NOTES

Limnol. Oceanogr., 35(6), 1990, 1356–1368

Bivariate and trivariate analysis in flow cytometry: Phytoplankton size and fluorescence

Abstract—A curve-fitting method was used to characterize bivariate and trivariate flow cytometric data (cell volume, chlorophyll *a* fluorescence, phycoerythrin fluorescence) from a mixture of cultured algae and from natural populations of phytoplankton. The frequency distribution of cells in two or three dimensions was approximated by a sum of *n* bivariate or trivariate lognormal distributions, where *n* was the number of subpopulations in the assemblage. Each subpopulation was parameterized by the number of cells, the mean and standard deviation for each of the two or three variables, and a correlation coefficient for each pair of variables.

Although it is relatively simple to obtain simultaneous measurements of cell size and pigment autofluorescence for individual phytoplankters by flow cytometry (Yentsch and Horan 1989), the methods used to extract information from each multivariate data set representing tens of thousands of cells are usually rudimentary (Legendre and Yentsch 1989). The so-called allometric approach (joint distributions of size and fluorescence) and ataxonomic approach (joint distributions of fluorescence from two wavebands) described by Yentsch et al. (1986) and Phinney and Cucci (1989) are useful bases for classification of natural phytoplankton assemblages (Li 1989*a*). Multivariate data from these analyses can be displayed in various ways (Andrews 1972; Valet 1980; Murphy 1985; Ormerod and Payne 1987; Shapiro 1988), all resulting in a so-called signature for the sample. Usually, subpopulations are identified by visual inspection of these displays and subpopulation size (number of counts) obtained by an interactive, graphics-windowing procedure. Recently, the problem of pattern recognition in phytoplankton size-fluorescence signatures has been addressed by neural network computation (Frankel et al. 1989).

Alternatively, flow cytometric distributions can be approximated by mathematical functions. In these cases, the object of the analysis is to adjust the parameters of the function so that a good fit is obtained between the function and the measured data (Gray and Dean 1980; Dean 1985). If the function is chosen with regard to some known or assumed fundamental biological process (e.g. Campbell and Yentsch 1989*a*,*b*; Campbell et al. 1989), then the analysis is termed model fitting and the results can be used for prediction outside the domain of the measured variables. On the other hand, if the function is chosen simply because it gives the right shape to the data, then the analysis is termed curve fitting and the results are most appropriately used only for interpolation within (or not too far removed from) the domain of measurement (Smith 1979).

In recent analyses of caryotypes via flow cytometry, the frequency distribution of *n* human chromosomes has been fitted with a sum of *n* bivariate Gaussian curves (Dean et al. 1989; van den Engh et al. 1990). In this paper, I show that phytoplankton size-fluorescence cytograms for assemblages consisting of *n* subpopulations can likewise be described empirically by a sum of *n* bivariate Guassian distributions. Further, I extend the parametric analysis to phytoplankton data consisting of three variables in order to describe the frequency distribution of cells in three dimensions by a sum of *n* trivariate Gaussian distributions.

In the bivariate case, the function relating

Acknowledgments

I thank Edward Horne for data from the CTD and Trevor Platt and George White for comments on the manuscript.

measured counts, $C_b(x,y)$, to Coulter volume (x) and Chl a fluorescence (y) is the sum of component bivariate Gaussian distributions (Dean et al. 1989):

$$C_b(x, y) = \sum_{i=1}^{n} F_i(x, y; \mu_x, \mu_y; \sigma_x, \sigma_y; \rho_{xy}; N) \quad (1)$$

$$F_i = \frac{N_i}{2\pi\sigma_{xi}\sigma_{yi}(1 - \rho_{xyi}^2)^{1/2}} \times \exp\left[\frac{\left(\frac{x - \mu_{xi}}{\sigma_{xi}}\right)^2 - 2\rho_{xyi}\left(\frac{x - \mu_{xi}}{\sigma_{xi}}\right)\left(\frac{y - \mu_{yi}}{\sigma_{yi}}\right) + \left(\frac{y - \mu_{yi}}{\sigma_{yi}}\right)^2}{-2(1 - \rho_{xyi}^2)}\right]. \quad (2)$$

Each component (i.e., cluster or subpopulation in the signature) is referred by its index i. The complete signature consists of n such components. Individual components are parameterized by the mean value of x (μ_{xi}), the mean value of y (μ_{yi}), the standard deviation of x (σ_{xi}), the standard deviation of y (σ_{yi}), the correlation coefficient (ρ_{xyi}) between x and y, and a scaling factor that is the number of counts in the bivariate Gaussian (N_i). Note that the explicit functional form for F_i given by Dean et al. (1989) (their equation 10) is missing the minus sign in the denominator in the exponential term: the correct form is given as Eq. 2 above. For an assemblage of phytoplankton with n subpopulations, the bivariate function C_b contains $6n$ parameters.

In the trivariate case, the function relating measured counts, $C_t(x,y,z)$, to Coulter volume (x), Chl a fluorescence (y), and phycoerythrin fluorescence (z) is the sum of component trivariate Gaussian distributions:

$$C_t(x, y, z) = \sum_{i=1}^{n} G_i(x, y, z; \mu_x, \mu_y, \mu_z; \sigma_x, \sigma_y, \sigma_z; \rho_{xy}, \rho_{xz}, \rho_{yz}; N). \quad (3)$$

The explicit expression for the trivariate Gaussian (Johnson and Kotz 1972) is:

$$G_i = \left[\frac{N_i}{(2\pi)^{3/2}\sigma_{xi}\sigma_{yi}\sigma_{zi}\Delta_i^{1/2}}\right] \times \exp\left[-\frac{1}{2}\sum_{j=1}^{3}\sum_{k=1}^{3} A_{jki}u_{ji}u_{ki}\right] \quad (4)$$

and

$$u_{1i} = \frac{x - \mu_{xi}}{\sigma_{xi}}$$

$$u_{2i} = \frac{y - \mu_{yi}}{\sigma_{yi}}$$

$$u_{3i} = \frac{z - \mu_{zi}}{\sigma_{zi}}$$

$$\Delta_i = 1 - \rho_{yzi}^2 - \rho_{xzi}^2 - \rho_{xyi}^2 + 2\rho_{yzi}\rho_{xzi}\rho_{xyi}$$

$$A_{11i} = (1 - \rho_{yzi}^2)\Delta_i^{-1}$$

$$A_{22i} = (1 - \rho_{xzi}^2)\Delta_i^{-1}$$

$$A_{33i} = (1 - \rho_{xyi}^2)\Delta_i^{-1}$$

$$A_{12i} = A_{21i} = (\rho_{xzi}\rho_{yzi} - \rho_{xyi})\Delta_i^{-1}$$

$$A_{13i} = A_{31i} = (\rho_{xyi}\rho_{yzi} - \rho_{xzi})\Delta_i^{-1}$$

$$A_{23i} = A_{32i} = (\rho_{xyi}\rho_{xzi} - \rho_{yzi})\Delta_i^{-1}.$$

The trivariate case includes four parameters not found in the bivariate case: the mean value of z (μ_{zi}), the standard deviation of z (σ_{zi}), as well as the correlation coefficients between x and z (ρ_{xzi}) and between y and z ($_{yzi}$). Thus, for a phytoplankton assemblage with n subpopulations, the trivariate function C_t contains $10n$ parameters.

