Limnol. Oceanogr., 52(2), 2007, 916–918

Response to: "Interpreting the results of oceanic mesoscale enrichment experiments: Caveats and lessons from limnology and coastal ecology"

In their comment, Hale and Rivkin (hereafter referred to as HR) criticize the general design of the open-ocean fertilization experiments carried out so far because of the lack of replication. The basic design involves fertilizing a large area of the surface ocean and then monitoring the biological responses inside the fertilized patch and those in unfertilized out-stations allocated in the vicinity of the fertilized area. Logistic constrains such as ship time and resource availability make it unfeasible to fertilize and continuously monitor more than one patch at a time. In addition, true replication in large-scale oceanic perturbation experiments may not be feasible at all because it would require finding several physically separated sites presenting similar initial conditions and subjected to similar environmental forcing for the length of the experiment. Therefore, replication efforts are likely to result in approximate repetition rather than exact replication. Several oceanic mesoscale enrichment experiments have reported increases in phytoplankton biomass or primary production inside the fertilized patches as compared with unfertilized surrounding waters (Boyd 2002; Boyd et al. 2004; Coale et al. 2004; De Baar et al. 2005; Thingstad et al. 2005; Le Clainche et al. 2006), which have been attributed to the effect of enrichment. HR rightfully argue that the use of unreplicated experimental designs means that it is not possible to unequivocally attribute the measured responses to the effect of the enrichment.

Because of this lack of replication of experimental units, HR criticize the interpretation and the validity of the statistical analyses performed in one of our publications in this journal (Arrieta et al. 2004), hereafter referred to as A2004 or "our paper." Testing for statistically significant treatment effects would require replicated experimental units and controls, but this is not what we did. As HR acknowledge in their comment, the Mann–Whitney *U*-test can be used to test whether there are differences between the values measured inside and outside of the fertilized patch (this is what we did). Note that every time that statistics are mentioned in A2004, they are also explicitly referred to "the values" of the bacterial properties found inside and outside the patch and not to the treatment. HR also find discrepancies between their calculations of the Mann–Whitney *U*-test and the values we reported in our paper. They conclude that the differences are due to the "unexplained and selective exclusion of data" in A2004. The decision of taking only the values after day 5 was not arbitrary, as suggested by HR, but was based on visual inspection of the data shown in figs. 1, 2, and 3 in A2004. These figures suggest that if there were any differences between the fertilized patch and the surrounding waters, the differences were not immediately measurable in the first days of the experiment but only became apparent after day 5. There are several instances in the results and discussion where the phrase "after day 5" is mentioned, to clarify that our claims refer only to that period. Therefore, we used the Mann–Whitney *U*-test only to evaluate whether the values of the different bacterial properties measured inside the patch after day 5 were different from those measured in the surrounding waters. Thus, the values reported in A2004 are meaningful within the context mentioned in our paper and useful to confirm the presence of increased values of some bacterial properties inside the fertilized patch as compared with unfertilized waters as suggested by graphical inspection.

An additional limitation of this design, not included in the criticisms formulated by HR, is the use of unlabeled out-stations. Open ocean surface waters are highly dynamic; the location and shape vary over time. Thus, it is necessary to use drifting buoys and conservative tracers like SF6 to ensure that consecutive samples come from the originally fertilized patch. However, out-stations are commonly chosen on the basis of the presence of background levels of SF6 in areas close to the fertilized patch, but there is no indication of whether or not the out-of-the-patch samples belong to a coherent experimental unit. Therefore, in the absence of a mechanism allowing consecutive sampling of a defined water mass, out-of-the patch measurements can only be used as an indication of the variability of "background" levels of the parameter studied in the area. Although we dedicated the first lines of the discussion to stress this point in A2004, HR do not mention this issue, which should be added to their list of concerns. This limitation affects directly the alternative analyses of our data proposed by HR. These authors suggest the use of an analysis of covariance (ANCOVA) including time as a covariate in order to assess whether the temporal trends in the Fe-fertilized patch did differ from those observed in the surrounding waters. However, because the out-of-the-patch stations are not necessarily linked to a coherent water mass, repeated sampling of the same experimental unit over consecutive days cannot be taken for granted. Therefore, it is impossible to determine whether any development observed in the variables measured outside the patch is the result of an ongoing process or that of randomly sampling different sites. This reasoning would preclude the use of ANCOVA for the out-of-the-patch data.

Although we are not confident about the usability of ANCOVA in this situation, we tried to repeat the analyses reported by HR. As stated by HR, we calculated the mean values of each parameter over the upper 40 m of the water column inside the patch and in the surrounding waters, excluding only those measurements of the edge stations, as in the original paper. ANCOVA analyses are based on a linear model. If this model fails, inferences of ANCOVA will be in error (Sokal and Rohlf 1997). Because most of the observed responses were not linear with time, we tried transforming the data (log or square root) as stated by HR to obtain a significant linear regression ($p < 0.05$) with time. The significance of different regression lines obtained by the use of untransformed values, log-transformed values, or square root–transformed values versus time is listed in Table 1. Most of the regression coefficients were not significant when either untransformed values or any of the proposed transformations were used, and no single case was observed where the regression coefficients were significant for both in- and out-of-the-patch values by the same transformation. Moreover, visual examination of residuals versus fitted data plots shows clear evidence of nonlinearity and/or high leverage for most of the regression lines by using either transformed or untransformed data. Although HR claim to have tested the linearity requirements of the ANCOVA analysis, we did not find any usable significant linear relationship in the data. Therefore, we must conclude that the results of their ANCOVA analysis are meaningless and should be disregarded. All analyses were conducted using R 2.2.0 (R Development Core Team 2005).

