The following appendix accompanies the article

Seasonal phytoplankton nutrient limitation patterns as revealed by bioassays over Baltic Sea gradients of salinity and eutrophication

Timo Tamminen[1,*], Tom Andersen[2]

[1]**Finnish Environment Institute (SYKE), PO Box 140, 00251 Helsinki, Finland**
[2]**Department of Biology, University of Oslo, PO Box 1066, Blindern, 0316 Oslo, Norway**

*Email: timo.tamminen@environment.fi

Marine Ecology Progress Series 340:121–138 (2007)

Appendix 1. Statistical properties of the data and quality control procedures

When applying bioassays in routine monitoring programs, establishing adequate quality control procedures is essential for obtaining reliable and comparable results from different sources, with differences in instrumentation and laboratory practices, in the selection of response parameters, and in the environments studied. The statistical analysis coupled to the model selection procedure provides powerful tools for quality control.

Statistical properties of data. The lognormal distribution is considered a useful statistical model for stochastic variables that are positive and that display a scaling relationship between mean and standard deviation (i.e. which have a constant relative error or a constant coefficient of variation). Since the present set of experiments has 2 replicated measurements at lowest design level (i.e. from the same experimental unit), we can easily demonstrate such a scaling relationship by plotting the first against the second replicate on a linear scale. Fig. A1 shows that there is a distinct tendency for increasing deviation between replicate chlorophyll *a* (chl *a*) measurements with increasing mean value, and that this heteroscedasticity disappears on a logarithmic scale – just as one should expect if the measurement errors were lognormal. The same pattern appeared for replicated samples of ^{14}C-uptake (data not shown), suggesting that the existence of such scaling phenomena might be independent of the actual measurement method (which are as conceptually different as spectrofluorimetry and liquid scintillation counting in the present context).

If we denote the 2 replicates from observation i by $y_{ij}(i = 1\ldots m, j = 1, 2)$ and assume that they are lognormally distributed with median γ_i and a constant coefficient of variation ω_R, then the logarithmic deviations $z_i = (\log y_{ij} - \log y_{i2})/\sqrt{2}$ should be identically, normally distributed with variance $S_R^2 = \log(1+\omega_R^2)$ and zero mean. Fig. A2 shows that the cumulative distributions of log deviations conform closely to normal distributions, both for ^{14}C-uptake and chl *a* measurements,

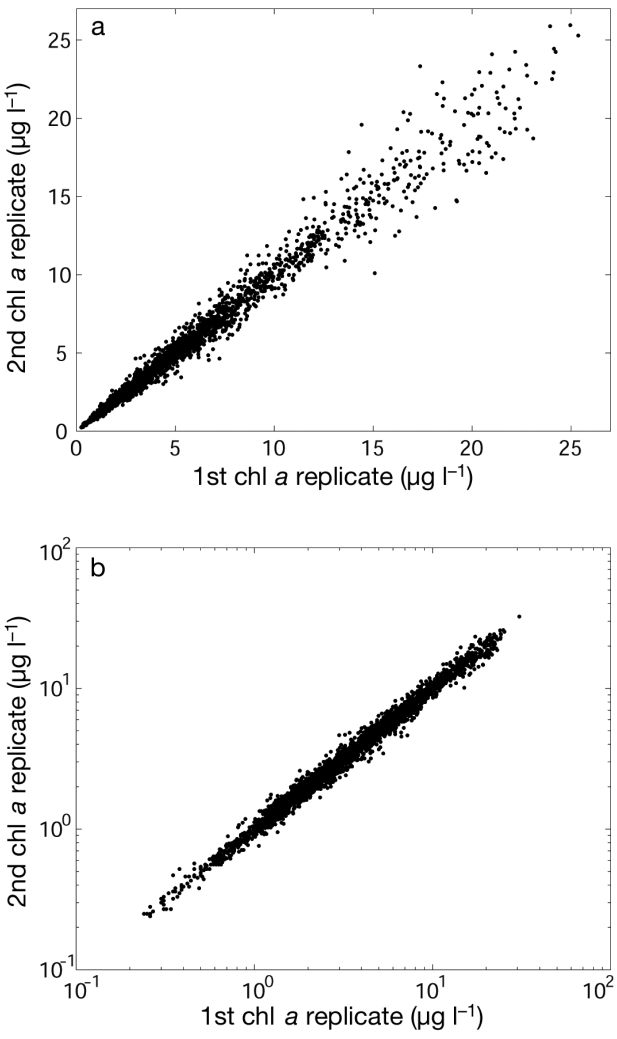

Fig. A1. Replicate samples of chl *a* (µg l^{-1}; n = 4025) on linear (panel a) and logarithmic (panel b) scales

