Review of Ho et al. : Chapter 6: Monitoring, Reporting and Verification for Ocean Alkalinity Enhancement

This paper is a contribution to a "Guide to Best Practices in OAE research", which is a timely and highly relevant endeavour. The chapter itself is rather short, maybe because it is sandwiched between the Fennel et al modelling chapter and the Schulz et al. chapter on carbonate chemistry measurements. Overall, the chapter is well readable and informative.

Besides a few specific comments, my major comment is on incomplete referencing. The full first section ("1. What is MRV") only includes one reference (plus two references to other chapters in this issue, which are, however, missing in the reference list). Even though I agree with most statements, the full chapter reads more like an opinion piece rather than a scientifically sound paper deeply rooted in the scientific literature. Given that the key message by the authors is that "early stage MRV research for OAE [...] carries a special obligation toward comprehensiveness, reproducibility, and transparency" (Abstract and section 5), I urge the authors to follow their own recommendation and to thoroughly reference the relevant literature. The links and references may be obvious to those deeply involved in this research, but this series of chapters should have the ambition to be a starting point to those new in the field. Thus, please point the reader to relevant literature, so that he/she/it can find further reading material. I give a few examples below, but this holds for the entire manuscript.
My first thought was that there might be a common decision/recommendation of a maximum number of references to be included in the chapters of the best practise guide. If this was the case, I would strongly urge the entire group to rethink their decision. However, looking at the Fennel et al. paper with a substantially longer (though maybe also not complete) reference list, this does not seem to be the case.

In the same spirit, the manuscript would benefit from sometimes sparring a few more words for an additional explanation when statements are rather vague and unspecific, examples are given below.

Specific comments:

Line 20: 'pCO2 and pH' : pH is not mentioned at all in section 3 on measurements. Please add.

Line 28-29 (and in general): MRV is defined to deliver 'the amount of additional CO2 removed from the atmosphere' and 'the durability of that removal'.
What I am missing in this paper is twofold:
- any ocean-centric MRV being through observations or modelling would overestimate the atmospheric CO2 draw-down by neglecting feedbacks from atmosphere and particularly land and thus overestimating atmospheric CO2 reduction by ~25% on decadal scales (Oschlies, 2009, www.biogeosciences.net/6/1603/2009/). There may be ways to account for this even if not using an emission-driven Earth System Model, but these approaches need to be developed, and awareness needs to be raised. I think this is mentioned in the Fennel et al chapter, but don't count on everyone reading all chapters.

- I think it is too simplistic to say that all carbon in the ocean counts equally as all alkalinity stays in the ocean for long time-scales. Isn't it still better (for humans) to get the carbon out of the surface mixed layer and ideally into water masses that will not be in contact with the atmosphere for centuries? I.e., what about the depth of carbon sequestration? Shouldn't this be included in MRV where the carbon goes (horizontally and vertically)?

Line 36: you here promise the reader some text on methane, N2O and DMS. I did not find this mentioned again… add text (preferred) or delete this in our outline.

Line 45: "CO2 that escapes removal": there will be a lot of CO2 that will not be removed, so should this be reformulated to be more specific?

Line 47: "such as": just wondering whether there is anything more to add to the list or if it is complete then drop "such as".

Line 52: The long time-scale of alkalinity cycling in the ocean of Middleburg was challenged by Köhler (2023), who found, for example, a glacial-interglacial amplitude in alkalinity of 100 μmol/kg, which, according to the paper, is equal to 100 ppm in atmospheric CO2. This may not be relevant to the timescales of CDR, but some caution/toning down might be worth considering. I am not sure we have a sufficiently good understanding of changes in river/erosion input of organic and inorganic matter to the coastal ocean and neither of how sedimentary processes in coastal regions change with climate change and other human impacts to be so certain that alkalinity cannot be lost (other than through secondary precipitation and calcification).

Line 64: "stored as CaCO3": CaCO3 formation is usually thought to increase ocean pCO2, so the formulation "CO2 stored as CaCO3" seems to be misleading, and might need a few extra words.

