COMPARISON OF CHEMOTAXONOMIC METHODS FOR THE DETERMINATION OF PERIPHYTON COMMUNITY COMPOSITION

by

Jamie L. Browne

A Thesis Submitted to the Faculty of

The Charles E. Schmidt College of Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, Florida

August 2010

COMPARISON OF CHEMOTAXONOMIC METHODS FOR THE DETERMINATION OF PERIPHYTON COMMUNITY COMPOSITION

by

Jamie L. Browne

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. J. William Louda, Department of Chemistry and Biochemistry, and has been approved by the members of her supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

J. William Louda, Ph.D.
Thesis Advisor

Erik G. Noonburg, Ph.D.

Scot E. Hagerthey, Ph.D.

J. E. Haky, Ph.D.

Rodney K. Murphey, Ph.D.
Chair, Department of Biological Sciences

Gary W. Perry, Ph.D.
Dean, The Charles E. Schmidt College of Science

Barry T. Rosson, Ph.D.
Dean, Graduate College

April 27, 2010
Date

ACKNOWLEDGEMENTS

I am deeply thankful for the encouragement and support of all my friends and my family. I would also like to thank the faculty of the Charles E. Schmidt College of Science, especially Dr. J. William Louda, Dr. Erik Noonburg, and Dr. Michael Salmon. The most important thing I have learned during my studies is that little can be truly accomplished without the help of one's colleagues, and so I gratefully acknowledge the support of all my fellow students, especially Bridget Armstrong, Simantha Ather, Cidya Grant, and Pam Mongkronsri.

ABSTRACT

Author: Jamie L. Browne

Title: Comparison of Chemotaxonomic Methods for the Determination of Periphyton Community Composition

Institution: Florida Atlantic University

Thesis Advisor: Dr. J. William Louda

Degree: Master of Science

Year: 2010

Pigment-based chemotaxonomy uses relative amounts of photosynthetic pigments (biomarkers) within algae samples to determine the algal class composition of each sample. Chemotaxonomy has been applied successfully to phytoplankton communities, but its efficacy for periphyton has not yet been established. This study examined the ability of simultaneous linear equations (SLE), CHEMTAX, and the Bayesian Compositional Estimator (BCE) to determine algal class composition in Florida Everglades periphyton. The methods were applied to artificial datasets, mixed lab cultures of known composition, and Everglades periphyton samples for which microscopic biovolume data was available. All methods were able to return accurate sample compositions for artificial data and mixed lab cultures. Correlation between pigment methods and microscopic results for natural periphyton samples was poor. SLE and CHEMTAX returned similar results for all samples while BCE performed less well.

DEDICATION

To Mary Grace, of course.

COMPARISON OF CHEMOTAXONOMIC METHODS FOR THE DETERMINATION OF PERIPHYTON COMMUNITY COMPOSITION

TABLES

FIGURES

INTRODUCTION

A truly efficient way to identify the constituents of a periphyton community has yet to be developed. Some categories of methods in use for algal communities in general are: traditional microscopic identification of preserved samples, in situ methods such as flow cytometry and fluorometry to count and identify cells, and chemotaxonomic methods (see, for example, Lee et al. 2005, Muylaert et al. 2006, and Greisberger and Teubner 2007; for summaries, see Millie et al. 1993 and Wright and Jeffrey 2006). Chemotaxonomic methods typically utilize high performance liquid chromatography and spectrophotometry to identify photosynthetic pigments and reconstruct the community based on the ratios of those pigments within particular algal groups. This reconstruction can be accomplished by several different methods including linear regression (Gieskes and Kraay 1983), simultaneous linear equations (Letelier et al. 1993), and numerical methods such as factor analysis (M. Mackey et al. 1996) and Bayesian/MCMC estimation (Van den Meersche et al. 2008). Each of these methods has its shortcomings (see, for example, Wright and Jeffrey 2006, Latasa 2007, and Van den Meersche et al. 2008). Numerically-based chemotaxonomic methods show great promise, and have proven especially useful for noncoastal, open ocean marine phytoplankton (see Wright and Jeffrey 2006 for recent examples). However, application of these methods to coastal marine phytoplankton (Lewitus et al. 2005; Louda 2008) is in the early stages, as is application to the relatively complex freshwater phytoplankton (although see Descy et al.

