

COMMENT

Limnol. Oceanogr., 52(2), 2007, 912–916
© 2007, by the American Society of Limnology and Oceanography, Inc.

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

Since the first mesoscale iron enrichment experiment was conducted in the equatorial Pacific (Martin et al. 1994), there have been more than nine open-ocean experiments investigating the ecological effects of iron or phosphorous additions on planktonic communities (*see* de Baar et al. 2005; Thingstad et al. 2005). In these experiments, a large area of the ocean is enriched with acidified iron sulfate or phosphoric acid solution, and the inert tracer, sulfur hexafluoride (SF₆). Samples are collected at different depths, inside and outside the enriched patch. The in-patch stations are normally chosen on the basis of their high SF₆ concentrations, whereas the out-patch stations, which are sampled to monitor changes in the unperturbed waters (i.e., reference site), typically have background levels of SF₆. Many mesoscale enrichment experiments have reported differences in the biological responses inside and outside the patch, such as increases in chlorophyll *a* concentration or primary production inside the patch (e.g., Martin et al. 1993; Boyd et al. 2004; Thingstad et al. 2005) and have attributed these differences to the effects of the enrichment. However, few studies have applied statistical analyses to determine whether responses inside and outside the patch are statistically significantly different, and none of the many published mesoscale iron or phosphorous enrichment experiments has discussed the limitations of the unreplicated experimental design in determining whether responses inside and outside the patch can indeed be unequivocally attributed to the effect of the enrichment.

Two papers published in this journal, Arrieta et al. (2004; hereafter referred to as A2004) and Oliver et al. (2004; hereafter referred to as O2004), used Mann–Whitney *U*-tests and two-sample *t*-tests, respectively, to statistically compare mean bacterial properties and processes inside and outside the patch during in situ mesoscale iron-enrichment experiments conducted in the Atlantic (A2004; EisenEx) and Pacific (O2004; SOFeX) sectors of the Southern Ocean. Although these studies make important contributions to our understanding of microbial processes in polar and subpolar high nutrient–low chlorophyll regions, the interpretations of the statistical analyses applied to the data are flawed, and therefore the conclusions made about the effects of iron on heterotrophic microbial processes are not supported by the data presented in these papers. The problems associated with the analysis and interpretation of data from unreplicated mesoscale enrichment experiments apply equally to the assessment of phytoplankton and zooplankton responses,

and differences in physicochemical properties. However, this comment focuses on the bacterial responses reported in A2004 and O2004 because these papers, unlike those that have reported responses of other ecosystem properties without statistical analyses, have incorrectly applied and interpreted inferential statistics to test for significant differences between responses inside and outside the iron-enriched patch. Here, we (1) discuss problems with the use, presentation, and interpretation of the statistical analyses in A2004 and O2004, and the possibility of confounding treatment effects due to the lack of spatial replication of experimental units; (2) present alternative analyses of the existing data and consider other explanations for any observed differences; and (3) recommend experimental designs, many with their origins in limnology and coastal ecology, appropriate for future ecosystem-scale perturbation experiments (e.g., in situ mesoscale iron or phosphorous addition experiments), where replication is not feasible.

A2004 and O2004 described the time course of changes in bacterial and other biological and physical properties within an iron-enriched patch and applied Mann–Whitney *U*-tests and two-sample *t*-tests, respectively, to statistically compare mean bacterial properties and processes inside and outside the patch. Neither paper clearly stated which data were included in the analyses or provided an a priori justification for the selection of the data used in the analyses. For example, O2004 applied two-sample *t*-tests to data for three time periods (days 5–22, days 8–22, and days 12–22), and compared data from inside the patch for each period with data from outside the patch over the entire observation period (days 5–22). In addition, neither paper clearly stated how the assumptions of the statistical analyses, in particular the independence of observations, were checked or validated. They also appeared to combine data from samples collected on different days with that of samples collected from different depths on the same day, to provide replication for their analyses.

