Review of Yoon *et al*. *Ocean Iron Fertilization Experiments*

I am sorry to hear that the funding for the proposed Korean Iron Fertilisation Experiment has fallen through, and I wish the authors good luck in trying to secure future funds to conduct this study.

While the initial justification of the manuscript was the presentation of this proposed experiment, I believe that publication of the manuscript even in the absence of secured funding is appropriate. The paper seeks to provide a useful synthesis of previous experiments, as well as to provide design guidelines for future experiments. A key point is that the previous experiments still leave us with many open questions as to the response of marine ecosystems to iron addition.

The authors have made quite extensive changes to the manuscript, which go some way towards addressing my previous comments. For example, the two new tables (now Table 2 and Table 5) help a lot to provide a better overview of the different iron fertilisation studies.

However, I still feel that the actual manuscript itself could be significantly improved if the authors could draw clearer, more specific conclusions to highlight the gaps in our understanding, and identify specifically what is needed to address these gaps. In its present form, I find that the manuscript is often rather long-winded (*e.g.* explaining very basic concepts such as the ^{234}Th method, or mentioning in the text which research ships were used for which experiments) without actually drawing clear and specific conclusions. Numerous reviews of iron fertilisation experiments have been published previously – albeit not of all 13 experiments that have now been conducted – and what I really miss in the present paper is a coherent argument and set of justified conclusions emerging from this review. In principle, I think this manuscript would be a good opportunity to examine specifically what key questions remain unanswered and what specifically would need to be done to address them. The manuscript at present still describes a lot of results, but I think without really drawing more than rather vague conclusions.

I will first go through the manuscript section-by-section to point out where and why I think the manuscript could be improved. I will end the review proposing a possible structure for the authors to re-work their manuscript into a more useful review that manages to move the field forward further. I hope that this is helpful for the authors – I do not intend to do this effort down, because I think that a review such as this, coupled with a proposal for a new experiment could be genuinely useful. However, I fear that the present manuscript misses a valuable opportunity, and I think that in its present form the manuscript will not achieve as much impact as it might if the authors attempt a more thorough discussion and evaluation, rather than just listing of results.

1. The abstract does not include any summary of the main conclusions that the authors have drawn from their synthesis.

2. It is not clear to me what the purpose of the entire Section 2.1 is. Ostensibly, the authors attempt to describe the objectives of each experiment, but I'm not convinced that this information is really important enough to merit a 2-page section, nor am I convinced that the authors explain these objectives terribly clearly. The section mixes general introduction to each experiment (*e.g.* descriptions of the oceanographic conditions of each site) with usually only a vague, general statement of purpose for the experiments (*e.g.* "To investigate the unexpected responses revealed in IronEx-1, a second OIF was conducted", or "To measure biologically-driven gas fluxes"). For some experiments, the authors only list briefly the main conclusions, but don't explain much about the objectives and design.

3. Section 2.2 is basically a long list of measurements of initial conditions for the various experiments. Many of these have been reviewed previously, and the authors don't draw any conclusions from this 1.5-page-long section. The authors have done an excellent job of summarising this information in their tables (which I think is very useful). But since the main message seems to be that OIF have been conducted under a diverse range of initial conditions, wouldn't it be better if the authors just made that point in, maybe, 1–2 paragraphs, and then focused on advancing an argument and drawing conclusions?

4. The subject matter of Section 2.3 is an important one: how many additions of iron should be made, and how should the fertilised patch be traced? What I miss in this section is to see the authors actually draw conclusions from these results. For example, are multiple additions the right way to design an experiment? At what intervals should they be made? The experiments listed vary a lot in duration, so simply listing the total number of additions, rather than the intervals at which they were made, seems uninformative. Likewise, it might be useful to see some discussion of the amount of iron added per square kilometre of patch – I believe that the additions for the various experiments varied a lot also on a per-area basis, but this is not discussed. Can we determine an optimum rate of application from the various experiments? If not, is this something that warrants further experimental work, and if so, of what kind? Likewise, the discussion of tracing methods just boils down to listing the fact that SF_6 and buoys can be used, without drawing conclusions that could be formulated as practical recommendations. For example, drifting buoys in previous experiments were found to ultimately leave the patch, I believe due to wind forcing, and one of the recommendations of this paper is that experiments should be conducted that last upwards of 30 days, under which conditions multiple SF_6 injections might become more necessary.