2000). A successful application of chemotaxonomy to periphyton has not yet been reported in the literature.

In this study, data from three years' worth of field samples of Everglades periphyton, computer-generated artificial data, and known mixes of laboratory-cultured algae were used to evaluate three different chemotaxonomic methods. A system of simultaneous linear equations currently in use in the FAU Environmental Biogeochemistry Laboratory, which has produced good results when applied to marine phytoplankton (Louda 2008), was compared to two computational methods: CHEMTAX (M. Mackey et al. 1996), a factor analysis application which has been increasingly used in the field of chemotaxonomy since its publication in 1996, and the Bayesian Compositional Estimator, a relatively new program which was designed in part to specifically address certain shortcomings in CHEMTAX (Van den Meersche et al. 2008). The comparison evaluated the three methods across several different levels to see which, if any, of the methods could adequately enumerate the periphyton community composition.

Background

In the Florida Everglades, a specialized group of algae called periphyton form one of the most important parts of the ecosystem's food web (Browder et al. 1981; Davis and Ogden 1994; McCormick and Stevenson 1998). Periphyton are a group of primarily phototrophic algae, including cyanobacteria, which live attached to structures and vegetation, sometimes in combination with other elements common to microbial mat communities such as sulfur-reducing bacteria and microconsumers. Over 1700 species of

algae currently comprise the Everglades periphyton community (S. Hagerthey, pers. comm. 2009). The study of periphyton can help resolve not only general questions regarding the ecosystem itself, but also specific questions about the area's water quality, nutrient shifts, anthropogenic impacts, and response to climate change. (Browder et al. 1994; McCormick and Stevenson 1998; Hagerthey et al. 2006; Iwaniec et al. 2006.)

The community composition of Everglades periphyton has long been of interest to researchers. Early on, it was discovered that the composition of this community was highly variable (Browder et al. 1981; Browder et al. 1994). The geographic area comprising the Everglades is subject to a variety of environmental gradients, such as wide ranges of total dissolved solids and total phosphorus in the water column, seasonal cycles, and various hydroperiod lengths. Variation has been observed along these gradients as different species are more or less tolerant of altered hydroperiods, eutrophication, absence or abundance of specific nutrients, and other environmental or anthropogenic stresses. Recently, it has also been suggested that algal evolutionary origin may be an additional important factor in the community composition shifts (S. Hagerthey, pers. comm.), as there are some areas where many algal species are of more recent marine origin. Algae can have very fast and varied responses to changes in growth conditions, and periphyton as a whole comprise many diverse organisms. This adds to a high spatial and temporal level of variability. (Browder et al. 1981; Browder et al. 1994; Vyzamal and Richardson 1995; McCormick and Stevenson 1998; Obeysekeera et al. 1999; Iwaniec et al. 2006.)

One of the consequences of the above for effective research and monitoring is the necessity for a fast, reliable method to determine periphyton community composition.

Traditionally, composition is determined by microscopic analysis: the identification of preserved specimens by various experts with an estimation of relative biomass, usually by cell counts or biovolume calculations. Although microscopy remains a reliable tool, it has certain drawbacks, notably the length of time required for multiple sample turnaround from collection to data output. This can often be several months. Sensitivity to different preservation methods, the necessity for experts in several subfields to make proper identifications, and potential underrepresentation of very small algae are other challenges of this method (Millie et al. 1993; Wright and Jeffrey 2006). Over the past four decades, a number of methods beyond the traditional microscopy have been developed to study algal and microbial communities (see summaries in Millie et al. 1993 and Wright and Jeffrey 2006). These methods have their own strengths and weaknesses, most typically a tradeoff of speed and resolution. For example, although many chemotaxonomic methods can return results within a few hours, the community constituents can typically be identified only to the class level, rather than to genus or species as is possible in microscopy. Most of these newer methods were first developed for marine phytoplankton inhabiting the open ocean. Although many of them can be extended to estuarine and freshwater phytoplankton, extension to periphyton presents more of a problem. The complications of substrates, mat structure, self-shading, and far more heterogeneous communities than are typical in phytoplankton present special challenges.

