the later sections will evaluate this possibility. The overall evolution in LDA's temperature during the period shown, from ∼26 to 30° C, represented a 4° change. LDB showed a slight cooling in December 2014, but this may have been an artifact of cloud cover. Station sampling for LDB occurred immediately after the maximum heating, though the values seen at LDB were relatively stable and slightly warmer than at LDA. The maxima in temperature for LDA and LDB seemed to overlap in time, in early March 2015. LDB's change in temperature, from ∼27 to 30° C, was a 3° change. LDC had the smallest change in temperature, from ∼27 to 29 °C for a 2° change. Sampling for LDC coincided with the warmest period observed in the satellite data, in late March ~~2015.~~ 2015, and was stable for the LD sampling period.

The timing of temperature maxima is important to note for biological reasons, since N_2 fixation by *Trichodesmium* spp. is known to occur in warm, stratified waters (specifically, a ∼25° C threshold, White et al., 2007), and one of the goals of OUTPACE was to observe this biogeochemical process. Since SST was above 25° C for all stations throughout this period, the thermal conditions during OUTPACE would not have limited N_2 fixation.

In between December 2014-January 2015, the region around LDA had higher chl-*a* concentrations than LDB. The period between February and May 2015 showed a remarkable increase in chl-*a* near the LDB site. This was due to advection of the surface bloom, which subsequently collapsed and advected away ~~(de Verneil et al., 2017).~~ , as documented in another study in this special issue (de Verneil et al., 2017). The downward trend of chl-*a* during this period is more indicative of in situ evolution, rather than advection of the bloom's boundaries, and does not invalidate subsequent use of our method. Near LDC, chl-*a* was systematically low, a reflection of the goals of OUTPACE to sample in the SPG.

Besides the increase in SST at the end of LDA and the decrease in chl-*a* during LDB, both SST and chl-*a* for the LD stations were stable, providing evidence that no surface gradients, physical or biological, immediately invalidate the application of our strategy. Though the change in chl-*a* at LDB has been argued to be due to endogenous dynamics in the aforementioned study,

application of our post-validation method provides an independent test of whether advection of gradients could be responsible. Likewise, the method will also determine whether the surface increase in SST was reflective of changes at depth.

3.2 In situ ~~tracers~~properties, statistical baseline, and timeseries analysis

The hydrographic variability during the three LD stations and surrounding SD stations are shown in the T-S diagrams of Fig. 4. All three stations followed a general pattern, where surface water near the 1022 $\mathrm{kg\ m^{-3}}$ isopycnal and 29° C temperature (though LDB had warmer surface water, Fig. 4b) dropped in temperature and increased in salinity until a subsurface salinity maximum near the 1025 $\mathrm{kg\ m^{-3}}$ isopycnal. The ~~salinity maximum reflects sampling along the latitudinal gradient of the~~ increase in salinity maximum from LDA to LDC reflects the high salinity tongue of the South Pacific (Kessler, 1999). The surface water in LDA (Fig. 4a) showed a bifurcation~~, where the heating event~~ . This change was manifest in the satellite data timeseries, as well. Whether this is due to a heating event or the arrival of new water at the end of LDA ~~sampling seen in satellite data was also manifest~~will be addressed in the timeseries analysis below. For LDA, neighboring stations SD2, 3, and 4 largely overlapped with the LDA profile. SD3, the station closest to LDA, almost entirely overlapped the LD profile, except for a subsurface salinity deviation below the 1024.5 $\mathrm{kg\ m^{-3}}$ isopycnal. SD2 and SD4 showed greater deviations, with SD4 being saltier than LDA for almost the entire profile. Similar overlaps occurred with LDB and its surrounding stations, SD12 and SD13 (Fig. 4b). SD12 showed lower salinity near the surface, with a kink in salinity at the 1025 $\mathrm{kg\ m^{-3}}$ isopycnal. The salinity offsets of SD4 and SD12 at depth are within climatological variability (Figs. S1 and S2). SD13 had similar surface structure to LDB, but higher salinities from the 1023.5 to 1025 $\mathrm{kg\ m^{-3}}$ isopycnal. The LDC, SD13, and SD14 (Fig. 4c) profiles nearly entirely overlapped except near the surface when the SD stations were at first less salty at the surface and then became more salty. Additionally, the saltier nature of LDC relative to LDA and LDB, especially between 1024 and 1025 $\mathrm{kg\ m^{-3}}$, was visible. The variability in T-S values between stations was within the range seen in the climatology of the region (Figs. S1 and S2).

The LD statistical baselines of spice in density space, with means and intervals of two standard errors, are shown in Fig. 5. These standard error intervals, representing the inherent variability in the baseline, show the values wherein a Z-score of ≤ 2 was achieved. LDB and LDC overlapped for essentially their entire profiles. ~~LDB is~~ All stations are missing observations near the surface and mixed layer due to the intense stratification which left several density bins with less than 50 observations, the threshold used in the spice analysis. LDA was noticeably less spicy than the other two LD stations for density less than 1024 $\mathrm{kg\ m^{-3}}$. ~~In the 1024-1025 range, where LDC had high variability in spice, some density layers showed LDC differing from LDA and LDB by more than two standard error intervals.~~ At the highest densities~~at depth~~, all three LD stations overlapped. The envelope of two standard errors, or Z-score ≥ 2, show that variability has some dependence on depth. The LDA baseline shows high variability near the surface, with a thin envelope below down to 1024 $\mathrm{kg\ m^{-3}}$, and widening at depth down to 1025 $\mathrm{kg\ m^{-3}}$ and beyond. LDB did not have high surface variability, but the envelope widens shortly below 1023 $\mathrm{kg\ m^{-3}}$. Variability in LDC shows similar widening as in LDB, with a noticeable pinch in the envelope near 1024.5 $\mathrm{kg\ m^{-3}}$.

The ~~Z-scores calculated from the other datasets with regard to the LD statistical baseline are plotted versus their distance from the LD station in Fig. 6. SedTrap Drifter Z-scores, centered on the left hand side of Fig. 6, showed a cloud of Z-scores~~

~~that does not depend on distance. Overall, these Z-Scores were below 2. The surface TSG data began to show a relationship between~~ Z-score ~~and distance. Starting near 15 ,~~ timeseries for the SedTrap Drifter sensors are shown in Figs. 6-8. During LDA, at 14 m depth (Fig. 6a), after the inertial period baseline determination the Z-score first descended, increased, and then leveled off after the first two and a half days. Afterwards, the Z-score increased, reaching 2, decreased, and surpassed Z=2 before falling again. The SedTrap drifter at 55 m showed no trend (Fig. 6b), and a single Z-score was seen below -2. Z-scores ~~were at ≤~~for 105, 137, and 197 m (Fig. 6d-f) showed no temporal trends and were always less than 2 in magnitude. The timeseries at 88 m showed no trend, but the variability in Z-score increased time, with some observations surpassing |Z|=2. ~~The LDA TSG~~

LDB SedTrap drifter Z-scores ~~remained below~~ (Fig. 7) showed simliar patterns to LDA. The surface sensor (Fig. 7a) decreased and increased over the first two days, then leveled with temporary departures below -2. The sensors at 55, 105, 137, and 197 m (Fig. 7b,d-f), similar to LDA, showed no trends and low variability. A few observations below -2 occurred at 55 m. The Z-scores at 88 m showed no trend, similar to LDA, with enhanced variability and some |Z|>2~~until~~ , but no time-dependence.

The LDC Z-scores were large at more depths than the other LD stations (Fig. 8). Values at 14 m started with Z>2, but that dropped before rising again after a day, before slowly dropping and eventually decreasing to ~~∼100 distance, after which point the~~ -2 at the end of LDC. Z-scores ~~quickly increasedwith distance until they reached the highest observed values of all the datasets near the 500 cut-off.~~ at 55, 137, and 197 m showed no trends in Z-score, and had limited observations with |Z|>2. At 88 m, no trend was seen, and for the first two days there were few observations with Z<-2. Toward the end of LDC, two spikes with Z>2, with Z∼4-5, occurred with returns back to |Z|<2. The Z-scores ~~from the TSG for LDB increasedfaster, passing~~ at this depth ended near Z=2. Observations at 105 m started around -2<Z<0, but spikes with Z∼2 ~~between 40 and 50 and continually increasing thereafter. The LDC TSG~~ occurred. Over time, Z-scores ~~started out high, having surpassed a~~ trended upward with more oscillations, with a shift to Z>2 becoming dominant during and following March 27, 2015.