5. Section 2.4 I think mostly re-hashes just the main findings of previous experiments, and again does so in an often rather long-winded way (*e.g.* the information about spatial pattern of the SOFEX blooms doesn't seem to contribute much). All in all, this section does not really build up an integrated understanding of the biogeochemical responses to iron addition, but reads rather more like a selective catalogue of results picked rather haphazardly from the different experiments: *e.g.* the difference in maximum chlorophyll between SEEDS-I and SEEDS-II is pointed out, but this isn't part of a broader discussion of chlorophyll levels across the different OIF and what their causes might be. In the next paragraph, SEEDS-I and EIFEX are compared with max chl-a concentration and integrated chl-a stock, but this is not compared with other experiments. My point here is not to ask the authors to add specifically a discussion of chlorophyll, I am just pointing this out as an instance of where the paper rather shies away from reaching interesting new insights. The entire section then

simply ends, without any attempt to conclude anything from all this information, aside from pointing out in the middle of the penultimate paragraph that integrated primary production ought to be monitored in future experiments.

6. I had previously requested that the authors make explicit definitions of terms like “efficiency” and “sequestration” in the context of carbon fluxes, but these terms are still used loosely without specific definitions. I think this is a significant problem when discussing the planning of future iron fertilisation work, as it is critical to achieving the stated objectives to ensure that the correct measurements are made, and this can only be done if we define clearly what we need to measure. I would suggest that the authors refer to the Lampitt et al. paper in the 2008 special issue of the Philosophical Transactions (same issue as the cited Smetacek & Naqvi paper), titled “Ocean fertilization: a potential means of geoengineering?”. This paper discusses explicit definitions of terms such as carbon sequestration with reference to the depth of winter mixing, and discusses how they can be measured.

7. Section 2.5 again leaves me rather unsatisfied: the authors present a list of findings, but don’t make any argument or properly discuss reasons for the divergent outcomes. As a result, by the end of the section, it is not clear to the reader why “the effectiveness of iron addition on this component of the biological pump remains a question”. In other words, the authors should be using this section to discuss why the previous 13 experiments have not managed to yield clear answers to the effectiveness of carbon sequestration. The answer lies in a combination of experiment duration, measurement methods (measuring only shallow or also deeper fluxes), patch size (with a very small patch, I suspect that deeper traps might miss export that may be occurring because the plume of sinking particles is confined to such a small area), and patch movement (tracking deep export is easy in a stationary patch, but very hard in a large patch). Although the authors do draw these conclusions in a very general sense, saying at the end that future experiments need to last long enough, fertilise a largeish area, and use multiple methods, the justification of publishing a review paper like this one must lie in a more detailed analysis of how which of the previous OIF were unable to achieve these requirements. To do this, the authors would need to delve more into the details of each experiment, rather than spending most of the time reviewing the basic biogeochemical responses of each experiment. Again, terms like “effectiveness of iron addition” could be defined in very specific ways, and using language as loosely as this does not help with achieving insights.

8. Section 2.6 starts as a promising paragraph (in fact, I think this could be a good introductory paragraph to the entire Section 2), but then falls flat for the same reasons pointed out above. What are we to conclude from these diverse outcomes? Which experiments may have missed an export event, or deep fluxes of sinking particles, and for what reasons? The issue of ecosystem responses and grazing is clearly an important one, but receives hardly any discussion.

9. After all this extensive review of previous OIF in Section 2, the review suddenly moves on to possible side-effects. What conclusions can be drawn from all of the data that the authors have reviewed in Section 2?

10. Section 3.1 is slightly better, as it at least ends with a proper attempt at conclusions. However, the section is again mostly a cataloguing of results from previous experiments. What would be more useful is if the authors attempted to synthesise the insights and discussions of the publications about the individual studies so that we can at try and find patterns in the large variability between studies.

11. Section 3.2 implies a discussion of the international legal situation, but the first paragraph is not about legal matters at all. The second paragraph does a decent job of summarising the legal situation. However, since the London Convention explicitly grants exemption only to scientific experiments in "coastal waters" I wonder where that would leave open-ocean experiments in practice. Of course, iron fertilising coastal waters is pointless, as everyone in the oceanographic community knows, but are the authors confident that legitimate research projects in the open ocean will get approval? Some clarification of this seems called for here, at least to the best of the authors' ability.