Chemotaxonomy

One potentially viable method for studying the makeup of a periphyton community is pigment-based chemotaxonomy. Chemotaxonomy is itself a combination

of several methods of analysis and results depend strongly upon which combination is chosen.

All chemotaxonomy depends first upon the existence of biomarkers. Here, a biomarker is defined as a molecule which occurs in a distinctive relationship to a class of organisms. A photosynthetic or photoprotective pigment that is found only in a certain type of algae is one common example of a biomarker. Presence of the biomarker is indicative of the presence of the organism or group of organisms. Further, the ratio of the biomarker to some common compound, such as chlorophyll-*a* in the case of photosynthetic organisms, can be used to estimate the relative amount of the organism in the sample.

For each chemotaxonomic problem, an appropriate set of biomarkers is quantified through some type of analysis of the sample. Once the biomarkers are determined for the samples, the mathematical aspect of chemotaxonomy begins. For this, we begin with 1) a set of known or estimated biomarker ratios and 2) the set of biomarker quantities we find within the sample. Ideally, the end result is some enumeration of the groups of organisms within the sample.

A wide literature exists on the use of photosynthetic pigments as biomarkers (esp. see Millie et al. 1993, Jeffrey et al. 1997, and Wright and Jeffrey 2006). Central to the development of pigment-based chemotaxonomy has been the implementation and refinement of high performance liquid chromatography (HPLC) (Mantoura and Llewelyn 1983) combined with realtime spectrophotometric detection (photodiode array, also known as PDA), along with improvements in pigment extraction procedures. This has allowed increasingly efficient quantification of a wide variety of pigments and thus an

increasingly accurate, and fast, picture of the presence and quantity of pigment biomarkers within field samples. (Gieskes and Kraay 1983; Wright et al. 1991; Kraay et al. 1992; Millie et al. 1993; Hagerthey et al. 2006.)

Along with a more accurate estimation of biomarker presence and quantity, more information is also being obtained by researchers about the biomarker ratios. Ratios of biomarker pigments to chlorophyll-*a* which are typical of certain algal classes have been researched for a variety of ecosystems (some examples are: M. Mackey et al. 1996; Descy et al. 2000; Lewitus et al. 2005; Lionard et al. 2008). These ratios can depend quite heavily on which species are predominant in a given algal class within an area. An open question remains as to how environmentally specific ratio sets need to be. For example, many pigment : chlorophyll-*a* ratios will change under different light conditions. This is a particularly intensive area of research in estuarine and freshwater ecosystems (see, for example, Grant 2006, Schlüter et al. 2006, Greisberger and Teubner 2007, and Guisande et al. 2008). Ratios are typically determined by analysis of laboratory cultures and by linear regression of field sample data.

Models

For clarity, in order to clearly distinguish them from the study methodology, mathematical chemotaxonomic methods will be termed models throughout this text. Although generally they are not designated as models in the literature, these methods all use mathematical relationships to describe real systems and, in this sense, use of the term is appropriate. The three chemotaxonomic models considered in this study were: simultaneous linear equations (SLE), CHEMTAX, and the Bayesian Compositional

Estimator (BCE). SLE is the system currently in use at the FAU Environmental Biogeochemistry Laboratory. SLE has shown promising results when applied to marine phytoplankton in Florida Bay (Louda 2008) and work is ongoing to extend the model to other algal/microbial communities present in South Florida ecosystems, including Everglades periphyton. CHEMTAX was chosen due its widespread use and increasing application to systems other than just marine phytoplankton (M. Mackey et al. 1996; Wright and Jeffrey 2006). BCE is a new, relatively untested model. It was chosen due to its open availability, newer methods taking advantage of current computing capabilities, and a different theoretical approach than CHEMTAX (Van den Meersche et al. 2008).