CTD Z-score ~~of~~ timeseries are shown for LDA, LDB, and LDC in Figs. 9, 10, and 11, respectively. LDA Z-scores were generally |Z|<2~~between 30 and 40 , and continued to increase in the same manner as the LDB TSG data. One should note, however, a large number of LDC TSG~~ , but Z-scores ~~below and near~~ for densities σ_θ < 1022 kg m^{-3} were greater than 2 ~~in the 100 to 300 range. SD station~~ starting March 1, and continued for the rest of the station. The increasing trend in Z-score near the surface was also reflected in the SedTrap drifter. LDB CTD Z-scores ~~replicated the trends seen in the TSG data, with increasing Zover greater distances. The only SD station within the canonical Rossby radius, SD3 near LDA, had a Z less than~~ showed almost no observations with |Z|>2. ~~Comparing between how the SD stations match up with LD baselines, LDA differed the most from the other stations, given its enhanced Z-scores. The SD~~ These occurred at the surface with low densities, and a few near σ_θ ∼1023.25 kgm^{-3}. All these observations occurred before or around March 19, and no temporal trend in |Z|>2 was seen. The Z-scores ~~for LDB and LDC were lower by comparison, and were largely grouped together, though they too displayed an increasing trend in Z-score with distance. The three LD station pairs, LDA-B, LDA-C~~for LDC showed similar trends to the SedTrap drifter. Near the surface close to $\sigma_\theta \sim 1022$ kg m^{-3}, |Z|>2 was seen early in the timeseries, but then dropped to |Z|<2 until another increase around March 27. This pattern was similar to the SedTrap drifter's observations at 14m. Regions of |Z|>2 appeared in the 1024-1025 kg m^{-3} range, primarily during March 27. A small density range near 1024.5 kg m^{-3} showed

|Z|>2 during March 26, but as time went on a larger swath of density had |Z|>2, and ~~LDB-C, also reflected a trend of increasing Z-score with distance. The closest station pair, LDB-C, had a Z-score less than 2. The other pairings, LDA-B and LDA-C, showed greater differences with~~ this change was largely permanent. Near 1025 kg m^{-3}, a separate series of large Z-scores ~~$\sim$ 10. The r^2 values for the model fits, and the corresponding distances where the models reach Z=~~appeared on March 27 and lasted for most of the rest of LDC.

3.3 Spatial scale and baseline context

The TSG Z-scores for the three LD stations are shown in Fig. 12. For LDA, Z>2 ~~, are summarized in Table 2. Model fits were the lowest when including all the data. The 'TSG+SedTrap' Drifter model fit was the highest, with the r^2 being in the top two for all model fits across all stations. The distances where the model fits cross Z~~occurred at 150 km. Z-scores were consistently large farther away from this point. The LDB TSG surpassed Z=2 at 55 km, though Z-score diminished again 300 km away. For LDC, TSG Z-score reached 2 ~~vary greatly depending upon the dataset. Negative distances occur for most fits with 'all data', possibly due to the influence of the TSG and SD/LD data having high Z-scores~~ at ~~two separate spatial scales, and that the model fits were not constrained to cross the origin. Negative and small distances occur with the 'TSG only' fits, again possibly due to the lack of constraint to cross the origin. Model fits for 'all but TSG' data result in large cut-off distances (all ≥400~~ 35 km ~~). This is possibly a result of Z-scores remaining low for some SD stations~~ distance, and |Z| oscillated between larger and less than 2 farther away. Therefore, at the surface layer, 150, ~~and the lack of data to introduce variability at the spatial scales between the SedTrap Drifter and SD/LD stations (normally spanned by the TSG). The 'TSG+SedTrap' model fits resulted in~~ 55, and 35 km were the ~~most consistent Z=~~spatial scales. Since at least some Z-scores were found to be greater than 2, the baseline was sensitive enough to determine gradients over a 500 km scale. Since Z-scores for LDB and LDC were not consistently |Z|>2, then the baseline's sensitivity was perhaps not as great as LDA. The Rossby radii for the three stations were 46.5, 48.8, and 60 km. The spatial scales for the TSG data at LDB and LDC matched up to the Rossby radii, whereas LDA's TSG data indicated a larger scale.

Z-scores from the SD stations are presented in Fig. 13. For LDA, the |Z|>2 distances ~~, between 30-90~~ demonstrated density dependence. Near 1022 kg m^{-3}, |Z|>2 immediately, at ~45km~~. This minimum of variability is possibly due to the SedTrap Drifter data forcing the model fits to be low at small scales, and that the large-scale TSG values are at spatial scales immediately adjacent to the SedTrap Drifter (TSG values begin at 15~~ , though this did not occur at the surface. Approaching 1000 km distance, |Z|>2 occurred from the surface to 1024 kg m^{-3}. By 3500 km, all density layers show |Z|>2. LDB showed large Z-scores in some density layers at the closest SD station located 189 km away. Past 750 km, Z from 1022-1024 kg m^{-3} was consistently high. For densities greater than 1024 kg m^{-3}, ~~as opposed to around 200~~ Z-scores were enhanced between 1000 and 1500 km~~for most SD/LD datapoints). Due to the fact that the model fits were consistently high, and the relative lack of variability in the cut-off distances, the Z~~, but then decreased farther away. LDC's Z-scores show that |Z| was greater than 2 from the first observations at 310 km. All density layers showed enhanced Z values, with the majority of all observations being larger than 2. Similar to the TSG data, the SD stations showed that the baselines were sufficiently sensitive to detect physical gradients on large scales, with some detecting changes immediately. Putting together the near-surface TSG data and CTD data

from the SD stations, LDA showed smaller |Z|=2 ~~distances for the 'TSG+SedTrap' models were chosen as the~~ scales at depth, whereas LDB and LDC both showed variability both near the surface and at depth at smaller scales. In order to be the most conservative in our velocity and trajectory estimates, we will use the smallest spatial scale of |Z|=2 to determine the spatial scale R_Z ~~spatial scale over which the physical environment was the same. Since there was no noticeable preference in r^2 for linear or exponential models, the cut-off distance is the mean weighted by r^2, resulting in R_Z values of: 66~~ and evaluate Lagrangian risk, namely 45 km for LDA, ~~47~~ 55 km for LDB, and ~~54~~ 35 km for LDC. ~~Therefore, with determination of R_Z, we have completed the first step in our protocol to evaluate whether the physical environment changed during the LD stations.~~

Having evaluated the ability of the statistical baselines to sense physical gradients over large scales, we are now ready to analyze the currents and trajectories.

