Vol. 144: 265–283, 1996 | MARINE ECOLOGY PROGRESS SERIES Mar Ecol Prog Ser | Published December 5

CHEMTAX — a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton

M. D. Mackey[1,2], D. J. Mackey[2,*], H. W. Higgins[2], S. W. Wright[3]

[1]University Chemical Laboratory, Lensfield Rd, Cambridge CB2 1EW, United Kingdom
[2]CSIRO Division of Oceanography, PO Box 1538, Hobart, Tasmania 7001, Australia
[3]Australian Antarctic Division, Channel Highway, Kingston, Tasmania 7050, Australia

ABSTRACT: We describe a new program for calculating algal class abundances from measurements of chlorophyll and carotenoid pigments determined by high-performance liquid chromatography (HPLC). The program uses factor analysis and a steepest descent algorithm to find the best fit to the data based on an initial guess of the pigment ratios for the classes to be determined. The program was tested with a range of synthetic data-sets that were constructed from known pigment ratios selected to be representative of samples of phytoplankton collected from the Southern Ocean and the Equatorial Pacific. Random errors were added both to the pigment ratios and to the calculated data-sets to simulate both uncertainties in the initial guess as to the pigment concentrations of each class and respectively experimental errors in the analysis of the pigments by HPLC. Provided that the analytical data is of good quality, the program can successfully determine the class abundances, even when the initial estimates of the pigment ratios contain large errors. Of particular interest is the observation that the program can provide good estimates of prochlorophytes, even in the absence of experimental data on the concentrations of divinyl-chlorophylls *a* and *b*. The program is not restricted to the estimation of phytoplankton and can be used whenever specific biomarkers exist that can be used as indicators of biological or chemical processes.

KEY WORDS: Biomarkers · Taxonomy · HPLC · Pigments · Phytoplankton

INTRODUCTION

The abundance and species composition of autotrophic marine microorganisms are important parameters in marine ecology, but this information can be difficult to obtain. Phytoplankton can be enumerated by light microscopy, but this requires extensive time for sample preparation and counting, especially if statistically valid counts of the less abundant plankton classes are required. Smaller phytoplankton, especially the picoplankton, can be difficult to identify since they lack taxonomically useful external morphological features; yet, they are now recognized as being significant contributors to the productivity of oceanic waters (Li et al. 1983, Platt et al. 1983, Iturriaga & Mitchell 1986, Chavez et al. 1990). In addition, many species are very fragile and do not survive sample fixation (Gieskes & Kraay 1983). The increased resolution of scanning or transmission electron microscopy allows identification of the picoplankton, but the sample preparation required renders electron microscopy extremely time-consuming for phytoplankton identification in large-scale surveys.

Identification and quantification of phytoplankton is often assisted by analysis of photosynthetic and photoprotective pigments: several pigments (the so-called 'marker' pigments) are restricted to 1 or 2 taxa and can be used as indicators for those taxa. The use of marker pigments in the identification of phytoplankton classes in seawater has increased in the past decade, mainly due to the development of high-performance liquid chromatography (HPLC) analytical techniques. Analysis of marine ecosystems by use of pigment concentra-

*Addressee for correspondence.
E-mail: denis.mackey@ml.csiro.au

tions has generally been qualitative (Jeffrey & Hallegraeff 1980, 1987, Hallegraeff & Jeffrey 1984, Ridout & Morris 1985, Gieskes & Kraay 1986, Klein & Sournia 1987), but more recently there have been attempts to estimate the abundances of various phytoplanktonic classes quantitatively from marker pigment concentrations (Gieskes & Kraay 1986, Gieskes et al. 1988, Everitt et al. 1990, Letelier et al. 1993). A robust method for such estimation would be invaluable, as it could lead to the development of fast semi-automated algal class identification using HPLC data. A fast method would allow a much more widespread investigation of phytoplankton abundances and distributions than is currently possible using cell counts and flow cytometry.

Gieskes et al. (1988) used HPLC analysis of pigments to estimate phytoplankton class abundances from chlorophyll (chl) *a* / marker pigment ratios (see also Gieskes & Kraay 1983). These pigment ratios were derived from a multiple regression analysis of the most important pigment markers. The analysis assumed that these ratios are constant within a sample group, and required a large data-set for statistical validity. However, this technique only established the contributions to the population from pigment-related groups. It showed, for instance, that fucoxanthin-containing species contributed 50% to a given sample, but could not differentiate between the diatoms, chrysophytes or prymnesiophytes which may have contributed this fucoxanthin. Minor groups are difficult to resolve from noise in the data and the technique fails if the concentration of all marker pigments co-vary. Shifts in algal pigment ratios with changes in light intensity, due to light adaptation (Gieskes & Kraay 1986, Demers et al. 1991), hinder the use of this technique for estimating the quantitative composition of natural phytoplankton. Gieskes et al. (1988) grouped their samples before analysis so as to take account of variations in the relative abundances of algal types defined in terms of a single pigment. This approach cannot be used if the relative abundances vary continuously across the data-set.

An alternative method, used by Everitt et al. (1990), involved dividing the plankton into classes on the basis of pigment types, and then determining the contribution of each class to the total chl *a* in the sample from measured pigment abundances and chl *a* / marker pigment ratios estimated from the literature. The abundance of classes without unique marker pigments were calculated by difference. The difference between the calculated and observed concentration of chl *a* was used to judge how well the predictions of the model matched experimental results, and an iterative procedure was used to minimise this difference by varying the chl *a* / marker pigment ratios. The drawback of this procedure was that the process of calculation by difference for those classes without clear marker pigments sometimes led to predictions of unrealistic or even negative concentrations for these classes.

Letelier et al. (1993) used a method based on a least squares solution of an overdetermined linear problem. Their method was not explained in great detail and it is not clear how they solve the problem that not all species have unique marker pigments. Some classes are calculated by difference, which can lead to negative chl *a* values, and the method does not seem to provide any way of optimising the auxiliary pigment ratios. No method was described for 'weighting' the pigment data to allow for different measurement errors in determining the individual pigments. Finally, if some of the pigment ratios are not well known, then their algorithm is not capable of providing a good answer. A similar approach was used by Bustillos-Guzman et al. (1995), Tester et al. (1995) and Andersen et al. (1996).

A more robust procedure is required if we are to make full use of the data from HPLC analyses. In this paper, we describe a new method for calculating plankton class abundances from measured pigment concentrations and estimated class pigment composition. The method was evaluated using a series of synthetic data-sets of HPLC pigment concentrations and corresponding algal class abundances. The application of this method to field samples is described in the accompanying paper (Wright et al. 1996).

We had no success with an alternative computational approach using factor analysis. It is described in Appendix 1 in the hope that others may find some way of overcoming the difficulties encountered and will not waste too much time repeating this work.

METHODS

Description of CHEMTAX program. The aim of the method outlined in this paper is to estimate the contributions of different phytoplankton classes to the pigment concentrations in various water samples. This is a factor analysis problem, where the data matrix S of pigment concentrations in a set of samples must be factorised into matrices F, giving the ratios of different pigments for each phytoplankton class, and C, giving the abundances of each phytoplankton class in each sample.

This problem is underdetermined and there are an infinite number of possible factorisations. In order to obtain a physically meaningful factorisation of S, an initial estimate of F, F_0, was made from literature values for pigment concentrations in various species (see Table 1—all pigment ratios normalised against chl *a* =

1.000). Estimates $\hat{C}$ and $\hat{F}$ for C and F were then determined such that $\hat{F}$ was as close as possible to F_0, subject to constraints on the positivity and normalisation of $\hat{C}$ and $\hat{F}$.

The initial guess for the phytoplankton class abundance matrix, $\hat{C}_0$, was directly calculated by solving the overdetermined least squares equation:

$$\text{minimise } \|S - \hat{C}_0 F_0\| \text{ subject to}$$

$$[\hat{C}_0]_{ij} \geq 0 \quad \forall i, j$$

$$\sum_i [\hat{C}_0]_{ij} = 1 \quad \forall j$$

The method outlined in Lawson & Hanson (1974) (least squares regression with inequality and equality constraints) was used to solve this equation, and the residua, ε_0, was calculated:

$$\varepsilon_0 = \|S - \hat{C}_0 F_0\|$$

A steepest descent algorithm was used to obtain a better factorisation of S. Each nonzero element $f_{i,j}$ of F_0 was varied in turn by a specified factor (typically 20%) and $\hat{C}$ and ε were recalculated each time. The variation causing the biggest decrease in ε was kept, giving a new ratio matrix F_1. Each element of F_1 was then varied in turn, with the variation giving the biggest decrease in ε being kept, and so on. Thus a series of matrices F_0, F_1, F_2, ... with corresponding $\hat{C}_0$, $\hat{C}_1$, $\hat{C}_2$, ... were determined, with $\{\varepsilon_i\} = \{\|S - \hat{C}_i F_i\|\}$ strictly decreasing with i. This series was determined until ε_i decreased below a preset limit, an iteration count was exceeded, or further iteration caused insignificant change in the value of ε_i. If the latter occurred, then the amount of variation on each step was reduced and the minimisation process continued.

In practice, it was found that variation of most of the elements of F_i in a particular iteration had little effect on either the residual or the calculated phytoplankton class abundance matrix C_i. Accordingly, rather than vary every element of F_i at each iteration, a small subset of the elements of F_i, which caused the largest decrease in the residual, was selected to be varied for a number of iterations. All the elements were then varied in order to select a new subset for downhill following (the pigments in this new subset were likely to be different from the the previous subset as a consequence of the continually decreasing residual during the iteration process). This procedure was several times faster than the full downhill following procedure and gave essentially the same results. In general, the calculation time for the procedure is proportional to the number of data samples and to the square of the number of plankton classes, but is largely independent of the number of pigments used.

The matrices F_n and $\hat{C}_n$ obtained at the end of the iterations are the final estimates of the pigment ratios within classes and class abundances within the samples, respectively. To avoid computational errors due to finite precision arithmetic, the data matrix S and the pigment ratio matrices F_i were normalised to unit row sum before the calculations (the program was designed to carry out this normalisation automatically allowing the user the freedom to enter the pigment ratios in any convenient form, e.g. as µg per 10^6 cells or as ratios to chl *a* as in Table 1). $\hat{C}_i$ was also forced to unit row sum, so that each row may be interpreted as giving the fraction of the total measured pigment due to each algal class. Before calculation, the data were weighted according to the reciprocal of the average pigment concentration in the data samples: this had the effect of making the residual a measure of relative rather than absolute fit to the data and increased the relative fit to the minor pigments at the expense of the major pigments.

The fraction of total chl *a* due to each phytoplankton class was also calculated from the fraction of total pigment due to each class and the elements of F_n; note that the direct comparison of the data obtained from this calculation with cell counts is complicated by the fact that the amount of pigment per cell in wild phytoplankton populations is usually unknown. This is especially important in samples from stratified waters, where the pigment content per cell of a given species may differ drastically between a surface sample and a deep water sample.

The calculations require that the pigment ratios within each phytoplankton class are constant across data samples, and hence that all of the data samples in any given calculation are from the same phytoplankton community and physiological state. A set of data samples which spans different physiological states, or communities of phytoplankton, should thus be split into groups to allow different optimum pigment ratio matrices to be used for each group (providing this does not reduce the sample size of the particular group below a critical value which will also introduce errors into the calculations—see below). For example, in the open ocean it is likely that the pigment 'fingerprint' for each class will change with depth, due to both light adaptation effects and the possibility that the species represented from a given algal class may vary with depth. A set of data samples from various depths along a transect should therefore be divided into a number of groups based on the depth at which each sample was taken, and optimum pigment ratio matrices for each of these groups calculated separately.

However, sample groups should not be too small. Although the calculation will work for small sets of data points, the more independent data points obtained from a particular phytoplankton community the better the estimate of the 'true' pigment ratio matrix F.

The regression procedure used is not overly robust to outliers, so pre-inspection of the data for obvious data errors is recommended.