A FACS analyzer (Becton Dickinson) was used to measure Coulter volume and autofluorescence from Chl a and phycoerythrin. The protocols have been described in detail elsewhere (Li and Wood 1988; Li 1989*a*, 1990). Different flow orifices (50, 75, 100 μm) were used in each of the examples

presented, but Coulter volume was calibrated with uniform microspheres so that equivalent spherical diameter (ESD) could be represented in every case. Flow cytometric signals were acquired in list mode and recorded in relative units (256 channels) on three-decade logarithmic scales. The logarithmic behavior of the cytometer amplifiers was checked (Schmid et al. 1988; Olson 1989; Gandler and Shapiro 1990) and found acceptable (Li 1989*b*). The variables x, y, and z were represented by the recorded channel numbers. Thus, the curve-fitting exercise assumed that for each subpopulation, Coulter volume, Chl *a* fluorescence, and phycoerythrin fluorescence were distributed lognormally.

The bivariate curve fit was first tested on a mixture of three cultured algae: *Dunaliella tertiolecta, Phaeodactylum tricornutum,* and an unidentified chlorophytelike, nearshore isolate (hereafter referred as BBasin). The search for best-fitting parameters began by inspection of the size-fluorescence contour plot (Fig. 1A). Initial parameter estimates (excepting correlation coefficients) were provided by CONSORT 30 software (Becton Dickinson); final estimates were obtained by nonlinear regression of Eq. 1 with the algorithm of Marquardt (1963) as implemented by Raycheba (1985). In order to comply with computer memory requirements, I reduced the data matrix from the original size of 256^2 to 64^2 by summing the number of counts recorded in successive 64^2 submatrices. When the Poisson weighting factor was used in the regression (cf. Dean et al. 1989), the fitted curve significantly underestimated the population sizes; regressions were therefore carried out on unweighted data. The fit of the function to the data was good (Fig. 1B): the reduced χ^2 statistic (χ^2_r), as modified by van den Engh et al. (1990), was 1.09, indicating no significant difference between the measured and fitted distributions (Taylor 1982).

Following Dean et al. (1989), I judge the quality of fit by collapsing the two-dimensional data (Fig. 1A,B) onto projections along the volume and Chl axes (Fig. 1C,D). The volume projection (U_{bx}) and Chl projection (U_{by}) for the bivariate function are defined respectively as

$$U_{bx} = \sum_y C_b(x, y) \tag{5}$$

$$U_{by} = \sum_x C_b(x, y). \tag{6}$$

To evaluate the fitting procedure for subpopulations represented by only a small number of cells, I analyzed the same algal mixture data in successive trials: starting with 2^{13} list-mode events, each ensuing analysis was performed with half the number of events. Successful convergence of the iterative fitting procedure was achieved even when only 2^7 (=128) events were used (Fig. 2). Estimates of μ_x and μ_y remained invariant over the range $2^7 \rightarrow 2^{13}$ (Fig. 2C,D) but N^* (defined here as N per event analyzed), ρ_{xy}, σ_x, and σ_y appeared unreliable as the total number of events analyzed decreased below 2^9 (Fig. 2A,B,E,F).

Conventional analysis of subpopulation size is based on summing the number of events circumscribed within bitmaps (usually rectangular) drawn around clusters on the size-fluorescence cytogram. Three such windows are indicated on Fig. 1A: they represent a subjective decision concerning separation of the three species. The number of events enclosed within each rectangle (N'_i) was compared to the corresponding parameter estimate N_i (Fig. 2G). The ratio $N'_i : N_i$ was largely invariant in the range $2^{10} \rightarrow 2^{13}$ events analyzed and was close to a value of 1.0, as expected in cases such as the algal mixture (Fig. 1A) where there is only slight overlap of subpopulations on the cytogram.

The bivariate function was next applied to the size-fluorescence signature of phytoplankton from Bedford Basin, Nova Scotia (63°38′30″W, 44°41′30″N; 30 March 1987; 5 m), a small (17 km^2), deep (70 m) eutrophic marine basin. In this example, the data were gated to remove cells exhibiting phycoerythrin fluorescence. The cytogram indicated a diverse assemblage of red-fluorescing cells with considerable overlap in the footprints (sensu van den Engh 1990) of the subpopulations (Fig. 3A). The matter of deciding how many components to include in the analysis (i.e. n) is difficult to