3.4 Velocities and lagrangian trajectories

~~The second step in our protocol requires the analysis of water velocities and their associated trajectories.~~ Timeseries of the 38 kHz SADCP and AQUADOPP data are presented in Fig. ~~7.~~ 14. The LDA timeseries of SADCP u and v components (Fig. ~~7a~~14a,d) showed strong near-inertial oscillations in the upper 200 m, with velocities reaching magnitudes of 0.6 m s^{-1}. A weaker tidal component was also present in this layer: below 200 m, vertical columns of alternating velocity sign indicated the semi-diurnal tide. These tidal signatures were also the dominant signal in the LDB and LDC timeseries (Fig. ~~7b-e~~14b-c,e-f). The mixed layer, which, for most of the cruise, was $\leq$20 m, was not resolved by either SADCP. So, the near-surface velocities were only captured by the 11 m AQUADOPP and the SVP drifters drogued at 15 m. Comparing the 55 m AQUADOPP timeseries with the 52 m SADCP, the two data sources displayed similar trends for LDA. The strong near-inertial oscillations led to correlations between the AQUADOPP and 38 kHz timeseries of 0.75 and 0.76 for the u and v components, respectively. During LDB and LDC, the weaker currents did not correlate as well, leading to u,v correlations of -0.0137, -0.0554 (LDB), and 0.30, 0.37 (LDC), respectively. For comparison, the 150 kHz 52 m timeseries produced u,v correlations with the AQUADOPP of 0.83, 0.80 (LDA), 0.00, 0.02 (LDB), and 0.68,0.68 (LDC). Vector correlations using the method of Crosby et al. (1993) for the three timeseries (not reported) similarly showed a maximum for LDA, minimum near-zero for LDB, and low values for LDC. These differences likely result from higher frequency fluctuations of the currents, at the inertial and tidal frequencies. The fact that a higher correlation is obtained at LDA is probably partly the consequence of the larger ~~spatial~~ horizontal scales of the near-inertial signal dominant at LDA compared to the baroclinic tidal signal, e.g. resulting from the dispersion relation (Alford et al., 2016). These oscillations, and their implications for turbulent mixing, are analyzed in greater detail in Bouruet-Aubertot et al. (this issue).

The disagreement between the two velocity data sources had an impact on the integrated trajectories. Take, for example, a closer inspection of the SADCP and AQUADOPP during LDA, which had the strongest currents. The initial positions of the ship and the SedTrap Drifter were 1.46 km apart. After 3 days and 2 hours, the AQUADOPP integration had traveled 67.75 km, the SADCP 60.71 km, with a final separation of 10.89 km. The result was a positional drift of $\sim$3 km day^{-1}, or an average increase in position difference of 147 m for each km traveled. A similar analysis for the LDB timeseries, with weaker currents but essentially no correlations over 4 days and 15 hours, resulted in a positional drift of 3.19 km day^{-1}, with an increased

position difference of 318 m for each km traveled. Thus, a lower correlation timeseries, but with lower magnitudes, resulted in similar misfit in the integrated trajectory.

The trajectories of the integrated velocities, as well as observations of SedTrap Drifter and SVP positions, are presented in Fig. ~~8.~~15. The average altimetry-derived currents suggested there should be recirculation around the positions of LDA and LDC, whereas LDB had a mean northward flow (Fig. ~~8~~15 a-c). The SedTrap Drifter trajectory for LDA did not follow the surface altimetry currents and their anticyclonic flow, but instead underwent several oscillations while cruising in a west-northwest direction (Fig. ~~8a~~15a). The SVP drifters for LDA (Fig. ~~8g~~15g), while also undergoing oscillatory loops, instead drifted to the south. The 38 kHz SADCP velocities showed a transition over depth with shallow water moving south-southwest, but with increasing depth the trajectories flowed northwest in a similar fashion to the SedTrap Drifter. During LDB, the SedTrap Drifter went north-northeast, in agreement with the altimetry currents (Fig. ~~8b~~15b). The SVP drifters moved in a similar fashion, north-northeast, though they ended up undergoing more oscillations and eventually advected more eastward (Fig. ~~8h~~15h). The 38 kHZ SADCP velocities demonstrated that shallow depths flowed east like the SVP drifters, but with depth this advection swung to a more northerly direction (Fig. ~~8e~~15e). The LDC SedTrap positioning was relatively uninformative, since the Iridium satellite fix was unavailable for the second half of the LD station and so showed little displacement (Fig. ~~8e~~15c). The SVP drifters for LDC (Fig. ~~8i~~15i), similar to LDB, underwent several oscillations and were advected the farthest, moving in a southeast direction. The SADCP data showed a shallow flow to the east, similar to the SVPs, but with depth the majority of trajectories oscillated near the station and even flowed southwest.

For all the LD stations, the SedTrap Drifter ~~and integrated velocities~~ stayed within a radius of R_Z and R_D centered at the LD starting position. Integrated velocities of the SADCP also stay within R_Z and R_D, except for the trajectory nearest the surface for LDA. The SVP drifters for LDA and LDB also stayed within the R_Z ~~radius. The SedTrap Drifter, integrated velocities, and SVP drifters also stayed within a Rossby radius~~ and R_D ~~of the starting positions of LDA and LDB~~distances, though the LDB SVP trajectories come close to R_Z. For LDC, however, the SVP drifters traveled farther than R_Z, but shorter than R_D away from the initial position. R_D was ~~smaller~~ larger than R_Z for ~~LDA~~all three stations. Since surface SADCP data for LDA crossed R_Z, we evaluate that the water during LDA might be from a different water mass. Likewise, the SVP trajectories for LDC crossed R_Z, meaning that surface water for LDC might be from a different water mass at the end of the station. As seen in the timeseries analysis, water at depth for LDC also changed part-way through the station, ~~almost equivalent in size for LDB, and larger for LDC . For LDC ,~~ though the ~~SedTrap Drifter and integrated velocities stayed within both R_Z and R_D, the SVP drifters reached R_Z but not R_D at the end of five days~~SADCP velocities show little displacement. As a result, according to our protocol ~~they may have moved into~~ the water observed in these density layers may have derived from a new physical environment, and biogeochemical observations at the end of ~~LDC~~ LDA (primarily the surface) and LDC (surface and at depth) may need to be examined in closer detail for changes~~associated with a change in water mass. However, for the greater majority of LDC, this should not be a concern.~~ . Thus, we conclude that ~~LDA and~~ LDB sampling was conducted in a single physical environment, ~~whereas LDC was mostly in a single environment except perhaps at the very end.~~ surface observations for LDA and LDC are suspect towards the end, with LDC water in the 1024-1025 $\mathrm{kg\ m^{-3}}$ range should be closely examined after March 27, 2015.

4 Discussion

4.1 Tracer analysis and spatial scale determination

The main goal of this study is to determine whether the quasi-Lagrangian sampling strategy during the LD stations of OUTPACE was successful, namely by remaining within a single physical environment. The motivation behind this exercise is to independently evaluate if the biogeochemical measurements of OUTPACE represented a single biogeochemical milieu, rather than the advection of the SedTrap ~~Drifter~~ drifter into a different area, as well as the advection of different water toward the drifter. Evaluation of the strategy is grounded in the variability of T-S and analysis of water velocities.

Before the spice analysis was conducted, the initial context of SST and chl-*a* variability at the surface in space and time was provided by satellite products. At the regional scale of the WTSP, the LD stations were roughly positioned within the zonal gradient of chl-*a* and meridional gradient in SST (Fig. 1). The gradients of surface chl-*a* around the LD stations were minimized in relation to the regional-scale gradients, partly by design in the process of choosing station locations (Fig. 2). The temporal trends of SST and chl-*a* largely reflected the seasonal cycle: chl-*a* was decreasing at the end of the summer in the MA and low values dominated in the SPG; SST reached its peak due to late summer timing (Fig. 3). The timing of temperature maxima is important to note, as mentioned in Sect. 3.1, since N_2 fixation by *Trichodesmium* spp. is known to occur in warm waters, and one of the goals of OUTPACE was to observe this biogeochemical process. While these satellite data are sufficient to identify large-scale structures or temporal trends, the LD stations by design were in regions where it is difficult to judge whether the SedTrap Drifter stayed in the same water mass from these surface data alone. ~~In light of this, these data are insufficient for our needs in validating the quasi-Lagrangian strategy~~However, they do justify the use of our methodology, ie there were no coherent mesoscale structures in the vicinity, and no nearby chl-*a* gradients were present. The use of remote sensing data to help identify small-scale structures during OUTPACE outside of the LD stations is further explored in Rousselet et al. (this issue). In order to continue with our post-validation strategy, in situ data are needed.