When significant differences were found in the analyses, A2004 and O2004 attributed them to effects of the iron enrichment on the respective property. For example, on the basis of reported significant differences ($p < 0.05$) in bacterial abundance, production, and the abundance of cells with high DNA content inside and outside the patch, O2004 concluded that the addition of iron “significantly enhanced phytoplankton and bacterial properties.” Similarly, on the basis of reported significant differences ($p < 0.05$) in bacterial production and cell-specific ectoenzy-

matic activity inside and outside the iron-enriched patch, A2004 concluded that bacterial activity was enhanced by iron enrichment. However, in both studies, the analyses used can only identify significant differences between bacterial properties and processes inside and outside the patch, and cannot distinguish whether differences were caused by the effects of iron enrichment or whether the bacteria were responding to some other factor, such as differences in temperature, salinity, mixed layer depth, nutrient availability, or grazer activity (Stewart-Oaten et al. 1986; Quinn and Keough 2002).

Mann–Whitney *U*- and *t*-tests assess treatment effects by comparing the variation between treatment groups with the variability between experimental units within each group. However, the variability between experimental units within each group can only be estimated if the treatment itself is replicated, and if the replicate experimental units are independent (Hurlbert 1984; Quinn and Keough 2002). Replicate experimental units are independent in that the outcome of a given replicate has no effect on the outcomes of any other replicates (Carpenter 1990), so that the replicates represent the total variability of the specified experimental condition. It is not intrinsically wrong to use a Mann–Whitney *U*- or two-sample *t*-test to test, for example, the null hypothesis that there is no significant difference in bacterial properties inside and outside an iron-enriched patch (Cottenie and De Meester 2003; Oksanen 2004). However, it is inappropriate to apply these statistical tests to data from an unreplicated mesoscale in situ iron enrichment experiment and to interpret low *p* values (i.e., $p < 0.05$) as evidence that the addition of iron caused a significant change in bacterial properties (Hurlbert 1984; Carpenter et al. 1998; Hurlbert 2004). In addition, because successive samples from a single experimental unit are highly associated and are quite likely to be correlated (Legendre et al. 1985), the potential for spurious treatment effects is very high. Thus, if one experimental unit (e.g., the iron-enriched patch) is repeatedly sampled through time, as is often done, and the samples are treated as if they were independent replicates, which they are not, the effects of the treatment (e.g., iron enrichment) are confounded with the inherent differences between the iron-enriched and unenriched waters (Hurlbert 1984; Eberhardt and Thomas 1991; Carpenter et al. 1998).

Interpretation of large-scale ecosystem manipulation experiments, such as iron enrichments, can be divided into two simple and interrelated questions: (1) did the system change, and if so, (2) did the manipulation cause the change? As discussed above, for unreplicated experiments, only the first question can be addressed statistically. To answer the second question, we must show that the manipulation is the most plausible explanation for the change (Carpenter 1993). To address these questions, we reanalyzed the published bacterial data reported for EisenEx (A2004) and SOFeX (O2004), and we explicitly incorporated the time component into the statistical model by using analyses of covariance (ANCOVA) to assess whether temporal changes in bacterial variables were significantly different inside and outside the iron-enriched patch. However, because the experiments were unreplicated and the observed values are not independent of each other,

we are still unable to unequivocally attribute significant differences between locations to the effects of iron. Instead, we must indirectly examine the question of causality by considering alternate explanations, as discussed below.

For EisenEx, we used the mean values from the upper 40 m, presented in figs. 2 and 3 of A2004. For SOFeX, we considered only data from the South Patch and used the depth-integrated mixed layer values, divided by mixed layer depth, as presented in figs. 1 and 2 of O2004. For our reanalyses, we used mean (A2004) and depth-integrated (O2004) values to avoid potential confounding influences from samples that were collected at different depths on the same day, and to allow direct comparison with the results presented in each of the studies. Although not stated explicitly in the manuscript, O2004 appear to have excluded data collected from transect stations from their analyses, despite presenting these data in their fig. 1; therefore, we did not include these data in our ANCOVA. Although A2004 appear to have excluded all bacterial ectoenzyme activity data before day 8 from their analyses, they provided no a priori justification for the selective use of data from inside the patch, and we therefore have included the entire data set in our reanalyses. To satisfy the requirements for ANCOVA, and to meet the assumptions concerning residuals, where necessary, the data for each parameter were transformed (log or square root) such that the data from inside and outside the iron-enriched patch were normally distributed and showed a linear relationship with time. For each variable, the same transformation was applied to data from inside and outside the iron-enriched patch (Quinn and Keough 2002). “Sampling day” was the covariate and “patch” (i.e., inside vs. outside the iron-enriched patch) was the fixed factor. A significant interaction between sampling day and patch shows that temporal changes in the variable were significantly different inside and outside the patch. The absence of a significant interaction means that the slopes of the regression of the parameter versus time were the same inside and outside the patch (Quinn and Keough 2002). All analyses were conducted by SPSS 12.0.