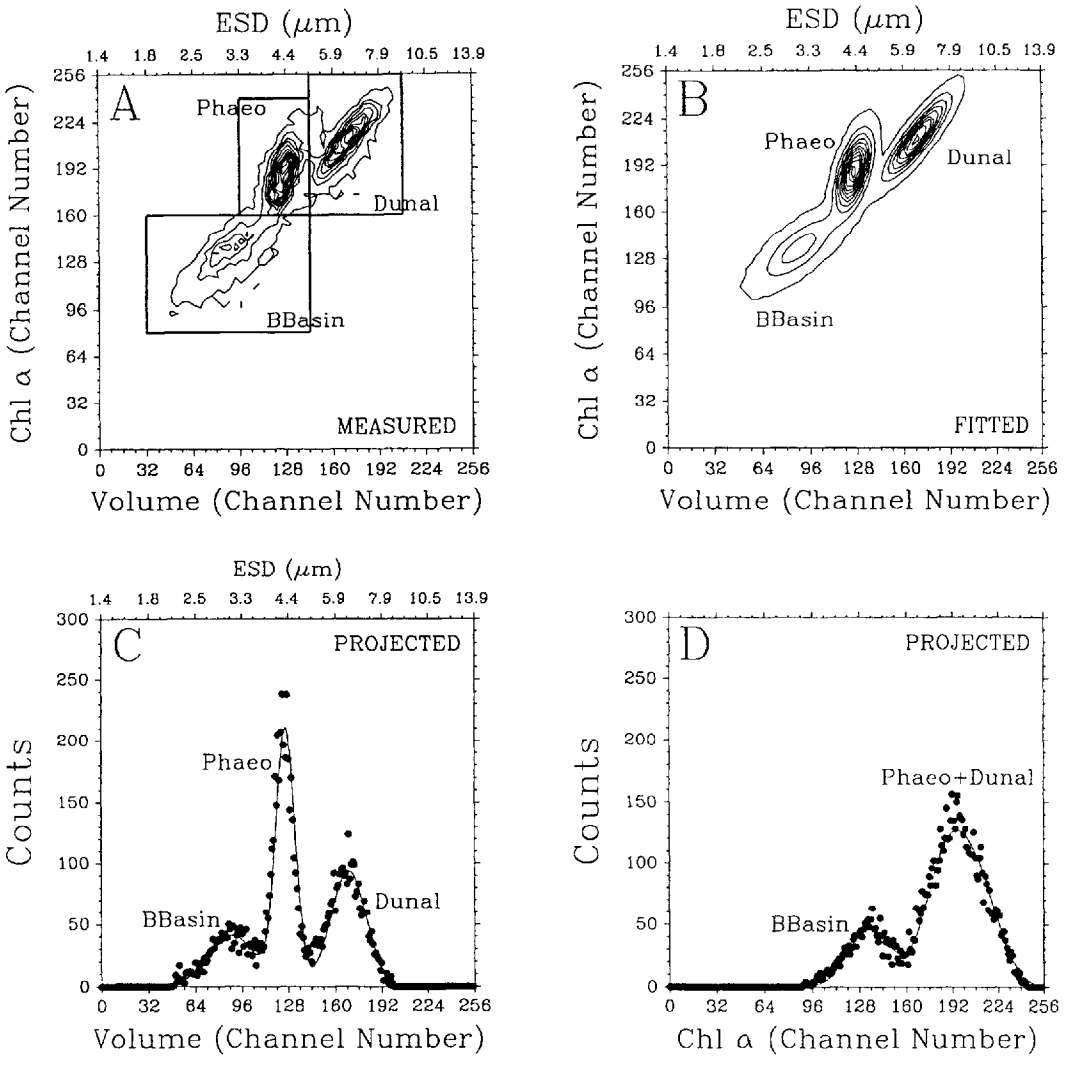
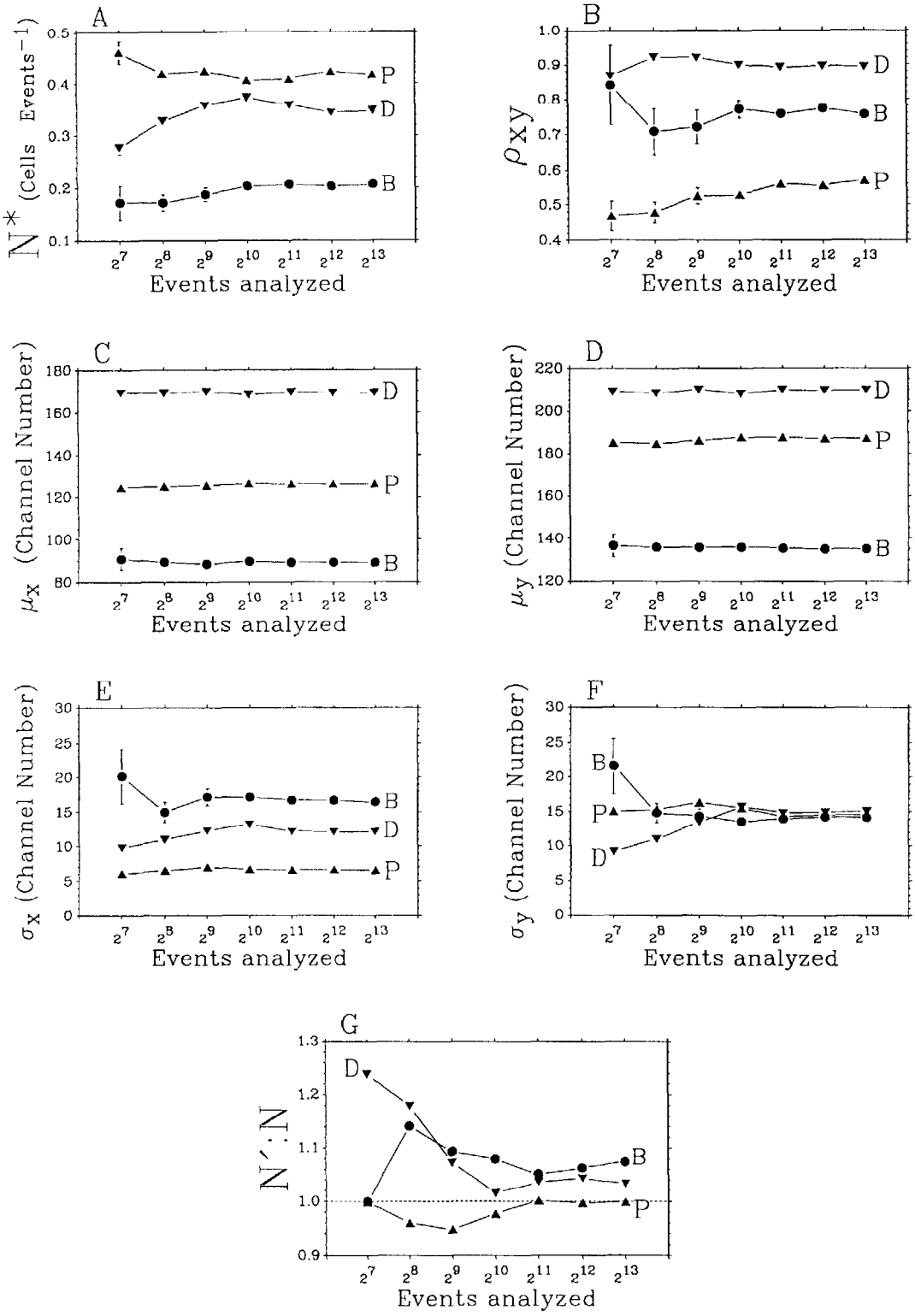


Fig. 1. Bivariate (Coulter volume and Chl *a*) analysis of an algal mixture of *Dunaliella tertiolecta, Phaeodactylum tricornutum,* and BBasin (an unidentified coastal isolate). A. Bivariate contour plot of measured data. Rectangles around each alga indicate domains for extracting N' (*see text and Fig. 2G*). B. Bivariate contour plot of data fitted according to Eq. 1. C. Univariate projection U_{bx} (Eq. 5) showing least-squares fit (curve) to measured counts (●). D. Univariate projection U_{by} (Eq. 6). Parameters of the fit are given in Fig. 2.

resolve. In principle, there should be as many components as there are distinct algal populations. Unfortunately, the clusters on many phytoplankton cytograms are not clearcut but merge with neighbors. In the extreme, it is possible that exactly the same position on the size-fluorescence domain is occupied by two or more taxonomically different populations. Furthermore, due to physiological plasticity, both size and fluorescence can change within populations. The recent application of machine algorithms for pattern recognition in flow cytometric signatures of phytoplankton (Frankel et al. 1989) may prove useful in designating components for bivariate and trivariate analyses.

