CSIRO MARINE LABORATORIES
Report 229

CHEMTAX User's Manual: a program for estimating class abundances from chemical markers – application to HPLC measurements of phytoplankton pigments

Mark D. Mackey, Harry W. Higgins, Denis J. Mackey, Simon W. Wright

1997

CSIRO Marine Laboratories
Report 229

CHEMTAX user's manual: a program for estimating class abundances from chemical markers — application to HPLC measurements of phytoplankton pigments

Mark D. Mackey[a,b], Harry W. Higgins[a], Denis J. Mackey[a], Simon W. Wright[c]

[a] CSIRO Division of Marine Research, GPO Box 1538, Hobart, Tasmania 7001, Australia
[b] University Chemical Laboratory, Lensfield Rd, Cambridge CB2 1EW, England
[c] Australian Antarctic Division, Channel Highway, Kingston, Tasmania 7050, Australia

Abstract / Scope of Document

This User's Manual describes the use of a new computer program, CHEMTAX (CHEMical TAXonomy), for calculating class abundances of any type of organism which has a sufficient number of suitable chemical marker compounds. Discussion is limited to the application of CHEMTAX to calculating algal class abundances from measurements of chlorophyll and carotenoid pigments determined by 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.

Guidelines for successfully using the CHEMTAX program are discussed and are illustrated by the results of our extensive testing of the program 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 of the pigment ratios for each class and, 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*.

National Library of Australia Cataloguing-in-Publication

CHEMTAX user's manual : a program for estimating class abundances from chemical markers — application to HPLC measurements of phytoplankton pigments.

ISBN 0 643 06040 5.

1. CHEMTAX (computer program).
2. Chemotaxonomy — Computer programs.
3. Chemotaxonomy — Laboratory manuals.
4. High performance liquid chromatography.
5. Pigments (Biology) — Measurement.
I. Mackey, Mark D.
II. CSIRO. Marine Laboratories (Series : Report (CSIRO. Marine Laboratories); 229).

578.0120285

This document or the parent paper entitled *CHEMTAX — A program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton* (M. D. Mackey, D. J. Mackey, H. W. Higgins and S. W. Wright, *Marine Ecology Progress Series*, **144**, 265-283, 1996) must be properly cited in any publication reporting results derived from this manual or parent paper, or obtained from the application and use of the software described in this document. The software is available from DJM and enquiries should be sent to Denis.Mackey@marine.csiro.au or the address on the title page.

This manual should be cited as: M. D. Mackey, H. W. Higgins, D. J. Mackey and S. W. Wright, 1997, *CHEMTAX user's manual: a program for estimating class abundances from chemical markers — application to HPLC measurements of phytoplankton pigments*, CSIRO Marine Laboratories Report 229, Hobart, Australia, ISBN 0 643 06040 5, 47 pp.

March 1997

Software Licence Agreement

1.0 Grant of Licence

1.1 The CSIRO Division of Oceanography (herein referred to as 'CSIRO') hereby grants you (hereinafter referred to as the 'Licencee'), subject to the Licence agreeing to comply with the terms and conditions of this Agreement, a non-transferable, non-exclusive licence to use the computer programs described in this document (hereinafter referred to as the 'Software') for the purpose of the Licencee's computing activity.

1.2 CSIRO hereby grants the Licencee the right to make copies of the Software for the purpose of the Licencee's computing activity only.

1.3 The benefit of the rights granted to the Licencee by the Licence and this Agreement generally shall be personal to the Licencee and the Licencee shall not mortgage, charge, assign, rent, lease, sell or otherwise dispose of or transfer the same or any part to any third party.

1.4 Unless otherwise agreed in writing or provided for in this Agreement, CSIRO shall be under no obligation or responsibility to provide the Licencee with any training, maintenance services, enhancements or updates of the Software or any services whatsoever.

2.0 Acknowledgment by the Licencee

2.1 The Licencee acknowledges and agrees that the Licencee shall not:

(i) sell, let for hire or by way of trade, offer or exhibit or expose for sale or hire or otherwise distribute the Software for the purposes of trade or any other purpose;

(ii) authorise or assist any third person to do any of the acts set out in (i) above;

(iii) modify the Software source code without advising CSIRO.

2.2 The Licencee agrees that:

(a) CSIRO is the owner of all copyright and other Intellectual Property Rights subsisting in the Software;

(b) this document must be properly cited in any publication reporting results derived from this document or obtained from application and use of this software. Any of the Licencee's documentation describing results generated by the Licencee's use of the Software will contain an acknowledgement of CSIRO's ownership of the Software;

(c) CSIRO reserves all rights in the Software other than the rights granted to the Licencee by this Agreement;

(d) each item of the Software will display a banner summarising the terms of this Agreement and acknowledging the source of the Software, and the contents of a banner will not be modified and its display will not be inactivated by the Licencee without the approval of CSIRO.

3.0 Indemnity

3.1 To the full extent permitted by law, CSIRO excludes any and all liability in respect of any loss or damage, whether personal (includes death or illness) or of property and whether direct, consequential or special (including consequential financial loss or damage) of the Licencee, its officers, agents and employees or any third party howsoever caused, which may be suffered or incurred or which may arise directly or indirectly in respect of or arising out of the Licencee's use or inability to use the Software or the failure or omission on the part of CSIRO to comply with the conditions and warranties under this Licence Agreement. Insofar as liability for loss or damages

under or pursuant to such legislation cannot be excluded, CSIRO's liability for loss or damages shall be limited to the amount of One Dollar ($1.00).

3.2 CSIRO makes no warranties, expressed or implied, and excludes all other warranties representations, terms or conditions, whether express or implied, oral or written, statutory or otherwise, relating in any way to the Software, or to this Agreement, including any implied warranty of merchantability or of fitness for particular purpose. To the full extent permitted by the law of the Commonwealth of Australia or the laws of any State or Territory of Australia, any conditions or warranties imposed by such legislation are hereby excluded. In so far as liability under or pursuant to such legislation may not be excluded, CSIRO's liability to the Licencee pursuant to this Agreement shall be limited as set out in clause 3.1 hereof.

3.3 The Licencee acknowledges and agrees that the Software was developed for CSIRO's research purposes and may have inherent defects, errors or deficiencies, and that it is the responsibility of the Licencee to make its own assessment of the suitability of the Software for the purpose of the Licencee's computing activity. The Licencee will use the Software, and advice, opinions or information supplied by CSIRO, its officers, employees or agents concerning the Software at the Licencee's own risk.

3.4 The Licencee hereby releases and indemnifies and shall continue to release and indemnify CSIRO, its officers, employees and agents from and against all actions, claims, proceedings or demands (including those brought by third parties) which may be brought against it or them, whether on their own or jointly with the Licencee and whether at common law, in equity or pursuant to statute or otherwise, in respect of any loss, death, injury, illness or damage (whether personal or property, and whether direct or consequential, including consequential financial loss) and any infringement of copyright, patents, trade marks, designs or other Intellectual Property Rights, howsoever arising out of the Licencee's exercise of its rights under this Agreement and from and against all damages, costs and expenses incurred in defending or settling any such claim, proceeding or demand.

3.5 The Licencee's obligation to indemnify CSIRO and its officers, employees and agents set out in clause 3.4 hereof is a continuing obligation separate from and independent of the Licencee's other obligations under this Agreement, and shall survive all expiration or termination of this Agreement.

4.0 Termination

4.1 The Licence shall terminate immediately upon the Licencee breaching any term or condition of this Agreement whether or not CSIRO is aware of the occurrence of the breach at the time that it happens.

4.2 CSIRO may terminate the Licence on reasonable grounds by notice in writing to the Licencee, and such notice of termination shall be effective immediately upon receipt by the Licencee.

Table of contents

1 Introduction

1.1 Applicability of CHEMTAX approach

The CHEMTAX (CHEMical TAXonomy) program described in this User's Manual can be used to calculate the class abundances of any type of organism which has a sufficient number of suitable chemical marker compounds. While the range of types of organisms and chemical markers which may be suitable for this type of analysis is diverse, obvious candidates include compounds that can be characterised by HPLC and GC such as pigments, fatty acids, sterols, amino acids and hydrocarbons. In this manual, however, we will restrict our discussion to the algae or phytoplankton and their pigments (both photosynthetic and photoprotective).

1.2 Pigments and algal taxonomy

The taxonomy of algae is dynamic and being continually revised (Andersen *et al.*, 1993; Fawley, 1992; Hooks *et al.*, 1998; Jeffrey and Wright, 1994; Simon *et al.*, 1994; Stauber and Jeffrey, 1989; Vaulot *et al.*, 1994; Wright and Jeffrey, 1987), and is usually based on a combination of a whole series of parameters including morphology and DNA/RNA ratios as well as pigments. Each of these complimentary techniques gives useful but often slightly different information.

Microscopy enables identification (via morphology and fluorescence characteristics) as well as direct cell counts for most algal classes (e.g. Chavez *et al.*, 1990) while flow cytometry provides better statistics and some data on cell size and pigment content but for a more limited range of algal classes (Simon *et al.*, 1994; Vaulot *et al.*, 1994).

HPLC pigment analysis provides quantitative data on 40-50 pigments from all algal classes (Wright *et al.*, 1991) including picoplankton and fragile cell types (Iturriaga and Mitchell, 1986; Li *et al.*, 1983; Platt *et al.*, 1983) which may be difficult to identify and count by microscopy or flow cytometry. The computer program CHEMTAX allows us to apportion the pigments between the various algal classes.

The biomass (carbon content) of the various algal classes can be estimated using these complimentary techniques if we have information on cellular pigment and carbon content as a function of physiological status and environmental parameters such as light and nutrient concentrations.

However, for the purposes of this discussion on CHEMTAX, algal taxonomy and quantitation will be based solely on the pigment groups specified and, in all but a few cases, this is completely adequate. Nevertheless, in certain situations phytoplankton from a number of taxonomic classes may be included in one pigment grouping. For example, a number of prasinophytes are indistinguishable from chlorophytes on the basis of pigments alone (Fawley, 1992; Ricketts, 1970) and hence the pigment contribution from these prasinophytes will be attributed to the 'chlorophyte' pigment class.

1.3 Limitations of previous pigment approaches

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 HPLC analytical

techniques. Analysis of marine ecosystems using pigments has generally been qualitative (Gieskes and Kraay, 1986; Hallegraeff and Jeffrey, 1984; Jeffrey and Hallegraeff, 1980, 1987; Klein and Sournia, 1987; Ridout and Morris, 1985), but more recently there have been attempts to estimate the abundances of various phytoplanktonic classes quantitatively from marker pigment concentrations (Andersen *et al.*, 1996, Bustillos-Guzman *et al.*, 1995; Everitt *et al.*, 1990; Gieskes and Kraay, 1983 and 1986; Gieskes *et al.*, 1988; Letelier *et al.*, 1993; Tester *et al.*, 1995).

While each of these quantitative approaches do indeed give useful information they also suffer from some quite severe limitations. Although Gieskes *et al.* (1988) could establish the contributions to the population from different pigment related groups they could not, with fucoxanthin for instance, differentiate between the diatoms, chrysophytes (some types of chrysophytes have also been called pelagophytes; Andersen *et al.*, 1993, 1996) or the haptophytes (prymnesiophytes) which may have contributed to this fucoxanthin.

Using an iterative method, Everitt *et al.* (1990) calculated the contribution of each algal class to the total chlorophyll *a* (chl *a*). The abundances of classes without unique marker pigments were calculated by difference. 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) and Andersen *et al.* (1996) used a method based on a least-squares solution of an overdetermined linear problem. Some classes were calculated by difference, which could lead to negative chl *a* values, and the method does not seem to provide any way of optimising the auxiliary pigment ratios, nor is it possible to 'weight' the pigment data to allow for different measurement errors in determining the individual pigments. Similar approaches were used by Tester *et al.* (1995) and Bustillos-Guzman *et al.* (1995). For a more detailed discussion of these previous attempts and drawbacks see Mackey *et al.* (1996).