The results of the ANCOVA, along with the results of Mann–Whitney *U*- and two-sample *t*-tests reported in A2004 and O2004, are shown in Table 1. In addition, where appropriate, we performed Mann–Whitney *U*- or two-sample *t*-tests on the same data analyzed by ANCOVA from all the variables measured, including variables for which no statistical results were presented in A2004 and O2004. This latter comparison revealed that there were several incongruities between the *p* values presented in A2004 and O2004, and the *p* values that we obtained from our analyses using the same statistical tests as the authors (Mann–Whitney *U*- or two-sample *t*-test; Table 1). On the basis of the results of Mann–Whitney *U*-tests, A2004 reported that there were a number of statistically significant differences in bacterial properties (i.e., bulk and cell-specific α - and β -glucosidase hydrolysis rates) inside and outside the patch reported. However, our analyses of the data by Mann–Whitney *U*-tests found that these relationships were not statistically significant (Table 1). These incongruities may be due to the A2004’s unexplained and selective exclusion of all data from inside the patch before day 8.

Table 1. Results of statistical analyses testing for significant differences in the temporal changes in the abundance of cells with high DNA content (HDNA), total bacterial abundance (BA), production estimated from thymidine (BP_{TdR}) and leucine (BP_{Leu}) incorporation rates, specific growth rate estimated from the incorporation of thymidine (SGR_{TdR}) and leucine (SGR_{Leu}), bulk and cell specific activity of bacterial ectoenzymes of α - and β -glucosidase, and aminopeptidase, temperature, and salinity inside and outside the iron-enriched patches during EisenEx and SOFeX. For EisenEx, ANCOVA and Mann–Whitney U -tests were conducted using the mean values from the upper 40 m (A2004). For SOFeX, ANCOVA and two-sample t -tests were conducted using the depth integrated mixed layer values, divided by mixed layer depth (O2004) (*see* text for details). Where available, p values from Mann–Whitney U - and two-sample t -tests reported in EisenEx (A2004) and SOFeX (O2004) and from our reanalysis of the data (HR) are presented. An asterisk (*) indicates where significant differences ($\alpha = 0.05$) inside and outside the iron-enriched patch were observed. The sample sizes inside (n_{in}) and outside (n_{out}) the patch are shown. Dashes indicate where the authors did not report p values. Blank cells indicate that measurements were not taken.

Parameter	EisenEx						SOFeX					
	n_{in}	n_{out}	ANCOVA		Mann–Whitney U		n_{in}	n_{out}	ANCOVA		Two-sample t -test	
			F statistic	p	p	p			F statistic	p	p	p
HDNA												
BA	7	4	$F_{1,7} = 4.52$	0.071	—	0.257	7	4	$F_{1,7} = 1.42$	0.272	>0.05	0.336
BP _{TdR}	8	4	$F_{1,8} = 0.04$	0.845	>0.05	0.062	7	4	$F_{1,7} = 9.96$	0.016*	>0.05	0.115
BP _{Leu}	8	4	$F_{1,8} = 0.02$	0.893	>0.05	0.497	7	4	$F_{1,7} = 9.12$	0.019*	0.028*	0.015*
SGR _{TdR}	7	4	$F_{1,7} = 1.15$	0.320	—	0.089	7	4	$F_{1,7} = 1.55$	0.253	—	0.422
SGR _{Leu}	7	4	$F_{1,7} = 1.97$	0.204	—	1.000	7	4	$F_{1,7} = 1.30$	0.292	—	0.016
Bulk α -glucosidase†	8	5	$F_{1,9} < 0.01$	0.987	<0.05*	0.105	7	4				
Bulk β -glucosidase†	8	5	$F_{1,9} = 0.71$	0.421	<0.05*	0.067	7	4				
Bulk aminopeptidase†	8	5	$F_{1,9} = 1.99$	0.192	—	0.005*	7	4				
Cell-specific α -glucosidase‡	6	4	$F_{1,6} = 0.21$	0.666	<0.05*	0.394	7	4				
Cell-specific β -glucosidase‡	6	4	$F_{1,6} < 0.01$	0.948	<0.05*	0.394	7	4				
Cell-specific aminopeptidase‡	6	4	$F_{1,6} = 0.71$	0.430	>0.05	0.517	7	4				
Temperature	9	5	$F_{1,7} = 0.24$	0.637	—	0.162	7	4	$F_{1,5} = 0.17$	0.694	—	0.172
Salinity	9	5	$F_{1,10} = 12.76$	0.005*	—	0.096	7	4	$F_{1,5} = 1.32$	0.302	—	0.806