Since the original problem of dividing the data matrix into pigment ratios and algal abundances was underdetermined, the choice of the initial pigment ratio matrix strongly affects the result obtained. The ratio matrix assumes that a 'typical' pigment composition is present in all members of 1 phytoplankton class. However, pigment compositions can vary widely even within a single species (Jeffrey & Wright 1994) and this introduces an unavoidable error into the estimates of class abundances produced by this method. If at all possible, the pigment ratios utilised should come from the major phytoplankton species native to the area where the data samples were obtained. It should be noted that the term 'class abundances' is slightly misleading: what is actually obtained is an estimate of the abundance of phytoplankton with the pigment type specified in the pigment ratio matrix, which may include phytoplankton from a number of taxonomic classes. For example, a number of prasinophytes are indistinguishable from chlorophytes on the basis of pigments alone (Ricketts 1970, Fawley 1992), and hence the pigment contribution from these prasinophytes will be attributed to the 'chlorophyte' pigment class. It should also be noted that the pigment ratios obtained from cultured phytoplankton may differ from the wild-type ratios.

The initial pigment ratio matrix F_0 must be set up with care if meaningful results are to be obtained from the calculation. The F_0 matrix must not be linearly dependent, and hence more pigments must be used than there are plankton classes to be calculated. However, using a highly overdetermined ratio matrix (i.e. many more pigments than plankton classes) can cause the iterative process to take an unduly long time. The best results are obtained when the number of pigments used is 2 or 3 greater than the number of pigment classes. It is important that each major phytoplankton pigment class likely to be present in the data samples is represented in the ratio matrix; for example, if a large number of chrysophyte-type phytoplankton are present in a sample but no close pigment type is available in the ratio matrix, then the results obtained will be unreliable.

Care should also be taken when selecting what pigments to use in the ratio matrix. Pigments that are present in nearly all phytoplankton are unlikely to give much useful information, while the use of pigments such as diadinoxanthin, which is converted rapidly to diatoxanthin in the light (Demers et al. 1991), or pigments which have wildly different abundances in different species within a class are also likely to give poor results. Each plankton class used should also prefer-

Table 1. Pigment to chl *a* ratios used in algal class composition calculation. The 2 rows of data under each algal class (or genus) represent the minimum and maximum, respectively, of pigment ratios found in the given references. All values are referenced against chl *a* = 1.000

Notes: [1]Using HPLC pigment analysis and flow cytometry Simon et al. (1994) have proposed 2 groups of prasinophytes, IIA and IIB, based on an earlier discrimination of Hooks et al. (1988). In part this discrimination was based on the PRAS:chl *a* ratio—about 0.18 for group IIA and 0.44 for group IIB. However, our literature survey revealed a continuum in the ratio over this range (S. W. Wright unpubl. results) and we consider these prasinophytes as 1 grouping (Type 3). Our proposed Types 1 and 2 prasinophytes had a much lower PRAS:chl *a* ratio with Type 1 having a high chl *b*:chl *a* ratio and Type 2 a low chl *b*:chl *a* ratio

[2]The haptophytes (prymnesiophytes) are sub-divided into 4 pigment groupings according to Jeffrey & Wright (1994). Type 1: FUCO but no chl c_3; Type 2: FUCO and chl c_3; Type 3: FUCO, HEX and chl c_3; Type 4: BUT and chl c_3 with or without FUCO and HEX and BUT/(FUCO+HEX) > 0.02. Note that pigment differences at the strain, species or genus level may mean that certain algae can be found in more than 1 pigment group (e.g. representatives of *Phaeocystis* may be found in both pigment Types 2 and 4)

[3]Andersen et al. (1993) have proposed that Type 2 be considered as a separate algal class, the Pelagophyceae

Abbreviations: Chl *c* = chlorophyll *c* unspecified; Chl*c1* = chlorophyll c_1; Chl*c2* = chlorophyll c_2; Chl*c3* = chlorophyll c_3; MgDV = Mg 3,8 divinyl pheoporphyrin a_5; SIPX = siphonoxanthin; PERI = peridinin; BUT = 19'-butanoyloxyfucoxanthin; FUCO = fucoxanthin; HEX = 19'-hexanoyloxyfucoxanthin; NEO = neoxanthin; PRAS = prasinoxanthin; MYXO = myxoxanthophyll; DINO = dinoxanthin; VIOL = violaxanthin; DDX = diadinoxanthin; ANTH = antheraxanthin; ALLO = alloxanthin; DIAT = diatoxanthin; LUT = lutein; ZEA = zeaxanthin; Chl*b1* = chlorophyll *b*; Chl*b2* = divinyl-chlorophyll *b*; Chl *b* = chlorophyll *b* or divinyl-chlorophyll *b*; Chl*a1* = chlorophyll *a*; Chl*a2* = divinyl-chlorophyll *a*; Chl *a* = chlorophyll *a* or divinyl-chlorophyll *a*; βγCA = βγ-carotene; βεCA = βε-carotene (α-carotene); ββCA = ββ-carotene (β-carotene); nd = not determined. These abbreviations are also used in Tables 2 to 6

	Chl *c*	Chl*c3*	Chl*c1*	Chl*c2*	MgDV	SIPX	PERI	BUT	FUCO	HEX	NEO	PRAS	MYXO	DINO	VIOL	DDX	ANTH	ALLO	DIAT	LUT	ZEA	Chl*b*	βγCA	βεCA	ββCA	Ref.
CYANOBACTERIA																										
Synechococcus																										a, b, p
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.077	0	0	0	0.021	
max	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.590	0	0	0	0.196	
Trichodesmium																										q
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.048	0	0	0	0	
max	0	0	0	0	0	0	0	0	0	0	0	0	0.034	0	0	0	0	0	0	0	0.175	0	0	0	0 139	
PROCHLOROPHYCEAE																										n, o
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.146	0.125	0	0	0	
max	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.321	1.074	0	0.135	0.171	
EUGLENOPHYCEAE																										b, o
min	0	0	0	0	0	0	0	0	0	0	0.015	0	0	0	0	0.230	0	0	0	0	0	0.406	0	0	0	
max	0	0	0	0	0	0	0	0	0	0	0.015	0	0	0	0	0.230	0	0	0.054	0	0	0.406	0	0	0.040	
CHLOROPHYCEAE																										a, j, k, l, m
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
max	0	0	0	0	0	0	0	0	0	0	0.074	0	0	0	0.055	0	0.048	0	0	0.283	0.118	0.569	0.009	0.002	4.667	
PRASINOPHYCEAE[1]																										
Type 1																										a, b, c, d
min	0	0	0	0	0	0	0	0	0	0	0.034	0	0	0	0.011	0	0	0	0	0	0.057	0.589	0.001	0.004	0.047	
max	0	0	0	0	0.034	0.099	0	0	0	0	0.067	0	0	0	0.055	0	0	0	0	0.086	0.093	1.034	0.031	0.035	0.150	
Type 2																										
min	0	0	0	0	0.001	0	0	0	0	0	0	0.005	0	0	0	0	0	0	0	0	0	0.009	0	0	nd	
max	0	0	0	0	0.006	0	0	0	0	0	0	0.043	0	0	0.015	0	0	0	0	0	0.006	0.252	0	0	nd	
Type 3																										
min	0	0	0	0	0	0	0	0	0	0	0	0.115	0	0	0	0	0	0	0	0	0	0.366	0	0	0	
max	0.448	0	0.001	0	0.384	0	0	0	0	0	0.210	0.606	0	0	0.229	0	0.006	0	0	0.051	0.283	1.410	0	0.027	0.261	
DINOPHYCEAE												f														a, b, e, f
min	0	0	0	0.037	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
max	0	0	0	0.532	0	0	1.063	0	0	0.384	0	0	0	0.005	0	0.241	0	0	0.201	0	0	0	0	0	0.023	
HAPTOPHYCEAE[2]																										g, h
Type 1																										
min	0.080	0	0	0	0	0	0	0	0.360	0	0	0	0	0	0	0.107	0	0	0.000	0	0	0	0	0	0.020	
max	0.306	0	0	0	0	0	0	0	0.948	0	0	0	0	0	0	0.254	0	0	0.014	0	0	0	0	0.031	0.123	
Type 2																										
min	0.175	0.062	0	0	0	0	0	0	0.465	0	0	0	0	0	0	0.055	0	0	0.001	0	0	0	0	0	0.016	
max	0.304	0.141	0	0	0	0	0	0	1.115	0	0	0	0	0	0	0.105	0	0	0.006	0	0	0	0	0.010	0.026	
Type 3																										
min	0.104	0.054	0	0	0	0	0	0	0.038	0.000	0	0	0	0	0	0.034	0	0	0	0	0	0	0	0	0.009	
max	0.330	0.170	0	0	0	0	0	0.023	1.205	1.360	0	0	0	0	0	0.280	0	0	0.035	0	0	0	0	0.015	0.030	
Type 4																										
min	0	0.097	0	0	0	0	0	0.023	0.080	0.010	0	0	0	0	0	0.060	0	0	0	0	0	0	0	0	0	
max	0.300	0.421	0	0	0	0	0	0.521	0.890	1.067	0	0	0	0	0	0.426	0	0	0.179	0	0	0	0	0.012	0.020	
CRYPTOPHYCEAE																										a, b
min	0	0	0	0.077	0	0	0	0	0	0	0	0	0	0	0	0	0	0.042	0	0	0	0	0	0.011	0	
max	0	0	0	0.174	0	0	0	0	0	0	0	0	0	0	0	0	0	0.229	0	0	0	0	0	0.057	0	
BACILLARIOPHYCEAE																										a, i
min	0	0	0	0	0	0	0	0	0.159	0	0	0	0	0	0	0.030	0	0	0	0	0	0	0	0	0	
max	0.183	0	0	0	0	0	0	0	0.755	0	0.003	0	0	0	0	0.448	0	0	0.269	0	0	0	0	0.154	0.027	
CHRYSOPHYCEAE[3]																										a, b, ı, r
Type 1																										
min	0	0	0	0	0	0	0	0	0.314	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.085	
max	0	0	0	0	0	0	0	0	0.678	0	0.013	0	0	0	0.174	0.246	0.012	0	0.094	0	0.119	0	0	0	0.085	
Type 2																										
min	0	0	0	0	0	0	0	0.303	0.275	0	0	0	0	0	0	0 018	0	0	0.001	0	0	0	0	0	0	
max	0.536	0.250	0	0.254	0	0	0	1.563	0.974	0	0	0	0	0	0	0.857	0	0	0.100	0	0	0	0	0	0.016	

References: (a) Jeffrey & Wright (in press), (b) S. W. Wright (unpubl. data), (c) Ricketts (1967) (d) Wilhelm et al. (1987), (e) Bjornland & Tangen (1979), (f) Jeffrey et al. (1975), (g) Berger et al. (1977), (h) Jeffrey & Wright (1994), (ı) Hager & Stransky (1970a), (j) Ben-Amotz et al. (1982), (k) Burczyk et al. (1981), (l) Hager & Stransky (1970b), (m) Fawley (1992), (n) Burger-Wiersma et al. (1986), (o) Stauber & Jeffrey (1989), (p) Stransky & Hager (1970), (q) Carpenter et al. (1993), (r) Andersen et al. (1993)

ably have at least 2 pigments in addition to chl *a*, and 'marker' pigments will give better results than more common pigments. If a given marker pigment is not present in a set of samples, then the plankton classes containing that marker pigment should be removed from the ratio matrix in order to reduce calculation time. To reduce computation time (and the likelihood of of unrealistic false minima) the initial values of the pigment to chl *a* ratios should also be as close as possible to expected values.

A MATLAB™ program, CHEMTAX, was developed to perform these calculations. The data files and options for the CHEMTAX calculations were set up by a preprocessor (PREPRO) program for the IBM PC. The user-defined CHEMTAX parameters selected in this study were based on our evaluation of the CHEMTAX program using the synthetic data-sets. Three matrices were required as input to the program: the data matrix S containing the HPLC pigment concentrations, the initial ratio matrix F_0, and the ratio limits matrix which controls the degree to which CHEMTAX was allowed to alter the initial pigment ratios. Unless stated otherwise, all the ratio limits were set to a default value of 500%, which allowed the initial pigment ratio, r, to vary from $r/5$ to $5r$.