In this Users Manual we describe the use of a new method, CHEMTAX, 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 also described.

An alternative computational approach using factor analysis was not successful. It is described in Mackey *et al.* (1996) in the hope that others may find some way of overcoming the difficulties encountered as the method has the potential of being much faster and is guaranteed to give the correct answer.

2 Methods

2.1 Description of the CHEMTAX program

The aim of the method described in this manual 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 estimate of F, F_0, was made from literature values for pigment concentrations in various species (see Table 1). 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:

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

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

$$\Sigma[\hat{C}_0]_{i,j} = 1 \quad \forall j$$

The method outlined in Lawson and Hanson 1974 (least squares regression with inequality and equality constraints) was used to solve this equation, and the residual 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 10%) 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 $F_0, F_1, F_2, \ldots$ with corresponding $\hat{C}_0, \hat{C}_1, \hat{C}_2, \ldots$ were determined, with $\{\varepsilon_0\} = \| S - \hat{C}_0 F_0 \|$ 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, then the amount of variation on each step was reduced and the minimisation process continued.

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 Southern Ocean and 9.25 h (89 iterations) for Equatorial data sets using a 486/50 PC.

In practice it was found that variation of most of the non-zero pigment ratios (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 (usually 5) of the elements of F_i, which caused the largest decrease in the residual, was selected to be varied for a number of iterations (again

Table 1

Pigment ratios[1] for use in algal class composition calculations

		Chl c	Chlc3	Chlc1	Chlc2	MgDV	SIPX	PERI	BUT	FUCO	HEX	NEO	PRAS	MYXO	DINO	VIOL	DDX	ANTH	ALLO	DIAT	LUT	ZEA	Chl b	BεCA	BεCA	ββCA	Ref
CYANOBACTERIA																											
Synechococcus																											a,b,p
	min [1]	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,i,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 [2]																											a,b,c,d
	Type 1																										
	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																											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 [3]																											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	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	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	
BACILLIAROPHYCEAE																											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																											a,b,i,r,s
	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 [4]																										
	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	

Notes:

(1) The two 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.

(2) 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 - unpublished results) and we consider these prasinophytes as one 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.

(3) The Haptophytes (Prymnesiophytes) are sub-divided into 4 pigment groupings according to Jeffrey and Wright (1994). Type 1: FUCO but no Chl*c3*; Type 2: FUCO and Chl*c3*; Type 3: FUCO, HEX and Chl*c3*; Type 4: BUT and Chl*c3* with or without FUCO and HEX but with the ratio 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 one pigment group (e.g. representatives of *Phaeocystis* may be found in both pigment Types 2 and 4).

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

References:

(a) Jeffrey and Wright (1997), (b) Simon W. Wright (unpublished data), (c) Ricketts (1967) (d) Wilhelm *et al.* (1987), (e) Bjornland and Tangen (1979), (f) Jeffrey *et al.* (1975) (g) Berger *et al.* (1977) (h) Jeffrey and Wright (1994), (i) Hager and Stransky (1970a) (j) Ben-Amotz *et al.* (1982), (k) Burczyk *et al.* (1981) (l) Hager and Stransky (1970b), (m) Fawley (1992) (n) Burger-Wiersma *et al.* (1986), (o) Stauber and Jeffrey (1989), (p) Stransky and Hager (1970), (q) Carpenter *et al.* (1993), (r) Vesk & Jeffrey (1987), (s) Andersen *et al.* (1993).

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 4.

usually 5). All the elements were then varied in order to select a new subset for downhill following (the pigments in this new subset are likely to be different from the previous subset as a consequence of the continually decreasing residual during the iteration process). This procedure was 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. $\hat{C}_i$ 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.

2.2 Hardware and Software requirements

CHEMTAX is a Matlab™ (ver. 4.2) program developed 'in house' to carry out these calculations. Matlab™ is a Mathworks Pty. Ltd. product (htpp:\\www.mathworks.com) for UNIX, PC (Windows™ 3.1+ or LINUX) and Power Macintosh platforms.

The input file for the CHEMTAX Matlab™ program is generated by a preprocessor, PREPRO (see next section) written for the PC in Pascal™.

2.3 PREPRO

2.3.1 Introduction

The CHEMTAX program takes as input a Matlab™ .m file which sets a number of variables required in the calculation process and also contains the data matrices. This input is simply an ASCII text file which may be generated or edited as required by any text editor (See Appendix A for a description of the contents of this file). To simplify data entry, however, an interactive program, PREPRO, is available.

PREPRO is a DOS program which will run on any 286 or later IBM-compatible PC under DOS 3.1 or later. It will also run under Windows 3.1 or Windows 95.

The program takes as input several ASCII files containing the data on which calculations are to be made, and outputs a Matlab™ .m-file suitable for input into the CHEMTAX program.

2.3.2 Overview

The interface to PREPRO is straight forward. At any stage the available keystrokes are summarised at the bottom of the screen, and help on the currently selected command may be obtained by pressing F_1.

The use of PREPRO can be divided into three stages: loading and examining the data files, setting the calculation options, and writing the output file. These stages are summarised below.

2.3.3 Loading the data files

PREPRO requires two data files, one containing the pigment data (S), the other containing the initial pigment ratios for each class (F_0). An optional pigment ratio limit matrix may also be specified. To load the files, enter their filenames in the appropriate boxes. If no file extension is specified then a default extension of '.DAT' for the data file, '.RAT' for the ratio file and '.RLM' for the ratio limits file will be assumed.

Once the files have been loaded their contents may be examined using the '*View Data*', '*View Ratios*' and '*View Limits*' buttons. It is recommended that all of the inputs be inspected to ensure that they have loaded correctly.

2.3.3.1 *Data file*

The data file is the file containing the HPLC data from which class abundances are to be estimated (the S matrix: see section 2.1). The required format is an ASCII (text) file containing an array of data values, with each row representing one sample and each column representing one pigment. The data values can be given in either normal or exponential notation (e.g. 0.013 and 1.3E-2 are both valid formats) and may be separated by any number of spaces or tab characters. Blank lines are ignored.

For example, a data set containing three data samples, with five pigments measured for each sample might look as follows:

```
0.00    0.23    1.82  0.00      1.24E-4
0.27    0.00    0.48  3.27E-3   4.54E-4
0.00    0.00    0.11  1.25E-4   2.44E-5
```

The actual data may be in any units, and the output files giving absolute pigment amounts for each class will be in the same units. Only the numerical data should be included in the file: do not include pigment names, headers, comments etc. The data file should be easy to generate from Lotus™ 1-2-3, Microsoft™ Excel or other spreadsheet (for example, in Lotus™ 1-2-3 print the matrix from the data spreadsheet to a file).

Note that the number and ordering of pigments must be the same as that used in the pigment ratio file (see next section). Also, each sample must have the same number of data values (e.g. in the example above each row must contain 5 and only 5 numbers).

2.3.3.2 *Pigment ratio file*

The ratio file contains the pigment ratios to be used in the calculation (the F_0 matrix: see Section 2.1). The required format is an ASCII (text) file containing a matrix with headers, with the pigments down the columns (i.e. each row represents one phytoplankton class).

The first line of the file should contain the pigment names. These must be separated by spaces or tabs and may not contain spaces themselves. Only the first 8 characters are significant, and the first 3 characters in each name should be unique, as these characters are used to label the output files e.g. 'cha' and 'chb'.

Each subsequent line should contain the class name (which may not contain any spaces) followed by the pigment ratios. Both decimal point and exponential notation are acceptable.

For example, if there were four classes which we were trying to distinguish on the basis of five pigments the file could look as follows:

	Pigment_A	Pigment_B	Pigment_C	Pigment_D	Pigment_E
Class_1	0.0	0.0	0.3	0.5	0.2
Class_2	0.2	0.2	0.2	0.2	0.2
Class_3	0.1	0.4	0.0	0.3	0.2
Class_4	0.4	0.0	0.0	0.0	0.6

The numbers represent the amount of the given pigment present in a 'typical' member of the given phytoplankton class. These values are normalised along each row by CHEMTAX, so the actual absolute values do not matter. For example, the first row above could read

Class_1 0 0 3 5 2

instead, and the result would be identical (in this case indicating that Class_1 has no Pigment_A or Pigment_B, 1.5 times as much Pigment_C as Pigment_E and 2.5 times as much Pigment_D as Pigment_E. Note that the number and ordering of pigments **must be identical** to that used in the data file.

In practice pigment ratios are usually entered as µg pigment per 10^6 cells or as ratios normalised against chl a as these are the most common representations in the literature.

2.3.3.3 *Ratio limits file*

This is an optional input file containing limits on how far CHEMTAX is to be permitted to modify the pigment ratio matrix. The file should contain an ASCII format matrix without headers, with the pigments down the columns. Only the numerical data should be included in the file: do not include headers, comments, pigment names etc.

The numbers give the maximum percentage by which CHEMTAX is to be allowed to modify the given ratio and is defined as follows: the pigment ratio is allowed to vary within the range initial ratio ÷ modifier to initial ratio × modifier where the modifier is 1 plus the percentage limit ÷ 100. A value of 50 thus allows a ratio matrix element of 0.4 to be increased to 0.6 (i.e. the initial value multiplied by 1.5) or decreased to 0.2666 (the initial value divided by 1.5), while a value of zero fixes the given ratio element. A value of 500 would allow a 36 fold variation of the pigment ratio matrix element e.g. 0.4 could be increased to 2.4 (i.e. the initial value multiplied by 6) or decreased to 0.0667 (the initial value divided by 6). If no limits matrix is specified then effectively no constraints are placed on changes to the pigment ratios. It is probably a good idea to constrain only those ratios known to be accurate and to allow others to be modified freely (by giving them large values in the ratio limits matrix).

For example, if the ratios for Class_1 in the example ratio matrix given earlier were known much more accurately than ratios for the other classes then the following ratio limit matrix might be appropriate:

```
10    10    10    10    10
500   500   500   500   500
500   500   500   500   500
500   500   500   500   500
```

which tells CHEMTAX that the values for Class_1 may be changed by at most 10% while the other values may be varied freely (between a sixth and six times their original values).

Note that the number and ordering of both pigments and classes must be the same as that used in the pigment ratio file (i.e. there should be a one-to-one relationship between the size of the pigment ratio matrix and that of the ratio limits matrix). Ratios whose initial values given in the pigment ratio matrix are zero will not be modified regardless of the value of the ratio limits matrix.

Unless stated otherwise, all the ratio limits for CHEMTAX calculations described in this manual were set to a default value of 500% which allowed the initial pigment ratio, r, to vary from r/6 to 6r.

2.3.4 Setting calculation options

A number of options are provided which allow the user to change the CHEMTAX calculation process. In most cases the default values, based on our extensive

evaluation of the CHEMTAX program using synthetic data sets (Section 2.5), will give good results: see Section 2.3.5 of this manual for a full discussion of the effects that changing various options has on the calculation process. A summary of the options available and their effects are given below.

2.3.4.1 *Diary file name*

The CHEMTAX program will produce a diary file containing details on the calculation process. By default this file will have the same root as the data file and the extension '.TXT', but this can be altered if desired. The amount of information to be written to the diary file is controlled by the '*verbosity*' option (see later). If this window is cleared then no diary file will be written.

2.3.4.2 *Weighting and weight bound*

The calculation process to determine the class abundances can be weighted to specify the distribution of the error in the original HPLC pigment concentrations. The weighting options are:

None: Each non-zero data point is assumed to have the same absolute error. The data are not weighted in the calculation. This option is not recommended for general use, since the abundances of minor pigments will not be well reproduced by CHEMTAX.

Relative error by pigment: Each set of pigment concentrations is assumed to have the same relative error, and hence is weighted proportional to the inverse of its sum. If one set of pigment concentrations is very small, however, this option is not recommended since very small values are given unrealistically large weights which can lead to problems in the calculations.