The basic problem for the models is the determination of the algal class composition of a sample given only biomarker amounts and expected ratios of biomarkers within classes. SLE models apply a series of straightforward equations to the problem, using one ratio of one biomarker to chlorophyll-*a* for each algal class in the sample. This method has only one possible answer. If the ratios and the algal classes used are accurate and complete for the sample, the answer will be correct.

In contrast to solving a simple equation with only one possible answer, numerical methods (such as CHEMTAX and BCE) are designed to sample a solution space in search of the best answer. These types of models are much more flexible in terms of handling errors in the ratio values, using what Omlin and Reichert (1999) term regional parameter estimation rather than point parameter estimation. However, they also run the risk of finding "false" solutions—solutions which, although they may be mathematically viable answers to the problem, are not the overall best answer. Another way to state this is that the model may locate a local best solution rather than a true global best solution.

One of the biggest challenges in designing numerical models is how to best sample the solution space: enough space should be sampled randomly to find the global best solution, without having to sample so much of the space that the problem becomes intractable in terms of computing resources and time. Although similar in many respects, CHEMTAX and BCE approach this particular question from different theoretical bases.

The general equation for chemotaxonomic SLE (where rat is the pigment : chlorophyll-*a* ratio for a given algal class, pig is the amount of pigment in a sample, and chl is the amount of chlorophyll-*a* attributable to a given algal class within the sample) is:

$$chl_1 = rat_1 * pig_1$$

$$chl_2 = rat_2 * pig_2$$

$$\ldots$$

$$chl_n = rat_n * pig_n$$

with the result that, for each sample:

$$\sum chl = rat_1 * pig_1 + rat_2 * pig_2 + \ldots + rat_n * pig_n$$

This method calculates the chlorophyll-*a* which is contributed by each algal class as a fraction of the total chlorophyll-*a.* Although not a direct measure of biomass, the taxon-specific chlorophyll-*a* can be used as an estimate of amount of specific taxa in each sample. If the SLE model has been correctly parameterized–that is, if the ratios are correct and the correct algal classes have been included–the calculated sum of

chlorophyll-*a* per sample should equal the amount of chlorophyll-*a* which was actually measured in that sample. (See Louda 2008 for an application of this method).

Another approach to chemotaxonomy is to formulate the problem as a linear inverse equation, which is what both CHEMTAX and BCE do. The general equation is set up as **AX** = **B**. **A** is a matrix of biomarker pigment : chlorophyll-*a* ratios within each algal class (the ratio matrix), where element a_{ij} is the ratio of biomarker *j* to chlorophyll-*a* in algal class *i.* **B** is a matrix of the amount of biomarker pigments within each field sample (the data matrix), where element b_{ij} is the amount of biomarker *j* in each sample *i.* **X** is a matrix of algal class compositions within each sample, where element x_{ij} is the percentage of algal class *j* in each sample *i.* This is an inverse problem because in order to find the solution the problem must be formulated as $\mathbf{X} = \mathbf{B}\mathbf{A}^{-1}$. The data matrix **B** will be known (the results of pigment measurements of samples), the ratio matrix **A** is also considered a known, and the composition matrix **X** (the "answer" to the problem) is entirely unknown. The problem is considered overdetermined if there are more biomarkers than algal classes, and in this case a best answer for **X** can be found. In contrast, underdetermined problems will have an infinite number of solutions. Chemotaxonomic linear inverse problems are subject to certain constraints. Notably, in the composition matrix **X**, there must be no negative elements and each row (the compositional elements of each sample) must equal 1.