Development of the method required an independent assessment of phytoplankton class abundances to compare with those calculated by CHEMTAX. While data-sets of HPLC-derived pigment concentrations and phytoplankton abundances estimated by microscopy or flow cytometry were available, they were known to be selective (for reasons outlined in the introduction) and there was no way of knowing the 'true' abundances of each algal class for assessment of the CHEMTAX results. Also, in most field data-sets there is usually some degree of co-variance where, for example, there are parallel increases in the abundances of several algal classes as a sub-surface chl *a* maximum is approached. While this co-variance could be adequately handled by the model, it complicated the initial development and evaluation. Therefore, the program was tested on a series of synthetic computer-generated random data-sets of algal class abundances and pigment concentrations.

Synthetic data-sets. The first data-set simulated a phytoplankton community from the Southern Ocean. Since pigment data for inclusion in pigment ratio matrices were not available for many Southern Ocean species, quantitative data from algal cultures grown under standard conditions from the SCOR-UNESCO Workshops (Jeffrey & Wright in press) were used for Bacillariophyceae (*Phaeodactylum tricornutum* CS-29), Prasinophyceae (*Pycnococcus provasolii* CS-185), Dinophyceae (*Amphidinium carterae* CS-212), Cryptophyceae (*Chroomonas salina* CS-174), Chlorophyceae (*Dunaliella tertiolecta* CS-175), Cyanobacteria [*Synechococcus* sp. (DC2) CS-197] and 2 species of Haptophyceae (*Emiliania huxleyi* CS-57 and *Phaeocystis pouchetii* CS-165). This enabled us to generate a known pigment ratio matrix F_0 (Table 2a) by using the values from the SCOR-UNESCO Workshop (Jeffrey & Wright in press). It should be noted that the CHEMTAX calculations are independent of the units used in the data matrix. In this study, pigment concentrations in the ratio matrix were specified in µg per 10^6 cells and the results were obtained both in terms of the absolute concentration of chl *a* due to each phytoplankton class and in terms of the relative contribution of each phytoplankton class to the total pigment.

A second data-set was constructed to simulate an equatorial phytoplankton community and used the pigment ratios given in Table 3a. The data-set included the following additional species: *Prochlorococcus marinus* (Chisholm et al. 1988), *Euglena* sp. (Hager & Stransky 1970a), *Pelagococcus subviridis* (Jeffrey & Wright in press) and *Trichodesmium theibautii* (Carpenter et al. 1993). *Phaeocystis pouchetii* was not used in this data-set (Table 3).

The pigment ratios for a real sample are unlikely to be known exactly and, therefore, we added random errors to the pigment ratio matrices to simulate deviations from the values due to regional variations of individual species, strain differences within a given species (e.g. Jeffrey & Wright in press) and local changes in algal physiology due to environmental factors such as temperature, salinity, light field, nutrient stress and mixing regimes. These errors were simulated by producing a set of normally distributed random numbers (mean = 0, variance = 1, using an algorithm derived from Zelen & Severo 1970) which were multiplied by the pigment concentration and a scaling factor and added to the original data to produce pigment ratios with standard errors of ±10%, ±25% and ±50%. These modified pigment ratio matrices are given in Table 2b, c & d for the Southern Ocean species. The individual matrix elements are given as percentages of the 'true' matrix elements (Table 2a) in Table 4a, b & c. For the Equatorial Pacific synthetic data-set, the 'true' matrix is given in Table 3a and the modified pigment ratios are given in Table 3b as percentages of the 'true' values after the addition of a normal-random error of ±25%.

As all CHEMTAX calculations first require normalization against total pigment, and all output is in this format, the synthetic ratio matrices and results of all CHEMTAX runs in this paper are also normalized against total pigment. Unless stated otherwise, all program runs were made on synthetic Southern Ocean and Equatorial Pacific data-sets with all non-zero pigment ratios of the matrix being allowed to vary. This

Table 2. Pigment ratios (normalized to total pigment) representative of Southern Ocean species. (a) Initial ratio matrix used to construct the synthetic data-set—'true' matrix and modified by the addition of random normalised errors of (b) ±10%; (c) ±25%; and (d) ±50%

Additional abbreviations: Pras (T3) = prasinophytes (Type 3); Dino = dinoflagellates; Cryp = cryptophytes; Hapt (T3, T4) = haptophytes (Type 3, Type 4); Chry = chrysophytes; Eugl = euglenophytes; Chlo = chlorophytes; Proc = prochlorophytes; Syne = *Synechococcus*; Tric = *Trichodesmium*; Diat = diatoms. These abbreviations also apply to Tables 3 to 6

	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b1*	Chl*a1*
(a)												
Pras (T3)	0	0	0	0	0.061	0.127	0.025	0	0.004	0	0.381	0.403
Dino	0.515	0	0	0	0	0	0	0	0	0	0	0.485
Cryp	0	0	0	0	0	0	0	0.186	0	0	0	0.814
Hapt (T3)	0	0	0	0.630	0	0	0	0	0	0	0	0.370
Hapt (T4)	0	0.104	0.247	0.227	0	0	0	0	0	0	0	0.422
Chlo	0	0	0	0	0.040	0	0.035	0	0.127	0.006	0.165	0.628
Syne	0	0	0	0	0	0	0	0	0	0.258	0	0.742
Diat	0	0	0.430	0	0	0	0	0	0	0	0	0.570
(b)												
Pras (T3)	0	0	0	0	0.057	0.144	0.026	0	0.003	0	0.369	0.402
Dino	0.574	0	0	0	0	0	0	0	0	0	0	0.426
Cryp	0	0	0	0	0	0	0	0.188	0	0	0	0.813
Hapt (T3)	0	0	0	0.691	0	0	0	0	0	0	0	0.309
Hapt (T4)	0	0.095	0.233	0.222	0	0	0	0	0	0	0	0.450
Chlo	0	0	0	0	0.045	0	0.038	0	0.153	0.006	0.174	0.584
Syne	0	0	0	0	0	0	0	0	0	0.306	0	0.694
Diat	0	0	0.418	0	0	0	0	0	0	0	0	0.582
(c)												
Pras (T3)	0	0	0	0	0.052	0.119	0.020	0	0.004	0	0.363	0.444
Dino	0.524	0	0	0	0	0	0	0	0	0	0	0.476
Cryp	0	0	0	0	0	0	0	0.166	0	0	0	0.834
Hapt (T3)	0	0	0	0.614	0	0	0	0	0	0	0	0.386
Hapt (T4)	0	0.093	0.281	0.207	0	0	0	0	0	0	0	0.419
Chlo	0	0	0	0	0.046	0	0.032	0	0.220	0.007	0.277	0.419
Syne	0	0	0	0	0	0	0	0	0	0.370	0	0.630
Diat	0	0	0.517	0	0	0	0	0	0	0	0	0.483
(d)												
Pras (T3)	0	0	0	0	0.031	0.131	0.049	0	0.005	0	0.464	0.364
Dino	0.401	0	0	0	0	0	0	0	0	0	0	0.599
Cryp	0	0	0	0	0	0	0	0.268	0	0	0	0.732
Hapt (T3)	0	0	0	0.490	0	0	0	0	0	0	0	0.510
Hapt (T4)	0	0.096	0.344	0.296	0	0	0	0	0	0	0	0.263
Chlo	0	0	0	0	0.069	0	0.077	0	0.346	0.003	0.373	0.132
Syne	0	0	0	0	0	0	0	0	0	0.241	0	0.759
Diat	0	0	0.633	0	0	0	0	0	0	0	0	0.367

gave a slight increase in accuracy albeit with longer computation times compared with calculations using a smaller subset.

A series of random data matrices were generated to simulate the Southern Ocean phytoplankton community. For each of up to 40 'samples', the 'cell number' of each class was set using a random number (between 0 and 1, mean = 0.5) divided by the chl *a* content per cell for that class. In this way, each class contributed, on average, 0.5 µg of chl *a* to each sample or 12.5% of the total chl *a* for the 8-class Southern Ocean data-set. These cell numbers were multiplied by the cellular content of each pigment to derive the contribution of each class to the population pigment content. These contributions were then summed for each sample to produce the basic synthetic field data-set *S*. For instance, the concentration of fucoxanthin represented the sum of contributions from *Phaeodactylum tricornutum* (diatom) and *Phaeocystis pouchetii* (haptophyte). For each test run, calculations were performed on 3 separate data matrices to ensure that no artifacts occurred during the computations. As for the pigment ratios, experimental error was simulated by producing a set of normally distributed random numbers (mean =

Table 3. Pigment ratios (normalized to total pigment) representative of Equatorial Pacific species. (a) Initial ratio matrix used to construct the synthetic data-set 'true' matrix; (b) modified by the addition of random normalised errors of ±25%. Matrix elements are expressed as a percentage of the 'true' matrix. Final ratio matrices, (c) and (d), after fitting by CHEMTAX with matrix elements expressed as a percentage of the 'true' matrix elements. Random normalised errors of ±25% were added to the pigment ratios and typical 'experimental errors' were added to the data-set. Calculations with: (c) divinyl-chl *a* and *b* and; (d) divinyl-chl *a* and *b* not distinguished from chl *a* and *b*

	PER	BUT	FUCO	HEX	NEO	PRAS	MYXO	VIOL	DDX	ALLO	LUT	ZEA	Chl*b2*	Chl*a2*	Chl*b1*	Chl*a1*
(a)																
Pras (T3)	0	0	0	0	0.061	0.127	0	0.025	0	0	0.004	0	0	0	0.381	0.403
Dino	0.462	0	0	0	0	0	0	0	0.104	0	0	0	0	0	0	0.434
Cryp	0	0	0	0	0	0	0	0	0	0.186	0	0	0	0	0	0.814
Hapt (T3)	0	0	0	0.608	0	0	0	0	0.036	0	0	0	0	0	0	0.356
Chry	0	0.152	0.400	0	0	0	0	0	0.037	0	0	0	0	0	0	0.411
Eugl	0	0	0	0	0.009	0	0	0	0.139	0	0	0	0	0	0.246	0.606
Chlo	0	0	0	0	0.040	0	0	0.035	0	0	0.127	0.006	0	0	0.165	0.628
Proc	0	0	0	0	0	0	0	0	0	0	0	0.134	0.449	0.418	0	0
Syne	0	0	0	0	0	0	0	0	0	0	0	0.258	0	0	0	0.742
Tric	0	0	0	0	0	0	0.015	0	0	0	0	0.092	0	0	0	0.893
Diat	0	0	0.399	0	0	0	0	0	0.072	0	0	0	0	0	0	0.529
(b)																
Pras (T3)	0	0	0	0	99.0	82.1	0	93.3	0	0	89.1	0	0	0	110.5	96.4
Dino	110.3	0	0	0	0	0	0	0	96.0	0	0	0	0	0	0	90.0
Cryp	0	0	0	0	0	0	0	0	0	92.9	0	0	0	0	0	101.6
Hapt (T3)	0	0	0	107.2	0	0	0	0	80.3	0	0	0	0	0	0	89.7
Chry	0	111.7	96.1	0	0	0	0	0	74.4	0	0	0	0	0	0	101.8
Eugl	0	0	0	0	76.7	0	0	0	86.8	0	0	0	0	0	90.4	107.3
Chlo	0	0	0	0	74.8	0	0	103.4	0	0	87.6	122.4	0	0	85.5	107.5
Proc	0	0	0	0	0	0	0	0	0	0	0	124.5	86.0	107.2	0	0
Syne	0	0	0	0	0	0	0	0	0	0	0	79.8	0	0	0	107.0
Tric	0	0	0	0	0	0	80.8	0	0	0	0	118.2	0	0	0	98.5
Diat	0	0	106.5	0	0	0	0	0	91.9	0	0	0	0	0	0	96.2
(c)																
Pras (T3)	0	0	0	0	102.8	104.1	0	104.6	0	0	95.0	0	0	0	102.8	95.4
Dino	99.1	0	0	0	0	0	0	0	100.3	0	0	0	0	0	0	100.9
Cryp	0	0	0	0	0	0	0	0	0	101.3	0	0	0	0	0	99.7
Hapt (T3)	0	0	0	100.9	0	0	0	0	87.5	0	0	0	0	0	0	99.7
Chry	0	117.8	99.8	0	0	0	0	0	75.2	0	0	0	0	0	0	95.8
Eugl	0	0	0	0	90.6	0	0	0	101.7	0	0	0	0	0	96.0	101.4
Chlo	0	0	0	0	102.7	0	0	98.8	0	0	99.4	138.9	0	0	102.2	99.1
Proc	0	0	0	0	0	0	0	0	0	0	0	99.5	100.1	100.1	0	0
Syne	0	0	0	0	0	0	0	0	0	0	0	102.3	0	0	0	99.2
Tric	0	0	0	0	0	0	95.6	0	0	0	0	88.6	0	0	0	101.3
Diat	0	0	101.2	0	0	0	0	0	102.0	0	0	0	0	0	0	98.9
(d)																
Pras (T3)	0	0	0	0	101.4	102.0	0	103.7	0	0	119.5	0	–	–	101.7	97.1
Dino	98.8	0	0	0	0	0	0	0	102.6	0	0	0	–	–	0	100.6
Cryp	0	0	0	0	0	0	0	0	0	98.6	0	0	–	–	0	100.3
Hapt (T3)	0	0	0	101.7	0	0	0	0	87.9	0	0	0	–	–	0	98.3
Chry	0	116.5	100.3	0	0	0	0	0	77.6	0	0	0	–	–	0	95.6
Eugl	0	0	0	0	94.2	0	0	0	104.7	0	0	0	–	–	94.8	101.1
Chlo	0	0	0	0	103.1	0	0	99.5	0	0	100.4	129.4	–	–	90.4	102.0
Proc	0	0	0	0	0	0	0	0	0	0	0	85.4	–	–	112.6	91.1
Syne	0	0	0	0	0	0	0	0	0	0	0	102.3	–	–	0	99.2
Tric	0	0	0	0	0	0	88.2	0	0	0	0	107.4	–	–	0	99.4
Diat	0	0	104.9	0	0	0	0	0	99.3	0	0	0	–	–	0	96.8