Bounded relative error by pigment: Each set of pigment concentrations is assumed to have the same relative error, except for very small values. The weights are calculated as for the previous option but an upper limit is placed on weights: if the calculated weight is higher than the '*weight bound*' value it is set to the '*weight bound*' value. This effectively means that small values are assumed to have the same absolute error, while large values are assumed to have the same relative error.

For example, suppose that the most abundant pigment on average is measured to have a concentration of 1 with a 1% measurement error. If the '*weight bound*' was set to 50 then pigments with average abundances down to 0.02 are assumed also to be measured with a relative error of 1%, while pigments with lower average abundances are assumed to an error of ± 0.0002 (i.e. 1% of 0.02).

The weights to be used in the calculation can be examined by using the '*view data*' button: the weights are appended to the end of the data.

2.3.4.3 *Iteration limit*

This option specifies the maximum number of iterations in the class composition calculation. Generally this number should be left large unless a short calculation time is desired. Use a small number (10-20) if a fast rough answer is required or if calculation time is limited.

2.3.4.4 *Residual limit*

The RMS difference between the measured and predicted pigment compositions is calculated during the iteration process. If S = the data matrix, F = the modified pigment ratio matrix, and C = the calculated composition matrix then if the RMS difference between S and $C \times F$ is less than the residual limit then the iteration process is assumed to be complete (i.e. a good enough fit to the data has been achieved). Essentially this value specifies how close a fit to the data is desired. Lower values may give a closer fit but can increase the number of iterations and hence the calculation time.

2.3.4.5 *Initial step size*

This value determines the initial amount by which each iteration will vary the pigment ratio matrix. A value of 5 will cause each iteration to vary one element of the ratio matrix by 1/5 (20%), a value of 10 will cause a variation of 1/10 (10%) and so on. This number should generally not be decreased below 5. However, high values can cause excessive calculation times. A value of 5-10 is probably optimal.

2.3.4.6 *Step ratio*

The *step size* number is multiplied by this value whenever the residual does not decrease on iteration with the original step size. This in effect decreases the size of changes made to the pigment ratio matrix with further iterations. For example if the '*initial step size*' was 10 and the '*step ratio*' was set to 2 the variation of *each non-zero element of the ratio matrix will be reduced from 10% (1/10) to 5% (1/(10×2) during iteration*. Values of about 2 are recommended (we use 1.3).

2.3.4.7 *Cutoff step*

The calculation is taken to be complete if the *step size* number increases past this limit. This indicates that variation of the ratio matrix is only having a slight effect (proportional to the reciprocal of the '*step size* number') on the residual. Values of 100 or so are recommended meaning that calculation is considered complete if the changes in the residual are less than 1/100 or 1%.

2.3.4.8 *Verbosity*

This controls the amount of information to be output to the diary file.

Minimum: only the header, pigment and plankton class types, and final RMS residual are output.

Low: adds initial data matrix, initial and final pigment ratio matrices, and final phytoplankton class composition matrices.

Normal: adds basic information on the iteration process, prints the absolute and relative differences between initial and final pigment ratio matrices

Verbose: adds extended information on the iteration process: gives the residual after modifying each element of the ratio matrix in each iteration.

Maximum: adds intermediate pigment ratio, class composition and weighted residual matrices after every iteration.

The *Verbose*, and especially the *Maximum* options will result in a very large diary file. The *Maximum* option is intended to be used only for debugging purposes.

2.3.4.9 *Elements varied*

The iteration process is quite slow if all the elements of the pigment ratio matrix are varied at each iteration step. Instead, the elements of the ratio matrix with the largest effect on the residual can be selected and these are varied for 5 or so iterations (the precise number being determined by the '*subiterations*' parameter). The whole matrix is then reexamined and a new set of elements selected for variation on the basis of having the largest effect on the residual.

This parameter indicates how many elements of the ratio matrix are selected for variation at each step.

2.3.4.10 *Subiterations*

This parameter indicates how many times the selected elements of the pigment ratio matrix are to be varied before the program reevaluates which elements of the ratio matrix are having the largest effect on the residual. Set this value to 1 to evaluate every nonzero ratio every iteration.

2.3.4.11 *Output options*

This button brings up the Output Options menu which will allow you to specify the type(s) of output to be produced by CHEMTAX. Use the arrow keys to navigate around the menu, and press the space bar or Enter to set or clear the check boxes.

Several output files may be produced:

Fraction total pigment: If selected, the class abundances as fractions of the total pigment in the data sample will be output to the file '<OUTNAME>.TOT'.

Final pigment ratio matrix: If selected, the final pigment ratio matrix for the iterative process will be output to the file '<OUTNAME>.RAT'. This file has the same format as the input ratio file.

Other pigments: Select the pigment to output by cycling through the available possibilities, and select one or both of the check boxes.

Absolute amount: The absolute amount of the selected pigment in each data sample due to each class is output to the file '<OUTNAME>A.<PIGMENT NAME>', in the same units as the original data.

Relative amount: The relative amount of the selected pigment in each data sample due to each class is output to the file '<OUTNAME>R.<PIGMENT NAME>'. (i.e. a value of 0.24 would indicate that 24% of the selected pigment in that data sample came from the indicated class).

2.3.4.12 *Comments*

Up to 5 lines of comments may be included in the diary file. The normal editing keys (arrow keys, backspace) work as usual, and pressing 'ins' toggles insert/overwrite mode.

2.3.4.13 *Save as default*

This button allows you to save all the current values of the CHEMTAX parameters as defaults. The next time PREPRO is run all parameters will be automatically set to the current values. However the user will still have to select the required output options (see 2.3.4.11), add comments (see 2.3.4.12) and write the output file (see 2.3.5).

To remove the current defaults and return to the original PREPRO settings delete the file 'PREPRO.DEF'.

2.3.5 Writing the output

Select the '*output*' button. Provided that the data and ratio files have been correctly loaded, PREPRO will prompt for the filename to give the command file (which must have the extension '.M'), and then for the filename that CHEMTAX will use when outputting pigment data (see the Output Options section, above). Once the command file has been written you can exit PREPRO and use the command file as input to the CHEMTAX program.

2.3.6 Effect of option selections on CHEMTAX calculations

The *'epsilon limit'*, *'initial step size'* and *'step ratio'* largely determine the accuracy of the final solution. An *'initial step size'* smaller than 5 (i.e. greater than 1/5 or 20%) or a *'step ratio'* greater than 2.0 may cause CHEMTAX to overshoot the optimum pigment ratios and cause the program to continually hunt for the correct set of ratios or even change ratios which were originally reasonable. With an *'initial step size'* greater than 10 (i.e. smaller than 1/10 or 10%), or a *'step ratio'* smaller than 1.3, CHEMTAX may home in on a false minimum. The use of a limited number of sub-iterations (usually 5) on a subset of pigments (again usually 5) can speed up the calculation process albeit with a small loss of accuracy. A quick rough answer may also be obtained by increasing the *'epsilon limit'*. The *'iteration limit'* and *'cutoff step'* are used to limit excess computation time and are usually only invoked if there is a problem in the data or pigment ratio matrices or in the setting of the PREPRO parameters.

2.4 Running CHEMTAX

To run CHEMTAX interactively, make sure that the CHEMTAX.M and associated M-files are on the Matlab™ PATH (type 'help.path' for details), then type 'chemtax'. You will be prompted for the name of the command file: enter this and CHEMTAX will run the calculation. Note that the calculation may require significant time (average 5-10 hours on a 486/50 for a medium size data set).

With Matlab™ installed on a PC, CHEMTAX calculations can be run in a batch mode using the following .m file:

```
t='input1'
chemtax
t='input2'
chemtax
t='input3'
chemtax
```

which runs 3 Matlab™ .m-files 'input1.m', 'input2.m', 'input3.m' which can be generated using PREPRO.

Under UNIX several CHEMTAX calculations may be run from the command line as follows:

```
matlab <<@ >/dev/null
chemtax
input1
chemtax
input2
chemtax
input 3
quit
@
```

which runs 3 CHEMTAX calculations on the command files 'input1.m', 'input2.m' and 'input3.m'. A sample script is provided which allows several calculations to be easily queued in this fashion.

2.5 Synthetic data sets

2.5.1 Pigment ratio matrices

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. This enabled us to evaluate the robustness of the program. Synthetic data sets are also useful when moving into new areas and one is trying to set up appropriate pigment ratio matrices.

The first data set simulated a phytoplankton community (with respect to taxa but not necessarily physiological state) from the Southern Ocean. Since pigment data from samples collected in the field 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 and Wright, 1997) were used for Bacilliarophyceae (*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 two species of Haptophyceae (the prymnesiophytes *Emiliania huxleyi* CS-57 and *Phaeocystis pouchetii* CS-165). This enabled us to generate a known pigment ratio matrix (Table 2(a)) by using the values from the SCOR-UNESCO Workshop (Jeffrey and Wright, 1997). 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 using the pigment ratios given in Table 3(a). The data set included the following additional species: *Prochlorococcus marinus* (Prochlorophyceae; Chisholm *et al.*, 1988), *Euglena* sp. (Euglenophyceae; Hager and Stransky, 1970a),

Table 2

Pigment ratios representative of **Southern Ocean** species: (a) Initial ratio matrix used to construct the synthetic data set ('true' matrix) and (b) modified by the addition of random normalised errors of ± 25% (matrix elements in (b), (c) and (d) are expressed as a percentage of the 'true' matrix elements shown in (a)). Final pigment ratios (c and d) after fitting by CHEMTAX. Calculations were for synthetic data sets where random normalised errors of ± 25% (b) were added to the pigment ratios and either experimental errors (c) or random normalised errors of ± 10% (d) were added to the data sets to simulate analytical errors.

(a)	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b*1	Chl*a*1
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)	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b*1	Chl*a*1
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)	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b*1	Chl*a*1
Pras (T3)	0	0	0	0	100.6	101.4	99.5	0	64.5	0	100.2	99.6
Dino	99.7	0	0	0	0	0	0	0	0	0	0	100.3
Cryp	0	0	0	0	0	0	0	100.7	0	0	0	99.9
Hapt (T3)	0	0	0	100.9	0	0	0	0	0	0	0	98.4
Hapt (T4)	0	96.93	103.2	96.7	0	0	0	0	0	0	0	100.7
Chlo	0	0	0	0	99.6	0	101.9	0	101.6	97.1	103.8	98.6
Syne	0	0	0	0	0	0	0	0	0	98.2	0	100.6
Diat	0	0	97.8	0	0	0	0	0	0	0	0	101.7

(d)	PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b*1	Chl*a*1
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

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 apply to Tables 2 to 4.

Table 3

Pigment ratios representative of **Equatorial** 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; (c-f) final ratio matrices 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 either typical 'experimental errors' (c and d) or ± 10% random normalised errors (e and f) were added to the data set. Calculations with: (c and e) divinyl chls *a* and *b* and; (d and f) divinyl chls *a* and *b* not distinguished from chls *a* and *b*.

