Review of Subhas et al. “A tracer study for the development of in-water monitoring, reporting, and verification (MRV) of ship-based ocean alkalinity enhancement”

This is a thoughtful and well-conducted study which attempts to take the challenge of quantifying the effects of marine carbon dioxide removal (mCDR) from the realm of theory, where it has mostly been confined, out into the real ocean. The authors convincingly use a tracer approach to simulate the addition of alkalinity to Atlantic waters and monitor the dispersal of the tracer over multiple days.

MAJOR COMMENTS

-The paper proposes an MRV framework, but it’s a little unclear what the framework is, exactly. Is it Equations 1a-1d? Is it the conceptual diagrams in Figure 1? Is it all the text in Lines 90-125? Is it all of these combined? I may have missed it, but after reading the manuscript I would struggle to describe the framework to someone else in a sentence or two. It would be good to clarify the basic points of the framework, maybe in some bullet points or a callout box.

-One major question I am left with regarding the dispersal of the tracer is the path of the patch and its north-south extent. Figure 5 shows that the team followed the tracer in generally east-west transects, which slowly moved from north to south over the experiment as the tracer was advected. But did the tracer move north to south as a fairly coherent patch, with the total surface area not changing much, or did it ‘smear’, so the tracer patch grew in total surface area over time? Essentially, it seems to me that the patch was well-sampled in one direction (east-west), but lightly-sampled in the north-south direction due to understandable time restrictions. Figure 3b shows a lot of heterogeneity in the patch concentrations: how can the authors know that they were always sampling ‘peak’ patch RT concentrations in later Figures (i.e. Figures 6, 10). This seems to have large implications for the estimate of CDR efficiency starting at Line 621, since the efficiency calculation is predicated on all the added TA remaining in the patch (i.e. no precipitation). Perhaps this is mostly unknowable, but Figure 3a shows that some drone imagery was taken, and perhaps maybe satellite imagery could be available as well, which might inform this question of the patch extent.

MINOR COMMENTS

-L51: “Research”

-L562-63: I don't think semicolons are the right punctuation here.

-L74: I was not familiar with the concept of ship wake dilution models. A sentence quickly explaining what these are and what they are used for might be helpful.

-L84: fCO_2 has not yet been defined

-Figure 1 and caption: most of the paper uses fCO_2, not pCO_2. The difference is minor, but consistency would be good.

-Figure 1: these plots show what would happen to the carbonate system in the surface ocean with infinite dilution, I think. My understanding is that for a fixed volume of seawater, OAE eventually results in a DIC that returns to near-baseline levels, but at a slightly higher concentration, and that pH also remains just slightly elevated above baseline levels. It's a little hard to tell from Fig 1, but I think DIC and pH return to baseline conditions. This is a subtle difference that might be worth mentioning.

-L109-110: Is it worth keeping the $\Delta TA(CDR)$ term in Equation 1a if it is always zero?

-L141-142: does storing CO_2 in biomass require a sustained, measurable fCO_2 gradient?

-L153: "(90-foot)"

-L163: strange citation format with all four authors listed

-L190: define "UHMW"

-L193: refer reader to Figure 2.

-L200: 38 knots is very fast for the *Connecticut*, unless it is a cigarette boat

-Equations 2 and 3: I'm curious if these are constructed to apply in straight-line travel conditions. How does the spiral dispersal pattern used in this experiment work in these models? It seemed like the swaths of RT sometimes dispersed into each other, which I imagine would have an effect.

-L217, it might be useful here to describe what the "in-patch"/"out of patch" thresholds were. It's later mentioned on L328.

L240-241: include metric units here (inches and feet used)

-L239: the TA was distributed at the surface, but the underway intake was at 1.5. Could this have affected anything in terms of peak RT (and thus estimated TA) concentration, patch edge detection, etc?

-L245: It seems the total time from the ship's hull to the fluorometer was about 30 seconds (mostly delayed in the debubbler). Does that lag time seem about right?