0, variance = 1, using an algorithm derived from Zelen & Severo 1970) which were multiplied by the pigment concentration and a scaling factor and added to the original data to produce data-sets with ±10% standard error.

More sophisticated data-sets were based on experimental observations and took into account 2 sources of experimental error, namely HPLC injection errors (which affect all peaks equally and do not alter the peak ratios) and errors of detection and integration

Table 4. Initial pigment ratios representative of Southern Ocean species used to construct synthetic data sets. Matrix elements are expressed as a percentage of the 'true' matrix elements (Table 2a) after the addition of random normalised errors of: (a) ±10%; (b) ±25%; and (c) ± 50%

	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b1*	Chl*a1*
(a)												
Pras (T3)	0	0	0	0	93.0	113.1	104.7	0	87.7	0	96.8	99.8
Dino	111.4	0	0	0	0	0	0	0	0	0	0	87.9
Cryp	0	0	0	0	0	0	0	100.6	0	0	0	99.9
Hapt (T3)	0	0	0	109.7	0	0	0	0	0	0	0	83.6
Hapt (T4)	0	91.6	94.4	97.6	0	0	0	0	0	0	0	106.7
Chlo	0	0	0	0	113.5	0	109.4	0	119.8	111.1	105.4	93.1
Syne	0	0	0	0	0	0	0	0	0	118.6	0	93.5
Diat	0	0	97.2	0	0	0	0	0	0	0	0	102.1
(b)												
Pras (T3)	0	0	0	0	84.9	93.6	78.5	0	101.4	0	95.3	110.1
Dino	101.6	0	0	0	0	0	0	0	0	0	0	98.3
Cryp	0	0	0	0	0	0	0	88.8	0	0	0	102.6
Hapt (T3)	0	0	0	97.4	0	0	0	0	0	0	0	104.4
Hapt (T4)	0	89.9	113.8	91.2	0	0	0	0	0	0	0	99.2
Chlo	0	0	0	0	114.7	0	92.1	0	172.7	120.8	167.9	66.7
Syne	0	0	0	0	0	0	0	0	0	143.5	0	84.9
Diat	0	0	120.2	0	0	0	0	0	0	0	0	84.7
(c)												
Pras (T3)	0	0	0	0	51.4	102.8	19.7	0	146.2	0	121.9	90.3
Dino	77.8	0	0	0	0	0	0	0	0	0	0	123.6
Cryp	0	0	0	0	0	0	0	143.5	0	0	0	90.0
Hapt (T3)	0	0	0	77.7	0	0	0	0	0	0	0	138.0
Hapt (T4)	0	92.7	139.4	130.5	0	0	0	0	0	0	0	62.4
Chlo	0	0	0	0	174.3	0	223.4	0	271.7	47.8	225.9	21.0
Syne	0	0	0	0	0	0	0	0	0	93.4	0	102.3
Diat	0	0	147.1	0	0	0	0	0	0	0	0	64.5

(which affect peaks individually and are proportionately greater for smaller peak areas). These were determined experimentally by repeated HPLC analysis of a solution of β-*apo*-carotenal (16.5 µg ml^{-1} in methanol, Sigma Chemical Co.). Ten injections of 100 µl were performed using a Gilson 231 autoinjector onto a Spherisorb ODS2 column (25 cm × 4.6 mm), eluted isocratically with methanol, detected at 405 and 436 nm (Waters 440 detector) or 435 and 470 nm (Spectraphysics detector), and integrated using Waters Baseline software. The solution was diluted by 50% and again analysed 10 times. The process was repeated until the peak was no longer detectable (10 dilutions). The covariance of the areas for the 2 channels was taken to be the injection error, which was independent of the peak area. The remaining error was taken to be quantitation error, for which a relationship with the reciprocal of log(peak area) was obtained (see 'Results'). This relationship was used to alter the scaling factor (used with the normally distributed random numbers described above) to generate a data-set in which the simulated experimental errors were related to peak area as in a real data-set.

RESULTS

Synthetic data-sets: Southern Ocean

For each simulated phytoplankton community, all 3 random data-sets gave essentially the same results, showing that there were no systematic errors introduced into the data-sets. We are therefore confident that the results presented below are representative of the real situations that were being simulated. In the following section, all the results are reported from a single data-set so that the results can be readily compared. Any changes to the data-set or conditions are explicitly mentioned.

Sensitivity to uncertainty in pigment ratios

In Table 2a we list the initial ratio matrix which was used to generate a synthetic HPLC data-set that would be representative of a sample from the Southern Ocean. This initial ratio matrix will be referred to as the 'true' matrix and all parameters derived from this

matrix (without the addition of errors) will be referred to as 'true' parameters. In Table 2, we also list ratio matrices to which a random error was added representing a normalised standard deviation of ±10%, ±25% and ±50%. Note that the actual error introduced in any particular pigment ratio can be considerably higher than these values. In the section following, all errors of ±x% are added as normal standard errors, which implies that errors of ±2x% and ±3x% will occur in 3% and 0.3% of the cases, respectively. For example, the addition of ±25% noise caused the ratio of lutein to total pigment in chlorophytes to increase to 173% of the 'true' value (from 0.127 to 0.222), while the ratio of chl *a* to total pigment in chlorophytes decreased to 67% of the 'true' value (from 0.628 to 0.419). In Table 4, we list the initial matrix elements used for the calculations as a percentage of the 'true' value which is given in Table 2a.

As expected, when there was no noise added to either the pigment ratio or the data files, the program went through 1 iteration and stopped. The calculated chl *a* was essentially distributed in proportion to the original algal class contributions. The program was then tested with errors of ±10%, ±25% and ±50% added to the pigment ratio matrix but with no noise added to the data matrix. This approximates the situation where there are no experimental errors involved in measuring the pigments by HPLC but where there is uncertainty as to the correct pigment ratios applicable to a given water mass.

The program should converge on a solution that produced a final ratio matrix identical to that used to generate the data-set. In other words, using the initial ratio matrices listed in Table 2b, c & d, the program should be able to derive the 'true' ratio matrix given in Table 2a. In Table 5, we list the final ratio matrices with each ratio listed as a percentage of the value in the 'true' matrix (Table 2a). If the program worked perfectly, all the matrix elements in Table 5 would be 0 or 100. While this is not the case, the ratios are generally much closer to the 'true' value than the starting value. For example, in the case of a ±25% error, the ratios of zeaxanthin and chl *a* to total pigment in chlorophytes changed from 121 and 67% of the 'true' value (Table 4b) to 89 and 100%, respectively (Table 5b).

Table 5. Final pigment ratios representative of Southern Ocean species after fitting by CHEMTAX. Matrix elements are expressed as a percentage of the 'true' matrix elements (Table 2a). Calculations were for synthetic data sets where random normalised errors of: (a) ±10% (Table 2b); (b) ±25% (Table 2c); and (c) ±50% (Table 2d) were added to the pigment ratios. No errors were added to the data sets

	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b1*	Chl*a1*
(a)												
Pras (T3)	0	0	0	0	99.1	97.4	101.1	0	193.2	0	98.6	101.4
Dino	100.0	0	0	0	0	0	0	0	0	0	0	100.0
Cryp	0	0	0	0	0	0	0	100.0	0	0	0	100.0
Hapt (T3)	0	0	0	100.0	0	0	0	0	0	0	0	100.1
Hapt (T4)	0	93.1	94.1	110.7	0	0	0	0	0	0	0	99.4
Chlo	0	0	0	0	100.2	0	100.3	0	100.3	92.9	100.1	100.0
Syne	0	0	0	0	0	0	0	0	0	100.2	0	100.0
Diat	0	0	99.9	0	0	0	0	0	0	0	0	100.1
(b)												
Pras (T3)	0	0	0	0	100.2	100.4	99.8	0	84.2	0	100.1	99.9
Dino	100.0	0	0	0	0	0	0	0	0	0	0	100.0
Cryp	0	0	0	0	0	0	0	100.0	0	0	0	100.0
Hapt (T3)	0	0	0	100.0	0	0	0	0	0	0	0	100.1
Hapt (T4)	0	101.4	104.2	93.2	0	0	0	0	0	0	0	100.9
Chlo	0	0	0	0	100.2	0	100.2	0	100.3	89.2	100.8	99.8
Syne	0	0	0	0	0	0	0	0	0	100.0	0	100.0
Diat	0	0	100.0	0	0	0	0	0	0	0	0	100.0
(c)												
Pras (T3)	0	0	0	0	97.5	103.6	30.2	0	166.0	0	97.1	105.7
Dino	89.2	0	0	0	0	0	0	0	0	0	0	111.5
Cryp	0	0	0	0	0	0	0	143.5	0	0	0	90.0
Hapt (T3)	0	0	0	72.6	0	0	0	0	0	0	0	146.8
Hapt (T4)	0	124.6	138.9	112.8	0	0	0	0	0	0	0	64.3
Chlo	0	0	0	0	218.5	0	326.3	0	181.7	59.9	242.6	26.3
Syne	0	0	0	0	0	0	0	0	0	72.2	0	109.7
Diat	0	0	127.6	0	0	0	0	0	0	0	0	79.1

When the initial ratio matrix had an error of only ±10%, the program was able to adjust the pigment ratios to within a few percent of the 'true' ratios with the exception of lutein in prasinophytes, where the final value was 193% of the 'true' value (Table 5a). However, for the data-set used here, lutein is only a minor pigment in these types of prasinophytes and the main source of lutein is from chlorophytes. When a perfect fit of the data is not possible (as with field data due to noise in the ratio or data files) the CHEMTAX program often optimises the major pigments at the expense of the minor pigments. However, for an initial ratio matrix with ±10% error, the program was still able to reproduce the abundances (as measured by chl *a*) of all phytoplankton classes (including prasinophytes) very well. In the analysis of a real sample, a large change in a pigment ratio could indicate a potential problem and, if the particular pigment ratio were well characterised, then the ratio limit matrix could be used to limit the amount that the ratio was permitted to vary.