(a)	PER	BUT	FUCO	HEX	NEO	PRAS	MYXO	VIOL	DDX	ALLO	LUT	ZEA	Chl*b2*	Chl*a2*	Chl*b1*	Chl*a1*
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)	PER	BUT	FUCO	HEX	NEO	PRAS	MYXO	VIOL	DDX	ALLO	LUT	ZEA	Chl*b2*	Chl*a2*	Chl*b1*	Chl*a1*
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)	PER	BUT	FUCO	HEX	NEO	PRAS	MYXO	VIOL	DDX	ALLO	LUT	ZEA	Chl*b2*	Chl*a2*	Chl*b1*	Chl*a1*
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)	PER	BUT	FUCO	HEX	NEO	PRAS	MYXO	VIOL	DDX	ALLO	LUT	ZEA	Chl*b2*	Chl*a2*	Chl*b1*	Chl*a1*
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

(e)	PER	BUT	FUCO	HEX	NEO	PRAS	MYXO	VIOL	DDX	ALLO	LUT	ZEA	Chl*b2*	Chl*a2*	Chl*b1*	Chl*a1*
Pras (T3)	0	0	0	0	113.4	105.2	0	106.9	0	0	102.0	0	0	0	101.7	94.3
Dino	109.7	0	0	0	0	0	0	0	101.1	0	0	0	0	0	0	89.5
Cryp	0	0	0	0	0	0	0	0	0	101.3	0	0	0	0	0	99.7
Hapt (T3)	0	0	0	107.2	0	0	0	0	80.3	0	0	0	0	0	0	89.7
Chry	0	109.4	99.2	0	0	0	0	0	72.9	0	0	0	0	0	0	99.8
Eugl	0	0	0	0	75.2	0	0	0	110.9	0	0	0	0	0	88.6	102.5
Chlo	0	0	0	0	92.0	0	0	114.4	0	0	110.3	140.1	0	0	97.8	97.8
Proc	0	0	0	0	0	0	0	0	0	0	0	114.9	97.0	98.5	0	0
Syne	0	0	0	0	0	0	0	0	0	0	0	92.3	0	0	0	102.7
Tric	0	0	0	0	0	0	84.1	0	0	0	0	54.0	0	0	0	105.0
Diat	0	0	101.2	0	0	0	0	0	91.9	0	0	0	0	0	0	96.2

(f)	PER	BUT	FUCO	HEX	NEO	PRAS	MYXO	VIOL	DDX	ALLO	LUT	ZEA	Chlb2	Chla2	Chlb1	Chla1
Pras (T3)	0	0	0	0	113.7	102.0	0	103.7	0	0	102.2	0	-	-	115.0	78.9
Dino	110.3	0	0	0	0	0	0	0	96.0	0	0	0	-	-	0	89.9
Cryp	0	0	0	0	0	0	0	0	0	112.2	0	0	-	-	0	97.2
Hapt (T3)	0	0	0	107.2	0	0	0	0	80.3	0	0	0	-	-	0	89.7
Chry	0	111.7	96.1	0	0	0	0	0	74.4	0	0	0	-	-	0	101.8
Eugl	0	0	0	0	95.0	0	0	0	115.2	0	0	0	-	-	111.9	91.7
Chlo	0	0	0	0	100.9	0	0	107.7	0	0	104.9	139.8	-	-	97.6	98.8
Proc	0	0	0	0	0	0	0	0	0	0	0	112.3	-	-	99.4	96.7
Sync	0	0	0	0	0	0	0	0	0	0	0	120.7	-	-	0	92.8
Tric	0	0	0	0	0	0	57.0	0	0	0	0	41.0	-	-	0	106.8
Diat	0	0	104.9	0	0	0	0	0	91.9	0	0	0	-	-	0	96.2

Pelagococcus subviridis (Chrysophyceae; Jeffrey and Wright, in press) and *Trichodesmium theibautii* (Cyanobacteria; Carpenter *et al.*, 1993). *Phaeocystis pouchetii* was not used in this data set.

The pigment ratios for a real sample are unlikely to be known exactly and we therefore added normal random errors of ± 10%, ± 25% and ± 50% 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 and Wright, 1994), 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 and Severo, (1970)) which were multiplied by the pigment concentration and a scaling factor added to the original data to produce pigment ratios with standard errors of ± 10%, ± 25% and ± 50%. The first two of these modified pigment ratio matrices for the Southern Ocean are given in Table 4(a) and Table 2(b) respectively and are expressed as percentages of the 'true' matrix elements (Table (2(a)). For the Equatorial Pacific data set, the 'true' matrix is given in Table 3(a) while the matrix modified by the addition of ± 25% random normal errors is given in Table 3(b) as a % of the 'true' value.

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 were also normalized against total pigment. Unless stated otherwise, all program runs were made on synthetic Southern Ocean (Equatorial Pacific) data sets with all non-zero ratios in the pigment ratio matrix being allowed to freely vary (ratio limits set at 500%). This gave a slight increase in accuracy albeit with longer computation times compared with calculations which allowed only a smaller subset to vary.

2.5.2 Data matrices

A series of random data matrices was 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, on average, 12.5% chl *a* for the 8 class Southern Ocean data set. These cell numbers were multiplied by the cellular content of each

Table 4

Pigment ratios representative of **Southern Ocean** species (matrix elements expressed as a percentage of the 'true' matrix elements shown inTable 2(a)): (a) initial ratio matrix used in the CHEMTAX calculations where ± 10% random normalised errors had been added to the 'true' ratio matrix and ± 10% random normalised errors were added to the data matrix; (b) final ratio matrix where the calculations were carried out on the complete data set of 40 samples and (c) the minimum (min) and maximun (max) values observed in the final ratio matrix where the calculations were performed on 8 subsets of 5 samples with the ratio limits matrices set at 50%.

(a)		PER	BUT	FUCO	HEX	NEO	PRAS	VIOL	ALLO	LUT	ZEA	Chl*b1*	Chl*a1*
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)		**PER**	**BUT**	**FUCO**	**HEX**	**NEO**	**PRAS**	**VIOL**	**ALLO**	**LUT**	**ZEA**	**Chl*b1***	**Chl*a1***
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
(c)		**PER**	**BUT**	**FUCO**	**HEX**	**NEO**	**PRAS**	**VIOL**	**ALLO**	**LUT**	**ZEA**	**Chl*b1***	**Chl*a1***
Pras (T3)	min	0	0	0	0	83.2	79.8	69.2	0	86.7	0	89.5	49.9
	max	0	0	0	0	137.2	136.7	137.7	0	119.0	0	132.3	118.0
Dino	min	67.8	0	0	0	0	0	0	0	0	0	0	76.6
	max	122.0	0	0	0	0	0	0	0	0	0	0	134.2
Cryp	min	0	0	0	0	0	0	0	93.9	0	0	0	77.0
	max	0	0	0	0	0	0	0	200.6	0	0	0	101.4
Hapt (T3)	min	0	0	0	78.0	0	0	0	0	0	0	0	47.6
	max	0	0	0	130.7	0	0	0	0	0	0	0	137.6
Hapt (T4)	min	0	72.1	44.1	76.9	0	0	0	0	0	0	0	100.0
	max	0	102.8	113.7	104.3	0	0	0	0	0	0	0	144.1
hlo	min	0	0	0	0	72.1	0	94.0	0	98.0	108.0	87.1	76.3
	max	0	0	0	0	170.6	0	151.8	0	135.9	147.9	138.5	97.4
Syne	min	0	0	0	0	0	0	0	0	0	92.0	0	48.0
	max	0	0	0	0	0	0	0	0	0	249.6	0	102.8
Diat	min	0	0	65.4	0	0	0	0	0	0	0	0	71.8
	max	0	0	137.4	0	0	0	0	0	0	0	0	128.8

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 three 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 = 0, variance = 1, using an algorithm derived from Zelen and 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 two 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 (which affect peaks individually and are proportionately greater for smaller peak areas). These were determined experimentally by repeated HPLC analysis of a solution of ß-*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 x 4.6 mm), eluted isocratically with methanol, detected at 405 nm and 436 nm (Waters 440 detector) or 435 nm and 470 nm (Spectraphysics detector), and integrated using Waters Baseline software. The solution was diluted by 50%, and again analysed ten times. The process was repeated

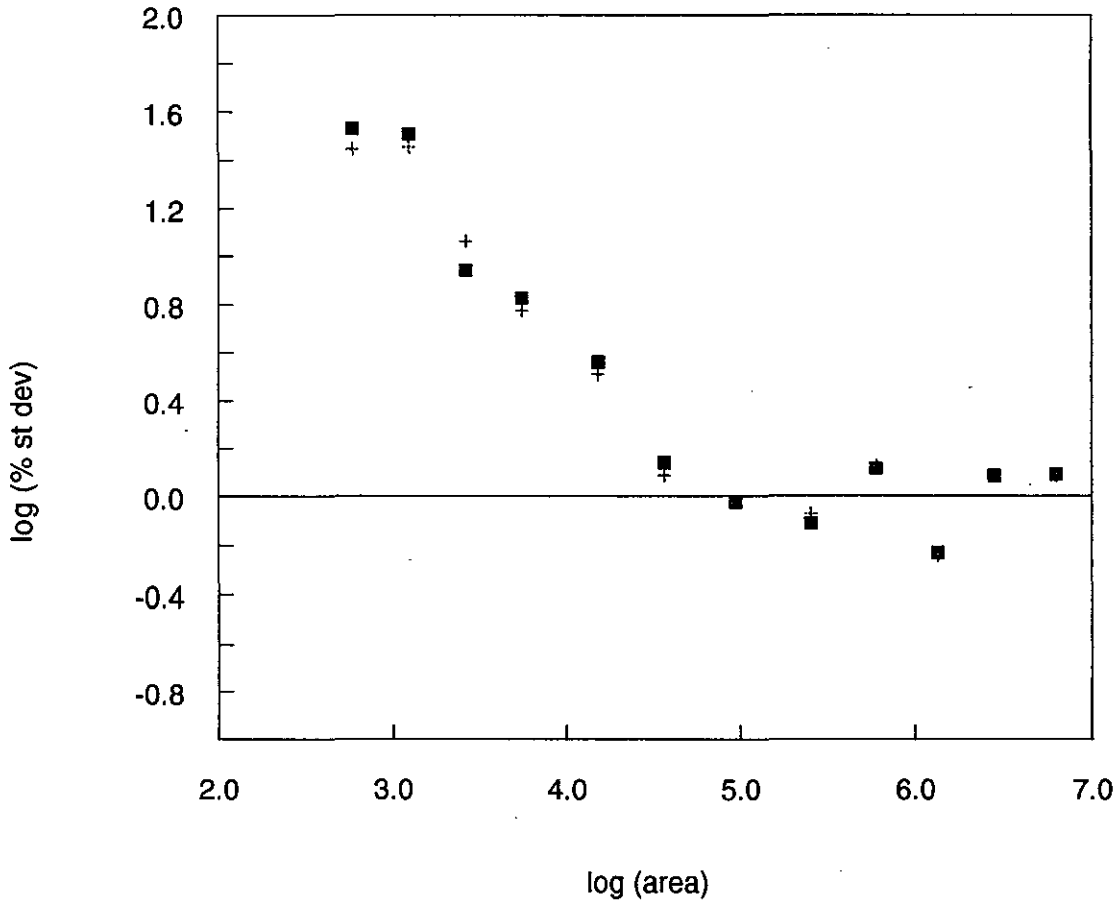


Figure 1

Plot of log (% standard deviation) for replicate (10) injections of ß-*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.

until the peak was no longer detectable (10 dilutions). The covariance of the areas for the two 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 Figure 1 and also the Results and Discussion). 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.

A second series of synthetic data sets to simulate the Equatorial Pacific was similarly constructed.

3 Results and Discussion

For CHEMTAX to produce realistic and reliable results, careful consideration must be given not only to the selection of optimum program parameters but also to the setting up of the data file and the selection of the appropriate pigment ratio matrix (and associated ratio limits matrix). These processes will be discussed in detail in the next sections.

3.1 HPLC data set

3.1.1 Errors in data matrix, *S*

As the factorisation of the data matrix, *S*, is underdetermined there is no unique solution and errors in the data set can have a significant effect on the final result. To obtain a meaningful solution these errors must be kept to a minimum. Careful attention should therefore be paid to the following:

3.1.1.1 *Outliers*

The CHEMTAX regression procedure used is not overly robust to outliers, so pre-inspection of the data for obvious data errors and their correction or removal is recommended.