Although mathematically both **A** and **B** are treated as knowns, a major consideration for chemotaxonomy in practice is that both matrices incorporate some uncertainty. The data matrix **B** will typically incorporate a small and relatively predictable amount of error. For example, this error may include the deviations possible

in the HPLC analysis. The ratio matrix **A**, in contrast, will be nearly entirely estimated, with far more confidence in some elements than in others. A major difference between the numerical models considered in this study is how they treat this uncertainty.

A straightforward approach to solving linear inverse problems is the use of least-squares methods. The equation **AX** = **B** will always incorporate an error term, since **AX** will never be exactly equal to **B** in this application of the problem. The least squares method finds the solution for **X** that minimizes this error term. In many least-squares methods, including CHEMTAX, **A** and **B** are allowed to vary as well. Although least-squares approaches to chemotaxonomic problems have been reported for over two decades (for example, see Letelier 1993), CHEMTAX was the first published development of a computational, numerical approach to the problem. In more general terms, the type of linear inverse problem which CHEMTAX sets up is actually a form of factor analysis, a method which has a long applied history in chemistry in particular, including chromatography (see Malinowski 2002 for a detailed description of factor analysis as applied to analytical data). CHEMTAX uses a form of gradient descent algorithm called steepest descent to determine each step in the search for a solution. Steepest descent is not considered a particularly efficient algorithm since it may take a large number of small steps to reach a solution (Press et al. 1992). Chemotaxonomic problems, however, are generally small and tractable enough, in terms of dimensions and matrix size, that the performance difference between this and more efficient algorithms is likely negligible.

In contrast to the straightforward least-squares methods used by CHEMTAX, BCE uses Bayesian inference to set up the problem, which is then solved using Markov

Chain Monte Carlo (MCMC). Unlike traditional approaches, which determine the probability of the data given the model, Bayesian methods are designed to determine the probability of the parameters (essentially, the model) given the data, treating both the observed data and the parameters as random variables. Note, however, that both traditional and Bayesian methods make the same assumption that a set of best, or true, parameters do exist. (O'Hagan and Forster 2004; Ellison 2004; Clark 2005.)

In practical terms, Bayesian inference is primarily distinguished by the formal inclusion of a prior. The prior is information the investigator has about some part of the probable results. In the case of chemotaxonomy, this prior information is in the form of 1) the expected possible presence of certain taxa in the sampling environment and 2) the ratio of biomarker pigment(s) to chlorophyll-*a* expected in specific algal classes. Any form of factor analysis, including the CHEMTAX program, must use this information, which is often thought of as a "training," "reference," or "expected" matrix (Malinowski 2002). Therefore, the difference between the BCE and the CHEMTAX methods is more in the mathematical treatment of this prior information than in its inclusion as such. Press et al. (1992, p. 808), writing in the classic *Numerical Recipes,* go so far as to state, in their section covering problems of this type, that: "An outsider looking only at the equations that are actually solved, and not at the accompanying philosophical justifications, would have a difficult time separating the so-called Bayesian methods from the so-called empirical ones, we think."

In BCE, the prior information is incorporated in the form of probability distributions for the sample data and the ratios. These are then used to determine a probability distribution for the algal compositions of the samples. For the exact

development of the method in terms of Bayesian probability, see Van den Meersche et al. (2008). A basic difference between BCE and CHEMTAX lies in the inclusion of these probability distributions. CHEMTAX makes de facto assumptions of a normal distribution for the data, and, most importantly, a uniform distribution for the ratios. Since CHEMTAX input specifies only an upper and lower limit for each ratio, all points between the limits are treated as having an equal probability of being the correct ratio. BCE authors make the point that most ratios will have a "most likely value" with some deviation, rather than a truly random value falling between two limits. The exact probability distributions can be altered in the BCE parameters, but the defaults for the data and ratios are a gamma-distribution. The gamma distribution is often used in numerical applications of this type. Given certain parameters, it approximates a log-normal distribution but avoids some of the numerical problems the log-normal distribution has with very low values. (Van den Meersche et al. 2008; Van den Meersche and Soetaert 2009.)