When the initial ratio matrix had an error of ±25%, the program was still able to adjust most of the pigment ratios to within a few percent of the 'true' values with the largest deviations being for zeaxanthin in chlorophytes and lutein in prasinophytes where the final values were 89 and 84% of the 'true' values, respectively (Table 5b). When the error in the initial ratio matrix was increased to ±50%, the program had great difficulty in estimating the 'true' pigment ratios (Table 5c).

Fig. 1 shows the correspondence between the concentrations of chl *a* calculated by CHEMTAX and the 'true' values used in determining the data matrix (*S*). In order to visualise the relationship, the 'true' values, which were originally randomly distributed, were re-arranged in increasing order for each class. They are plotted with a solid line against sample number, while the calculated values (where ±25% and ±50% error were added to the ratio matrix) are plotted as points. Note that because of the re-arrangement, the sample numbers do not correspond between graphs for different classes. In agreement with the observation that the program was able to closely reproduce the correct pigment ratios when a ±25% error had been added, there was excellent agreement between the calculated and 'true' values (Fig. 1), even for the prasinophyte and chlorophyte classes where the largest errors in pigment ratios were found (Table 5b).

However, with an error of ±50% added to the pigment ratio matrix, there was good agreement only for prasinophytes (Fig. 1a) with acceptable agreement for dinoflagellates (Fig. 1b). For the other phytoplankton classes, an indication of the goodness-of-fit can be

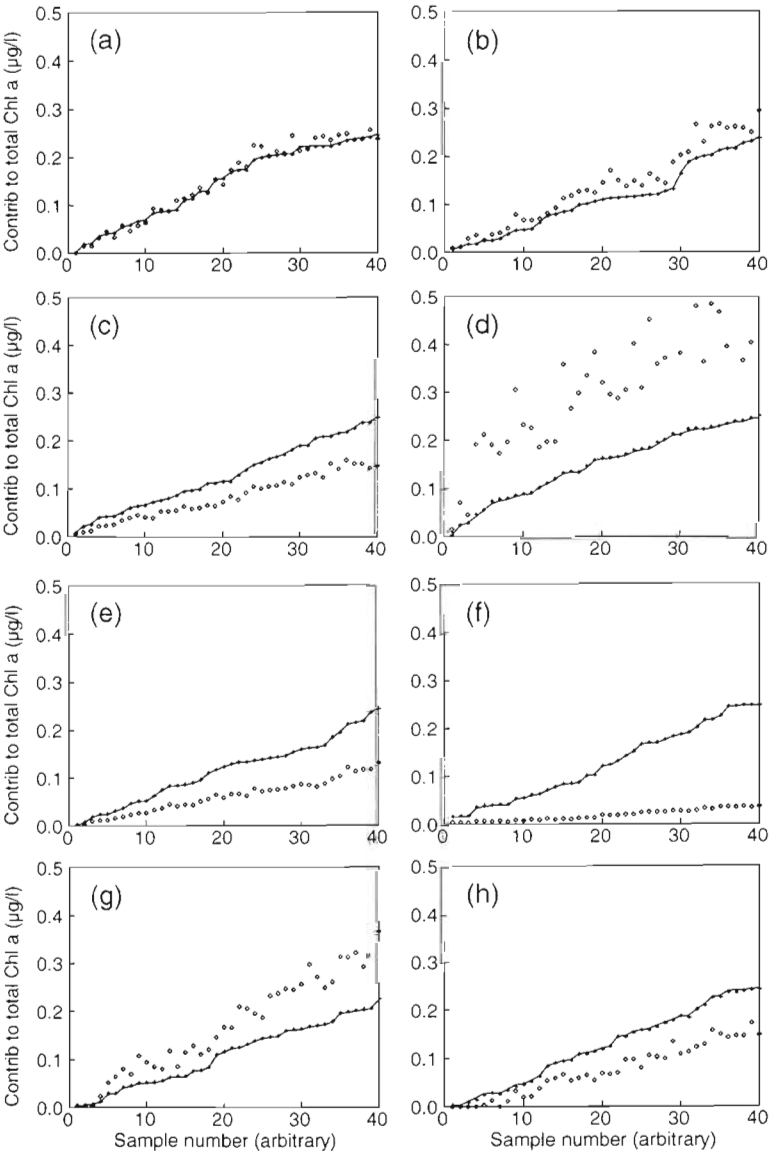


Fig. 1 Contribution to total chl *a* in the synthetic HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class: (a) prasinophyte (T3), (b) dinoflagellate, (c) cryptophyte, (d) haptophyte (T3), (e) haptophyte (T4), (f) chlorophyte, (g) cyanobacteria and (h) diatom. The solid line is the 'true' value. The calculated values are given for the case where there were no errors added to the data and with random normal standard errors of (+) ±25% and (◇) ±50% added to the pigment ratio matrix

Table 6. Final pigment ratios representative of Southern Ocean species after fitting by CHEMTAX. Matrix elements are expressed as a percentage of the 'true' matrix elements (Table 2a). Calculations were for synthetic data sets where random normalised errors of: (a) ±10% (Table 2b); (b) ±25% (Table 2c); and (c) ±50% (Table 2d) were added to the pigment ratios. Random normalised errors of ±10% were added to the data-sets to simulate analytical errors

	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b1*	Chl*a1*
(a)												
Pras (T3)	0	0	0	0	98.9	97.4	101.2	0	94.1	0	96.2	104.5
Dino	96.8	0	0	0	0	0	0	0	0	0	0	103.4
Cryp	0	0	0	0	0	0	0	109.7	0	0	0	97.8
Hapt (T3)	0	0	0	96.3	0	0	0	0	0	0	0	106.3
Hapt (T4)	0	93.8	86.7	100.0	0	0	0	0	0	0	0	109.3
Chlo	0	0	0	0	122.0	0	110.8	0	113.8	108.6	115.8	91.0
Syne	0	0	0	0	0	0	0	0	0	114.7	0	94.9
Diat	0	0	100.8	0	0	0	0	0	0	0	0	99.4
(b)												
Pras (T3)	0	0	0	0	97.7	96.1	100.3	0	99.3	0	93.3	107.9
Dino	101.6	0	0	0	0	0	0	0	0	0	0	98.3
Cryp	0	0	0	0	0	0	0	111.1	0	0	0	97.5
Hapt (T3)	0	0	0	86.1	0	0	0	0	0	0	0	123.8
Hapt (T4)	0	97.6	88.8	99.0	0	0	0	0	0	0	0	107.7
Chlo	0	0	0	0	135.9	0	125.4	0	126.7	143.2	143.7	79.0
Syne	0	0	0	0	0	0	0	0	0	100.9	0	99.7
Diat	0	0	118.9	0	0	0	0	0	0	0	0	85.8
(c)												
Pras (T3)	0	0	0	0	101.9	110.4	23.2	0	165.8	0	98.1	102.4
Dino	83.0	0	0	0	0	0	0	0	0	0	0	118.1
Cryp	0	0	0	0	0	0	0	143.5	0	0	0	90.0
Hapt (T3)	0	0	0	69.6	0	0	0	0	0	0	0	151.9
Hapt (T4)	0	128.5	130.1	116.8	0	0	0	0	0	0	0	66.4
Chlo	0	0	0	0	228.9	0	310.9	0	161.3	62.8	254.2	27.6
Syne	0	0	0	0	0	0	0	0	0	77.2	0	108.0
Diat	0	0	147.1	0	0	0	0	0	0	0	0	64.5

obtained from the changes in the ratio of chl *a* to total pigment (assuming that none of the other ratios are grossly inaccurate). The agreement is particularly poor for chlorophytes, which are underestimated, reflecting the fact that the chl *a* ratio has decreased to 26% of the initial value. Despite the poor agreement, the concentrations of all classes tend to follow the correct trend.

In general, this observation was found to apply in nearly all the tests that we ran and indicates that the program is particularly good at predicting relative concentrations within a given phytoplankton class even under conditions where the pigment ratios may not be known with a great deal of certainty. However, in no case where an uncertainty of ±50% was added to the pigment ratio matrix was the program able to satisfactorily reproduce the class abundances.

Sensitivity to random errors in data

When errors are added to the data matrix, there is no longer an exact solution to the problem. With errors of ±10% added to the synthetic HPLC data-set, and errors of ±10% added to the pigment ratio matrix, the program was still able to give a reliable estimate of the class distribution and the final pigment ratio matrix was in reasonable agreement with the 'true' ratios (Table 6a). Even when the errors in the ratio matrix were increased to ±25%, the scatter in the class distribution was of the same order as the errors that were added to the data, i.e. ±10% (Fig. 2), while the calculated pigment ratios were generally within 10 to 20% of the 'true' values (Table 6b). With errors of ±50% added to the ratio matrix, it made little difference to the calculated class distribution whether the data was correct (Fig. 1) or had errors of ±10% added to the data-set (Fig. 2).

The large number of samples (40) chosen in the tests above ensured that the program was able to reproduce the 'true' ratio matrix (Table 5b) and class distribution (Fig. 1), even if there was considerable uncertainty in the starting matrix, provided that there were no errors in the data-set. With the inclusion of errors, we needed to establish the minimum number

of samples in a data-set required before the program could no longer provide a reasonable estimation of the class distribution. This was readily tested by selecting subsets of the data-set corresponding to the analysis of 30, 26, 20, 10 and 5 samples. No significant difference in the distribution of chl *a* between algal classes was noted when the number of samples was reduced to 20. For a sample size of 10, the trends were as expected but the distribution of chl *a* between algal classes showed more scatter than with larger sample sizes.

When the sample size was reduced to 5, the recoveries of class specific chl *a* was unsatisfactory even with an error of only ±10% added to the data-set. The fit was improved by altering the ratio limit matrix so that the program did not allow any pigment ratio to vary by more than 50%. In Fig. 3, we compare the 'true' class distributions with those calculated using all 40 samples and calculated as 8 sets of 5 samples. It is clear that, in this case, 5 samples are insufficient to provide good estimates of class composition. However, it is also clear by comparing Figs. 2 & 3 that for 40 samples the ability of the program to calculate the class composition is more dependent on the errors in the data (±10%) than on the errors in the ratio matrix (±10% or ±25%).