3.1.1.2 *Analytical errors*

Apart from errors in sampling and filtration there are 2 main sources of analytical errors (Figure 1; from Mackey *et al.*, 1996) associated with the HPLC determination of pigment concentrations:

volumetric errors of the autosampler which at large peak areas correspond to ~1% of the total peak area.

quantitation errors of the detector and integrator which approach 100% standard deviation for peak areas near the limit of detection.

These analytical errors should be kept to an absolute minimum by paying particular care to the setting up and optimization of the HPLC system.

Our observations from the many runs we have carried out using synthetic data sets to simulate either a Southern Ocean or an Equatorial phytoplankton community suggest that a well optimised HPLC can generate a data set with experimentally derived errors of the order of ± 5% random normal standard deviation.

3.1.2 Complexity of the data set

With normal random errors of ± 25% standard deviation in the pigment ratio matrix and experimental errors in the data set good recoveries of chl *a* were obtained using CHEMTAX for both the Southern Ocean (Figure 2; adapted from Mackey *et al.*, 1996) and the Equatorial (Figure 3; from Mackey *et al.*, 1996) phytoplankton communities. The final pigment ratios for the Southern Ocean samples (Table 2(c)) were near the ratios used to generate the data set (Table 2(a)). The final pigment ratios for the Equatorial samples calculated assuming divinyl chls *a* and *b* were resolved from chls *a* and *b* (Table 3(c)) and assuming no resolution (Table 3(d)) were also near the 'true' ratios (Table 3(a)).

With normal random errors of ± 10% in the Southern Ocean data set (about twice the error of a well optimised HPLC system) and ± 25% normal random noise in the pigment ratio matrix recoveries of chl *a*, although not as tight as was obtained using experimental errors, was nevertheless reasonable for all algal classes. The final pigment ratios (Table 2(d)) were also near to the 'true' ratios (Table 2(a)). On the other hand, although correct trends and good absolute recoveries of chl *a* were observed for most algal classes for the more complex Equatorial data set, the trends and absolute recoveries of chl *a* (Figure 4) for some classes were not good. The corresponding final pigment ratios (Table 3(e) and 3(f)) also showed large deviations from the 'true' ratios (Table 3(a)). This underscores the fact that optimisation of the HPLC system is very important and is much more critical for more complicated data sets.

3.1.3 Homogeneity of the data set

The calculations require that the pigment ratios within each phytoplankton class are constant across the data samples, and hence that all of the data samples are from the same phytoplankton community with the same physiological status.

A set of samples which spans several phytoplankton communities should thus be split into groups and calculated separately, each group with its own optimum but different pigment ratio matrix. As phytoplankton communities or physiological status are likely to vary within a stratified water column or between different water masses this should be taken into account when splitting the data set into sub-groups and these points are discussed further below:

3.1.3.1 *Stratification*

In a stratified water column pigment ratio fingerprints are likely to change with depth due to:

light adaptation (Demers *et al.*, 1991)

changes in the species composition of a given algal class (Gieskes and Kraay, 1983)

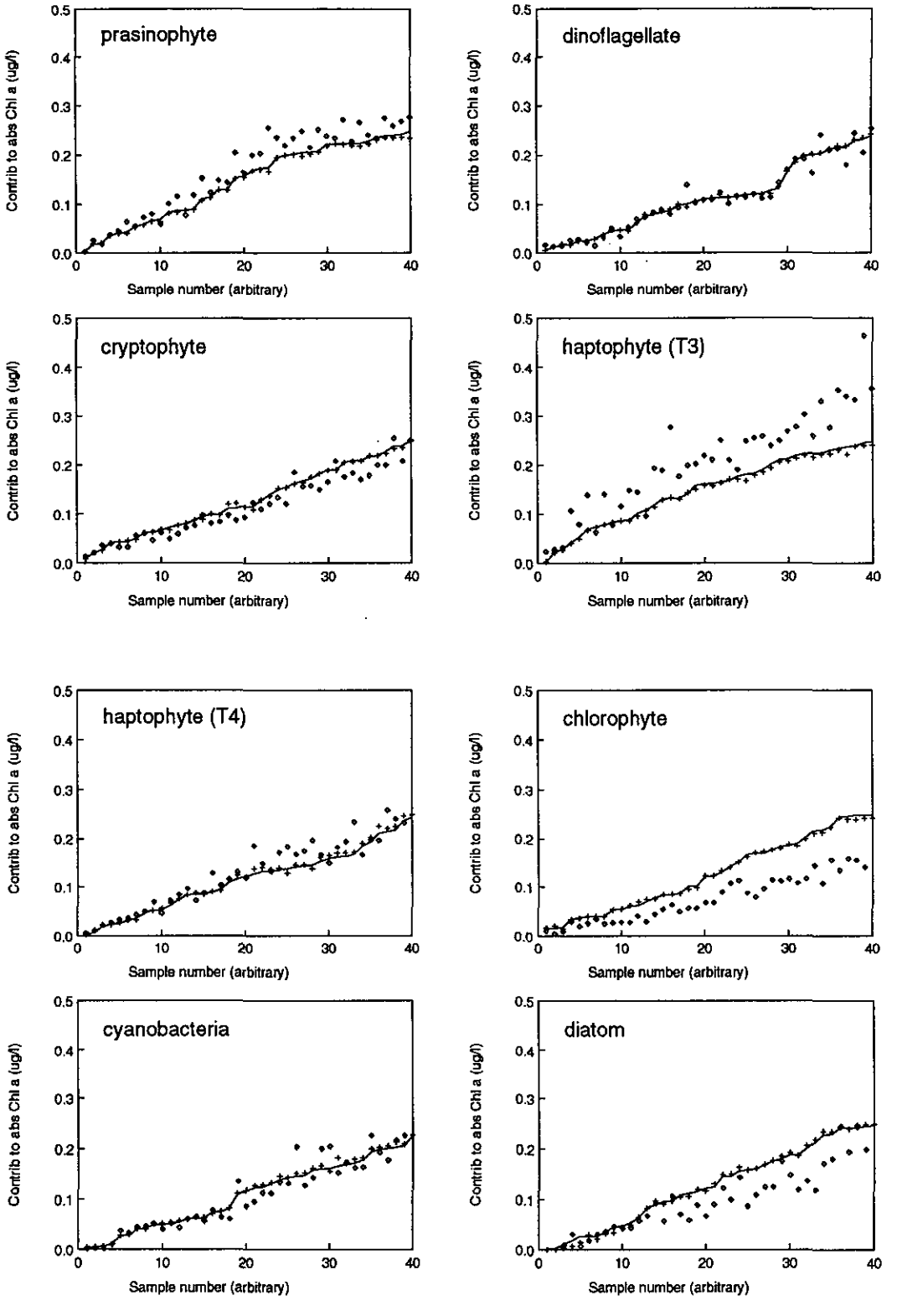


Figure 2

Plots of contribution to total Chl *a* in the synthetic **Southern Ocean** HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class. The solid line is the 'true' value. The calculated values are given for the cases where random normal standard errors of ± 25% were added to the pigment ratio matrix and there were either simulated experimental errors (+) or ± 10% random normal standard errors (◊) added to the data.

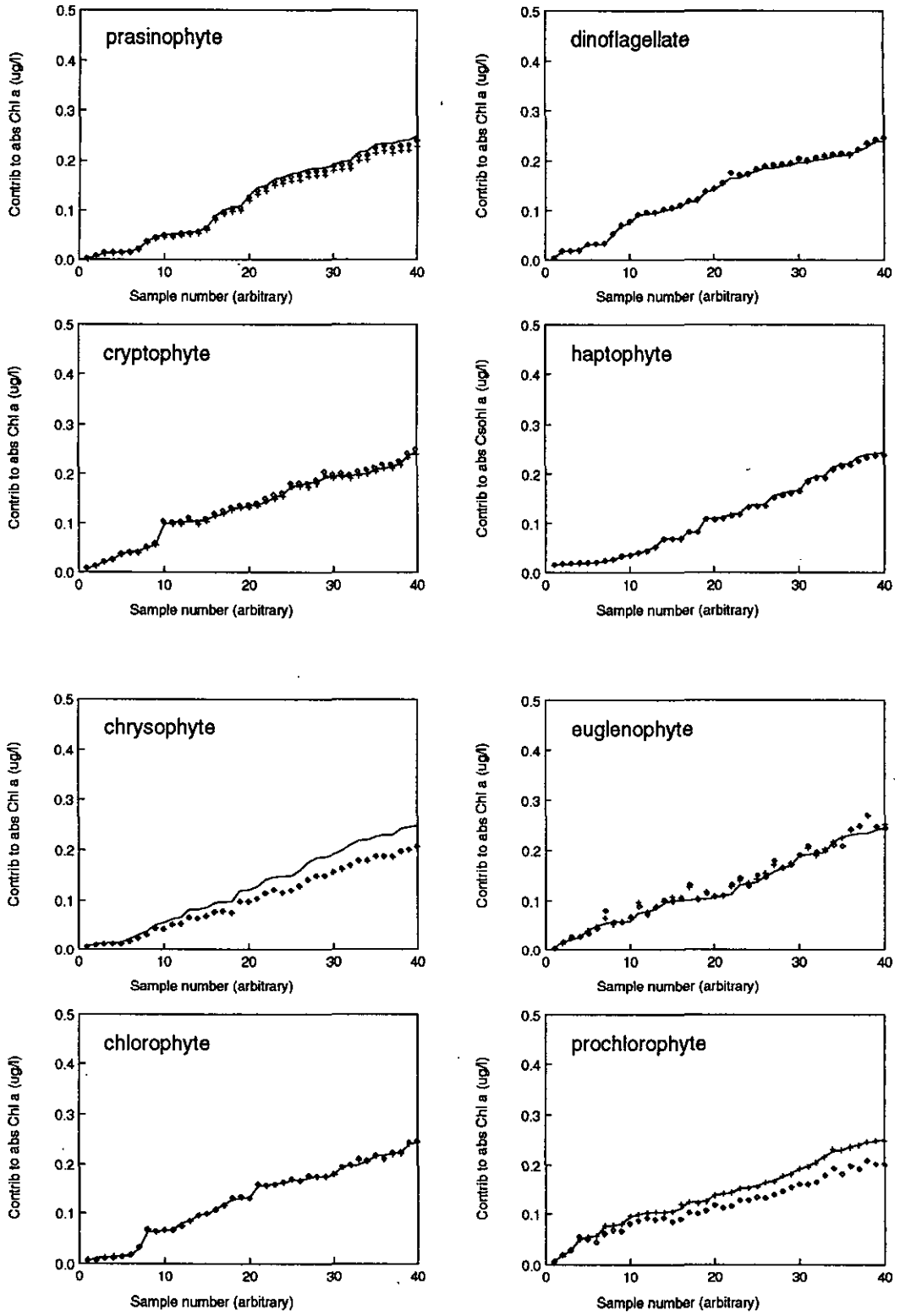


Figure 3

Plots of contribution to total Chl *a* in the synthetic **Equatorial** HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class. 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 (◊).