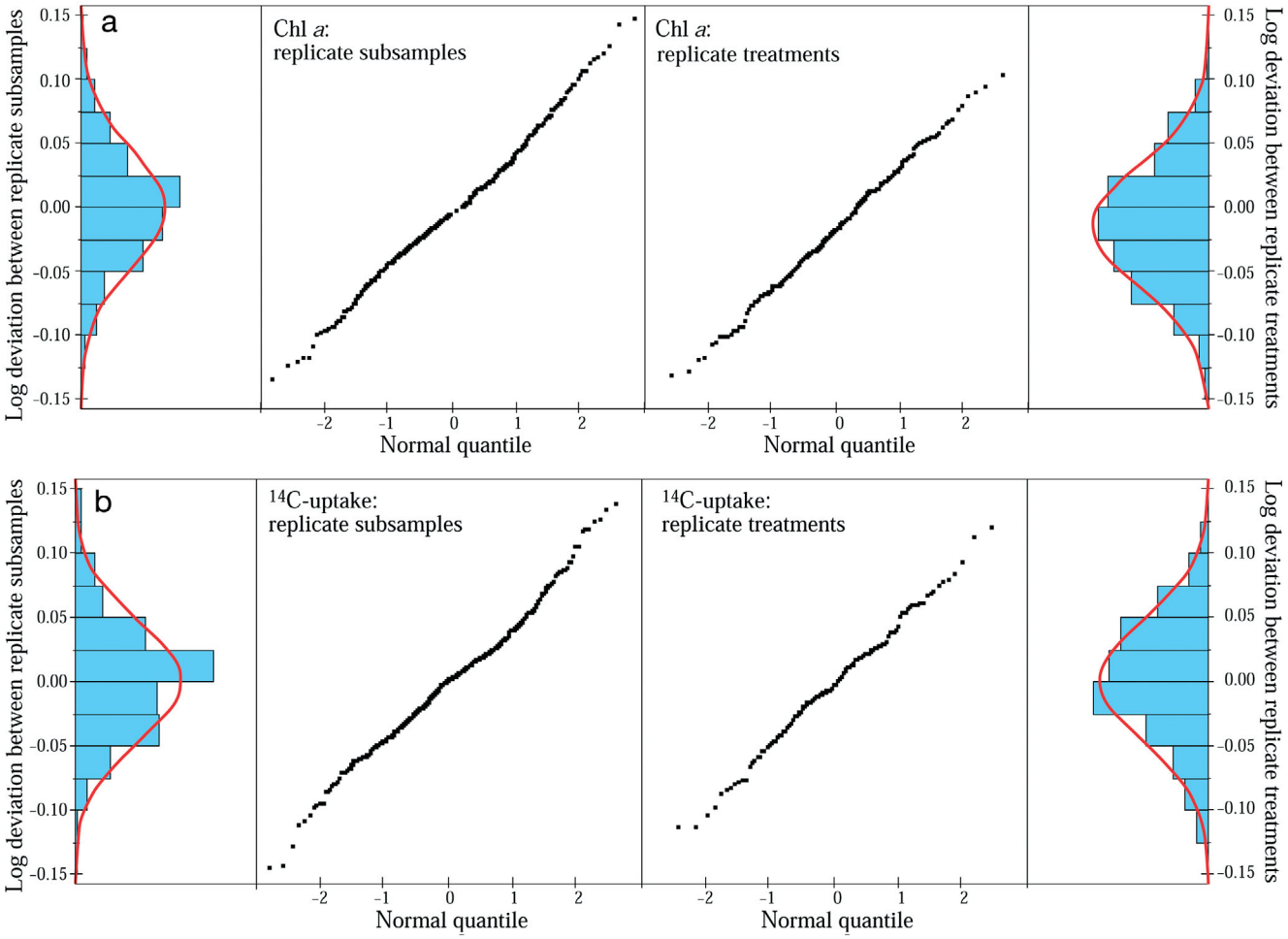

Fig. A2. Frequency distributions and quantile–quantile plots of deviations between log-transformed, measurement-level (left) and treatment-level (right) replicates of chl *a* (panel a) and ^{14}C-uptake (panel b). Frequency distribution histograms are compared to scaled standard normal distributions (red lines)

thus strongly supporting that errors are lognormally distributed with a constant coefficient of variation. In other words, there seems to be substantial evidence that statistical inferences based on log-transformed measurements from this material could be safely founded on normal theory.

When comparing coefficients of variation across the whole material (see Table 2 in main article), we find a striking similarity between chl *a* (5.7 % relative error, based on 4025 sample pairs) and ^{14}C-uptake (5.4 % relative error, based on 2771 sample pairs), perhaps suggesting some common factor having a dominant contribution to the relative measurement noise of both parameters. If we group the material by district (Table 2), we find laboratory-specific relative errors to vary from 4.4 to 6.9 %; a pattern that seems consistent with more informal impressions of the relative qualities of the different laboratories.

Since the present experimental design contains 2 levels of replication (both replicate samples from the same experimental unit and replicated units given the same treatment), we can repeat the same analysis to investigate the noise structure associated with treatment replicates. In order to do this, we first take the geometric mean of the replicate samples from each treatment unit (which will be an unbiased estimator of the median under the assumption of lognormal errors), and then proceed as above. The variance of the resulting log deviations (S_G^2) will have contributions both from within-treatment errors (S_T^2) and between-sample errors ($\frac{1}{2}S_R^2$; the factor of 2 stems from using geometric means). The within-treatment coefficient of variation can thus be estimated as:

$$\omega_T = \sqrt{\exp\left(S_G^2 - \tfrac{1}{2}S_R^2\right) - 1}$$

The within-treatment noise appears to be strikingly similar to the between-sample noise estimated above (see Table 2), with the relative error of ^{14}C-uptake measurements being slightly higher (5.3 %, based on 1304 treatment pairs) than chl *a* measurements (4.6 %, based on 1932 treatment pairs). If we interpret the within-treatment noise as deviations induced by biological factors and the between-sample noise as mostly due to subsampling and analytical errors, we might conclude that the present experimental design appears to be well balanced with respect to these major noise sources. If we