-L260: "analyzed"

_L293: "what does "subsequently routinely" mean? Were all samples not poisoned?

-L294: the 13-C DIC data are only presented in the supplementary, and not discussed anywhere that I can see. These data and associated methods probably don't need to be included. If they are needed for some reason, how was the 13-C isotope ratio calibration done (Dickson CRM is not certified for 13-C).

-L317: were the four drifters all deployed in the same place, at the same time?

-L339: what concentration of liquid NaOH is assumed?

-L345-349: this section shifts into present tense.

-L423-424: this gets back to the convening swaths of RT dispersal. It would be good to discuss the implications of this some more, in terms of OAE.

-L439: the decrease in RT concentration is hard to see with the color scale used in Fig. 5a. Also can mention here the decrease is seen in Fig. 6a as well.

-L460: It's hard to parse the three exceptions noted here. Perhaps number or identify them individually somehow.

-Figure 6: Seelmann et al. (2019 p. 526) note the occurrence of outliers in underway TA data from the Contros HydroFIA instrument. Just going on looks, the high TA points in Figure 6d of this manuscript seem to match those criteria. The Seelmann paper cites a method to filter out these outliers as well.

-Figure 7: the salinity 31.8 contours are very faint, and will be hard to see in the final paper

-L501: Is "interpolation" the right word here? I usually think of interpolation as filling in gaps in lower-frequency data to match higher-frequency measurements. As in, the TA data in this experiment could be linearly interpolated to match the higher-frequency T, S etc. Were the higher-frequency data subsampled, or averaged somehow, to match the 10-minute TA data?

-L508: describe the statistical test used to compare the fCO_2 data categories.

-L515:516: describe the statistical tests and coefficients which indicate correlation between fCO_2 and TA.

-L530: should be Fig 3c.

-L555 same question regarding interpolation

-Figure 9: it is really hard to tell the two boxes/sets of horizontal lines apart. The open box and shaded box are easy to mix up. One option is to use dashed horizontal lines for the mean and +/- standard deviation of TA_bl, and solid lines for mean +/- standard deviation for TA_bl,interp, or some other similar scheme. The shading doesn't really add anything for me.

-L578-587: I thought this section was really well done.

-L595, fCO_2 returns to near-baseline values

-Figure 10: do all the points in this Figu correspond to points in Figures 8 or 9? Are they all matched up to the underway TA measurements?

-L606: should be uatm, not ppm.

-Figure 10 caption, Line 2: should reference Figure 8 I think, not Figure 7. Also, do you mean "RT concentration" instead of "RT signal"?

-L612: what is a "small but mean" offset?

-L623: should reference Figure 10h. Line 593 days the TA enhancement was >20 umol/kg)

-L624: should reference Figure 10e

-L627-630: Isn't this a big deal? How do the authors suggest the spatial heterogeneity of the patch should be addressed?

-L627: this focused on the peak RT concentration measured in the patch. Could it not have been higher in another spot? Wasn't the patch quite patchy (i.e. Figure 3b)?

-L634: how likely is it for the water mass to remain at the surface? The mass in this experiment advected quite a bit, and presumably mixed as well.

-L643-644: this returns to my question of patch behavior. Did the whole mass travel south, or did it spread out south with some concentration remaining at the original northerly dispersal location?

-L678: "may advect may disperse"

-L705-710: I thought this section was well done.

-Final thought: I suspect that some will read this work, see the 8% CDR efficiency number, and be disheartened for the prospect of OAE. Perhaps the authors can speculate about this some more. How much longer, or how much more broadly spatially, would

observations need to be carried out to observe a higher efficiency? What might be the highest detectable efficiency, given uncertainties in the baseline estimates and analytical instrumentation? More broadly, what even represents a successful CDR? One that is measureable? One that is profitable? Presumably the CDR effects would have continued after the monitoring efforts in this experiment were done, but how might someone realistically do MRV for these ongoing CO2 removals?