For now, a pragmatic approach is taken: the model should have enough components so that the best fit is statistically acceptable but not so many as to incur overspecification redundancy (Gunst and Mason 1980)


A
N^* (Cells Events^{-1})
0.5
0.4
0.3
0.2
0.1
P
D
B
2^7 2^8 2^9 2^{10} 2^{11} 2^{12} 2^{13}
Events analyzed
B
ρ_{xy}
1.0
0.9
0.8
0.7
0.6
0.5
0.4
D
B
P
Events analyzed
C
μ_x (Channel Number)
180
160
140
120
100
80
D
P
B
Events analyzed
D
μ_y (Channel Number)
220
200
180
160
140
120
D
P
B
Events analyzed
E
σ_x (Channel Number)
30
25
20
15
10
5
0
B
D
P
Events analyzed
F
σ_y (Channel Number)
B
P
D
Events analyzed
G
N'/N
1.3
1.2
1.1
1.0
0.9
D
B
P
Events analyzed

or that the computation becomes unacceptably slow. It must be assumed that there are no abrupt changes in signature from one sampling time to the next (or from one depth to another in the case of vertical profiles). In other words, a component identified in a certain region of the size-chlorophyll domain at one time (or at one depth) is assumed to be the same component seen in the same or nearby domain region at a later time (or adjacent depth). It is important to point out that the subjective nature of choosing n is not a weakness inherent in the curve-fitting method itself. The same considerations apply when one draws bitmaps by inspection or when one establishes training sets for neural networks that perform machine pattern recognition (Frankel et al. 1989).

On the basis of visual inspection (Fig. 3A), a value of $n = 7$ was chosen for the Bedford Basin assemblage. The bivariate function provided a good fit to the data (Fig. 3B,C, D; $\chi^2_r = 1.02$). An examination of residual errors (=measured value − fitted value) showed a distribution that was positively biased about zero (not shown), indicating that the function, even with random errors included, did not provide an exact match to the data. This point is also made clear in a comparison between ΣN_i and the total number of events actually recorded: the ratio of ΣN_i to actual number of events was 0.94.

For most practical purposes, this slight underestimate would be of little significance. It is useful, however, to point out the reasons for the discrepancy. First, the function did not take into account the isolated counts in regions of the cytogram away from the main clusters: they represented rare cells in the sample. Second, the bivariate function (as used) may have been an imperfect description of the data: e.g. underspecified with respect to the number of subpopulations (n); invalidity of the assumption of lognormal distributions; existence of an underlying background function (Dean et al. 1989). Although the apparent random appearance of rare events on the cytogram would be difficult to model, it may be possible to adjust for the latter problem. With respect to a background function, Dean et al. (1989) used a bivariate second-degree polynomial that included a term to account for rotation with respect to the orthogonal (measurement) x- and y-axes. Even so, the extended model failed to give an unbiased fit to the data: the selection of a proper background function, the shape of which can change greatly between samples and over different regions of the cytogram remains problematic (Dean et al. 1989).

The trivariate curve-fitting scheme was tested on phytoplankton samples collected from various depths at a station on the continental slope of Nova Scotia (42°49.96′N, 62°31.08′W, 19 May 1988, 0800 hours) with a rosette of Niskin bottles. Water temperature, salinity, σ_t, and in situ Chl *a* fluorescence were profiled with a CTD (Guildline 8705) and a submersible fluorometer (Aquatracka). The density structure (σ_t) at this station was largely a reflection of the salinity structure (Fig. 4A): both indicated discrete layering of the water column. The vertical distribution of cells counted by flow cytometry closely matched that of bulk Chl measured by in situ fluorometry (Fig. 4B): both indicated a broad subsurface maximum centered between 20 and 30 m.

From the ataxonomic perspective of color groups (Yentsch et al. 1986; Phinney and Cucci 1989), a dichotomy exists between phytoplankters that contain phycoerythrin and those that do not. Most cells at this station did not contain phycoerythrin. Only 2–6% (depending on depth) of all cells in any sample had measurable phycoerythrin fluorescence. These cells were the ones chosen for the trivariate analysis described below. This small subset of cells was extracted

←

Fig. 2. Bivariate analysis of algal mixture: variation of the fitted parameters according to the total number of list-mode events used for analysis. A. The fitted parameter N is normalized to the number of events analyzed. B. ρ_{xy}. C. μ_x. D. μ_y. E. σ_x. F. σ_y. G. $N' : N$ where N' is the number of events enclosed within each rectangle shown in Fig. 1A and N is the corresponding fitted parameter. Species designations: D—*Dunaliella tertiolecta*; P—*Phaeodactylum tricornutum*; B—BBasin.