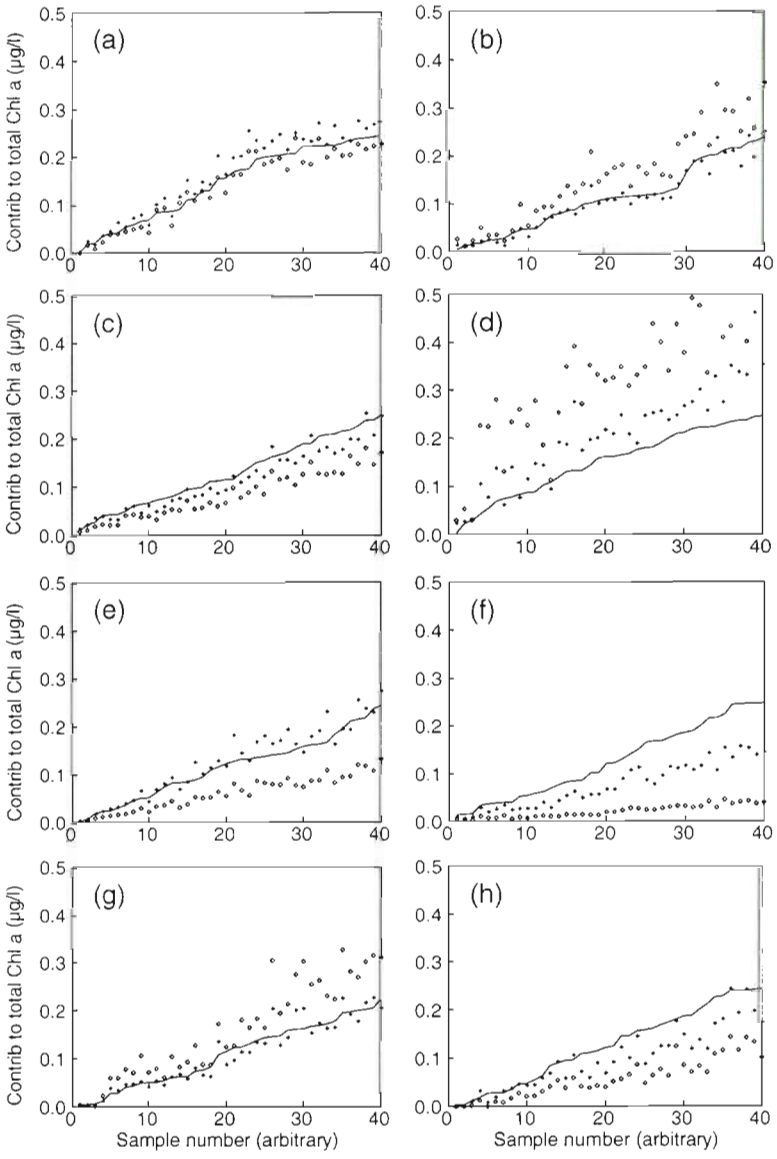


Fig. 2. Contribution to total chl *a*. Plots as in Fig. 1 The calculated values are given for the case where there were random normal standard errors of ±10% added to the data and with random normal standard errors (+) ±25% and (◇) ±50% added to the pigment ratio matrix

Sensitivity to experimental errors in data

Fig. 4 shows the relationship between experimental error and peak area for the experiment on repeated injections of β-*apo*-carotenal. The experimental deviation of the area measurements increased dramatically at smaller peak areas and was very similar for the 2 channels of the detector. At large peak areas (>10^5 µV·s where, for the detector used, 1 V = 1 Absorbance Unit) the standard deviation asymptoted to 1%. In this range, approximately 90% of the standard deviation of replicate injections was accounted for by covariance between the 470 and 435 nm channels, and hence resulted from real differences in the size of the peaks integrated. This 1% error was taken to be the volumetric error from the autoinjector. The remaining error, which reached 100% standard deviation when the peak size was reduced to the limits of detection, was taken to be the quantitative error from the detector and integration. This relationship was used to compute the error appropriate to peaks of different size in the synthetic data-sets.

This simulated estimate of experimental error was generally less than the lowest error of ±10% that was used in previous calculations. When these simulated errors were added to the 'true' synthetic data-set, the program gave excellent agreement between the 'true' and calculated class abundances for all the phytoplankton classes considered (Fig. 5).

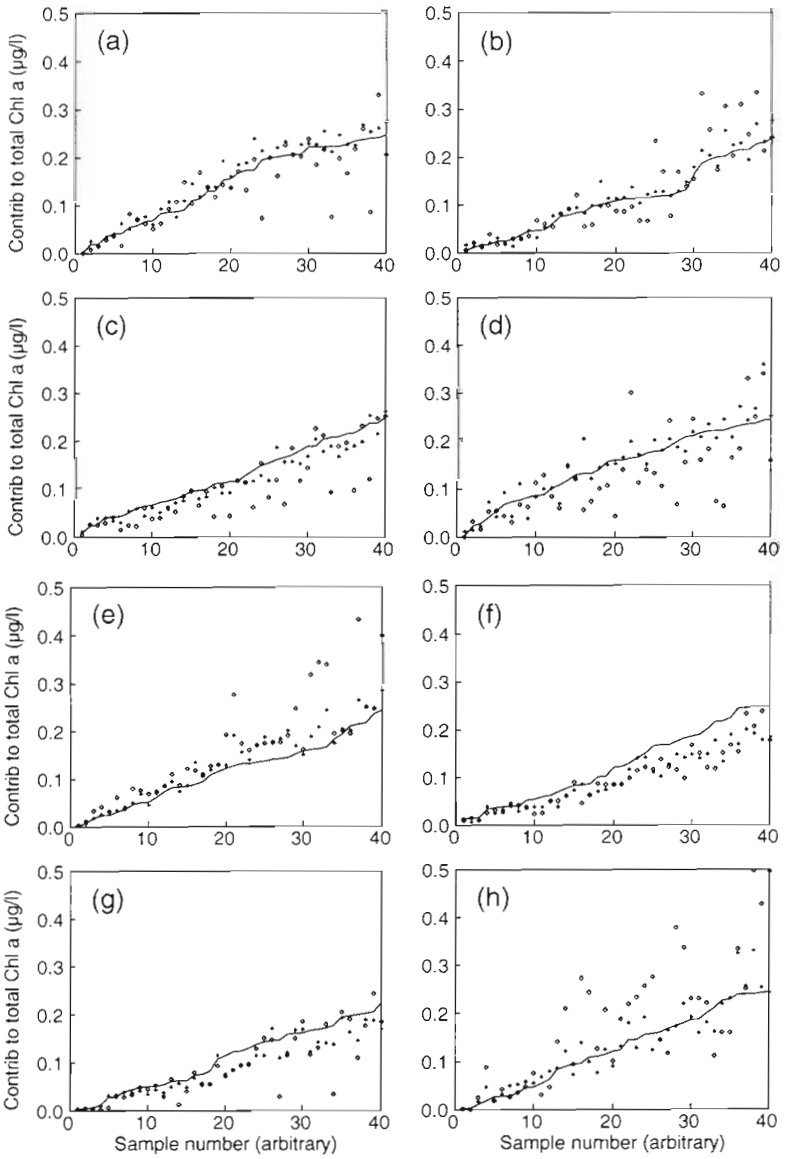


Fig. 3. Contribution to total chl *a*. Plots as in Fig. 1. The calculated values are given for the case where there were random normal standard errors of ±10% added to the data and with random normal standard errors of ±10% added to the pigment ratio matrix. The data-set was analysed with (+) all 40 samples simultaneously and (◇) as 8 groups of 5 samples

included additional classes such as the prochlorophytes. The latter contain divinyl-chl *a* and *b* (instead of chl *a* and *b*) and many HPLC separations are unable to distinguish these chlorophylls from chl *a* and chl *b*, respectively. In order to determine the necessity of separating these compound by HPLC, the class abundances were estimated (1) with the inclusion of divinyl-chl *a* and *b* as separate entities; and (2) by assuming that the divinyl-chl *a* and *b* were included in the determination of chl a and chl *b*, respectively.

The ratio matrix used for constructing the synthetic data-set is given in Table 3. Despite the increased number of classes considered in the Equatorial Pacific data-sets, the ability of CHEMTAX to calculate the class abundances was very similar to its performance with the Southern Ocean data-sets. As before, the analysis of 3 separate synthetic data-sets confirmed that there were no systematic errors introduced. The following comments apply to a single representative data-set.

With the inclusion of divinyl-chl *a* and *b*, with simulated experimental errors

Synthetic data-sets: Equatorial Pacific

After establishing the ability of the program to estimate phytoplankton class abundances for synthetic data-sets chosen to be representative of the Southern Ocean, the whole procedure outlined above was repeated for 3 data-sets representative of waters from the Equatorial Pacific. The Equatorial Pacific data-sets differed from those of the Southern Ocean in that they

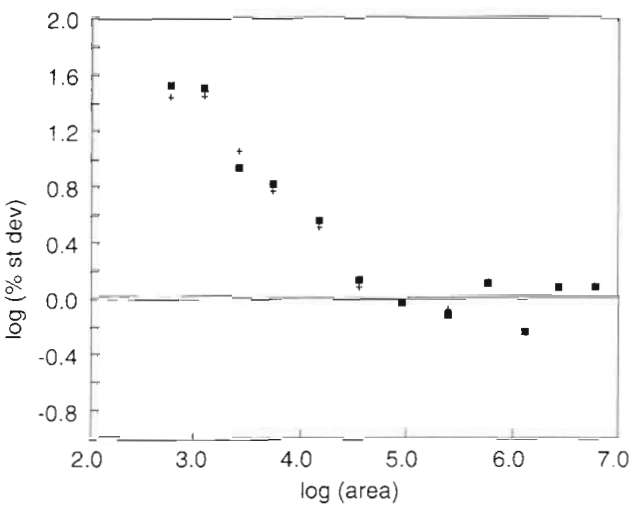


Fig. 4. Plot of log(% standard deviation) for replicate (10) injections of β-*apo*-carotenal as a function of log(peak area) measured at (■) 435 nm and (+) 470 nm. The peak areas are in units of µV·s where, for the detector used, 1 µV = 1 Absorbance Unit

added to the data-set, and ±25% error added to the ratio matrix, there was excellent agreement between the 'true' and calculated abundances for nearly all of the phytoplankton classes (Fig. 6). The calculated abundances were about 15% too low for chrysophytes (Fig. 6e) although the trend was produced very well, and there was some scatter in the fit for euglenophytes (Fig. 6f), *Trichodesmium* (Fig 6j) and diatoms (Fig. 6k). Even more important is the fact that the fit was almost as good when divinyl-chl *a* and *b* were treated as if they were chl *a* and chl *b*, respectively (Fig. 6).