12. Section 4 finally attempts to put forward some practical suggestions. However, I think that this needs to be expanded on and made more specific. Simply concluding that "multiple additions of iron are more efficient" is not the same as formulating specific guidelines about fertilisation levels (e.g. minimum addition per km^2) and application intervals. Likewise, the third paragraph in this section ("What") is much too brief a discussion to be genuinely useful and to move the field forward. A long list of measurements is put forward without discussion of the relative merits of each, or a real discussion of the difficulty of tracking a fertilised patch, and whether we've learnt how to do this better over the 13 experiments that have been conducted. Likewise, the discussion of how to measure carbon fluxes is too indiscriminate a list to be of much use. What would really help would be an evaluation of what the various measurements contribute to our understanding, what their pitfalls are (especially, for example, surface-tethered sediment traps, which are probably better avoided in favour of their neutrally buoyant counterparts), and, again, where in the water column we really need to be measuring. The fact that the sections concludes in saying that carbon fluxes need to be monitored using "both trap fluxes and/or ^{234}Th deficiency" is somewhat worrying: surely, one of the lessons from previous experiments is that both measurements really are needed, and that traps must be at multiple depths so that we can track both the export out of the surface as well as the sequestration below the depth of winter mixing? In discussing possible candidate regions, it would be useful if the authors proposed specific regions, rather than just making general observations such as recommending "regions with high silicate concentrations and low copepod abundances".

13. The manuscript ends with the design proposal for KIFES. Even though KIFES currently has no funding, it is an experiment that the authors clearly hope to conduct in more or less this form in the future. Therefore, I think it is valuable to have this section in the manuscript. What might be useful is if the authors used this opportunity to explain in slightly more detail how the design of KIFES will avoid the problems encountered in earlier experiments, i.e. explain the specifics of duration, patch size, measurement methods, etc.

There were additionally numerous minor points about specific phrasing that I felt could be improved throughout the manuscript. However, I think there is little point in listing all of these, since I suspect that the authors will need to re-write this manuscript in a relatively major way anyway. I actually believe that this might not necessarily involve too much work: what I would recommend, however, is that the authors think carefully about what their main message is from reviewing all of the previous studies, and that they seek to systematically craft these arguments instead of rather indiscriminately listing previous results.

I would suggest that a revised structure might look something like this:

1. Overall introduction to OIF and statement of the purpose of the present paper.

2. A brief overview of each of the experiments, maybe written as a historical narrative (1–2 pages) that describes how the various experiments built on each other and what the main hypotheses were that each was designed to address. This section can be used to highlight the different physical and biogeochemical conditions of each experiment, and also highlight the main biogeochemical responses and findings. This experiment-by-experiment approach might help to give the reader are more integrated understanding of each experiment than the current parameter-by-parameter approach used throughout Section 2. The section could then end with a paragraph highlighting the key outstanding questions that future experiments (including KIFES) need to address. These include questions relating to the amount of carbon sequestration, trace gas production, and plankton community shifts.

2. A detailed discussion of each of these questions in turn, highlighting why the previous 13 experiments have failed to reach consensus, and what needs to be done to move our understanding forward. This needs to be an issue-based discussion, for example going into real detail about how carbon fluxes need to be measured, which experiments managed to take these measurements (but failed to find effects), which experiments were maybe hampered by size and/or duration, and also a discussion of the uncertainties relating to each method (*e.g.* methodological problems with sediment traps, issues with the thorium technique, problems with the estimation of net community production from O_2:Ar ratios, etc.).

3. A section that reaches specific conclusions and makes recommendations about the design of future experiments.

4. Specific proposal for KIFES and description of the logic for conducting this in the Bransfield Strait. Do the authors have some preliminary altimetry images to show stable eddies in this region?

Finally, I think the authors might not want to neglect the ecological arguments for conducting OIF experiments: terrestrial ecologists have learnt a large amount about the functioning of terrestrial ecosystems from conducting nutrient manipulation experiments. OIF are effectively our only way of manipulatively testing similar hypotheses about marine pelagic ecosystems, and I believe that OIF could be a useful tool purely for scientific purposes, regardless of geoengineering feasibility.

End of review.