The depth-resolved in situ T-S data (Fig. 4) helped to capture some of the variability present during the LD and surrounding SD stations. The T-S structures showed consistent values during LD stations with deviations observed in the neighboring SD stations. As with the satellite data, the in situ data in this form, although informative, provide qualitative interpretation. ~~The~~ In order to identify water masses, traditional quantitative methods ~~for breaking down tracer data~~ require identified water masses (Mackas et al., 1987; Poole and Tomczak, 1999). In some regions of the ocean, these methods are difficult to apply. ~~Sometimes well-defined water masses are not found or difficult to distinguish.~~ Also, maybe the full complement of tracer data (dissolved O_2, nutrients, etc.) cannot be used (like in this study) because they are liable to rapidly change in the euphotic zone due to biogeochemical processes. Additionally, local mesoscale activity can contribute to variability, as has been seen in the WTSP (Rousselet et al., 2016). In that study, O_2 measurements were the distinguishing tracer, which we are precluded from using. As a result of all these factors, another quantitative method that works within the dataset of a single cruise in the WTSP is needed.

The quantitative approach used in this study leverages the large quantity of in situ T-S data available from multiple platforms to condense the physical variability present in the WTSP during OUTPACE over 4000 km distance during austral summer 2015. In order to do this, the statistical baseline in spice was defined (Fig. 5). In effect, as opposed to an absolute

measure (i.e. specific water mass determination), this provides a relative measure of variability that can be used. The ~~spice baselines for all the LD stations yielded similar relationships~~ timeseries analysis for the latter portion of the SedTrap drifter and CTD timeseries data showed density layers where variability was enhanced, and caution should be applied to the analysis of biogeochemical data. Since the baseline is a relative measure, observations from outside the LD station were used to see at what scales physical gradients appear. The relationship between distance and variability (summarized as Z-scores) ~~when compared to complementary datasets (Fig. 6).~~ provided a method by which to establish this scale. Overall, variability increased with distance, as one would logically expect~~. The rapid~~ , but this was not monotonic across all datasets. Therefore, the first increase in Z-score above 2 (using an $\alpha = 0.5$ criterion) ~~led to the determination of the~~ was used, and the smallest of these scales across datasets and density layers was conservatively chosen in determining the cut-off scale R_Z ~~to define a cut-off point at which the physical environment began to change~~. These distances were of the same order of magnitude as the Rossby radius R_D.

~~The limitation in this approach is that the measure is now dependent on how applicable a statistical baseline is to the given dataset, in our case the LD CTD timeseries. At some level, this has to be a subjective decision based upon the data at hand. For OUTPACE, creating the statistical baseline appears most reasonable for stations LDA and LDB due to the concentration of data over the ensemble of profiles (Fig. 4a-b). During LDC, it is clear that the mid-water spice variation, mostly due to salinity, may make this assumption problematic. Unless there is a reason to a priori remove such data (e.g. if there were visible changes in the deep chl-a maximum during LDC), then this variability must be incorporated into the baseline definition, and it will affect the Z-scores. The combination of multiple density layers, however, should provide a more robust comparison of how similar an entire T-S profile is, and not just the variability present in a single density range. A similar phenomenon where variability is introduced to the baseline occurs with the surface heating at the end of LDA. The resulting spice baseline has enlarged standard error near the surface (Fig. 5), which possibly lowered the TSG Z-scores~~ Despite the testing of the baseline on complementary data outside of LD sampling, there is still the question of its generality. In some instances, the variability can be partly attributed to an in situ process, such as the near-surface changes in LDA coinciding with surface heating (though advection possibly played a role, as seen in the SADCP trajectories). While seeing increased variability with distance was reassuring, the non-monotonic nature of some Z vs. distance relationship raises some questions. Is the water on the other side, where |Z| goes back to below 2~~for up to 150 . This produced an R_Z larger than R_D for LDA, the only of the three LD stations to do so. Measuring only the surface instead of multiple density layers, the TSG data are probably the most sensitive to this phenomenon~~, truly the same water mass? Is |Z|>2 truly a change in water mass relative to traditional methods? Is the Z-score approach based on the spice-density relationship more, or less, sensitive than these methods? These questions merit further study, and will have to be explored using both methods simultaneously.

The example of surface heating at LDA introducing unwanted variation brings up another assumption in our analysis: we used spice hypothesizing that there was no diapycnal forcing. Clearly, at the surface, atmospheric forcing can influence the water's T-S (and spice) characteristics, and so will impact TSG measurements as well as observations in the upper mixed layer. Future applications of this method will have to take this variability into account, and perhaps make greater use of survey data to fill in the spice variability below the surface at these spatial scales. At depth, however, the greatest source of

along-isopycnal gradients in T-S, i.e. density-compensated features, is mesoscale stirring (Smith and Ferrari, 2009), and so, generally, the assumption should be applicable. ~~Mesoscale eddies and fronts are the physical structures~~ Part of our assumptions in developing this method requires that sampling is not near mesoscale fronts and eddies. However, their residual effects are the most likely to ~~impact biogeochemical observations~~ affect along-isopycnal variability in a given field campaign, and it is probably the reason that both Z-scores ~~across the board~~ (especially near the surface) begin to increase and the R_Z's were found to be at or near the Rossby radius R_D for each LD station. Granted, if ship sampling happens to be placed immediately next to a strong eddy or filament, Z-scores could increase over much shorter distances, though ~~this was not observed during OUTPACE~~as previously stated this situation was expressly avoided. Therefore, in situations where our starting assumptions are met, at first order ~~,~~the Rossby radius R_D serves as a default scale at which the integrated history of previous mesoscale stirring will on average manifest itself. In situ data and further analysis is needed, however, to verify whether smaller-scale variability is large through determination of R_Z, as evidenced by the appearance of salty, spicy water in the mid-water column at the end of LDC. One recommendation that emerges from these results is that, if future field measurements are to take place in an area devoid of obvious mesoscale structures, the Rossby radius R_D quickly calculated from a deep CTD cast may be useful in determining at which point a quasi-Lagrangian drift array should be recovered as a precautionary measure.

The choice of spice as a variable, though useful, is not a magical transformation in itself. The similarity in T-S between ~~LDB and LDC~~ LD stations is still manifest in spice-density space (Fig. 5)~~. In fact, by our own metrics, since stations LDB and LDC had a Z-score less than 2, left with only this measurement the two stations were indistinguishable from each other. Only through comparison~~ , especially at depth with some density layers overlapping. This means that for these density ranges, the stations are not distinguishable from one another. The overlap in statistical baselines further emphasizes the need to compare with other datasets ~~does it become clear that variability does exist between and around the two stations, namely the SD stations surrounding LDB and LDC. This observation makes clear that independent data sources are a key requirement in this method.~~ to highlight and determine the scales at which variability exists. Having enough data to span a sufficient spatial range is what will determine what differences in spice are relevant or not. ~~Indeed, having enough data can help validate whether application of the spice baseline method in itself is insightful. For instance, if the SedTrap Drifter CTD data~~The fact that the SedTrap drifter, positioned close to the CTD timeseries, mirrored the same trends, corroborated both the presence or lack of small-scale gradients shown through Z-score. If the SedTrap Drifter timeseries, representing the smallest scales, had displayed much larger Z-scores, then the conclusion would have been that there was large variability right next to the ship somehow missed by the CTD baseline. Instead, by ~~having limited variability (Z≤2) that did not visibly depend on distance~~showing similar trends, the SedTrap ~~Z-score distribution validated the~~ data validated the baseline and the spatial scales at which the LD CTD timeseries was sampled~~and its use in creating a statistical baseline~~, even if statistical baselines between the far-flung LD stations overlap.

Having considered some of the caveats and assumptions implicit in our present approach, we feel that its application for the OUTPACE campaign was warranted and subsequently validated by the consistency of the results, both in consideration of the multiple data sources concerned but also ~~of~~ the theory of mesoscale circulation. In our application, a conservative R_Z was used based on surface ~~, and not depth-resolved,~~ data at intermediate scales. Ideally, future applications of this method could

processes, can be used to quantify the effect of the physical circulation upon the biological environment encountered in a field campaign.

