Major Comments

Internal waves (IW) – Authors appear to shy away from presentation and discussion of IW and any spatial variability due to them. I recommend at least presenting the near-bed baroclinic currents at both moorings for comparison, as these are where sediment concentrations are highest. If the near-bed baroclinic currents are very similar then the glider current verification can remain focused on total currents. If not, focus the verification on barotropic currents, but still present RMSD for total and near-bed baroclinic.

Explanation of chosen analyses – There is often quite a few different statistics and measures presented without much explanation as to what the reader should take away from this? I would recommend thinking a bit more about what is the aim of each statistic / analysis you include. For calibration things like R2 are key. For verification of currents things like (total or normalised) RMSD are key. Regression slope could be helpful to show a bias, but regression is not the aim of the verification (I assume) and R2 provides an overly confident estimate to the reader. Comparing statistics such as range or mean speed generally aren't very useful for the verification either.

LISST section - What was the aim of the LISST section? LISST data is very interesting but I wasn't sure what it adds to the paper in its current form. The only takeaway for me was that on 14 Feb when acoustics greatly underestimate optical SPMC near the bed there appeared to be less fines present. I would still include the profiles of $b_{bp}(z)$ if you remove the LISST data.

Schematic – at some points its not exactly clear what is being compared. You could write this clearer in the text but a schematic or photo of the glider with associated measurements might be useful. Given the calibration and current verification is a main aim of the paper referencing another paper is not sufficient.

Minor comments

Line 7: 'satisfying' – understandable but probably not quite the right word for a physical sciences manuscript

Line 10: 'from the acoustic' … instrument?

Line 49: More of a limitation for verifying models and not an inherent limitation of our ability to model?

Line 60: Section 2.~~1~~

Line 60: 'enumerates', describes or details maybe?

Line 90: Remove submesoscale and mesoscale. I would remove these terms from the whole paper and just state distances. I would also give the box dimensions in km.

Line 100: Probably worth extending this sentence about why this BU shape is important.

Line 127: delete notably

Line 152: delete even

Line 158: Change yo to profile or downcast or something, unless this is an accepted glider term.

Line 159: Was stationarity assessed for each 20 min period? Either generally to get a sense if this is appropriate or for the calibration using CIAM mooring data?

Line 175: Is this interpolation step necessary? There is missing data near the surface for the moorings also, why not just stick with observations and not include 'inferred' data?

Line 195: Sentence is a bit confusing, I would say something like 'the ~~submesoscale'~~ MELANGE area, the instantaneous water level and associated tidal currents are considered…'. I think meso-tidal range is a thing but not mesoscale tidal range.

Line 198: spatially varying bathymetry. It's only varying in time because of the gilder movement.

Line 204: I would expect more than one sentence on this extrapolation, or a highly appropriate reference. In general, I just wouldn't include the extrapolation. Data in Fig5 suggests that current observations do not extend to the sea floor either so that also requires explanation if fluxes are going to be calculated to the floor.

Line 212: Bit strange to throw in specific numbers like 30 um and 1mm for these instruments given it's a continuous function. Could say the 1 MHz scattering response peaks at ~1 mm (10.1016/j.margeo.2010.11.002 Fig 1). But in my experience optical sensors will happily measure coarse sand and the 1 MHz will pick up suspended sediment populations with d50 down to ~20 um (maybe even smaller, I haven't tried). The key assumption you want to make here is that the observed sediment population is not changing much with time, horizontal space, and height above seabed. Maybe this is an opportunity to connect the LISST data?

Line 239: I'm not quite sure how you got the 2.5 cm/s error? Maybe just a standard method? Could compare this to error estimates from similar studies?

Line 254: What does match mean, you assume cells were vertically co-located if the vertical separation was <1m? Or you are interpolating?

Line 255: What does temporally matched mean? Extracting equal / synchronised time points from the datasets or doing some interpolation?

Line 259 and 239: I got a bit confused by the GLI overlap. Does the GLI ADCP record 4 pings per second over 5 seconds (20 pings)? Maybe clearer to say it sampled at 4 Hz continuously and recorded a 5 second average. Then you are computing the expected overlaps based on glider speed correct? Not from raw data? Why less than 3 overlaps at 239 and less than 20% at line 259?

Line 268: delete obviously

Line 280: Were both NTU sensors recently calibrated? Uncalibrated 'identical' instruments can read differently.

Figure 3: Looks like some unresolved dependency on site, i.e. bottom sediment type. Worth mentioning that you expect some of the variability in the relation was probably due to seabed sediment types variation but not included in the regression.

Section 4.9: It is not clear exactly what is being compared here. Are you using the backscatter from just the first ACDP cell to compare to optical? Make clear and state the estimated vertical separation of the optical and acoustic measurements, if any.

Line 304: Doing what?

Line 306: Why are you multiplying the error values by the slope of Eq 9? If the idea to convert acoustic data to SPMC using the 2 regressions in succession then we would expect an increase in uncertainty as error propagates (assuming independent regression models), no?

Line 307: Maybe worth noting that uncertainty will "increase" when converting from log10 units back to normal units. Found example of this here (10.1029/2021JC017538 Fig 12).