again split the material by district, we find laboratory-specific relative errors to vary from 3.7 to 7.0 % – with more or less the same ranking among laboratories as above.

Data quality management. The present amount of experimental data, involving 6 laboratories producing 5–8000 samples on several parameters over a 3-year period, requires a serious consideration of issues related to data quality management. The existence of scaling laws as implied by lognormally distributed errors provides us with powerful tools for general performance monitoring and quality assurance. Since we have indications that relative errors are constant and characteristic for each laboratory, we can actually then compute the probability of obtaining a given deviation between a pair of replicates (from the corresponding cumulative log-normal distribution function). Deviations that are unlikely to occur at least once within a given subset (for example, all experiments from one district in a certain year) can then be easily flagged for closer examination. Candidate outliers were checked against notes made during the experiment or during the analysis to see if there were reasons to discard the sample (for example, due to filtration error or instrument malfunction). With one major exception (see below), this procedure generally revealed 0 to 4 definitive outliers per laboratory and year. Since clear outliers were rare (less than one per 100 samples), and since all of them involved only one of the experimental units on the same day, they were simply replaced by their treatment replicate in further analysis.

Log deviations can also be plotted sequentially to produce quality control charts based on sample or treatment replicates. Such charts are useful for monitoring temporal changes in laboratory performance due to changes in staff or implementation. Fig. A3 shows an example of such control charts based on treatment replicates of ^{14}C-uptake. In the first year of experiments, this particular laboratory produced a distinctly nonrandom amount of positive deviations between treatment replicates compared to the other laboratories. Closer investigation revealed that uneven light distribution in the primary productivity incubator, which gave consistently increased irradiance on the same block of treatment replicates, was the source of this problem. After an intervention leading to appropriate readjustments of the incubation procedures, the problem disappeared completely for the remaining experimental years.