Lines 64-70, 73-74, 76-78, 81, 85-86, 91-93!, 95-100, 114-116, 137: essentially, each sentence needs a reference.

Line 82: one does wonder: there are a couple of global pCO2-products estimating the ocean carbon sink from pCO2-observations. How does MRV relate to these estimates? Do you also consider them unreliable or what makes the difference to quantifying air-sea CO2 flux after OAE from observations? Is it the small scales? It would be welcome, if you commented on that for the reader.

Line 89: hmm, if a long-term climatology from observations would be used as a 'counterfactual scenario', what if the climate state/weather would be so different in the year of OAE (climate variability) that the different climate state/weather can explain all the differences in CO2 uptake. Should a company get credit for this?

Section "2: Specificities of ocean CDR for MRV": it would be nice to embed this for the reader in how this differs from land MRV (just a thought).

Line 122: 'coastal stations, where OAE is likely to be deployed': needs a ref or explanation

Line 134: "ionization fraction": is this not a very similar measure to the Revelle factor?

Line 136ff: not only alkalinity, also carbon will be transported (is that included in your 'OAE signal'?)

Line 142ff: "Lessons learned from OIF": it would be useful to extend this part and explain better. Currently, I do not take any lesson home from this text.

Line 147: "in the context of M (Measurement)": I thought M is Monitoring, but also I don't get the meaning of this phrase.

Line 156-158: not true, at least not always, see Smetacek et al 2012: "A large diatom bloom peaked in the fourth week after fertilization. This was followed by mass mortality of several diatom species that formed rapidly sinking, mucilaginous aggregates of entangled cells and chains."

Line 159: I'm not sure these are 'experimental artefacts' and 'do not represent the ocean C cycle', but yes, even well-defined experiments take place in a complex Earth System and have limitations. The experiments took place in the real world though and do represent the carbon cycle, though maybe not in its entirety or the aspects that you implicitly refer to (please specify).

Line 160: please explain why OAE would be less impacted by physical effects. Alkalinity and carbon are subject to circulation and mixing, only they may not be subject to the same potentially large amount of biological feedbacks.

Section 3: Observation-based techniques:
- as the abstract mentions pH, some text on pH sensors and their limitations should be added. I warmly recommend Wimart-Rousseau et al (BGD): "In the context of converting surface ocean pH measurements into pCO2 data for the purpose to derive air-sea CO2 fluxes, we conclude that the minimum accuracy requirement of 0.01 pH units (equivalent to the minimum pCO2 accuracy of 10 µatm for potential future inclusion into the SOCAT database) is not systematically achieved in the upper ocean."
- I also miss text on discrete samples of pCO2, DIC and TA (and their limitations)

Line 164: 'namely TA and pCO2' and 'DIC throughout the perturbed volume': please spare a few more words to explain, this is not obvious from the previous sentence.

Line 166: 'a carbon budget could theoretically be closed': add evidence or reference? (okay you say theoretically, but still)

Line 184: "restricted to the upper ocean (50m)": please check. I recently read that a SAMI-2 CO2 sensor has a maximum depth of 600 m, and CONTROS HydroC CO2 sensor has a maximum depth of 1000 m.

Line 184: "this is potentially important": does "this" refer to the depth restriction or to CO2 sensors being effective?

Line 186: "measurements of pCO2": according to the title, the Wanninkhof paper is about discrete measurements of pCO2, whereas this paragraph is about autonomous sensors. Are accuracies identical?

Line 188: "little involvement": isn't this a bit too optimistic, what about calibration? How long can they be deployed? A bit more explanation would be useful, as this is a central part of the manuscript.

Line 210: "fit-for-purpose models are not available": even if there were a reference to the Fennel et al chapter (which is not), this statement needs some more explanation, what are the major limitations?

Line 216: "ensemble": what kind of ensemble do you refer to? Perturbed parameters or perturbed initial conditions or different models?

Line 221: "quantifying uncertainty": how does DA help with quantifying uncertainty?

Line 222: Can you find a better title for this section? It is not about adding alkalinity to models.