†A2004 excluded data before day 8 from Mann–Whitney U -tests ($n_{in} = 5$).

‡A2004 excluded data before day 8 from Mann–Whitney U -tests ($n_{in} = 3$).

On the basis of ANCOVA, there were no significant differences in the temporal changes inside compared with outside the patch for any bacterial properties during EisenEx (Table 1). This was in contrast to the results of Mann–Whitney U -tests (A2004), which showed significantly higher bulk and cell specific α - and β -glucosidase hydrolysis rates inside than outside the patch. The results of our analyses contrast markedly with those of A2004, and we suggest that the reported significance of the trends in bacterial responses reported in this paper should be viewed with caution. ANCOVA results for bacterial properties during SOFeX were generally similar to those reported by O2004 using two-sample t -tests. Time-dependent changes in bacterial production estimated from leucine (BP_{Leu}; ANCOVA; $F_{1,7} = 9.12$, $p = 0.019$) and thymidine (BP_{TdR}; ANCOVA; $F_{1,7} = 9.96$, $p = 0.016$) were significantly different inside and outside the patch, whereas only BP_{Leu} was significantly different inside and outside the patch according to the results of two sample t -tests reported by the authors (Table 1).

For the reasons explained above, we cannot unequivocally attribute the differences in BP_{Leu} and BP_{TdR} to the effects of iron. Instead, we must indirectly examine the question of casualty by considering the following in turn:

(1) whether the magnitude of changes in bacterial properties were ecologically meaningful; (2) whether the changes observed were in the direction predicted by the theory; and (3) whether it was likely that the addition of iron, as opposed to other factors, caused the response (Carpenter 1993). Although below we address these aspects separately, their effects are closely interrelated.

First, the approximately threefold increase in bacterial production observed inside the patch during SOFeX is similar to the seasonal range in Antarctic waters (Rivkin 1991; Ducklow 1999). This increase in bacterial production would lead to large changes in ecosystem processes as a result of increased bacterial carbon demand and remineralization of biogenic carbon produced in surface waters. Second, if rates of bacterial production were resource limited (i.e., organic or mineral substrates), then we would predict that the addition of these limiting substrates would increase bacterial productivity, as was observed during SOFeX.

Third, differences in the temporal patterns of bacterial production inside compared with outside the patch could be due to factors unrelated to the addition of iron, such as differential environmental forcing inside and outside the patch. To specifically address this, we used ANCOVA to

determine whether there were significant differences in the temporal changes in selected physical factors (i.e., temperature and salinity) inside and outside the iron-enriched patches. Mean mixed layer temperature and salinity values were compiled for days when bacterial variables were measured. For EisenEx, we obtained data from the Publishing Network for Geoscientific & Environmental Data (<http://www.pangea.de/>). ANCOVA identified a significant difference in the temporal change in salinity ($F_{1,10} = 12.76$, $p = 0.005$), but not temperature, suggesting that there were potentially different water column structures inside and outside the patch (Table 1). For SOFeX, data were extracted from Bishop et al. (2004). ANCOVA showed no significant differences in the temporal changes in temperature and salinity inside and outside the patch (Table 1). Therefore, the increase in bacterial production inside the iron-enriched patch during SOFeX was probably due to factors other differences in water column structure between the two water masses. Although it is attractive to conclude that differences in bacterial production inside and outside the patch were directly or indirectly due to the availability of iron, this is an inferred conclusion and not made based on statistical analyses of the treatment effects. Moreover, because insufficient data are available for dissolved organic carbon or nitrogen, we are unable to statistically distinguish between potential organic and mineral limitation of bacterial production in this system.

