

## ***Interactive comment on “Ideas and Perspectives: When ocean acidification experiments are not the same, reproducibility is not tested” by Phillip Williamson et al.***

**M. Solan**

m.solan@soton.ac.uk

Received and published: 11 November 2020

Some observations that might help improve the narrative:

1. The opinion reinforces the argument about the need for proper experimental design, both within and across experiments, but doesn't explicitly provide solutions or minimal requirements. I have emailed you the recent Haddaway perspective (<https://doi.org/10.1038/s41559-020-01295-x>), maybe some of the styling from that would be good, but either a Boxed flow diagram or take home message at the end of each section of what readers should do would be good.

Printer-friendly version

Discussion paper



2. Three things are missing for me. (i) one glove doesn't fit all, and what is high variability in one system (e.g. pelagic) is within the noise in another system (e.g. benthic), and that needs to be recognized, especially in review. Variability in carbonate chemistry between systems has not been summarised anywhere, although could be now that there are so many papers, and then these need to be couched within temporal variation (e.g. diurnal) for the same system. (ii) the detection of treatment effects is valid, even if the carbonate chemistry is different to other experiments. You hint at this, but this would benefit from some elaboration. (iii) reading most ocean acidification papers, the narrative is that OA is 'bad' which is not necessarily true. I take the point that calcifiers are affected, but would challenge the statement that they are most sensitive as this is only one parameter and there is a bias in the literature (people have picked calcifiers). You highlight other examples, including behaviour, in non-calcifying species which could be equally devastating to that species. I key message is that this literature base needs to move on from documenting effects, and think about what the consequences of those effects are for species interactions, fitness, reproduction/growth and the rest of the foodweb etc etc.

3. is there a standard checklist, or could you provide one in this article, of what authors should be reporting in every paper, e.g. in a table in supplementary? The carbonate chemistry, but what else? Which of these are necessary, and which are nice to have?

4. It would be good if you could add some commentary about being pragmatic. Alkalinity, in particular, is expensive so an experiment with hundreds of replicates cannot hope to achieve regular samples from all units on a daily basis. You need some, but there are ways to achieve something sensible (e.g. a set number of aquaria within each treatment, once a week or something). There has to be some common sense, but also some indication of what the acceptable minimum is. The point I am making, and have witnessed at several OA meetings, is that the conversation about chemistry can go way further than is needed when discussing accuracy, repeatability and reproducibility. All of these depend on the system you are in, what the question you are asking

[Printer-friendly version](#)[Discussion paper](#)

is, and what is practically possible, i.e. the requirements are context dependent. A related point is that what works in one system should not dictate what is acceptable in another system. An analogy is the US water quality standards – when they were brought in some states were automatically above threshold as the ground composition was markedly different to the areas where the standards were formulated.

5. Variability – there are more sophisticated statistical methods available to look at variability and outliers. There is a danger that trying to make everyone conform to a very controlled set of conditions means that you lose the insights from the variability that you have factored out. Part of the answer has to be embracing variability and using appropriate statistical approaches to account for and/or explore them. Meta-analysis is one way to look at multiple experiments, but not the only way and much could be done with mixed modelling, GLMM, GAMM and then specific analyses that analyse outliers (rather than account for them).

Hope this helps

---

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2020-394>, 2020.

Printer-friendly version

Discussion paper