Line 326: I don't think you want to be computing the R2 from a linear regression between the 2 moorings (or the glider), if that is what you are doing? You aren't interested in how to translate observations from one thing to another like you were with the calibration, you are interested in the difference (error) between the datasets. This method of computing R2 does not account for magnitude differences between the datasets because it is already accounted for in the slope of the regression. I think if you want a similar metric to your regression R2 you could compute 1 - normalised MSD in a similar way to how you have computed the RMSD (one mooring minus the other). Or maybe you could compute R2 when the regression slope is fixed to 1 and the intercept is fixed to 0. Or just use RMSD and leave out R2 here.

Line 338: Shipwreck seems tenuous. If this is an issue why isn't the RMSD between GV1 and CIAM higher? Seems only an issue with the GV1 – GLI comparison which is difficult to explain when they were so close during the VM period.

Line 343: Systematic bias between GLI and moorings probably not due to spatial variability, rather the shear method as you mention or something else.

Figure 5: black line is free surface – 70 m? Need to somehow mention you have adjusted it for plotting.

Figure 5: This figure isn't really discussed in the text. What is the aim of including it? We can see differences in strength between GLI and moorings, especially at the surface. The white gridded lines make it a bit tough on the eye. Also why the low resolution colormap, was the data too noisy for a good visualisation with a continuous colormap?

Figure 356: Tide was expected to be the main… or show the total + baroclinic currents

Line 358: How should the reader interpret this ratio mean physically? Maybe better to include a bit less here but explain it a bit more.

Line 362: Satisfactory? Also using 'very' is usually avoided for more specific words / numbers.

Line 369: Add figure number again, 'from Figure 7'

Line 375: GLI and CIAM clearly match better in Figure 5. Are you selecting this pair because of this? It is a bit confusing to compare GLI to CIAM when GV1 looks like it was supposed to be the virtual mooring calibration. Is the higher error between GLI and GV1 due to stronger baroclinic currents at GV1 that weren't picked up by the glider very well? If it's because you suspect magnetic interference, or some other error maybe restate it here. But if most of the difference in total currents between GV1 and GLI is found in the baroclinic component there is more to add to the discussion here. If the differences in baroclinic currents are mainly in the BU period, then internal waves are just another source of uncertainty due to spatial separation. If the difference persists through the VM period, then you need to discuss the shortcomings of the glider in observing baroclinic currents.

Line 376: Is +-0.15 for u and +-0.1 for v? Maybe replace first 2 sentences here with GLI u and v std, then CIAM u and v std

Line 378: Stick with the usual RMSD

Line 380: These regions also coincide with the strongest baroclinic currents. If observations from GLI are more uncertain where we want to observe strong baroclinic currents this is a limitation that needs stating.

Line 382: Dispersion of data? Higher uncertainty maybe? Could you add subplot to Fig 8 showing the overlap counts to see if it lines up with errors?

Line 393: Delete obviously

Line 408: Define deep, middle surface here in m as you have in Fig9 caption.

Line 412: You can see the relation deviate from linear in Fig 4 which I think explains why ADCP SPMC is under-estimated at high values. This could be due to a shift to finer particles in theory, but it doesn't look like the LISST supports this. And we would expect a shift towards larger particles as we near the bed, if any change.

Line 414: Need to mention you are moving from muddy to sandy to gravel bottoms over this period. Very important for what you are seeing. The VM phase is on gravel as tides get smaller so maybe not surprising we don't see much then?

Line 424: What is the median grain size? Say it here don't reference it

Line 430: I would remove all the speculation about erosion and advection. This is usually done much better using a bottom lander mooring with turbulence measurements and direct sampling of the seabed. Here you are moving over different sediment types of unknown grain size (at least in this paper) and (I think) extrapolating near bed currents from higher in the water column. Pretty tricky to make accurate comments on erosion and hence the origin of observed sediment. You have observed sediment in suspension and you can calculate some horizontal fluxes, that is the strength of this paper.

Line 435: 'emphasize the role of fine particles in nepheloid layers' – you are bringing back earlier speculation as fact here. This seems to be contradicted in your next section (Fig10 b shows lower volume of fines as you approach the bed).

Line 460: I'm confused as to whether you are saying this is bubbles or zooplankton. I wouldn't speculate, just say it wasn't detected by the NTU sensor and was not considered to be sediment. Could then add that bubbles and biological particles are known to effect LISST.

Fig10b: suggest adding a line that tracks the d50 at each depth going from the surface to the sea floor. Don't include the spikes in the largest bin.

Fig 10d: I wouldn't recommend taking the median of such data that is clearly not grouped. Maybe remove the median and match scatter colors to subplot f?

Fig 10e: Again you have distinct differences in the PSD in subplot f so why take the median here? Maybe show the station with high fines as an example?

Fig 10 caption: Can trim this down. LISST info can be in text or just reference the paper for details. I'd remove the large bins and not discuss values from them here. It is a common issue with the LISST.

Line 478: Because it underestimates SPMC at higher concentrations?

Line 487: I wouldn't expect a 25-hr running mean to remove tides very well. Did you try a lowpass filter like Butterworth?

Line 489: the case for barotropic currents

Line 526: across relevant spatial scales? Lots of gliders to do a whole shelf

Fig 12: Could add filtered currents so reader can see if that's driving the fluxes

Line 539: RMSD is the key metric