An appropriately replicated experiment allows the causal relationship between treatment and effect to be identified and tested by classical inferential statistics and should be used whenever possible. However, when designing ecosystem-scale experiments, researchers are often forced to compromise and sacrifice replication, spatial extent, and/or duration in order to conduct experiments at scales that are appropriate for the process being studied (Schindler 1998). Large-scale experiments are necessary to assess ecosystem responses to perturbations, such as iron (Watson et al. 1992) or phosphorous enrichment (Thingstad et al. 2005), and replication of experimental units with randomly assigned treatments is not always possible (Stewart-Oaten et al. 1992; Carpenter et al. 1998). When replication is not feasible, future mesoscale enrichment experiments should use sampling designs and statistical methods, many with their origins in limnology and coastal ecology, that allow treatment effects to be assessed. These include time-series analyses, such as transfer functions (Carpenter 1993), intervention analysis (Eberhardt and Thomas 1991), randomized intervention analysis (Carpenter 1993), and variants of the before-after-control-impact (BACI) design (Bence et al. 1996; Underwood 1996). For example, in a before-after-control-impact-pairs design (BACIP), samples are collected simultaneously on several dates before and after the perturbation (e.g., the addition of iron to a patch of water) at both the impact site and a nearby control site. The samples are paired in that they are sampled simultaneously (as nearly as possible), and replication comes from collecting paired samples at a number of times (dates) both before and after the perturbation. The difference between the impact and control sites before the perturbation is taken as an estimate

of the difference between the two sites in the period after the perturbation if it had not occurred (Stewart-Oaten et al. 1992). Furthermore, if the perturbation itself cannot be replicated, spatial replication can be achieved by including multiple control locations in the experimental design (Underwood 1996). Unfortunately, methods for assessing the effects of unreplicated perturbations are less well known to, or appreciated by, oceanographers, and as a result, many large-scale enrichment experiments, such as in situ mesoscale iron (summarized by Boyd 2002; de Baar et al. 2005) and phosphorus (Thingstad et al. 2005) additions, have been conducted with inadequate attention to appropriate experimental design or analysis.

Where inferential statistics are applied to data from nonreplicated experiments, such as in A2004 and O2004, it is crucial to recognize that significant differences between treatment groups cannot be unequivocally attributed to the effect of the treatment. The authors of such papers must do the following: (1) clearly describe the sampling regime and the statistical analyses applied to the data, including justifications of the statistical analyses used and details of how the assumptions of the tests were validated, and provide an a priori justification for the exclusion of data from the analyses; (2) explicitly discuss the limitations of the experimental design, clearly state the null hypotheses being tested (i.e., H_0 : there is no significant difference between locations, rather than H_0 : there is no significant treatment effect), and carefully consider alternative explanations; and (3) recognize that in the absence of proper randomization, replication, and interspersing of experimental units, extrapolating their results to a wider population of systems is tenuous. By ignoring the large body of literature on ecosystem-scale experiments conducted by our colleagues in limnology and coastal ecology, the oceanographic community has been handicapped in the design, execution, and interpretation of mesoscale nutrient enrichment experiments.

Michelle S. Hale¹
Richard B. Rivkin

Ocean Sciences Centre
Memorial University of Newfoundland
St. John's, Newfoundland and Labrador
Canada
A1C 5S7

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.
- BENCE, J. R., A. STEWART-OATEN, AND S. C. SCHROETER. 1996. Estimating the size of an effect from a before-after-control-impact paired series design, p. 133–149. *In* R. J. Schmitt and C. W. Osenberg [eds.], *Detecting ecological impacts: Concepts and applications in coastal habitats*. Academic.
- BISHOP, J. K. B., T. J. WOOD, R. E. DAVIS, AND J. T. SHERMAN. 2004. Robotic observations of enhanced carbon biomass and export at 55°S during SOFeX. *Science* **304**: 417–420.