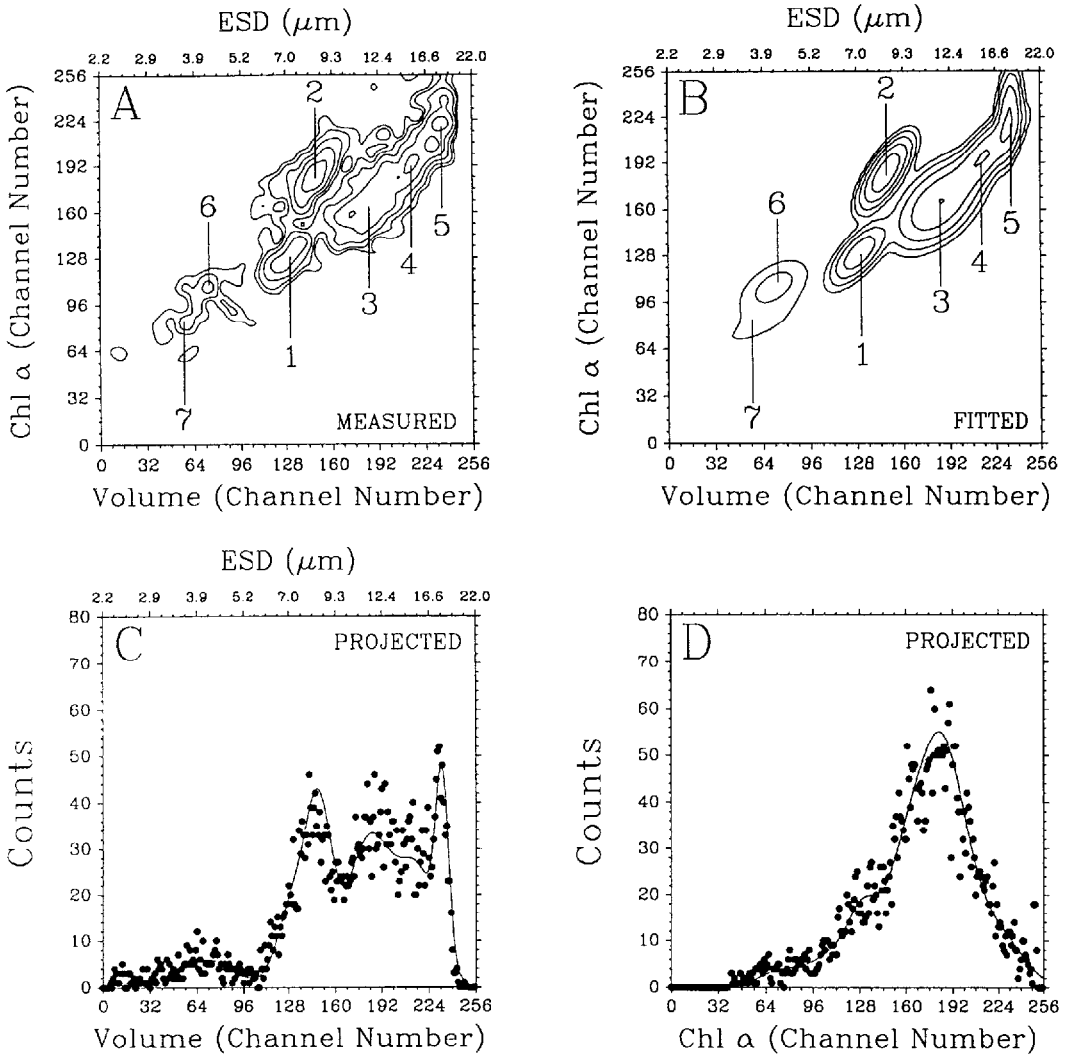


Fig. 3. As Fig. 1, but for phytoplankton assemblage in Bedford Basin. Parameter values as follows: N_i (counts), μ_{xi} (channel number), μ_{yi} (channel number), μ_{xi} (channel number), σ_{yi} (channel number), ρ_{xyi}. For i = 1: 406, 129.9, 129.5, 10.0, 10.8, 0.71; for i = 2: 851, 147.9, 184.3, 9.2, 12.4, 0.74,; for i = 3: 1,271, 183.9, 163.5, 15.6, 15.6, 0.53; for i = 4: 754, 216.7, 197.1, 12.8, 14.8, 0.79; for i = 5: 423, 233.1, 225.5, 4.4, 16.8, 0.28; for i = 6: 159, 72.3, 107.9, 15.2, 12.0, 0.53; for i = 7: 181, 58.7, 77.5, 38.4, 18.4, 0.78.

from list-mode data by rejecting all events not exceeding a phycoerythrin threshold of 20 channels. Based on measurements of Coulter volume, it is estimated that these phycoerythrin-containing cells were 4–10 μm in equivalent spherical diameter, suggesting that they may have been cryptomonads (Thomsen 1986). Phytoplankters at and below 40 m were present at 10^4 ml^{-1} (Fig. 4B): the number of them that contained phycoerythrin was so small that trivariate analysis was restricted to samples above these depths.

As with bivariate data, the first step in trivariate analysis was to select a value for n, the number of subpopulations. A visual inspection of the data indicated clearly that $n > 1$ (Fig. 5): subjectively, a value of n = 2 was chosen. To facilitate nonlinear regression, I reduced the data matrix from a size of 256^3 to 16^3 by summing the number of counts recorded in successive 16^3 sub-

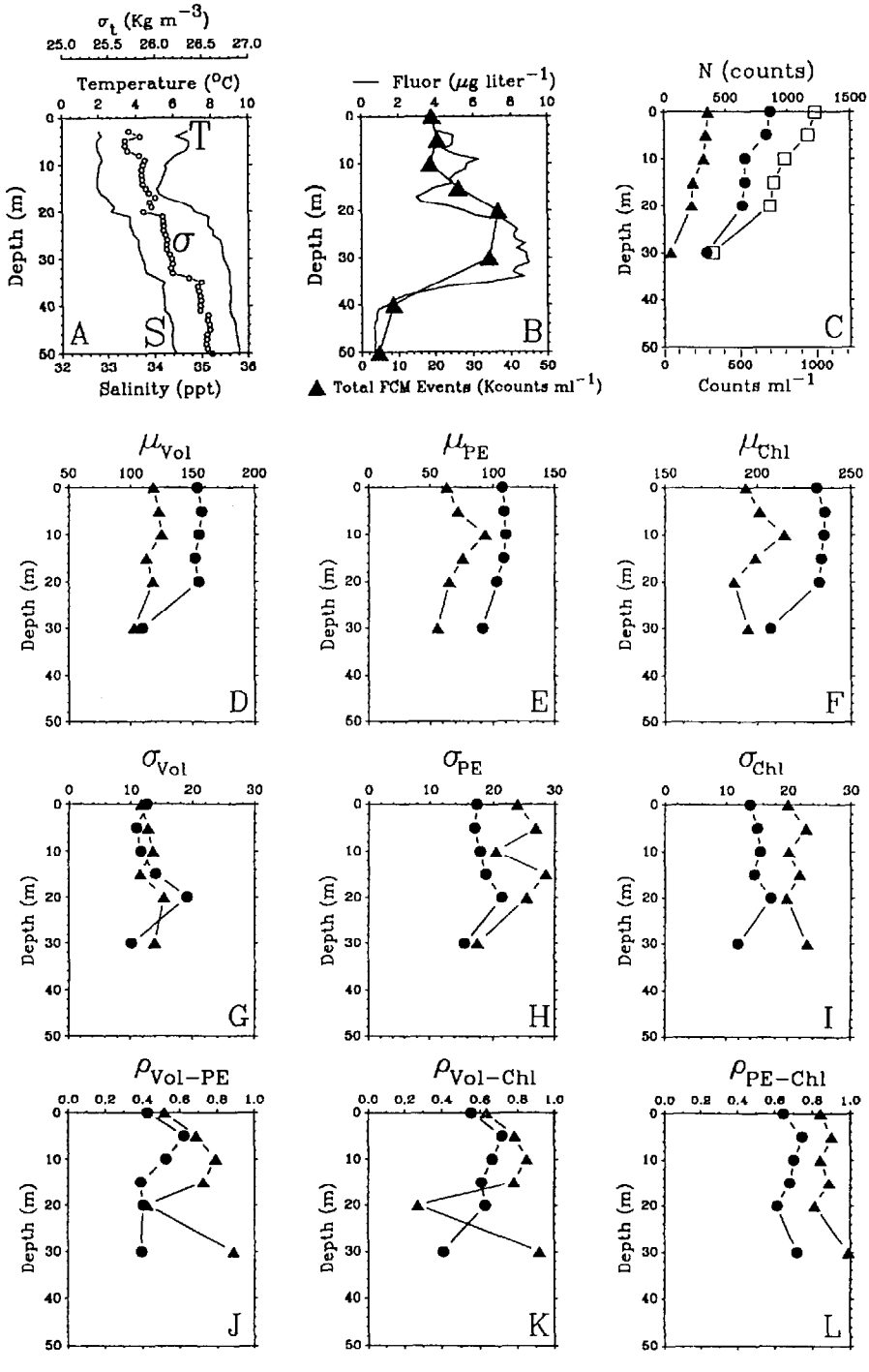


Fig. 4. Nova Scotia slope station. A. Temperature, salinity, and σ_t. B. In situ Chl *a* fluorescence measured by submersible fluorometer (——) and total autofluorescent particles counted by flow cytometry (▲). C–L. Parameters of trivariate fit. The two different subpopulations presumed to constitute the phycoerythrin-containing cells—●, ▲; □ (panel C)—sum of the two presumed subpopulations.