Line 225: "air sea CO2 exchange": I'm missing "because…"

Line 228-230: please explain/elaborate and add references. You might also need to introduce and explain the term 'near-field'.

Line 230-233, 246-247: add references

Line 234: biological responses will have many more impacts than just on 'efficacy'. Expand and add refs.

Line 253-254: "further empirical research, "this aspect": can you be more specific, what precisely is missing? What do you recommend? What kind of research do we need? What about the inability of the models to accurately represent the background CaCO3 cycle (because we lack a good understanding from field observations; Planchat et al., 2023; Hinrichs et al., 2023)

Section 4.3: I don't understand what this section is about. Why would one neglect feedbacks of biology on circulation? This is included in many models anyway (see e.g. Seferian et al., 2020). It probably doesn't matter much (as you discuss), but in principle this should be included in the OAE simulation (and not in the CTRL/counterfactual simulation). What sort of models does this relate to?

Line 266: Oschlies reference missing in reference list.

Line 268-269: explain your expectation or add a reference.

Section 5:
- It is not quite clear why some recommendations are given as bullet points and others in plain text. Are they qualitatively different (then please explain)?
- I kind of expected a clear recommendation on what measurements to take, or at least that two carbonate system variables should be monitored continuously. I think in the Schulz et al chapter, there is a statement that pH and pCO2 together have higher uncertainty than any other pair of carbonate system variables. Would be useful to refer to this (it's a pity that they are the only ones with autonomous sensors available). If optimal monitoring is still to be figured out, state this.
- What about the overestimation of atmospheric CO2 reduction from ocean-centric MRV alone (see above)?

Line 277-279: add references (at least to Fennel et al.). Data-assimilation wasn't really discussed in this manuscript, it comes a bit as a surprise that this is a key conclusion.

Line 288-289: similarly, 'ecosystem effects and sediment-water exchange' were not discussed in this ms, why is it a key conclusion? If it stays, please elaborate.

Line 301-305: add references. (what is meant with "chapter 2"?)

Line 307: "should be done" → should ALSO be done

Line 315-317: this list of measurements is similar to what I expected as a major recommendation. The recommendation is well hidden though, and not all (not even most) of these measurements are explained in the manuscript.

Line 317: "measurements of carbonate chemistry parameters": too unspecific. How many? Which ones? Isn't this so central to OAE MRV that it should be mentioned more prominently?

Line 318ff: references needed

Line 326: "validated": not even evaluated

Line 329-330: well, the models exist and simulations are being produced. What is the problem? Is it not economically feasible for MRV to run them on supercomputers, or are not enough supercomputers or … please specify.

Line 333: "establish that models": how? More/other computers? More efficient code? Missing processes? Please specify your recommendations.

Line 344: which observations? With regard to what? Please specify

Line 350: Shouldn't the company of the lead author be mentioned as competing interest?

Technical comments:

Line 14: "reporting so the" → reported

Line 106: OGCM → OGCMs

Line 118: mCDR was not introduced

Line 133: ""time scale" → time to reach equilibrium

Line 134: "Revelle buffer factor" → Revelle factor (buffer factor has the opposite behaviour, pick one)

Line 251ff: feedback → feedbacks

Line 315: ADCPs: spell out

References:

Köhler, P. (2023). Atmospheric CO2 concentration based on boron isotopes versus simulations of the global carboncycle during the Plio-Pleistocene. Paleoceanography and Paleoclimatology, 38, e2022PA004439. https://doi.org/10.1029/2022PA004439

Planchat et al., 2023: https://doi.org/10.5194/bg-20-1195-2023

Seferian et al., 2020: https://doi.org/10.1007/s40641-020-00160-0

Smetacek et al 2012, doi:10.1038/nature11229

Wimart-Rousseau, C., Steinhoff, T., Klein, B., Bittig, H., and Körtzinger, A.: Technical note: Enhancement of float-pH data quality control methods: A study case in the Subpolar Northwestern Atlantic region, Biogeosciences Discuss. [preprint], https://doi.org/10.5194/bg-2023-76, in review, 2023