Adequately and efficiently sampling the posterior probability distribution, or in other words, searching the solution space for the answer, has historically been one of the biggest difficulties in applying Bayesian methods. Markov Chain Monte Carlo techniques have now become a standard strategy. Although these methods have been in use since the 1940s, increases in computing efficiency have allowed a far more widespread application (Hammersley and Hanscomb 1964, Gilks et al. 1996). A Markov chain searches a sample space by choosing a new point (X_t) based only on the immediately previous point (X_{t-1}). The chain therefore has no memory prior to the one previous step. Some method will be used to determine whether or not each point is accepted as the chain moves. If the point is

not accepted it is simply disregarded, or "forgotten," and the previous point remains X_{t-1}. In BCE, acceptance of new points is determined by applying the classic Metropolis-Hastings algorithm. Metropolis-Hastings uses a proposal distribution, which does not necessarily have to match the target distribution, to sample candidate points. The points are accepted according to a probability which depends on the proposal and target distributions (see Gilks et al. 1996). As a default value, BCE uses the initial least-squares solution to the problem as the starting point for the chain. As with CHEMTAX, BCE results include revised ratios and various measures of uncertainty. BCE results also include estimates of covariance.

The following sections discuss in detail the datasets used to test the models, as well as outline how the models were set up, parameterized, and applied to the data. Results were evaluated to determine which, if any, model could adequately describe particular algal classes as well as the periphyton community composition as a whole. Models were also evaluated across several scales for relative strengths and weaknesses. An attempt was made to find both the best overall ratio set as well as the best overall model.

METHODS

Field samples of Everglades periphyton, computer-generated artificial data, and artificial mixes of algal cultures were used to evaluate the ability of different models to determine algal class composition in periphyton. Three models were chosen for comparison: one fixed-coefficient method (simultaneous linear equations) and two numerical methods (CHEMTAX and the Bayesian Compositional Estimator). The numerical methods were tested first on artificial datasets, then on a series of data from mixed lab cultures of known composition. All three methods were used to analyze periphyton samples from five areas in the Everglades, including a subsample for which microscopic results were available. Pearson's product-moment coefficient (r), was used to test how closely the different methods matched the algal class composition.

Datasets

Field Sample Data

Periphyton samples were gathered by the South Florida Water Management District (SFWMD) over a three-year span and forwarded to the FAU Environmental Biogeochemistry Laboratory for analysis. The samples were from a set of sites which spanned most ecological gradients within the Everglades ecosystem. Sites were sampled as close to trimonthly as possible, depending on the hydrology of the site. The bulk of the

samples were taken from periphytometers, which are glass slides acting as artificial substrate placed in situ and allowed to accumulate algal growth for a certain period of time. Periphytometers are advantageous in their minimization of confounding factors such as a previous substrate and unknown growth period. There is, however, the possibility that they may not completely accurately represent actual growth in the field. A certain number of "grab samples" were also taken. These were true samples of the algal mat which is often much thicker than the relatively short-term growth on periphytometers. Grab samples therefore included cells from more of the lower-light layers and thus had the potential for higher variability in pigment ratios within algal groups per sample.

Periphytometer samples used in this study were collected between March 2005 and December 2008. Grab samples were collected in September, October, and November of 2006. Areas represented were, from north to south, Water Conservation Areas (WCA) 1 (Arthur R. Marshall Loxahatchee National Wildlife Refuge), 2A, and 3A, and both Shark River and Taylor Sloughs. Within each of these areas, sampling sites were located along several transects. Table 1 shows a breakdown of samples per area. Data from a total of 475 periphytometer samples and 29 grab samples were used in the present analysis.

After collection, samples were light-protected and were frozen as soon as possible. Samples were stored for varying lengths of time depending on the date collected, and were transported back to Florida Atlantic University for analysis on an ongoing basis.