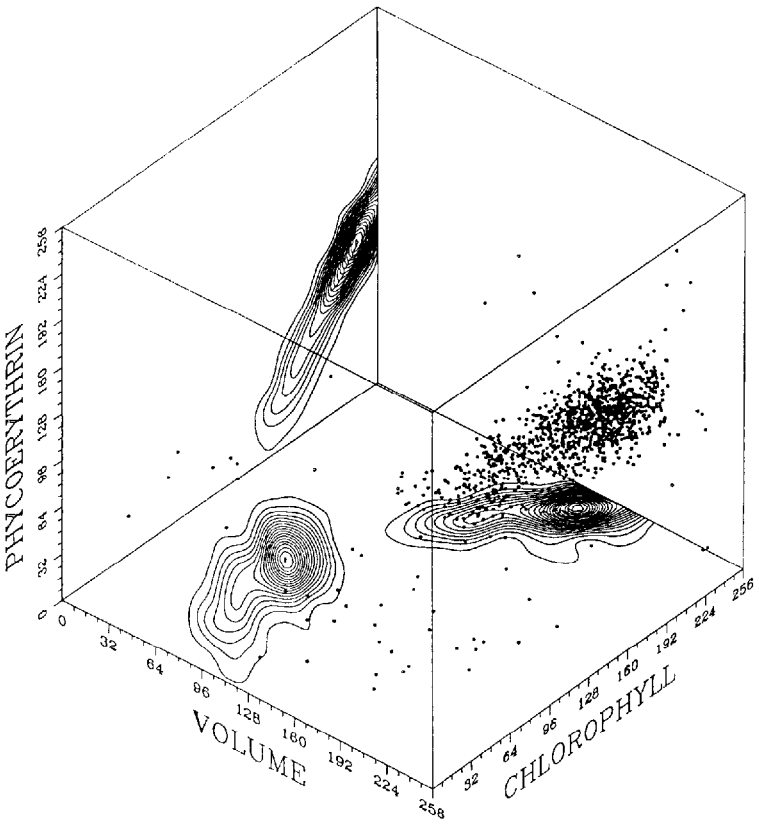


Fig. 5 Phycoerythrin-containing cells of the Nova Scotia slope. The dots represent a scatterplot of 1,232 cells in three dimensional space (x—Coulter volume; y—Chl *a* fluorescence; z—phycoerythrin fluorescence). The three contour plots represent the bivariate frequency distributions of cells projected onto the planes where x = 0 or y = 0 or z = 0. Each axis is given in relative units (256 channels) and spans three decades of a logarithmic scale.

matrices. Initial estimates of the parameters were obtained from bivariate analyses (Eq. 1) of the data projected onto each of the three bivariate coordinate planes (cf. faces of the trivariate cube in Fig. 5). Valet (1980) indicated that this would be a good method of obtaining initial parameter estimates for fits of trivariate Gaussian distributions.

Results of fitting unweighted data are shown for the 0-m sample in Fig. 6A,B. The major characteristics of the two dominant subpopulations were faithfully represented ($\chi^2_r = 1.03$) but, as expected with $n = 2$, not all of the fine details of the measured data could be reproduced by the mathematical function. The quality of fit was judged by collapsing both the measured and fitted values onto projections along the three separate axes. The volume projection (U_{tx}), chlorophyll projection (U_{ty}), and phycoerythrin projection (U_{tz}) for trivariate data are defined respectively as

$$U_{tx} = \sum_y \sum_z C_t(x, y, z), \tag{7}$$

$$U_{ty} = \sum_x \sum_z C_t(x, y, z), \tag{8}$$

$$U_{tz} = \sum_x \sum_y C_t(x, y, z). \tag{9}$$

In each case (Fig. 7A,B,C), the overall fit of the curves to the data appeared to be good. An examination of residuals according to Draper and Smith (1966) indicated no apparent gross abnormality in the least-squares analyses (not shown). The runs test was used to examine randomness of grouping of results arranged in order of ascending

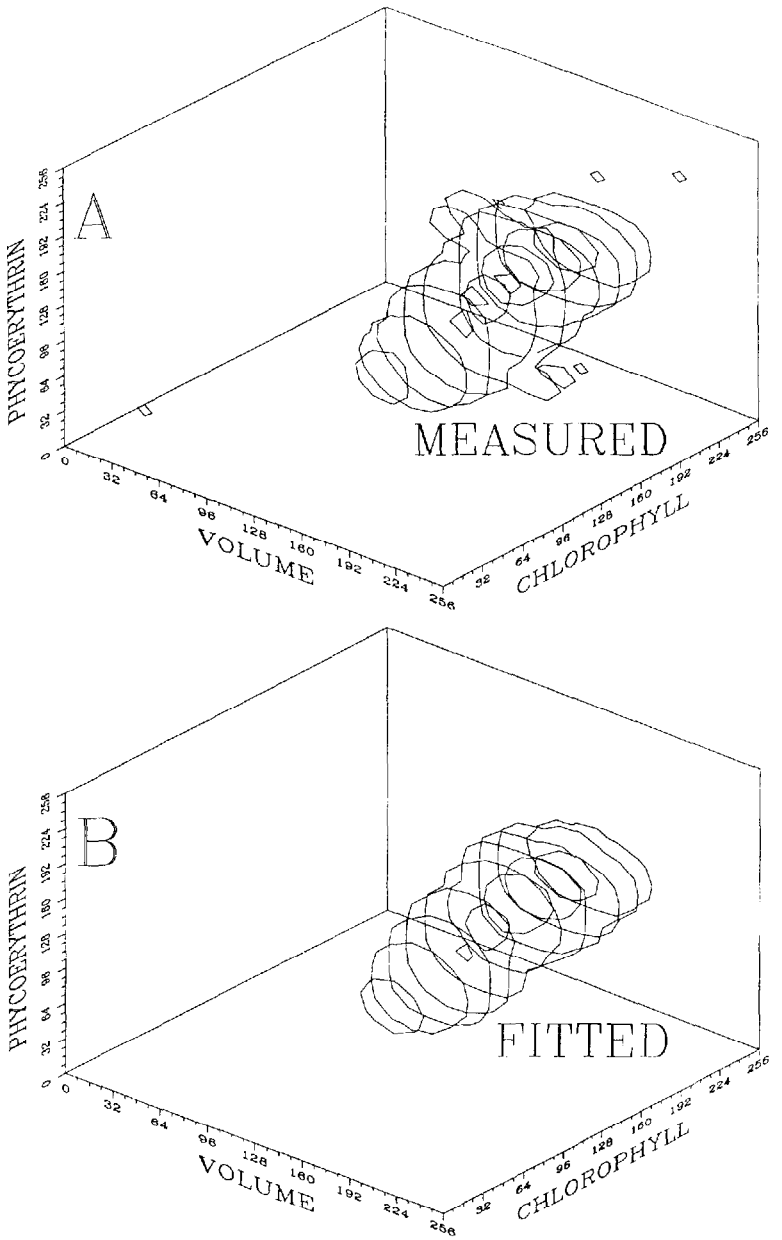


Fig. 6. Trivariate (Coulter volume, Chl *a*, and phycoerythrin fluorescence) analysis of Nova Scotia slope phytoplankton collected at the sea surface. A. Trivariate contour plot of measured data. Contour slices are taken along the Chl axis. B. Trivariate contour plot of data fitted according to Eq. 3. Parameters of the fit are given in Fig. 4 (0-m sample).

fitted value (Draper and Smith 1966). The z statistic (unit normal deviate) calculated for the runs in U_{tx}, U_{ty}, and U_{tz} was -0.42, 0.08, and -1.60, respectively—indicating that in no case should one reject the hypothesis of random arrangement in the algebraic sign of the residuals.

