Biogeosciences Discuss.,
doi:10.5194/bg-2015-641-SC2, 2016

Biogeosciences Discussions

Interactive comment on "Effects of *in situ* CO_2 enrichment on structural characteristics, photosynthesis, and growth of the Mediterranean seagrass *Posidonia oceanica*" *by* T. E. Cox et al.

J. Havenhand

jon.havenhand@bioenv.gu.se

Received and published: 16 February 2016

This manuscript has already received two insightful reviews and one very useful comment. All of those have rightly praised Cox et al for doing a difficult job well: long-term, well-documented, and well-controlled FOCE studies are very difficult and the authors should be congratulated for the work they present here.

I am writing this comment in response to the acknowledged lack of true replication in this study, while noting that I too am not an expert in these issues but have some observations that may be pertinent to others.

The authors use the term "pseudo-replication" (sensu Hurlbert, 1984, Ecol. Monogr.) to

describe their repeated sampling within the same mesocosms. This term is commonly used, but misleading in that the replication is not "false" (the true meaning of "pseudo"), but rather is at the wrong level for the question we are asking. Cox et al measured many parameters on multiple occasions in their two FOCE chambers and their Ambient plot, but in all cases these measurements were made on the same plants in the same 3 plots (each of which received a different treatment). Consequently, while the replicate observations provide a very accurate picture of what happened in those three plots, they provide no information on the variability in response among plots within the same treatment. Consequently, Tom Arnold is correct in his comment to state that the authors have only one replicate per treatment.

For most of the experiments we run this is a problem: we typically wish to estimate population-wide responses from (hopefully) many "replicate" random samples that we have manipulated in our experiments. However, in some instances these true replicates are difficult to obtain, and this is particularly so when the equipment we need to generate them is scarce and/or expensive. Therefore we have to resort to alternative approaches. This doesn't mean the data we obtain are not valuable (they are!) but it does mean we may have to use different approaches to evaluate them.

Unreplicated designs in ecology are not new, and several seminal papers have been written on unreplicated systems. Among the best known are the papers of Carpenter and co-workers on trophic cascades in lakes (see e.g. SR Carpenter et al. 1987, Ecology). In these experiments each treatment had only one replicate (lake), but the authors had the benefit of many years of historical data on the ecology of those lakes. By monitoring the lakes over a long period, quantifying the natural variation over that period, then imposing manipulations and monitoring the responses, they could observe the degree of change in the different treatments and relate those changes to their hypotheses. This approach is analogous to that of Cox et al, who monitored the seagrasses in their FOCE chambers for an extended period before imposing their treatments. This approach is clearly not as reliable as multiple FOCE units per treatment

(something Cox et al point out), but it is far better than simply imposing the treatments from the outset.

Importantly, these sorts of design often do not need complex statistical testing to show the effects of the treatment. Simple graphical plots of responses over time showing divergence between treatments and controls are often sufficient. This was the case in the Carpenter et al paper cited above, but not in Cox et al – who found relatively little effect of treatment (their Figs 2 & 3).

This raises the last important point: the absence of a statistically significant effect in a formal statistical test means the experiment was inconclusive (not that there was "no effect"). It is possible that the results Cox et al describe (i.e. only small effects of OA on Posidonia) are correct, however we cannot discriminate between that possibility and the possibility that the experiment had too little statistical power to be able to detect effects. This is an unfortunate outcome for such a comprehensive piece of work. Rather than using statistical testing to reduce the complexities and nuance of experiments like this to a series of binary "probably" vs "don't know" answers, I suggest it would be more useful to only refer to the Figures and interpret the scale and magnitude of the effects the authors observed. This argument applies irrespective of whether there are many – or no – replicates at the level of interest.

In closing I echo Tom Arnold's point that lack of true replication in many complex experiments (not only FOCE experiments) limits our conclusions. We need appropriate replication – not only within treatments, but also across time, populations, and habitats – in order to understand the variability of responses that marine organisms may show to future marine climate change. By showing how we can measure relevant responses "in-situ", Cox et al have made a highly valuable step down that road.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2015-641, 2016.