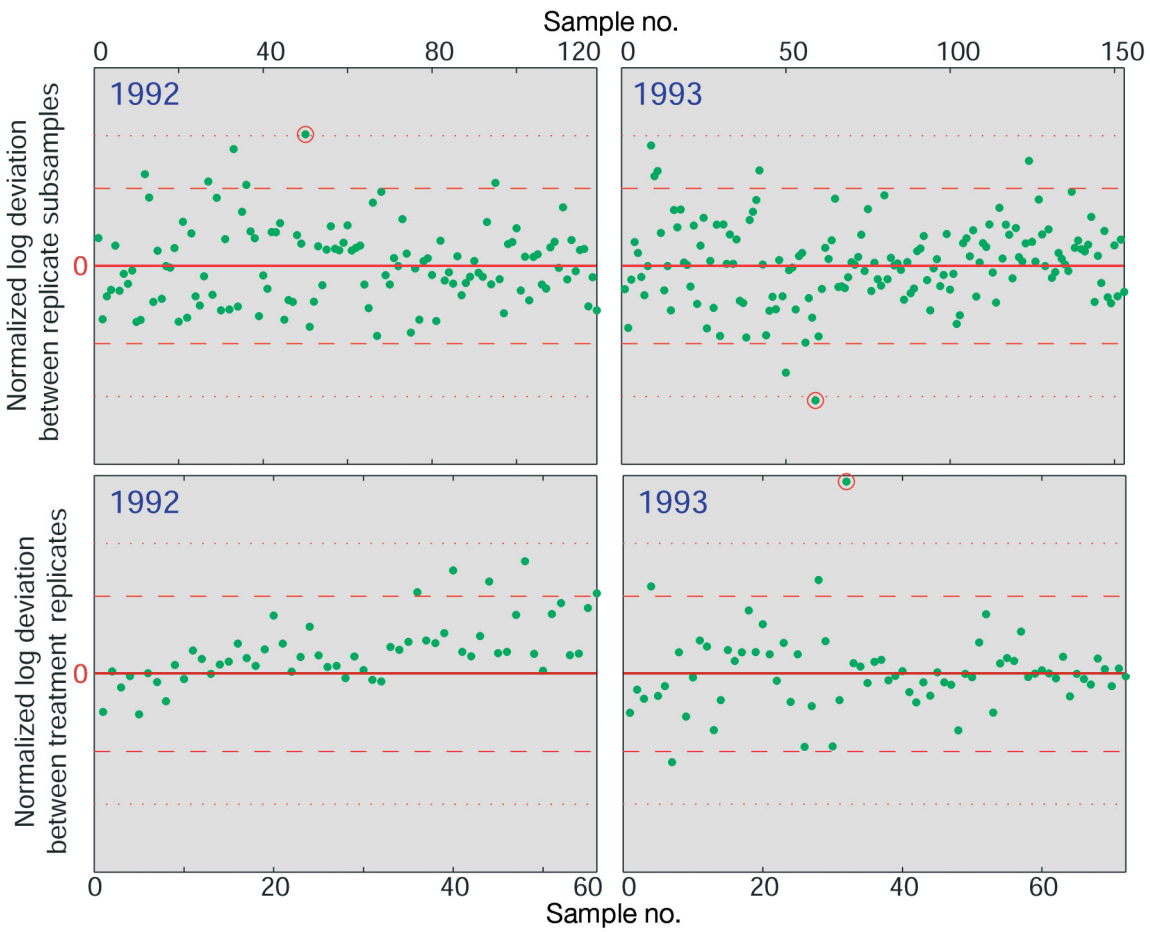

Fig. A3. Sequential quality control charts for 2 consecutive years, based on treatment replicates of ^{14}C-uptake, for one particular district laboratory (see text). Dashed and dotted lines are 95 and 99 % confidence limits, respectively