The multiple in situ data sources compiled during the LD stations of the OUTPACE cruise allowed for the determination of whether the ship, and its associated quasi-Lagrangian drifting array, sampled the same physical environment. The procedure procedures used to do this consists of two several steps. First, the one needs to look for large gradients (physical or biogeochemical) and structures of circulation (fronts or eddies) that would both impact the trajectory of the drifter and bring different water masses into close contact. Then, a statistical baseline is created from the T-S variability from a single source is variability seen during the station's occupation and transformed into spice-density coordinates and used to establish an adequate statistical baseline. Comparison of independent T-S data to the baseline is used to calculate Z-scores and , first to analyze the in situ timeseries of the drifter deployment, and then to establish a spatial scale R_Z beyond which the T-S differences amount to a change in the physical environment. For the OUTPACE cruise, this scale was found to be close to the Rossby radius R_D, in the 45-65 35-55 km range. The second last step is to then use all available data regarding currents and drifter positions to evaluate whether any water parcel could ostensibly have traveled farther than R_Z. During OUTPACE, while some density layers were shown to be variable enough to represent a change in water mass, we conclude that largely this did not occur.

The methodology used in this study provides a framework wherein readily available T-S data can be used to answer the same question (whether a single physical environment was sampled following a quasi-Lagrangian drifting mooring) for other oceanographic cruises. More traditional methods, such as absolute water mass determination, or using alternative tracers such as dissolved oxygen, require prior knowledge of a given region or are not applicable in the euphotic zone where biogeochemical measurements are made. While sampling in a Lagrangian manner is preferable to not attempting to follow a water parcel at all, the inevitable failure to be truly Lagrangian with these platforms should be recognized so that experiments are not allowed to either last too long or be deployed in an inappropriate flow regime. Regarding the use of this methodology, we give a few recommendations for future cruise sampling:

- Maximize use of remote sensing data during the cruise to identify possible mesoscale features to either avoid or sample inside of. This can be achieved with software such as SPASSO (Petrenko et al., 2017).
- Upon arrival at the selected site, a deep CTD cast below the thermocline can be used to quickly calculate the local R_D Rossby radius in real time and produce a rough estimate for maximum spatial scale. In patchier environments, this scale might be too large, and must be reduced.
- Before and after each station, sample with a surveying instrument such as ISIIS, SWIMS, SeaSoar, or MVP beyond R_D to get depth-resolved data at intermediate scales.
- If possible, mount CTDs and current meters on the quasi-Lagrangian drifting array (perhaps a sediment trap that does not need to be removed constantly). Multiple independent observations over a large range of spatial scales are essential to calculate robust R_Z estimates.

~~r^2 values and Z=2 distances for the linear and exponential model fits of each LD station.~~ **~~STATION TSG only All but TSG TSG+SedTrap All data~~**

~~LDA~~ ~~Linear r^2 0.79 0.85 0.81 0.13 $Dist(Z=2)$ () 96 437 46 -344 Exponential r^2 0.79 0.87 0.83 -0.32 $Dist(Z=2)$ () 0.3 809 85 -7.4×10^3~~

~~LDB~~ ~~Linear r^2 0.78 0.60 0.90 0.18 $Dist(Z=2)$ () 21 658 37 89 Exponential r^2 0.77 0.56 0.81 0.04 $Dist(Z=2)$ () -30 900 57 -990~~

~~LDC~~ ~~Linear r^2 0.54 0.68 0.73 0.25 $Dist(Z=2)$ () 43 726 56 -355 Exponential r^2 0.50 0.72 0.67 0.10 $Dist(Z=2)$ () -49 1.05×10^3 51 -3×10^3~~

~~Supplementary Figure 2. T-S Diagrams like in Fig. S1, but with the CSIRO Atlast of Regional Seas (CARS) climatology.~~

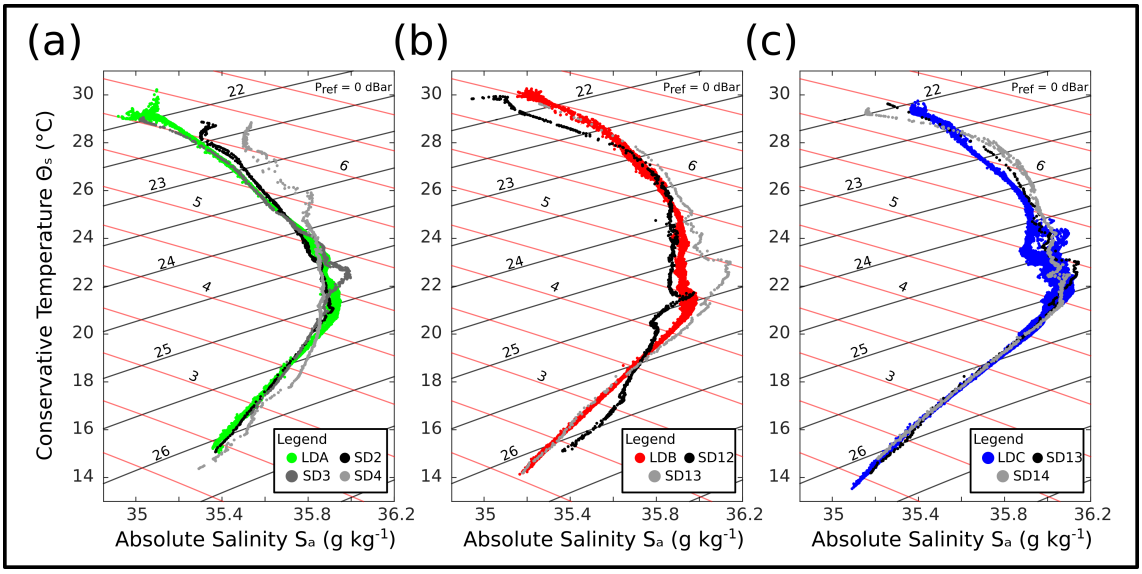


Figure 4. T-S diagrams of (a) LDA, (b) LDB, and (c) LDC and surrounding stations. LD stations are color-coded, and SD stations different shades of gray. Isopycnals are displayed in black, with isopleths of spice shown in red.

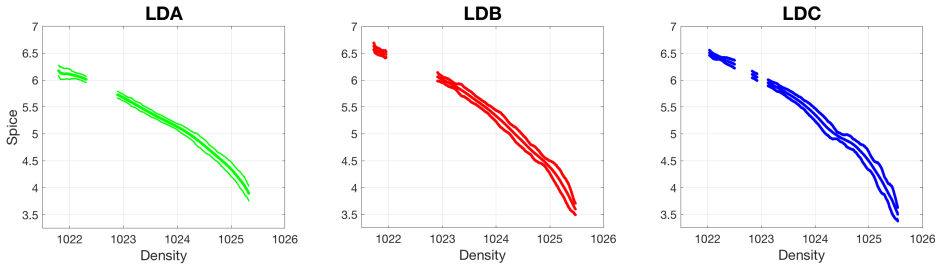


Figure 5. Statistical LD baseline of spice versus potential density for (a) LDA, (b) LDC, and (c) LDC. Mean values ~~(dots)~~ plotted ~~with intervals~~ in between envelope of ~~two standard error~~ $\pm 2\ SErr_{obs}$.