The most important valid criticism formulated by HR in their comment is that we concluded that the observed induction of bacterial activity was caused directly or indirectly by iron fertilization. We admit that it is not possible to completely rule out alternative explanations with an unreplicated experiment. This is an important fact that has been ignored in previous reports where the observed biological responses have been attributed to the effect of iron enrichment (Boyd 2002; De Baar et al. 2005; Le Clainche et al. 2006), including those signed by HR (Boyd et al. 2004; Le Clainche et al. 2006). Although each of these unreplicated experiments alone may provide weak evidence in favor of the claimed effect of iron fertilization, the fact that different teams have found an enhancement of primary production in different regions of the ocean cannot be ignored. The repetition of these unreplicated experiments by different teams that used sometimes different methodologies adds weight to our inference (Hawkins 1986) and may be more informative than mere replication (Carpenter 1990). Thus, we would like to stress

Table 1. Significance (p values) of the regression lines of untransformed and transformed bacterial properties versus time. Values in bold indicate a significant ($p < 0.05$) slope of the regression line.

	Untransformed		Square-root transformed		Log transformed	
	Out	In	Out	In	Out	In
Bacterial abundance	0.64	**0.02**	0.67	**0.02**	0.70	**0.02**
Leucine incorporation	0.28	0.16	0.27	0.18	0.27	0.20
Thymidine incorporation	0.24	0.36	0.22	0.40	0.21	0.44
α-glucosidase	0.82	0.90	0.77	0.81	*	0.52
β-glucosidase	0.96	0.49	0.95	0.42	0.94	0.40
Aminopeptidase	0.82	0.11	0.81	0.12	0.81	0.12
Cell-specific leucine incorporation	0.67	0.19	0.65	0.22	0.63	0.24
Cell-specific thymidine incorporation	0.54	0.39	0.50	0.37	0.46	0.35
Cell-specific α-glucosidase	0.95	0.58	0.85	0.72	*	0.99
Cell-specific β-glucosidase	0.85	0.80	0.90	0.96	0.94	0.85
Cell-specific aminopeptidase	0.71	0.17	0.71	0.17	0.71	0.17

*α-glucosidase and cell-specific α-glucosidase contained "zero" values, and therefore the logarithmic transformation was not used.

that despite the design limitations outlined by HR, the open-ocean mesoscale enrichment experiments performed so far have made very important contributions to our understanding of biological processes in high nutrient–low chlorophyll (HNLC) regions of the ocean. The repetition of unreplicated experiments has resulted in compelling evidence in favor of the hypothesized effects of iron fertilization on phytoplankton in HNLC regions of the ocean. We are confident that similar contributions from recent unreplicated experiments will be extremely useful in elucidating the dynamics of other compartments of the food web, which have received comparatively less attention. Future reports on the outcome of unreplicated perturbation experiments in the open ocean should discuss the limitations of the experimental design. However, these limitations should not be an excuse to disregard these data and those not yet published, because accumulating evidence will be an invaluable source of information for future meta-analyses.

Finally, we agree with HR in that the design of future open-ocean fertilization experiments should be improved. Although the logistic constraints of open ocean research are very different from those in limnology and coastal ecology, we may be able to adapt some of the designs that have been successfully applied by our colleagues.

Jesús M. Arrieta
Instituto Mediterráneo de Estudios Avanzados (CSIC-UIB)
Miquel Marqués 21
07190 Esporles, Mallorca, Spain

Markus G. Weinbauer
Laboratoire d'Océanographie de Villefranche (LOV)
B.P. 28
06234 Villefranche-sur-mer, Cedex-France

G. J. Herndl
Department of Biological Oceanography
Royal Netherlands Institute for Sea Research (NIOZ)
P.O. Box 59
1790 AB
Den Burg, The Netherlands

References

Arrieta, J. M., M. G. Weinbauer, C. Lute, and G. J. Herndl. 2004. Response of bacterioplankton to iron fertilization in the Southern Ocean. Limnol. Oceanogr. **49:** 799–808.

Boyd, P. W. 2002. The role of iron in the biogeochemistry of the Southern Ocean and equatorial Pacific: A comparison of in situ iron enrichments. Deep Sea Res. II **49:** 1803–1821.

———, and others. 2004. The decline and fate of an iron-induced subarctic phytoplankton bloom. Nature **428:** 549–553.

Carpenter, S. R. 1990. Large-scale perturbations—opportunities for innovation. Ecology **71:** 2038–2043.

Coale, K. H., and others. 2004. Southern Ocean iron enrichment experiment: Carbon cycling in high- and low-Si waters. Science **304:** 408–414.

De Baar, H. J. W., and others. 2005. Synthesis of iron fertilization experiments: From the Iron Age in the Age of Enlightenment. J. Geophys. Res. **110:** C09S16.

Hawkins, C. P. 1986. Pseudo-understanding of pseudoreplication: A cautionary note. Bull. Ecol. Soc. Am. **67:** 184–185.

Le Clainche, Y., and others. 2006. Modeling analysis of the effect of iron enrichment on dimethyl sulfide dynamics in the NE Pacific (SERIES experiment). J. Geophys. Res. **111:** C01011.

R Development Core Team. 2005. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing [accessed July 2006]. Available from http://www.R-project.org

Sokal, R. R., and F. J. Rohlf. 1997. Biometry, 3rd ed. W. H. Freeman.

Thingstad, T. F., and others. 2005. Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean. Science **309:** 1068–1071.

Received: 7 April 2006
Amended: 2 August 2006
Accepted: 8 August 2006