¹ Present address: School of Earth and Environmental Sciences, University of Portsmouth, Portsmouth, Hants, United Kingdom.

- 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.
- . 1993. Statistical analysis of the ecosystem experiments, p. 26–42. *In* S. R. Carpenter and J. F. Kitchell [eds.], *The trophic cascade in lakes*. Cambridge Univ. Press.
- , J. J. COLE, T. E. ESSINGTON, J. R. HODGSON, J. N. HOUSER, J. F. KITCHELL, AND M. L. PACE. 1998. Evaluating alternative explanations in ecosystem experiments. *Ecosystems* **1**: 335–344.
- COTTENIE, K., AND L. DE MEESTER. 2003. Comment to Oksanen (2001): reconciling Oksanen (2001) and Hurlbert (1984). *Oikos* **100**: 394–396.
- DE BAAR, H. J. W., AND OTHERS. 2005. Synthesis of iron fertilization experiments: From the Iron Age to the Age of Enlightenment. *J. Geophys. Res.* **110**: C09S16.
- DUCKLOW, H. W. 1999. The bacterial component of the oceanic euphotic zone. *FEMS Microb. Ecol.* **30**: 1–10.
- EBERHARDT, L. L., AND J. M. THOMAS. 1991. Designing environmental field studies. *Ecol. Monogr.* **61**: 53–73.
- HURLBERT, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* **54**: 187–211.
- . 2004. On misrepresentation of pseudoreplication and related matters: A reply to Oksanen. *Oikos* **104**: 591–597.
- LEGENDRE, P., S. DALLOT, AND L. LEGENDRE. 1985. Succession of species within a community: Chronological clustering, with applications to marine and freshwater zooplankton. *Am. Nat.* **125**: 257–288.
- MARTIN, J. H., AND OTHERS. 1994. Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**: 123–129.
- OKSANEN, L. 2004. The devil lies in the details: Reply to Stuart Hurlbert. *Oikos* **104**: 598–605.
- OLIVER, J. L., R. T. BARBER, W. O. SMITH, AND H. W. DUCKLOW. 2004. The heterotrophic response during the Southern Ocean Iron Experiment (SOFEX). *Limnol. Oceanogr.* **49**: 2129–2140.
- QUINN, G. P., AND M. J. KEOUGH. 2002. *Experimental design and data analysis for biologists*. Cambridge Univ. Press.
- RIVKIN, R. B. 1991. Seasonal patterns of planktonic production in McMurdo Sound, Antarctica. *Am. Zool.* **31**: 5–16.
- SCHINDLER, D. W. 1998. Replication versus realism: The need for ecosystem-scale experiments. *Ecosystems* **1**: 323–334.
- STEWART-OATEN, A., J. R. BENICE, AND C. W. OSENBERG. 1992. Assessing effects of unreplicated perturbations: No simple solutions. *Ecology* **73**: 1396–1404.
- , W. W. MURDOCH, AND K. R. PARKER. 1986. Environmental impact assessment: “Pseudoreplication” in time? *Ecology* **67**: 929–940.
- THINGSTAD, T. F., AND OTHERS. 2005. Nature of phosphorous limitation in the ultraoligotrophic eastern Mediterranean. *Science* **309**: 1068–1071.
- UNDERWOOD, A. J. 1996. On beyond BACI: Sampling designs that might reliably detect environmental disturbances, p. 151–175. *In* R. J. Schmitt and C. W. Osenberg [eds.], *Detecting ecological impacts: Concepts and applications in coastal habitats*. Academic.
- WATSON, A. J., P. S. LISS, AND R. A. DUCE. 1992. Design of a small-scale iron fertilisation experiment. *Limnol. Oceanogr.* **36**: 1960–1965.

Acknowledgments

This research was supported in part by grants to R.B.R. from Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Surface Ocean Lower Atmosphere Study (C-SOLAS) Research Network. Funding was provided to C-SOLAS by NSERC and the Canadian Foundation for Climate and Atmospheric Sciences. We thank H. Evans for laboratory assistance, and Dr. D. Schneider and T. Avery for advice on statistics. We also thank Drs. M. R. Anderson and L. Legendre, and two anonymous reviewers for useful comments.

Received: 6 January 2005
Accepted: 19 December 2005
Amended: 6 February 2006