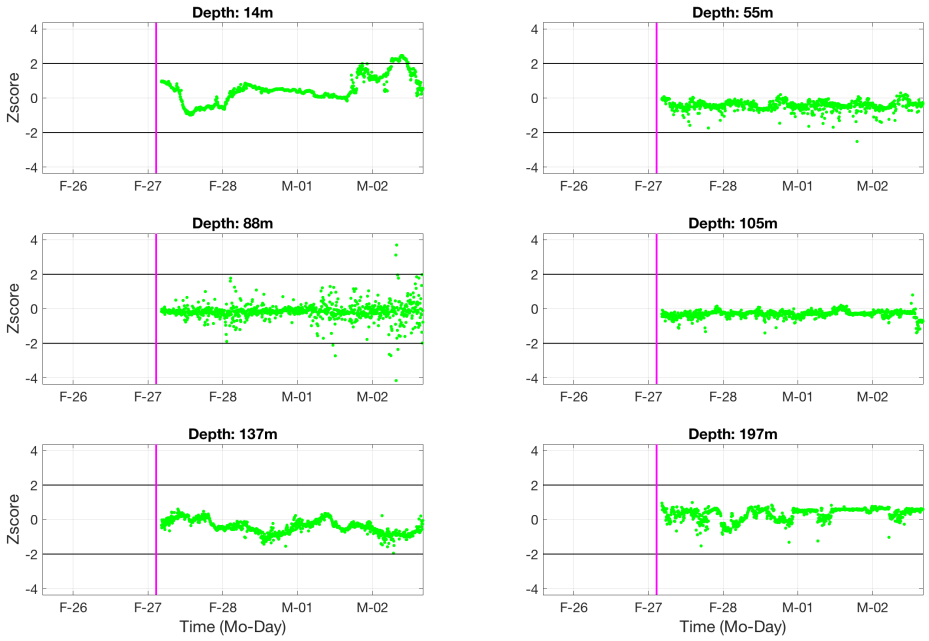


Figure 6. SedTrap drifter Z-score ~~plotted over distance~~ timeseries for ~~the SedTrap Drifter~~ (~~large empty circles~~a) 14, ~~TSG~~ (~~small filled circles~~b) 55, ~~SD~~ (~~+'s~~c) 88, (d) 105, (e) 137, and ~~LD~~ (~~x's~~f) ~~datasets~~197 m depth. ~~Colors~~ End of inertial period timeframe for ~~each symbol correspond with the same color code as previous figures, except the LD stations~~ baseline definition plotted in ~~black~~magenta. ~~Distance axis is log-transformed.~~ Z=-2 and 2 ~~reference line~~ plotted ~~with a~~ in black~~horizontal dashed line~~.~~Rossby radii R_D calculated from deep CTD casts plotted in vertical dashed lines following the color code.~~

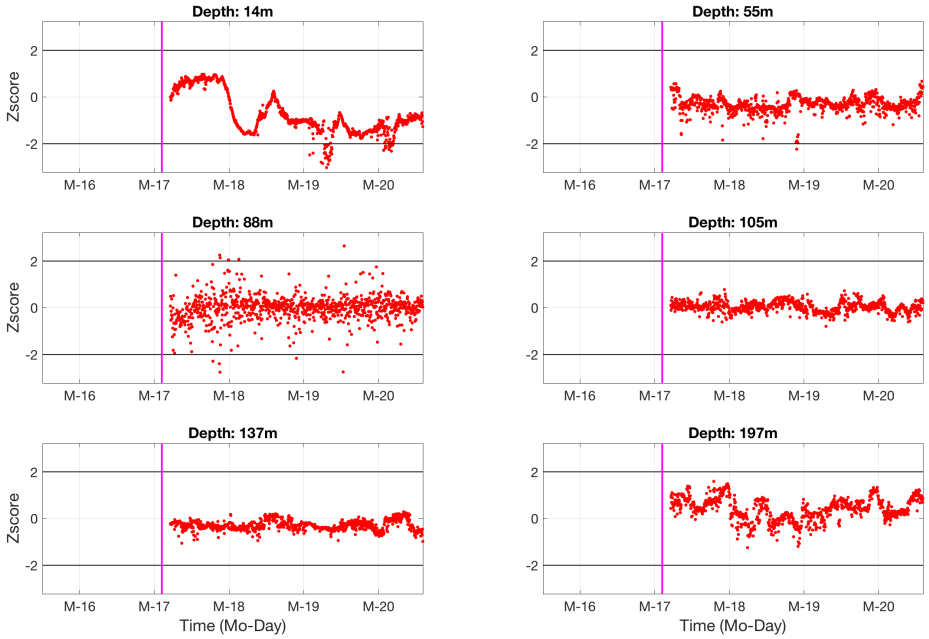


Figure 7. Same as Fig. 6 but for LDB.

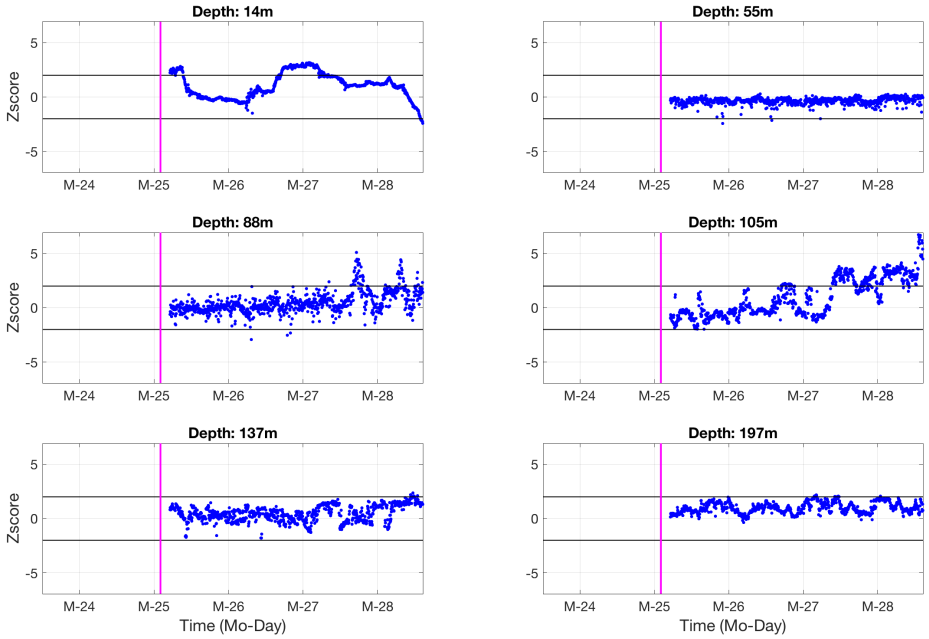


Figure 8. Same as Fig. 6 but for LDC.

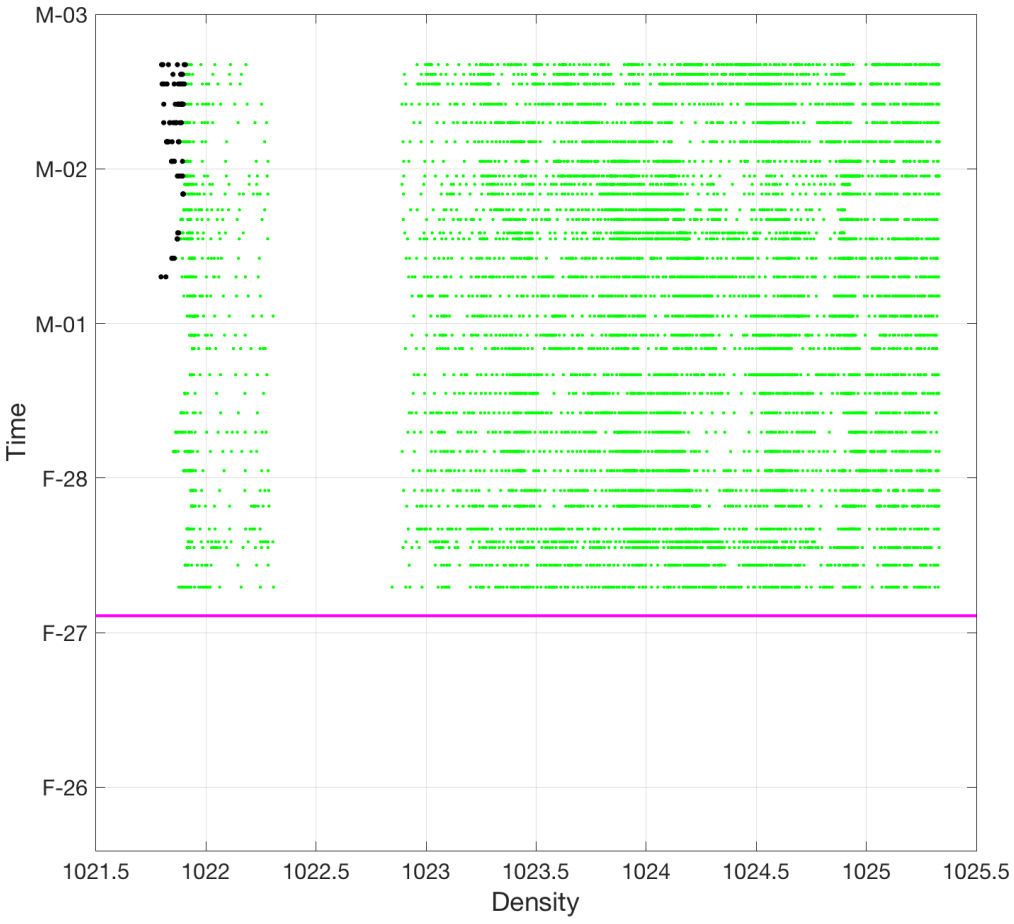


Figure 9. Timeseries of CTD observations for LDA. Observations with |Z|<2 plotted in green, |Z|>2 in black. End of statistical baseline definition period shown in magenta.