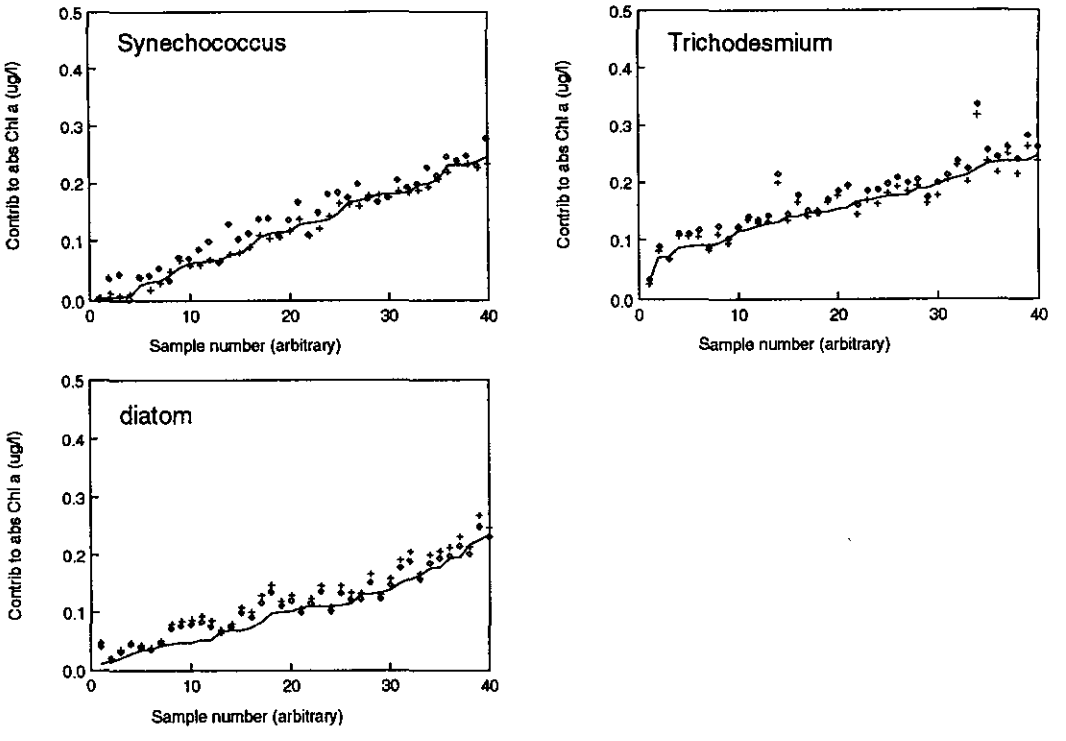


Figure 3 (continued)

Plots of contribution to total Chl *a* in the synthetic **Equatorial** HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class. 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 (◊).

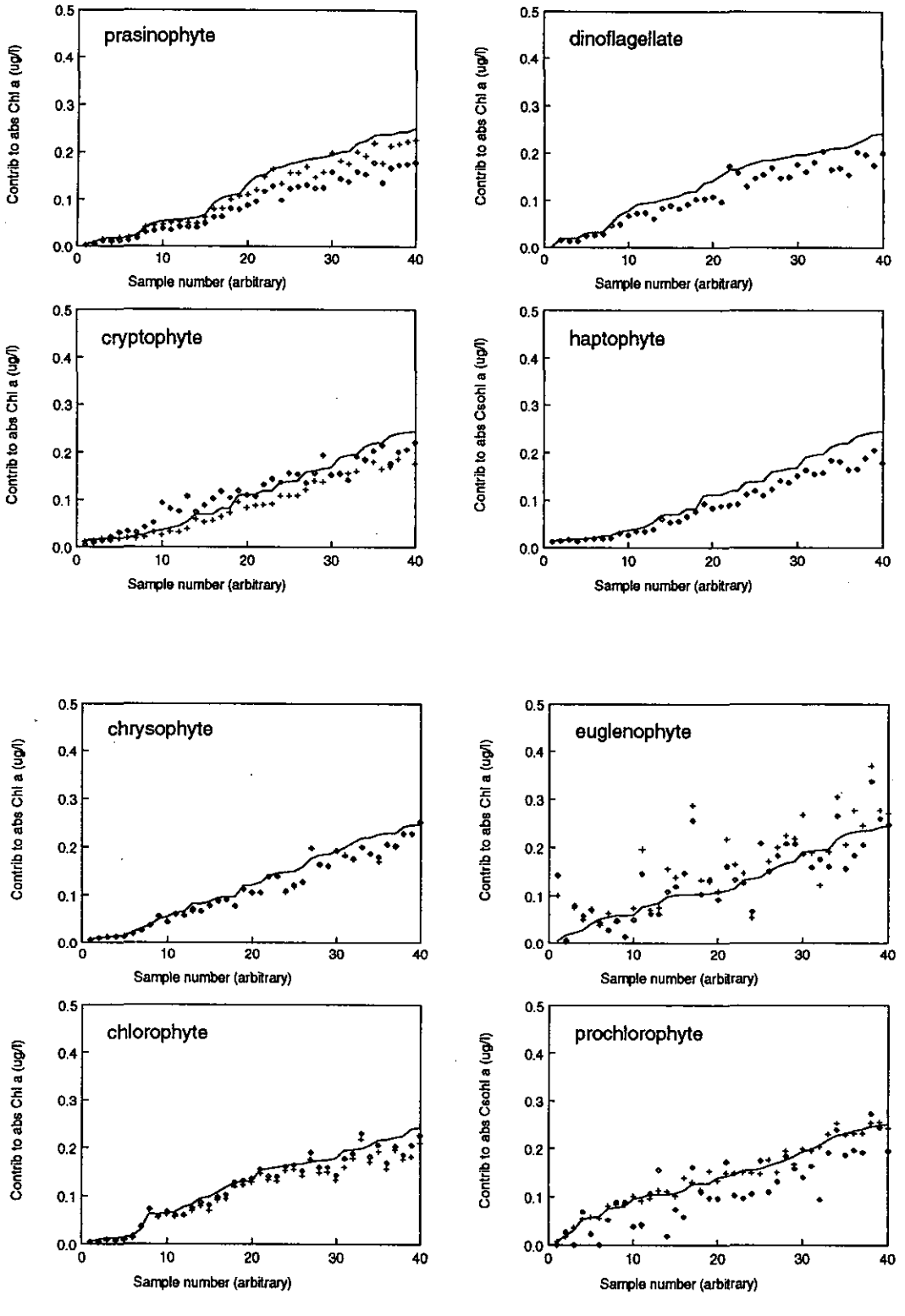


Figure 4

Plots of contribution to total Chl *a* in the synthetic **Equatorial** HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class. The solid line is the 'true' value. The calculated values are given for the case where there were random normal standard errors of ± 10% added to the data and 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 (◊).

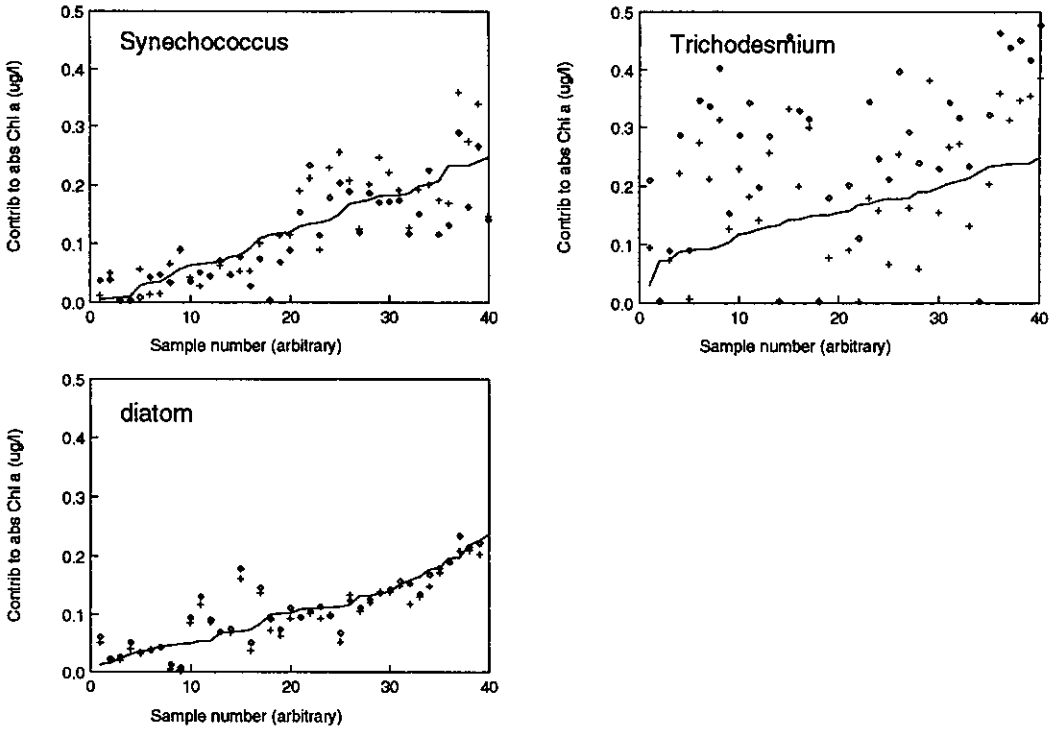


Figure 4 (continued)

Plots of contribution to total Chl *a* in the synthetic **Equatorial** HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class. The solid line is the 'true' value. The calculated values are given for the case where there were random normal standard errors of ± 10% added to the data and 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 (◊).

3.1.3.2 *Different water masses*

Dissimilar phytoplankton communities may occur in different water masses as a result of:

latitudinal variations (e.g. north and south of the Sub-Antarctic Convergence; Wright *et al.*, 1996).

depth variations (e.g. the surface Equatorial Current versus the deeper Equatorial Undercurrent).

estuarine gradients (fresh versus brackish versus marine waters).

3.1.3.3 *Homogeneity test*

To test the homogeneity of a particular grouping of samples it is useful to allow CHEMTAX to calculate class abundances using all the samples of the group and then as separate sub-groups (provided the sub-group size is not too small: see

below). If the original sample grouping is homogeneous there should be no significant difference between the series of CHEMTAX calculations.

3.1.4 Group size

Despite the neccessity of dividing a disparate data set into homogeneous groupings, the groups must not be so small so as to adversely affect the accuracy of the CHEMTAX calculations. The effect of group size can be illustrated by a series of CHEMTAX calculations using the Southern Ocean synthetic data set.

The large number of samples (40) present in the complete data set ensured that the program was able to adequately reproduce (Table 4(b)) the 'true' ratio matrix (Table 4(a)) and class distribution (Figure 5; from Mackey *et al.*, 1996), even if there was considerable uncertainty (± 10% random standard errors) in the data set and the initial pigment ratio matrix. 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 we selected 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 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 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 Figure 5 (from Mackey *et al.*, 1996) and Table 3, we have compared the 'true' class distributions with those calculated using the whole 40 samples and calculated as 8 sets of 5 samples. It is clear that, in this case (Figure 5), 5 samples are insufficient to provide good estimates of class composition. This was reflected in the range of the final pigment ratios (Table 4(c)). However, it is also clear by comparing Figure 3 and Figure 5 that for 40 samples the ability of the program to calculate the class composition is more dependent on the errors in the data (± 10% random standard errors) than on the errors in the ratio matrix (± 10% or ± 25% random standard errors).

3.1.5 Relative algal class contribution

In field samples the proportion of a given phytoplankton class can vary dramatically (e.g. with depth in a stratified water column). With the Southern Ocean data set, CHEMTAX could adequately resolve the contribution of a given algal class when the average contribution to total chl *a* was set at various levels between 5% and 33% (with the random sample values within each data set ranging from 0 to 30%). Contributions to the total chl *a* less than this may also be resolved but this will be dependent on the quality of the data set and the accuracy of the pigment ratio matrix.