The trivariate parameters are concise statistical descriptors that clearly show differences between subpopulations at any depth

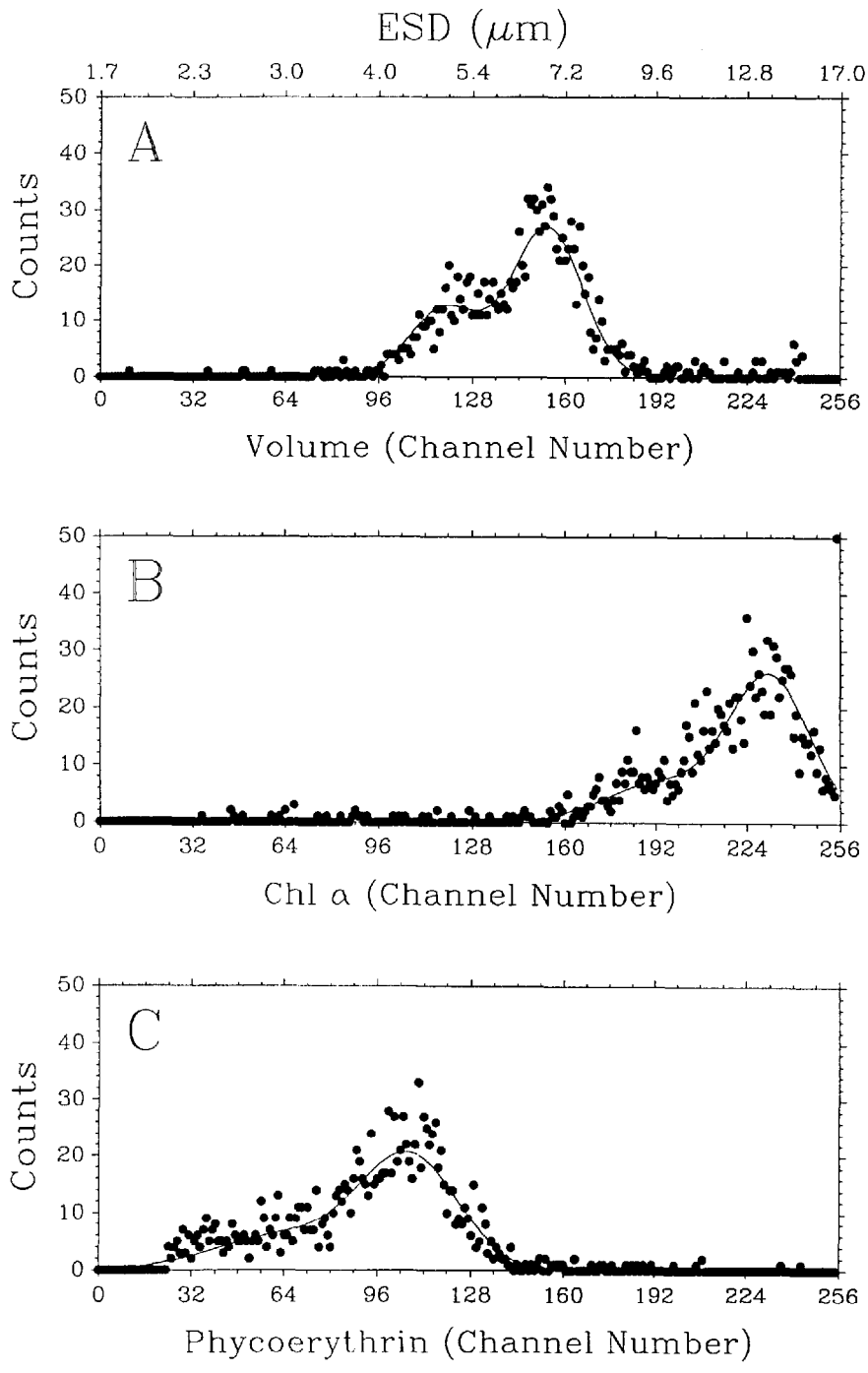


Fig. 7. Trivariate analysis of Nova Scotia slope phytoplankton collected at the sea surface. A. Univariate projection U_{tx} (Eq. 7) showing least-squares fit (curve) to measured counts (●). B. Univariate projection U_{ty} (Eq. 8). C. Univariate projection U_{tz} (Eq. 9). Parameters of the fit are given in Fig. 4 (0-m sample).

and also changes in these subpopulations over depth (Fig. 4C–L). They demonstrate, for example, that the phycoerythrin-containing cells did not peak in abundance at a subsurface depth as did most other phytoplankton (compare Fig. 4B and C); that the two presumptive subpopulations occurred in numerical ratios of about 2 : 1 to 3 : 1, depending on depth (Fig. 4C); that a subsurface peak in fluorescence intensity (μ_{PE}, μ_{Chl}) occurred in one but not both subpopulations (Fig. 4E,F); and that the variance in fluorescence (σ_{PE}, σ_{Chl}) differed significantly between the subpopulations (Fig. 4H,I), although the variance in volume (σ_{vol}) did not (Fig. 4G). The last point is especially interesting because it suggests variability in fluorescence beyond that accounted for by variability in cell size.

In conclusion, the examples in this paper indicate that a sum of bivariate or trivariate lognormal distributions can provide good fits to cytograms of phytoplankton size-fluorescence. This approach achieves a formal mathematical description of the so-called allometric approach (sensu Yentsch et al. 1986; Phinney and Cucci 1989): as such, it does not, by itself, indicate anything about the taxonomic identity of the component subgroups. In fact, the analysis (as presently applied) depends on a subjective determination of the number of components. This weakness of the analysis may have to be addressed by other methods (e.g. verification of subpopulations by sorting).

Proceeding on the basis that the allometric approach is desirable even without knowing the identity of components, the analyses described here allow extraction of ecologically useful information. Bivariate and trivariate analyses are potentially useful tools for the quantitative exploration of data derived from depth profiles, horizontal transects, time series, or any other observational or experimental progressions. The analyses are especially important in cases where the components are not clearly separated from each other. The models assume that each component population is distributed in a lognormal fashion in all variables; although the models have no clear theoretical bases, their parameters are easily interpretable (descriptively) and useful for the testing of hypotheses. The empirical nature of the formulation does not allow a physiological interpretation of the data, but the formulation is able to accommodate any number of subpopulations (limited by computational considerations) and therefore is valuable in applications involving natural phytoplankton assemblages. In the future, an alternative approach might extend theoretically based univariate analyses into two and three dimensions.