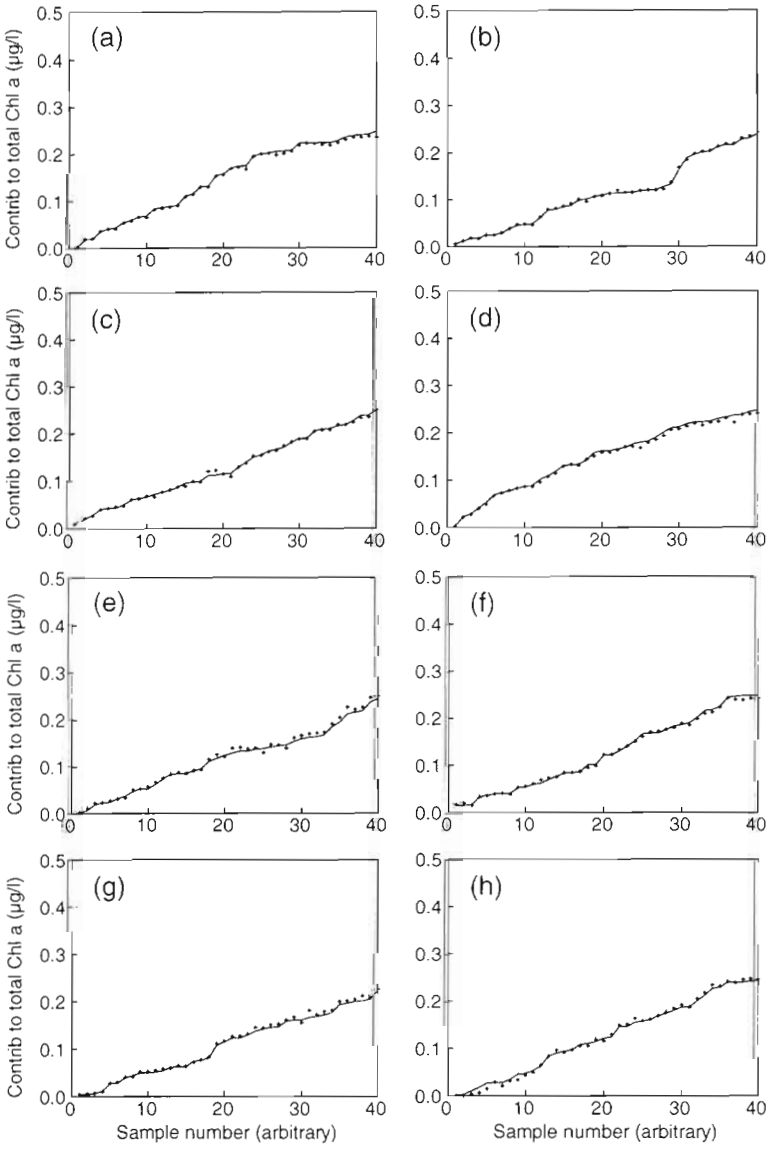


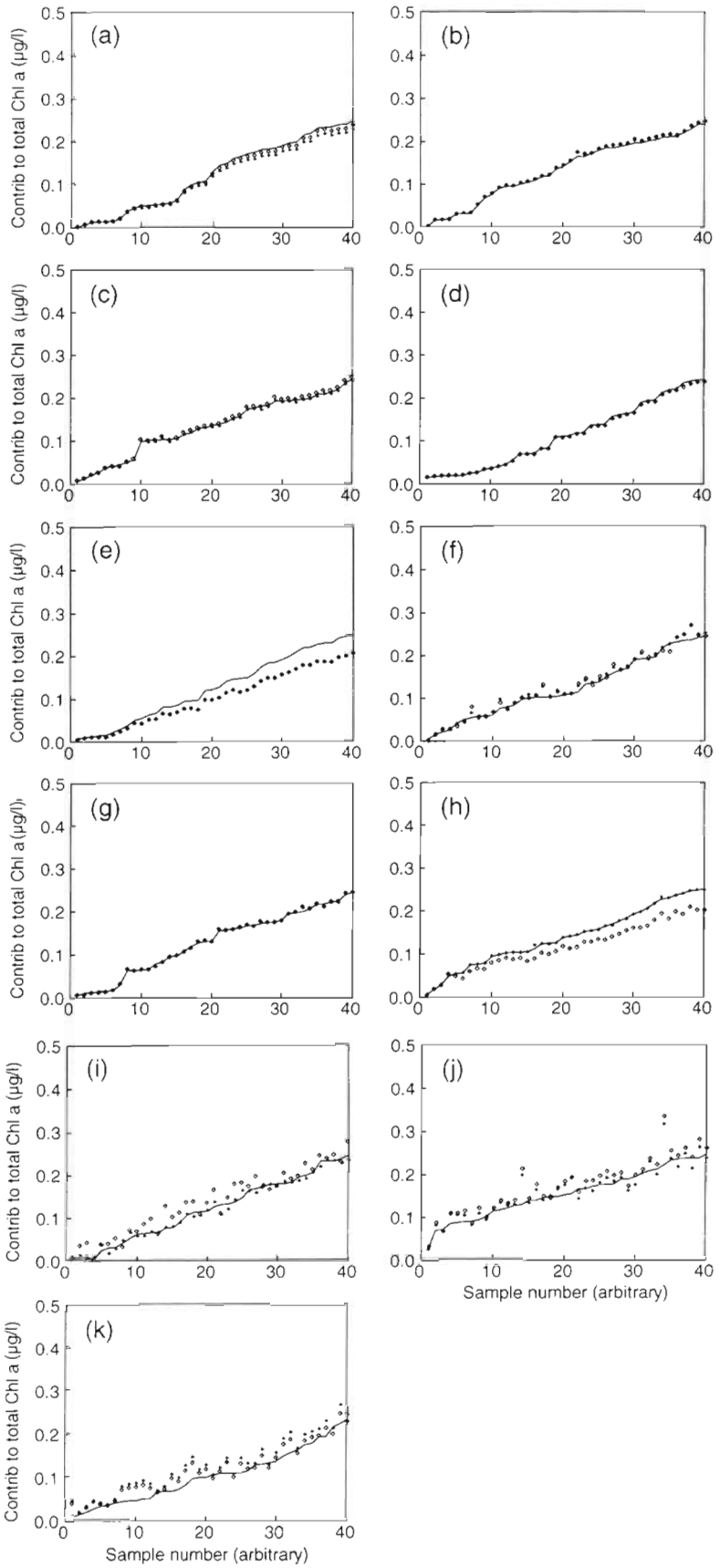
Fig. 5. Contribution to total chl *a*. Plots as in Fig. 1. The calculated values (+) are given for the case where there were simulated experimental errors added to the data and with random normal standard errors of ±25% added to the pigment ratio matrix

DISCUSSION

The most accurate optimisation of class abundances was achieved when all pigment ratios (including chl *a*) were varied. However, this required the longest computational times, which were typically 4.75 h (106 iterations) for the Southern Ocean (Fig. 5) and 9.25 h (89 iterations) for the Equatorial Pacific (Fig. 6) data-sets using a 486/50 PC. To reduce this time, without seriously compromising the optimisation, a small subset of the pigments (usually 5) could be chosen and these varied for a given number of subiterations (again usually 5). The pigments selected were those that caused the largest decrease in the residual.

Although, from a mathematical perspective, it is preferable to have at least 2 pigments in addition to chl *a* for each algal class, it is sometimes not experimentally feasible. In fact, for our Southern Ocean data-set there were 5 algal classes which only had one pigment other than chl *a*. Although we considered chl c_1, c_2 and c_3 and Mg 3,8 DVP, these pigments were not included in the ratio matrix because of poor chromatographic resolution using our HPLC system and a confusing taxonomic distribution at the class level (Jeffrey 1989, Jeffrey & Wright 1994). Diadinoxanthin, although chromatographically well resolved, was not included in the Southern Ocean data-sets since it is widely distributed, is involved in the xanthophyll cycle (Demers et al. 1991) and sample concentrations can vary substantially. Nevertheless, diadinoxanthin was included in the Equatorial Pacific data-sets so as to adequately resolve the additional algal classes and, in particular, the euglenophytes. The pigment β,ε-carotene, while useful from a taxonomic perspective, is generally a very small peak that is not well resolved chromatographically from β,β-carotene. These pigments were not included in the ratio matrix because of the large errors involved in estimating areas of shoulders on HPLC peaks.

Nevertheless, the CHEMTAX program was able to adequately cope with 5 of the 8 algal classes of the Southern Ocean data-set having only 1 extra pigment in addition to chl *a* providing that the initial ratios were not too far away from the 'true' ratio. With more pigments per algal class (say 2



to 4 in addition to chl *a*) there would presumably be more flexibility in the choice of initial ratios.

For the calculation with no errors added to the Southern Ocean data-set and ±25% added to the ratio matrix, the final ratio for lutein in prasinophytes was calculated to be only 84.2% of the value expected (Table 5b) while the fit of chl *a* was good (Fig. 1a) as the program could adequately optimise the remaining 5 pigments. An even better fit was obtained with only ±10% error in the ratio matrix, even though the final ratio for the minor pigment lutein in the prasinophytes was estimated to be as high as 193% of the 'true' value (Table 5a).

For tropical waters, we were surprised that CHEMTAX was able to estimate the abundance of prochlorophytes in the absence of data on the concentrations of divinyl-chl *a* and *b*. This is particularly gratifying for the experimental scientist since these compounds are not usually separated from chl *a* and *b* by HPLC. Prochlorophytes have been shown to contribute up to 35% of carbon biomass in tropical waters (Campbell & Nolla 1994) and, given the size of the Equatorial Pacific, they therefore play a major role in the global carbon cycle.

In this paper, we have only presented the results of a small selection of the many runs that we have used to test the ability of CHEMTAX to calculate the contribution of various phytoplankton classes to the total concentration of chl *a* using

Fig. 6. Contribution to total chl *a* in the synthetic HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class: (a) prasinophyte, (b) dinoflagellate, (c) cryptophyte, (d) haptophyte, (e) chrysophyte, (f) euglenophyte, (g) chlorophyte, (h) prochlorophyte, (i) cyanobacteria (*Synechococcus*), (j) cyanobacteria (*Trichodesmium*) and (k) diatom. The solid line is the 'true' value. The calculated values are given for the case where there were simulated experimental errors added to the data and with random normal standard errors of ±25% added to the pigment ratio matrix. The data-set was analysed with (+) the inclusion of divinyl-chl *a* and *b* as separate entities, and (◇) by assuming that divinyl-chl *a* and *b* were included in the determination of chl *a* and chl *b* respectively

simulated data-sets chosen to represent waters typical of the Southern Ocean and the western Equatorial Pacific.

The CHEMTAX results reported in this paper used data-sets in which all algal classes, on average, contributed equally to the total concentration of chl *a*. For 8 classes, each class contributed, on average, 12.5% of the chl *a* even though individual values ranged from about 0 to 30% (e.g. see Fig. 2). For field samples, some classes would always be expected to be minor or major components of the total phytoplankton population. We, therefore, constructed several other data-sets in which the average weighting of the various classes was changed to 5 or 33.3% of the total chl *a* and tested these as described above. In all cases, the behaviour of CHEMTAX was similar to the data-sets where the average class weighting was equal.

As more data become available on species compositions of different water masses, pigment compositions of algal species and pigment ratios for cultured and wild species, we should be able to continually improve our initial estimates of the pigment ratio matrix for CHEMTAX. Nevertheless, it must be remembered that pigment ratios may vary for any phytoplankton species within a given data-set due to differences in light regimes, nutrient concentrations, physiological status, etc. If enough samples are available, the data-set should be divided into more homogeneous subsets. In particular, samples from different depths should be analysed separately since the pigment concentrations of individual cells are known to be strongly dependent on ambient light intensity.

Even if the pigment ratio matrix were constant for a given set of samples and even if the ratios were known exactly, there would be no unique solution to the general problem of calculating class abundances since there will always be experimental errors in the HPLC data-set. Our calculations suggest that these errors can be more important than occasional, much larger, uncertainties in pigment ratios. While we have no control over the natural variability in pigment ratios, we do have some control over the way we collect the experimental data and it is essential to minimise the errors involved in the HPLC analyses.

We determined the conditions under which CHEMTAX can calculate class abundances for synthetic samples selected to represent typical waters of the western Equatorial Pacific and the Southern Ocean. If other classes, or unusual pigment ratios, were suspected to be important, it would be a simple matter to modify the relevant ratio matrix and construct synthetic data-sets to study whether these changes led to computational problems. For use in other waters, it would be straightforward to set up appropriate synthetic data-sets to assess the performance of the program.

CONCLUSIONS

The program CHEMTAX has been tested with synthetic data-sets representative of samples taken from the Equatorial Pacific and the Southern Ocean. These synthetic data-sets have identified some potential problems that may occur but, in general, have shown that the program can successfully calculate phytoplankton class abundances from HPLC chromatograms of chlorophyll and carotenoid pigments. This is possible for the algal class prochlorophyta, even in the absence of measurements of its major pigments divinyl-chl *a* and *b*. This is particularly significant since prochlorophytes are suspected of being widely abundant and are difficult to count using conventional methods.

It is also notable that good fits were obtained in the absence of other major pigments such as chl c_1, c_2, c_3 and the many other related pigments that are being identified as improved chromatographic techniques become available. As more data become available for the abundances of these and other carotenoid pigments, programs such as CHEMTAX should be able to provide ever more reliable estimates of the phytoplankton class abundances from a wide range of water bodies including freshwater systems.

The procedure described in this paper is general and can therefore be used to calculate the abundances of any other classes of organism where there are sufficient specific chemical marker compounds. While this paper has discussed only photosynthetic marker compounds that are quantitated by HPLC, there are obviously many more chemical markers that have been characterised by HPLC and, particularly, GC. Suitable candidates would include compounds such as fatty acids, sterols, amino acids and hydrocarbons.

The CHEMTAX program described in this paper can be run on any PC, Macintosh or UNIX based workstation that has access to MATLAB software. The program PREPRO, which constructs the matrices used by CHEMTAX, is a DOS based program written for a PC. However, the relevant matrices can also be constructed as an ASCII file using any text editor. The software is available from D.J.M. and enquiries should be sent to the e-mail or postal address given at the head of the article.

Acknowledgements. We thank W. de la Mare (Australian Antarctic Division) for suggestions for the use of experimental errors and provision of the random normal distribution subroutine and S. W Jeffrey, J K. Volkman and R. F. C. Mantoura for helpful discussions.