3.2 Pigment ratio matrix

Since the original problem of dividing the data matrix into pigment ratios and algal class abundances was underdetermined the choice of the initial pigment ratio matrix largely determines the results obtained (i.e. if the pigment ratios used are 'too far' removed from the

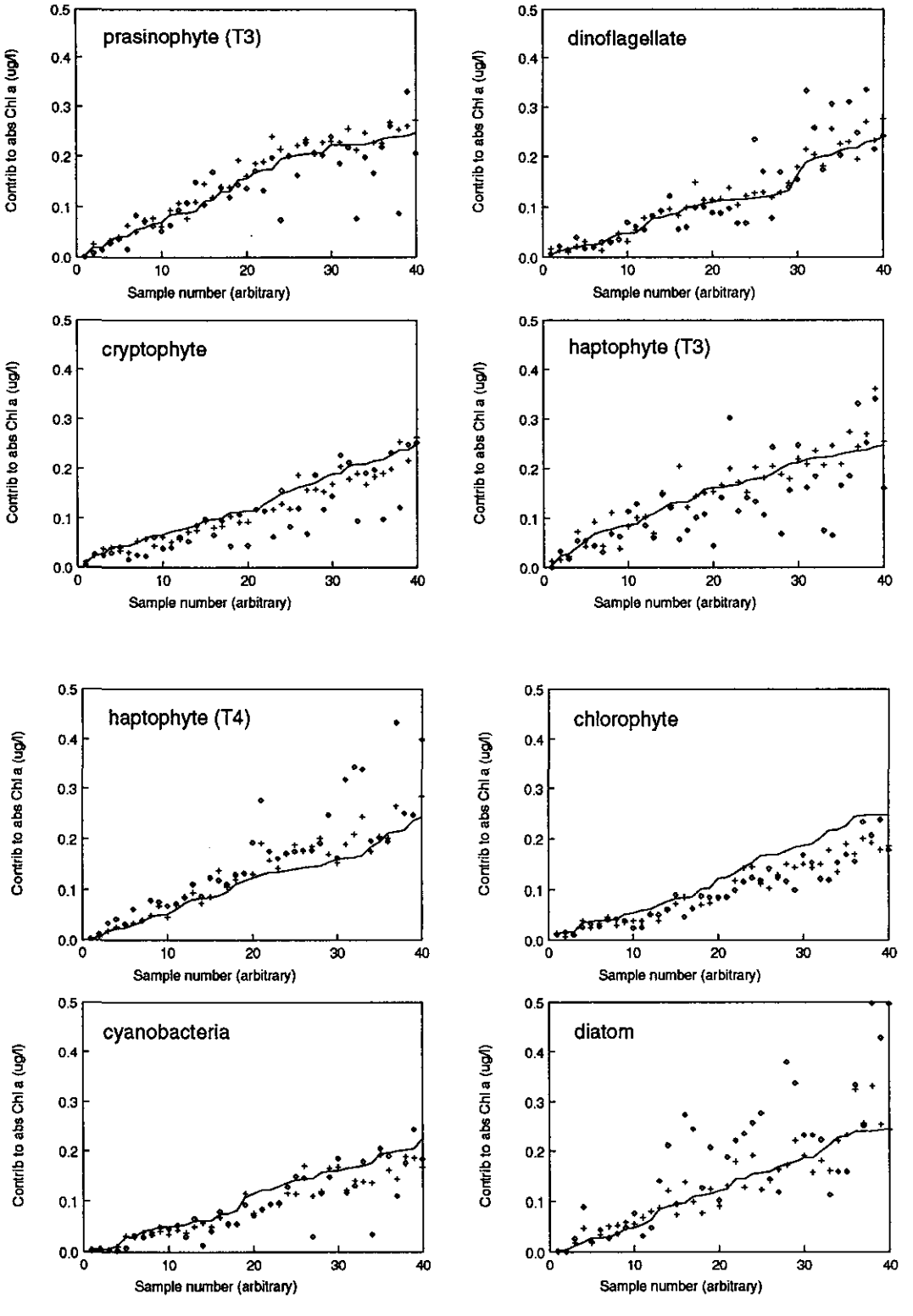


Figure 5

Plots of contribution to total Chl *a* in the synthetic **Southern Ocean** HPLC samples against sample number (arbitrary) ordered according to increasing contribution within each phytoplankton class. The solid line is the 'true' value. 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 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 (◊).

actual pigment ratios found in the sample then CHEMTAX may fail to resolve the algal classes or converge to an unrealistic solution).

In selecting the optimum pigment ratio matrix for a given set of samples, careful consideration should be given not only to the selection of appropriate species to represent the different algal classes, but also to variations of cellular pigment concentrations with changes in physiological status as well as the number and characteristics of pigments used for each class. Each of these points will be discussed further below:

3.2.1 Selection of representative species for a given algal class

The CHEMTAX program assumes that all members of a given algal class have the same 'typical' set of pigment ratios. Care should therefore be taken to ensure that the pigment ratios used are from the dominant species of the given algal class or that the species chosen is indeed representative of the remainder of the algal class for the particular group of samples under study. In other words, an *a priori* knowledge of the community under study is extremely useful in determining the range of algal classes likely to be found, the dominant or representative species of each class, and the effect of forcing factors including mixed layer depth and stratification on various environmental parameters such as light and nutrients and hence cellular pigment composition.

Each major phytoplankton class likely to be present in the samples of the data set should be represented in the pigment ratio matrix. Conversely any algal classes which would not be expected to occur in a group of samples should be removed from the pigment ratio matrix.

3.2.2 Variation of individual pigment ratios

Pigment ratios can vary dramatically between different species within a given algal class (Table 1, from Mackey *et al.*, 1996) or indeed between different strains of a particular species (Jeffrey and Wright, 1994) and this introduces an unavoidable error into the estimates of class abundances produced by this method.

Variation in the pigment ratios of any phytoplankton species within a given data set with changes in the light regime, nutrient concentration and physiological status must be taken into account and it should be noted that there may be significant differences between data obtained from cultures and field samples.

3.2.3 Pigment selection

Care should also be taken when selecting what pigments to include in the pigment ratio matrix and the following should be noted:

3.2.3.1 *Universal distribution*

Pigments which are present in nearly all algal classes are unlikely to yield useful information with 'marker' pigments giving better results than the more common pigments.

3.2.3.2 *Cellular abundance*

Pigments which have vastly different cellular concentrations between different species or strains within the same algal class are likely to give poor results.

3.2.3.3 *Complex taxonomic distribution*

Pigments which have a complex taxonomic distribution (e.g. chl c_1, c_2, c_3 and 3,8-Mg divinyl pheoporphyrin *a*; Jeffrey, 1989; Jeffrey and Wright, 1994) or have only been recently identified with improved chromatographic techniques, may also be of limited use until more quantitative data becomes available on cellular content, species distribution and ecological range.

3.2.3.4 *Pigment conversion*

Pigments such as diadinoxanthin, which is rapidly converted to diatoxanthin in the light (Demers *et al.*, 1991), may be of limited use.

3.2.3.5 *Pigment resolution*

Pigments which may not be adequately resolved by HPLC (e.g. ββ-carotene and βε-carotene) can provide problems with quantitation and should be avoided as they add noise to the data set.

3.2.3.6 *Divinyl chlorophyll a and b*

Prochlorophytes (Campbell and Nolla, 1994; Chisholm *et al.*, 1988) are best estimated using divinyl chl *a* and *b*. However, when these pigments are not resolved from, and are therefore included in the estimation of, chl a and b CHEMTAX can still provide a reasonable estimate of prochlorophytes, although there may be some underestimation of their contribution (Table 3 and Figure 3).

3.2.4 Number of pigments per algal class

The initial pigmemt ratio matrix, F_0, must be set up with care if meaningful results are to be obtained from the calculations. The 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 CHEMTAX program needs a minimum of 2 or 3 pigments more than the number of algal classes and it is preferable that each plankton class in the pigment ratio matrix have at least 2 pigments with non-zero abundance in addition to chl *a*.

3.2.5 Uncertainties in the pigment ratio matrix, F_0

The CHEMTAX program is more tolerant of occasional uncertainties in the pigment ratio matrix than errors in the data matrix. The following points have been observed during our extensive testing of the CHEMTAX program with synthetic data sets:

3.2.5.1 *Major pigments*

Generally uncertainties of up to about 30-40% in the major pigments can be tolerated.

3.2.5.2 *Minor pigments*

Larger uncertainties are often tolerated if a pigment is a minor proportion of a given species pigment complement, or if the source of the pigment is minor compared to the total concentration of that pigment in the sample (e.g. lutein is a small proportion of the pigment complement of the prasinophyte pigment type defined in our pigment ratio matrix, and lutein from this prasinophyte source is minor when compared to the chlorophyte pigment type even if, as was the case with our synthetic data sets, the average contribution of each class to the total chl *a* is about equal. The recovery of chl *a* in the prasinophyte class is good despite large potential errors in lutein (e.g.Table 2(c) and Figure 2(a)) and is largely due to the good fit of the major pigments).

3.2.5.3 *Trends*

With uncertainties of the order of 40-80% in the major pigments (Table 3(e) and (f)) correct trends are usually observed across a series of samples but the absolute abundances of the classes (in terms of chl *a*) are often in error (Figure 4).

3.2.5.4 *Group size*

When the numbers of samples per group is small CHEMTAX is less tolerant of errors in the pigment ratio matrix.

3.2.6 Ratio limits matrix

If there is a large degree of confidence in a particular pigment ratio, its optimisation by CHEMTAX can be restricted to a narrow range by reducing r in the ratio limits matrix from the default of 500% to say 20-30%. This is particularly relevant if the number of samples in the data set is small.

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. We also thank P. P. Morgan and P. Campbell for computing advice.

Bibliography

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

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

Ben-Amotz, A., Katz, A., Avron, M. (1982). Accumulation of ß-carotene in halotolerant alga: purification qand characterization of ß-carotene-rich globules from *Dunaliella* bardawil (Chlorophyceae). Journal of Phycology 18: 529-537

Berger, R., Liaaen-Jensen, S., McAlister, V., Guillard, R. R. L. (1977). Carotenoids of Prymnesiophyceae (Haptophyceae). Biochemical Systematics and Ecology 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. Journal of Phycology 15: 457-463

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

Burger-Wiersma, R., Veenhuis, M., Korthals, H. J., Van de Wiel, C. C. M., Mur, L. R. (1986). A new prokaryote containing chlorophyll *a* and *b*. Nature 320: 262-264

Bustillos-Guzman, J., Claustre, H., Marty, J-C. (1995). Specific phytoplankton signatures and their relationships to hydrographic conditions in the coastal northwestern Mediterranean. Marine Ecology - Progress Series 124: 247-258

Campbell L., Nolla H. A. (1994). The importance of *Prochlorococcuus* to community structure in the central North Pacific Ocean. Limnology and Oceanography 39: 954-961

Carpenter, E. J., O'Neil, J. M., Dawson, R., Capone, D. G., Siddiqui, D. J. A., Orenneberg, T, Bergnan, B. (1993). The tropical diazotrophic phytoplankter *Trichodesmium* : biological characteristics of two common species. Marine Ecology - Progress Series 95: 295-304

Chavez, F. P., Buck, K. R., Barber, R. T. (1990). Phytoplankton taxa in relation to primary production in the equatorial Pacific. Deep-Sea Research 37: 1733-1752

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., Welschmeyer, N.A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature 334: 340-343

Demers, S., Roy Suzanne, Gagnon, R., Vignault, C. (1991). Rapid light-induced changes in cell fluorescence and in xanthophyll- cycle pigments of *Alexandrium excavatum* (Dinophyceae) and *Thalassiosira pseudona* (Bacillariophyceae): a photo-protection

mechanism. Marine Ecology - Progress Series 76: 185-193

Everitt, D. A., Wright, S. W., Volkman, J. K., Thomas, D. P., Lindstrom, E. (1990). Phytoplankton community compositions in the western equatorial Pacific determined from chlorophyll and carotenoid pigment distributions. Deep-Sea Research 37: 975-997

Fawley, M. W. (1992). Photosynthetic pigments of *Pseudocourfielda* marina and select green flagellates and coccoid ultraplankton: implications for the systematics of the micromonadophycaea (Chlorophyta). Journal of Phycology 28: 26-31

Gieskes, W. W., Kraay, G. W. (1983). Dominance of Cryptophyceae during the phytoplankton spring bloom in the central North Sea detected by HPLC analysis of pigments. Marine Biology 75: 179-185

Gieskes, W. W., Kraay, G. W. (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. Marine Biology 91: 567-576

Gieskes, W. W. C., Kraay, G. W., 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. Netherlands Journal of Sea Research 22: 123-137