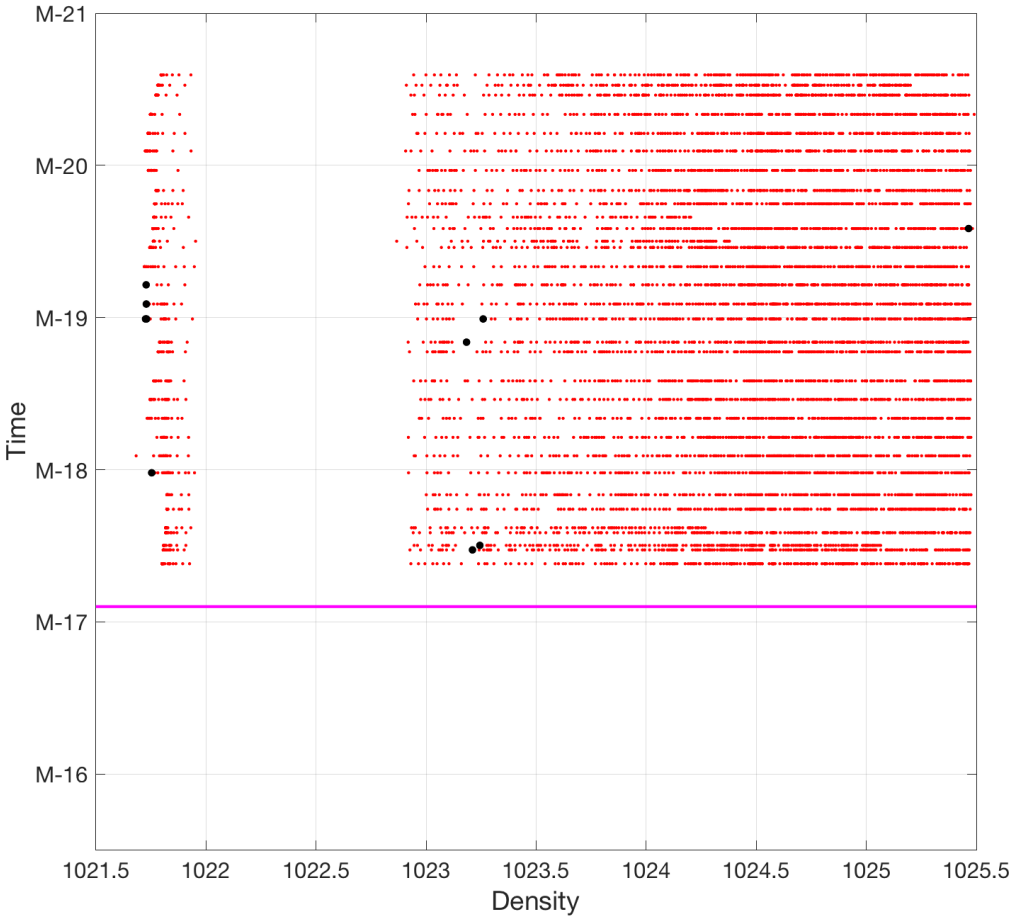


Figure 10. Same as Fig. 9 for LDB, with |Z|<2 observations plotted in red.

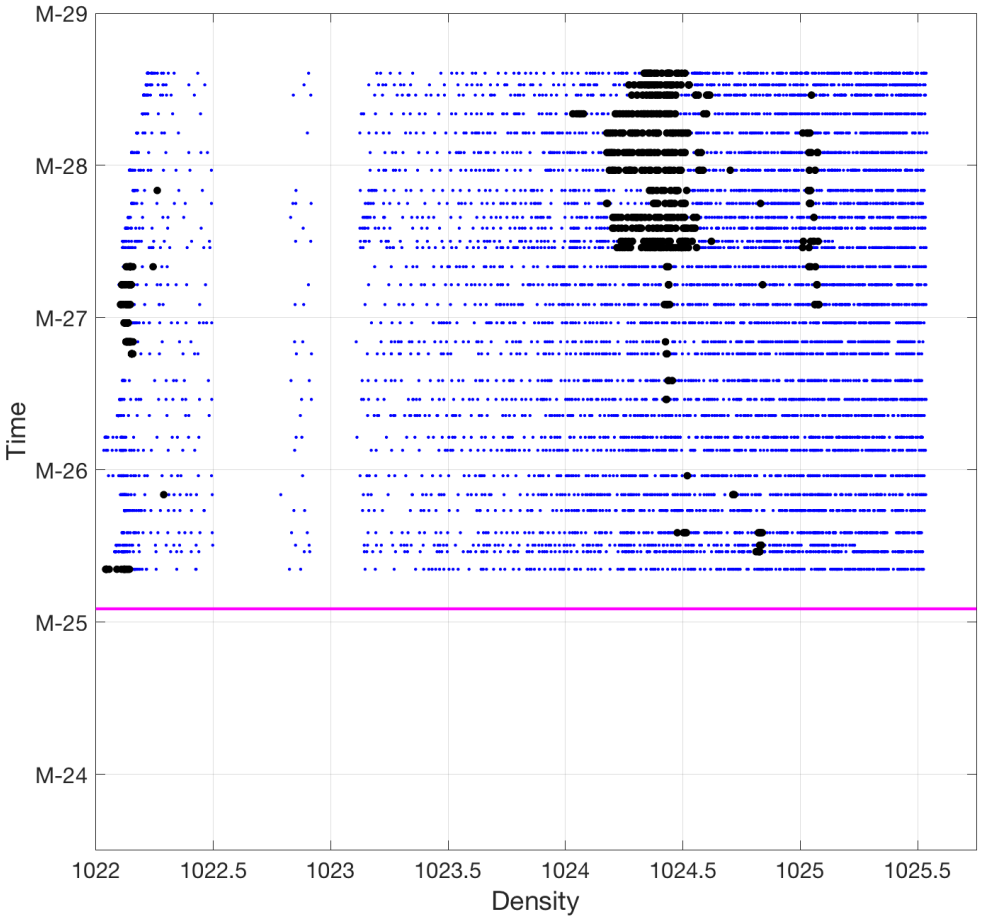


Figure 11. Same as Fig. 9 for LDC, with |Z|<2 observations plotted in blue.