W. K. W. Li

Biological Oceanography Division
Bedford Institute of Oceanography
Dartmouth, Nova Scotia B2Y 4A2

References

ANDREWS, D. F. 1972. Plots of high-dimensional data. Biometrics **28:** 125–136.

CAMPBELL, J. W., AND C. M. YENTSCH. 1989*a, b.* Variance within homogeneous phytoplankton populations. 1. Theoretical framework for interpreting histograms. 2. Analysis of clonal cultures. Cytometry **10:** 587–595, 596–604.

———, ———, AND T. L. CUCCI. 1989. Variance within homogeneous phytoplankton populations. 3. Analysis of natural populations. Cytometry **10:** 605–611.

DEAN, P. N. 1985. Methods of data analysis in flow cytometry, p. 195–221. *In* M. A. Van Dilla et al. [eds.], Flow cytometry: Instrumentation and data analysis. Academic.

———, S. KOLLA, AND M. A. VAN DILLA. 1989. Analysis of bivariate flow karyotypes. Cytometry **10:** 109–123.

DRAPER, N. R., AND H. SMITH. 1966. Applied regression analysis. Wiley.

FRANKEL, D. S., R. J. OLSON, S. L. FRANKEL, AND S. W. CHISHOLM. 1989. Use of a neural net computer system for analysis of flow cytometric data of phytoplankton populations. Cytometry **10:** 540–550.

GANDLER, W., AND H. SHAPIRO. 1990. Logarithmic amplifiers. Cytometry **11:** 447–450

GRAY, J. W., AND P. N. DEAN. 1980. Display and analysis of flow cytometric data. Annu. Rev. Biophys. Bioeng. **9:** 509–539.

GUNST, R. F., AND R. L. MASON. 1980. Regression analysis and its application. Dekker.

JOHNSON, N. L., AND S. KOTZ. 1972. Distributions in statistics: Continuous multivariate distributions. V. 4. Wiley.

LEGENDRE, L., AND C. M. YENTSCH. 1989. Overview of flow cytometry and image analysis in biological oceanography and limnology. Cytometry **10:** 501–510.

LI, W. K. W. 1989*a.* Shipboard and analytical flow cytometry of oceanic ultraphytoplankton. Cytometry **10:** 564–579.

———. 1989*b*. Logarithmic amplification of optical signals of BIO BD-FACS®. Signal Noise **2:** 5

———. 1990. Particles in "particle-free" seawater: Growth of ultraphytoplankton and implications for dilution experiments. Can. J. Fish. Aquat. Sci. **47:** 1258–1268.

———, AND A. M. WOOD. 1988. Vertical distribution of North Atlantic ultraphytoplankton: Analysis by flow cytometry and epifluorescence microscopy. Deep-Sea Res. **35:** 1615–1638.

MARQUARDT, D. W. 1963. An algorithm for least-squares estimation of non-linear parameters. J. Soc. Ind. Appl. Math. **11:** 434–441.

MURPHY, R. F. 1985. Automated identification of subpopulations in flow cytometric list mode data using cluster analysis. Cytometry **6:** 302–309.

OLSON, R. J. 1989. A note of caution regarding logarithmic amplifiers. Signal Noise **2:** 1.

ORMEROD, M. G., AND A. W. R. PAYNE. 1987. Display of three-parametric data acquired by a flow cytometer. Cytometry **8:** 240–243.

PHINNEY, D. A., AND T. L. CUCCI. 1989. Flow cytometry and phytoplankton. Cytometry **10:** 511–521.

RAYCHEBA, J. M. T. 1985. Fitall. MTR Software, Toronto.

SCHMID, I. P., P. SCHMID, AND J. V. GIORGI. 1988. Conversion of logarithmic channel numbers into relative linear fluorescence intensity. Cytometry **9:** 533–538.

SHAPIRO, H. M. 1988. Practical flow cytometry, 2nd ed. Liss.

SMITH, W. R. 1979. Parameter estimation in nonlinear models of biological systems. Can. Fish. Mar. Serv. Tech. Rep. 889. 90 p.

TAYLOR, J. R. 1982. An introduction to error analysis. Univ. Sci. Books.

THOMSEN H. A. 1986. A survey of the smallest eucaryotic organisms of the marine phytoplankton, p. 121–158. *In* Photosynthetic picoplankton. Can. Bull. Fish. Aquat. Sci. 214.

VALET, G. 1980. Graphical representation of three-parameter flow cytometer histograms by a newly developed Fortran IV computer program, p. 125–129. *In* Proc. 4th Int. Symp. Flow Cytometry. Acta Pathol. Micro. Scand. Sec. A. Suppl. 274.

VAN DEN ENGH, G., D. HANSON, AND B. TRASK. 1990. A computer program for analyzing bivariate flow karyotypes. Cytometry **11:** 173–183.

YENTSCH, C. M., T. L. CUCCI, D. A. PHINNEY, AND J. A. TOPINKA. 1986. Real-time characterization of individual marine particles at sea, p. 414–448. *In* M. Bowman et al. [eds.], Tidal mixing and plankton dynamics. Springer.

———, AND P. K. HORAN [EDS.]. 1989. Cytometry in aquatic sciences. Cytometry **10:** 672 p.

Submitted: 27 October 1989
Accepted: 24 April 1990
Revised: 7 June 1990

Limnol. Oceanogr., 35(6), 1990, 1368–1375

Cross-flow filtration of colloids from aquatic environments

Abstract—In recognition of problems associated with traditional ultrafiltration techniques, cross-flow ultrafiltration has been adapted to isolate colloidal material from natural aquatic environments, both freshwater and marine. The procedure is capable of delivering filtration rates on the order of tens of liters per hour without introducing significant quantities of organic or trace-metal contaminants, and thus is suitable for large-volume and trace-component applications. Mass balance experiments in which concentrations were measured in ultrafiltrate and retained fractions show that both organic material and several trace metals are recovered without significant losses on the filters.

Biogeochemical investigations of inorganic and organic material in aquatic environments generally require knowledge of the physicochemical distribution of the material. In this regard the colloidal fraction is potentially significant. Colloids reside in the 1-μm–1-nm size range and may be dispersed in water as liquids (emulsions) or solids (suspensions) (Everett 1988). Thus the common environmental science procedure of using a filter to separate particulate and dissolved fractions does not distinguish between dissolved and colloidal material. Common flow-through ultrafiltration procedures can isolate colloids, but they have limited application to investigations of trace components in the aquatic colloidal fraction. In part this situation reflects the fact that sample volumes in the range of tens to hundreds of liters may be required to obtain sufficient colloidal material for chemical analyses. Sample processing times with

Acknowledgment

This work was partially funded by the Department of Supply and Services (Canada) unsolicited proposal program, contract 09SC.FP901-5-A514.