Hager, A., Stransky, H. (1970a). The carotenoid pattern and the occurence of the light induced xanthophyll cycle in various classes of algae V. A few members of Cryptophyceae, Euglenophyceae, Bacilliarophyceae, Chrysophyceae, and Phaeophyceae. Archiv fuer Mikrobiologie 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. Archiv fuer Mikrobiologie 72: 68-83

Hallegraeff G. M., Jeffrey, S. W. (1984). Tropical phytoplankton species and pigments of continental shelf waters of North and North-West of Australia. Marine Ecology - Progress Series 20: 59-74

Hooks, C. E., Bidigare, R. R., Keller, M. D., Guillard, R. R. L. (1988). Coccoid eukaryotic marine ultraplankters with four different HPLC pigment signatures. Journal of Phycology 24: 571- 580

Iturriaga, R., Mitchell, B. G. (1986). Chroococoid cyanobacteria: a significant component in the food web dynamics of the open ocean. Marine Ecology - Progress Series 28: 291-297

Jeffrey, S. W. (1976). The occurence of chlorophyll c_1 and c_2 in algae. Journal of Phycology 12: 349-354

Jeffrey, S. W. (1989). Chlorophyll *c* pigments and their distribution in the chromophyte algae. In: J. C. Green, B. S. C. Leadbeater, and W. L. Diver (ed.) The Chromophyte Algae: Problems and Perspectives. Clarendon Press, Oxford, p. 11-36

Jeffrey, S. W., Hallegraeff G. M. (1980). Studies of phytoplankton species and photosynthetic pigments in a warm core eddy of the East Australian Current. I. Summer populations. Marine Ecology - Progress Series 3: 285-294

Jeffrey, S. W., Hallegraeff G. M. (1987). Phytoplankton pigments, species and light climate in a complex warm-core eddy of the East Australian Current. Deep-Sea Research 34: 649-673

Jeffrey, S. W., Sielicki, M., Haxo, F. T. (1975). Chloroplast pigment patterns in the dinoflagellates. Journal of Phycology 11: 374-384

Jeffrey, S. W., Wright, S. W. (1987). A new spectrally distinct component in preparations of chl *c* from the micro-alga *Emiliania huxleyi* (Prymnesiophyceae). Biochimica et Biophysica Acta 894: 180-188

Jeffrey, S. W., Wright, S. W. (1994). Photosynthetic pigments in the Haptophyceae. The Haptophyte Algae. Green, J. C., Leadbeater, B. S. C. eds, Systematics Association Special Volume No. 51, Claredon Press, Oxford, p. 111-132

Jeffrey, S. W., Wright, S. W., (1997). Qualitative and Quantitative analysis of SCOR reference algal cultures. In: S. W. Jeffrey, R. F. C. Mantoura, and S. W. Wright (eds.) Phytoplankton pigments in oceanography: guidelines to modern methods. SCOR- UNESCO, Paris, p. 343-360.

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. Marine Ecology - Progress Series 37: 265-275

Li, W. K. W., Subba-Rao, D. V., Harrison, W. G., Smith, J. C., Cullen, J. J., Irwan, B., Platt, T. (1983). Autotrophic picoplankton in the tropical ocean. Science 219: 292-295

Letelier, R. M., Bidegare, R. R., Hebel, D. V., Ondrusek, M., Winn, C. D., Carl, D. M. (1993). Temporal variability of phytoplankton community structure based on pigment analysis. Limnology and Oceanography 38: 1420-1437

M. D. Mackey, M. D., Mackey, D. J., Higgins, H. W., Wright, S. W. (1996). CHEMTAX - A program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. Marine Ecology - Progress Series 144: 265-283.

Platt, T., Subba Rao, D. V., Irwan, B. (1983). Photosynthesis of picoplankton in the oligotrophic ocean. Nature 301: 702-704

Ricketts, T. R. (1967). Further investigations into the pigment composition of green flagellates possessing scaly flagella. Phytochemistry 6: 1375-1386

Ricketts, T. R. (1970). The pigments of Prasinophycae and related organisms. Phytochemistry 9: 1835-1842

Ridout, P. S., Morris, R. J. (1985). Short-term variations in the pigment composition of a spring phytoplankton bloom from an enclosed experimental ecosystem. Marine Biology 87: 7-11

Simon, N., Barlow, R. G., MariAe, D., Partensky, F., Vaulot, D. (1994). Characterization of oceanic photosynthetic picoeukaryotes by flow cytometry. Journal of Phycology 30: 922-935

Stauber, J. L., Jeffrey, S. W., (1989). Photosynthetic pigments in 51 species of marine diatoms. Journal of Phycology 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. Archiv fuer Mikrobiologie 72: 84-96

Tester, P. A., Geesey, M. E., Guo, C., Pearl, H.W., Millie, D. F. (1995). Evaluating phytoplankton dynamics in the Newport River estuary (North Carolina, USA) by HPLC-derived pigment proliles. Marine Ecology - Progress Series 124: 237-245

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. Photosynthesis Research 13: 101-111

Vaulot, D., Birrien, J., Marie, D., Partensky, C. R., Veldhuis, M. J. W., Kraay, G. W., Chrétiennot-Dinet, M. (1994). Morphology, ploidy, pigment composition, and genome size of cultured strains of *Phaeocystis* (Prymnesiophyceae). Journal of Phycology 30: 1022- 1035

Vesk M., Jeffrey, S. W. (1987). Ultrastructure and pigments of two strains of the picoplanktonic alga *Pelagocccus subviridis* (Chrysophyceae). Journal of Phycology 23: 322-336

Wright, S. W., Jeffrey, S. W. (1987). Fucoxanthin pigment markers of marine phytoplankton analysed by HPLC and HPTLC. Marine Ecology - Progress Series 38: 259-266

Wright, S. W., Jeffrey, S. W., Mantoura, R. F. C., Llewellyn, C. A., Bjornland, T., Repeta, D., Welschmeyer, N. (1991). Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. Marine Ecology - Progress Series 77: 183-196

Wright, S. W., Thomas, D. P., Marchant, H. J., Higgins, H. W., Mackey, M. D. and Mackey, D. J. (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. Marine Ecology - Progress Series 144: 285-298.

Wright, S. W., Shearer, J. D. (1984). Rapid extraction and high performance liquid chromatography of chlorophylls and carotenoids from marine phytoplankton. Journal of Chromatography 294: 281-295

Zelen, M., Severo, N. C. (1965). Probability functions. Handbook of mathematical functions. Abramowitz, M., Stegun, I. A. eds. Dover Publications, New York, p. 925-995

Appendix A CHEMTAX input file format

The input file for the CHEMTAX Matlab™ program is simply a text file (an '.m file') which sets the variables that the program uses. The best way to generate this file is to use the preprocessor program PREPRO.EXE, but if this is not possible or desired then the input file may be produced or modified manually.

To set up a calculation on d data points, with c classes and p pigments, the following variables must be set:

S: $(d \times p)$: matrix containing the data points. Any units may be used.

F_0: $(c \times p)$ matrix containing the initial pigment ratios. Any units may be used: the matrix will be normalised so that the sum of values along each row is 1.

ratiolimit: $(c \times p)$ matrix containing the ratio limits as percentages. This matrix controls the extent of variation of the matrix allowed. For example, if one element of the ratiolimit matrix was 50, then the corresponding element r of the matrix could vary between r/1.5 and r×1.5. Use a large value (eg 500) to allow the values to vary freely.

pignames: $(p \times 8)$ text matrix containing the names of the p pigments.

speciesnames: $(c \times 20)$ text matrix containing the names of the c classes.

comments: text matrix containing zero or more lines of comments to be printed in the diary file.

diaryname: text vector, contains the name of the diary file to write. If empty, no diary file is written.

outname: text vector, the root of the names of the output files.

weights: $(1 \times p)$ vector containing the weightings for each pigment (usually determined from the data: see the PREPRO documentation for more details).

maxiter: the maximum number of iterations before the program will stop.

errlimit: the cutoff value for the residual. When the residual is less than this value the calculation is deemed to be complete.

divisor: the fraction by which the elements of will be varied at each step. For instance, a value of 20 indicates a 5% variation.

stepratio: after each calculation stage is complete the divisor is multiplied by this value.

steplimit: the calculation is deemed to be complete if the divisor exceeds this value.

verbose: integer, controls the verbosity of the diary output. 1 is the least verbose, 5 is the most. Levels 4 and 5 are generally useful only for debugging purposes.

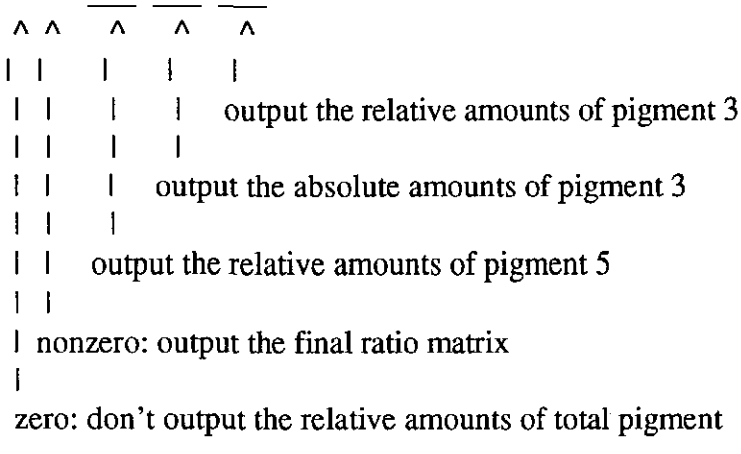
numvaried: integer, the number of entries to vary in each subiteration stage.

numsubiters: integer, the number of subiterations to perform before reexamining the gradients of all of the elements of the F matrix.

outputtypes: vector. The first two elements of the vector control whether CHEMTAX is to output the relative amounts of total pigment (to file '<OUTNAME>.TOT') and the final ratio matrix (to file '<OUTNAME>.RAT') respectively: if the vector elements are nonzero the corresponding file is written.

After these first two values are one or more pairs of integers. The first integer of each pair denotes the pigment to write (if zero this pair is ignored). The second integer denotes whether the absolute or relative amount of that pigment is to be output (zero means absolute, nonzero means relative).

For example, if the outputtypes vector equalled (0,1,5,1,3,0,3,1), this would be interpreted as follows:



and so the following files would be output:

<OUTNAME>.RAT (the final ratio matrix)
<OUTNAME>R.<PIG5> (the relative amounts of pigment 5 due to each class)
<OUTNAME>A.<PIG3> (the absolute amounts of pigment 3 due to each class)
<OUTNAME>R.<PIG3> (the relative amounts of pigment 3 due to each class)

Appendix B Distribution and installation

The PREPRO and CHEMTAX software can be obtained by arrangement with the authors who can be contacted at the address below:

Dr. D. J. Mackey
CSIRO
Division of Marine Research
GPO Box 1538
Hobart, TAS 7001
AUSTRALIA
Fax: +61 3 6232 5123
Email: Denis.Mackey@marine.csiro.au

PREPRO can be installed in any directory on a PC but can only run successfully if it has access to the appropriate data, pigment ratio and ratio limits files.

CHEMTAX and the Matlab™ .m files can likewise be installed in any directories providing the appropriate paths are set in Matlab™; see the Matlab™ command *help path* for instructions.

CSIRO Marine Laboratories

Division of Marine Research
(formerly the Division of Fisheries and the Division of Oceanography)

Headquarters
Castray Esplanade, Hobart, Tasmania 7001
GPO Box 1538, Hobart, Tasmania 7001, Australia

Queensland Laboratory
133 Middle Street, Cleveland, Queensland 4163
PO Box 120, Cleveland, Queensland 4163, Australia

Western Australia Laboratory
Leach Street, Marmion, W.A.
PO Box 20, North Beach, W.A. 6020, Australia

ISBN 0 643 06040 5
ISSN 0725–4598