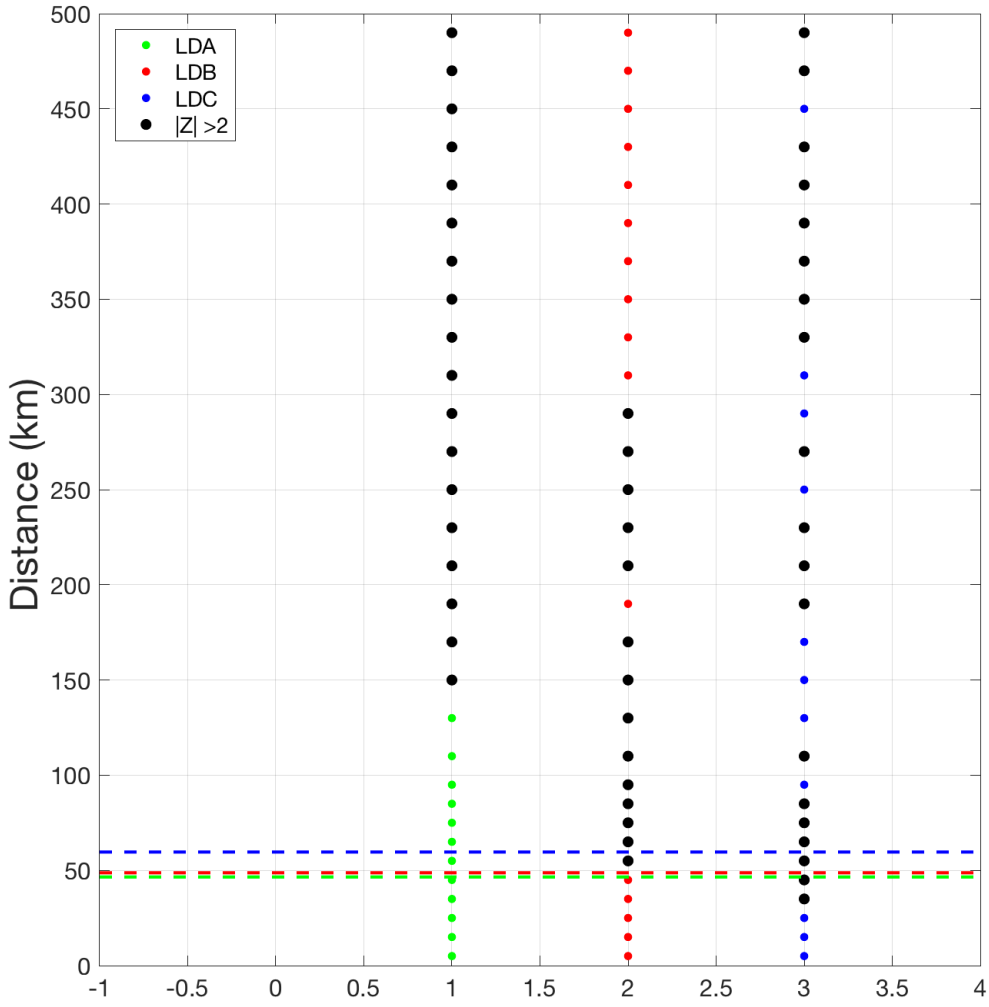


Figure 12. TSG Z-score over distance for LDA(left, green), LDB (center, red), and LDC (right, blue). |Z|>2 shaded black. Rossby radii R_D distance plotted in horizontal dashed lines, color-coded to the LD stations.

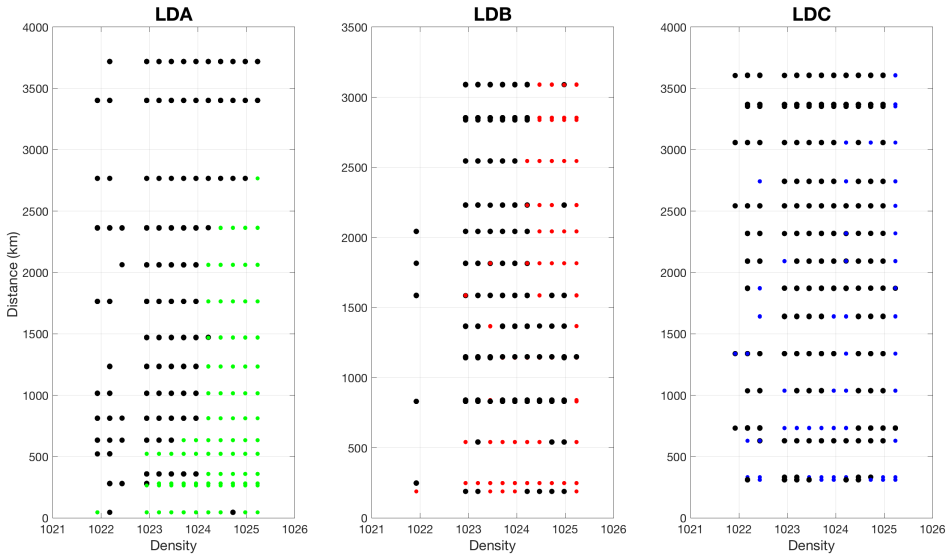


Figure 13. SD station Z-score over distance for (a) LDA, (b) LDB, and (c) LDC. |Z|>2 shaded black. Rossby radii R_D distance plotted in horizontal dashed lines, color-coded to the LD stations.

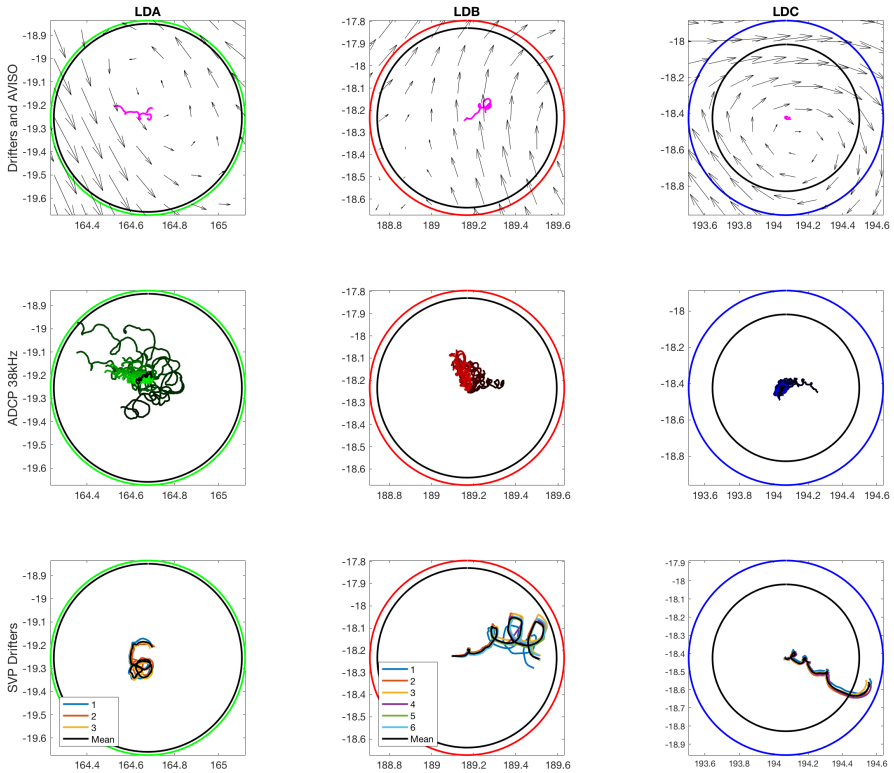


Figure 15. ~~SedTrap Drifter Trajectories, SADCP 38hodographs,~~ Observed and ~~SVP~~ calculated trajectories for ~~, respectively, (a,d,g)~~ currents analysis. Data from LDA, ~~(b,e,h)~~ LDB, and ~~(c~~LDC, ~~f~~shown in left, ~~i~~center, and right columns, respectively. Top row (a-c)~~LDC~~: Observed trajectory of SedTrap drifter plotted in magenta. Time-averaged altimetry-derived surface currents shown with black arrows. Rossby ~~radii R_D plotted following the color code~~radius RD traced as a color-coded circle, and RZ, the calculated spatial scale~~found through tracer analysis traced~~ , in black. Starting position of SedTrap ~~Drifter Trajectories plotted in magenta, with starting positions~~ drifter shown with a ~~yellow~~ star~~, superimposed on time-averaged altimetry-derived surface currents provided by CLS/CNES~~. ~~Mean~~ Middle row (d-f): Calculated SADCP 38 kHz trajectories of water at each depth down to 600 m. Bottom row (g-i): Observed SVP drifter ~~position plotted~~ trajectories, with ~~black line. Domains~~ mean trajectory plotted ~~are the 120 × 120 squares shown~~ in ~~Fig~~black.~~2~~