Appendix 1

An alternative approach to the problem of obtaining reasonable pigment ratios and algal class abundances, involving factor analysis techniques, was also investigated. Initially, the weighted data matrix $S' = SW$ was factorised into 2 matrices $\hat{C}$ and $\hat{F}$. Although any arbitrary factorisation could have been used, in this case the singular value decomposition was used for ease of data analysis.

$$S' = (U\Lambda)V^{T}$$
$$= \hat{C}\hat{F}$$
$$\text{i.e. } S = \hat{C}\hat{F}W^{-1}$$

where W is chosen so that the elements of S' have approximately equal variance. From this initial factorisation a new factorisation was sought using an arbitrary transformation matrix T, to give

$$S = (\hat{C}T^{-1})(T\hat{F}W^{-1})$$

Choosing T to minimize $\|T\hat{F}W^{-1} - F_0\|$ subject to the conditions

$$\sum_i [T\hat{F}W^{-1}]_{ij} = 1 \qquad \forall\ i \qquad (1)$$
$$[\hat{C}T^{-1}]_{ij} \geq 0 \qquad \forall\ i, j \qquad (2)$$
$$[T\hat{F}W^{-1}]_{ij} \geq 0 \qquad \forall\ i, j \qquad (3)$$

gave the estimates $\hat{C} = \hat{C}T^{-1}$ and $\hat{F} = T\hat{F}W^{-1}$. Note that since T is not necessarily square, T^{-1} denotes the Moore-Penrose pseudoinverse. See Menke (1984) for a fuller discussion.

This procedure finds the matrices $\hat{F}$ and $\hat{C}$ with $\hat{F}$ closest to F_0, such that the pigment ratios are positive and normalised and the phytoplankton class abundances are non-negative. In practise, matrix T was evaluated by solving the weighted least squares equation

$$W_e\hat{F}(W^{-1})^{T}T^{T} = W_eF_0^{T}$$

subject to constraints (1) and (3) above. The weighting matrix W_e was chosen so that the nonzero elements of F_0 were of approximately equal weight in the calculation, regardless of absolute magnitude.

This approach had several drawbacks. The first was that the number of data samples was required to be greater than or equal to the number of classes used in the calculation, and that all these samples were assumed to have the same pigment ratios. Unsurprisingly, since the data matrix S was usually composed of sets of measurements taken in near-identical conditions, it was usually near-singular which adversely affected the robustness of the solution. *R*-mode analysis (factor analysis of the deviations of the data from the mean) could not be applied in this case.

The second drawback was due to the fact that constraint (2) above was not implemented. This constraint is nonlinear in the elements of T and proved extremely difficult to include in the calculations. Without this constraint, the factor loading matrix $\hat{C}$ obtained was sometimes physically unrealistic, giving negative or overly large phytoplankton abundances. Several approaches, including transformation of variables, singular value analysis and various weighting schemes were attempted in order to alleviate this problem, but were unsuccessful. Reasonable abundances were sometimes obtained for the major classes present in the samples, but the abundances obtained for the minor classes were often clearly unrealistic. However, if techniques were developed to allow the inclusion of constraint (2) into the calculation, then this factor analysis method would be preferable to the iterative least squares solution, both because it is guaranteed to give the best solution and because it is much faster to calculate.

LITERATURE CITED

Andersen RA, Bidigare RR, Keller MD, Latasa M (1996) A comparison of HPLC signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. Deep Sea Res 43:517–537

Andersen RA, Saunders GW, Paskind MD, Sexton JP (1993) Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the Pelagophyceae classis nov. J Phycol 29:701–715

Ben-Amotz A, Katz A, Avron M (1982) Accumulation of β-carotene in halotolerant alga: purification and characterization of β-carotene-rich globules from *Dunaliella bardawil* (Chlorophyceae). J Phycol 18:529–537

Berger R, Liaaen-Jensen S, McAlister V, Guillard RRL (1977) Carotenoids of Prymnesiophyceae (Haptophyceae). Biochem Syst Ecol 5:71–75

Bjornland T, Tangen K (1979) Pigmentation and morphology of a marine *Gyrodinium* (Dinophyceae) with a major carotenoid different from peridinin and fucoxanthin. J Phycol 15:457–463

Burczyk J, Szkawran H, Zontek I, Czygan FCh (1981) Carotenoids in the outer cell-wall layer of *Scenedesmus* (Chlorophyceae). Planta 151:247–250

Burger-Wiersma R, Veenhuis M, Korthals HJ, Van de Wiel CCM, Mur LR (1986) A new prokaryote containing chlorophyll *a* and *b*. Nature 320:262–264

Bustillos-Guzman J, Claustre H, Marty JC (1995) Specific phytoplankton signatures and their relationships to hydrographic conditions in the coastal northwestern Mediterranean. Mar Ecol Prog Ser 124:247–258

Campbell L, Nolla HA (1994) The importance of *Prochlorococcuus* to community structure in the central North Pacific Ocean. Limnol Oceanogr 39:954–961

Carpenter EJ, O'Neil JM, Dawson R, Capone DG, Siddiqui DJA, Orenneberg T, Bergnan B (1993) The tropical diazotrophic phytoplankter *Trichodesmium*: biological characteristics of two common species. Mar Ecol Prog Ser 95: 295–304

Chavez FP, Buck KR, Barber RT (1990) Phytoplankton taxa in relation to primary production in the equatorial Pacific. Deep Sea Res 37:1733–1752

Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature 334: 340–343

Demers S, Roy S, Gagnon R, Vignault C (1991) Rapid light-induced changes in cell fluorescence and in xanthophyll-cycle pigments of *Alexandrium excavatum* (Dinophyceae) and *Thalassiosira pseudonana* (Bacillariophyceae): a photoprotection mechanism. Mar Ecol Prog Ser 76:185–193

Everitt DA, Wright SW, Volkman JK, Thomas DP, Lindstrom E (1990) Phytoplankton community compositions in the western equatorial Pacific determined from chlorophyll and carotenoid pigment distributions. Deep Sea Res 37: 975–997

Fawley MW (1992) Photosynthetic pigments of *Pseudocourfielda marin* and selected green flagellates and coccoid ultraplankton: implications for the systematics of the Micromonadophycaea (Chlorophyta). J Phycol 28:26–31

Gieskes WW, Kraay GW (1983) Dominance of Cryptophyceae during the phytoplankton spring bloom in the central North Sea detected by HPLC analysis of pigments. Mar Biol 75:179–185

Gieskes WW, Kraay GW (1986) Floristic and physiological differences between the shallow and the deep nanoplankton community in the euphotic zone of the open tropical Atlantic revealed by HPLC analysis of pigments. Mar Biol 91:567–576

Gieskes WWC, Kraay GW, Nontji A, Setiapermana D, Sutmono (1988) Monsoonal alteration of a mixed and a layered structure in the phytoplankton of the euphotic zone of the Banda Sea (Indonesia): a mathematical analysis of algal fingerprints. Neth J Sea Res 22:123–137

Hager A, Stransky H (1970a) The carotenoid pattern and the occurrence of the light induced xanthophyll cycle in various classes of algae V. A few members of Cryptophyceae, Euglenophyceae, Bacilliarophyceae, Chrysophyceae, and Phaeophyceae. Arch Mikrobiol 73:77–89

Hager A, Stransky H (1970b) The carotenoid pattern and the occurrence of the light induced xanthophyll cycle in various classes of algae III. Green algae. Arch Mikrobiol 72: 68–83

Hallegraeff GM, Jeffrey SW (1984) Tropical phytoplankton species and pigments of continental shelf waters of North and North-West of Australia. Mar Ecol Prog Ser 20:59–74

Hooks CE, Bidigare RR, Keller MD, Guillard RRL (1988) Coccoid eukaryotic marine ultraplankters with four different HPLC pigment signatures. J Phycol 24:571–580

Iturriaga R, Mitchell BG (1986) Chroococcoid cyanobacteria: a significant component in the food web dynamics of the open ocean. Mar Ecol Prog Ser 28:291–297

Jeffrey SW (1989) Chlorophyll *c* pigments and their distribution in the chromophyte algae. In: Green JC, Leadbeater BSC, Diver WL (eds) The chromophyte algae: problems and perspectives. Clarendon Press, Oxford, p 11–36

Jeffrey SW, Hallegraeff GM (1980) Studies of phytoplankton species and photosynthetic pigments in a warm core eddy of the East Australian Current. I. Summer populations. Mar Ecol Prog Ser 3:285–294

Jeffrey SW, Hallegraeff GM (1987) Phytoplankton pigments, species and light climate in a complex warm-core eddy of the East Australian Current. Deep Sea Res 34:649–673

Jeffrey SW, Sielicki M, Haxo FT (1975) Chloroplast pigment patterns in the dinoflagellates. J Phycol 11:374–384

Jeffrey SW, Wright SW (1994) Photosynthetic pigments in the Haptophyceae. In: Green JC, Leadbeater BSC (eds) The haptophyte algae. Syst Assoc Special Vol 51, Clarendon Press, Oxford, p 111–132

Jeffrey SW, Wright SW (in press) Quantitative analysis of SCOR reference algal cultures. In: Jeffrey SW, Mantoura RFC, Wright SW (eds) Phytoplankton pigments in oceanography: guidelines to modern methods. SCOR-UNESCO, Paris

Klein B, Sournia A (1987) A daily study of the diatom spring bloom at Roscoff (France) in 1985. II. Phytoplankton pigment composition studied by HPLC analysis. Mar Ecol Prog Ser 37:265–275

Lawson CL, Hanson RJ (1974) Solving least square problems. Prentice-Hall, Englewood Cliffs, NJ

Letelier RM, Bidigare RR, Hebel DV, Ondrusek M, Winn CD, Carl DM (1993) Temporal variability of phytoplankton community structure based on pigment analysis. Limnol Oceanogr 38:1420–1437

Li WKW, Subba-Rao DV, Harrison WG, Smith JC, Cullen JJ, Irwin B, Platt T (1983) Autotrophic picoplankton in the tropical ocean. Science 219:292–295

Menke W (1984) Geophysical data analysis: discrete inverse theory. Academic Press, Orlando, FL

Platt T, Subba Rao DV, Irwin B (1983) Photosynthesis of picoplankton in the oligotrophic ocean. Nature 301: 702–704

Ricketts TR (1967) Further investigations into the pigment composition of green flagellates possessing scaly flagella. Phytochemistry (Oxf) 6:1375–1386

Ricketts TR (1970) The pigments of Prasinophyceae and related organisms. Phytochemistry (Oxf) 9:1835–1842

Ridout PS, Morris RJ (1985) Short-term variations in the pigment composition of a spring phytoplankton bloom from an enclosed experimental ecosystem. Mar Biol 87:7–11

Simon N, Barlow RG, Marie D, Partensky F, Vaulot D (1994) Characterization of oceanic photosynthetic picoeukaryotes by flow cytometry. J Phycol 30:922–935

Stauber JL, Jeffrey SW (1989) Photosynthetic pigments in 51 species of marine diatoms. J Phycol 24:158–172

Stransky H, Hager A (1970) The carotenoid pattern and the occurrence of the light induced xanthophyll cycle in various classes of algae IV. Cyanophyceae and Rhodophyceae. Arch Mikrobiol 72:84–96

Tester PA, Geesey ME, Guo C, Pearl HW, Millie DF (1995) Evaluating phytoplankton dynamics in the Newport River estuary (North Carolina, USA) by HPLC-derived pigment proliles. Mar Ecol Prog Ser 124:237–245

Wright SW, Thomas DP, Marchant HJ, Higgins HW, Mackey MD, Mackey DJ (1996) Analysis of phytoplankton of the Australian sector of the Southern Ocean: comparisons of microscopy and size frequency data with interpretations of pigment HPLC data using the 'CHEMTAX' matrix factorisation program. Mar Ecol Prog Ser 144:285–298

Wilhelm C, Lenarz-Weiler I (1987) Energy transfer and pigment composition in three chlorophyll *b*-containing light-harvesting complexes isolated from *Mantoniella squamata* (Prasinophyceae), *Chlorella fusca* (Chlorophyceae) and *Sinapis alba*. Photosynth Res 13:101–111

Zelen M, Severo NC (1970) Probability functions. In: Abramowitz M, Stegun IA (eds) Handbook of mathematical functions. Dover Publications, New York, p 925–995

This article was submitted to the editor

Manuscript first received: April 23, 1996
Revised version accepted: September 9, 1996