Pigment Analysis. Pigment analysis was performed in the FAU Environmental Biogeochemistry Laboratory by Panne Mongkronsri upon receipt of each batch of field samples. Each sample was analyzed according the following protocol. Samples were vacuum-filtered through Whatman glass-fiber filters which were then folded, blotted, frozen, and subsequently freeze-dried. Freeze-dried samples were extracted with a solution consisting of an internal standard and a 90% aqueous methanol / acetone / dimethylformamide (MAD) solvent by grinding, sonicating for 1-2 minutes in an ice bath, steeping for 1 hour, and sonicating again. The mixture was then centrifuged, decanted, and filtered using a 0.45μm filter. All extraction procedures were performed in dim yellow light as close to freezing temperatures as possible. The filtered extract was combined with an ion-pairing solution and analyzed using reversed-phase high performance liquid chromatography combined with photo-diode array detection (HPLC/PDA). For the first two years of data collection, chromatographic detection was performed with Waters 990 or 996. All analyses were performed under dim yellow light. (For a detailed description of this protocol, see Hagerthey et al. 2006 and Louda 2008.)

Data returned by each analysis is a list of the presence and amount of up to 47 photosynthetic pigments. This total includes biomarkers and potential biomarkers as well as chlorophyll-*a* and its breakdown products (Table 2). Total chlorophyll-*a* is determined by summing the values for chlorophyll-*a* and its measured derivatives. Also calculated is the percentage of chlorophyll-*a* and derivatives such as the pheopigments. Analytical data are maintained in printed hard-copy and digital form. Each analysis was stored digitally on an individual Excel spreadsheet. Initial input data, prior to any Excel calculation, was

single-precision as dictated by the HPLC computational output. All subsequent calculations were performed at double precision.

For this study, sampling data and the pigment weights from the analysis for all measured pigments were pulled from each spreadsheet and stored in a separate Excel spreadsheet. (For details of this process, please refer to Appendix A, Field Sample Data Preparation.) This spreadsheet was used to construct the data matrices for all models as needed.

Microscopy. Microscopic identification of a subset of the field samples (n=211) was outsourced to the Florida Department of Environmental Protection (FDEP) laboratory in Tallahassee by the South Florida Water Management District (SFWMD). The resulting identification data was then supplied to the FAU Environmental Biogeochemistry Laboratory. This included periphytometer cell counts (which count all cells and 10-micron filament lengths individually) and unit counts (which count each colony or filament as one) for 17 categories of algal groups. These groups were combined to provide relative abundance estimates for chlorophytes, cryptophytes, cyanobacteria, diatoms, and dinoflagellates. For the grab samples (n=122), relative abundance in relationship to biovolume, rather than cell/unit counts, was provided. These estimates were based on microscopic identification and listed the relative biovolume per sample for seven categories of algae (diatoms, desmid and non-desmid greens, and four categories of cyanobacteria), which were then combined into total cyanobacteria, chlorophytes, and diatoms. Cryptophytes and dinoflagellates were either not present or not included.

Artificial Data

In order to check that the numerical programs were set up correctly and to determine the best parameterization for them using our ratio matrix prior to any evaluation of real data, an R (R Development Core Team 2009) function was constructed which would generate artificial datasets. (Although it is relatively easy in R to construct functions which will generate matrices of random numbers, generation of random compositional data is more difficult, and the assistance of committee member Dr. Erik Noonburg in building this function is greatly appreciated.) Our function, "build.x," uses a ratio matrix, a coefficient of variation (cv), and a value for desired number of samples as input. The output of the function is a data matrix of biomarker pigment weights per sample and a matrix of algal composition per sample. The output matrix dimensions are determined by the input ratio matrix dimensions (pigments x algal classes) and by the number of samples specified. The original ratio matrix and the generated data matrices were input into CHEMTAX and BCE and the resulting sample compositions were compared to the sample composition generated by the R function.

The function (see Appendix B for the basic code) first creates a new ratio matrix which incorporates the error factor specified by the input cv. To do this, the rnorm function is used to generate random numbers under a normal distribution, with a mean of 1 and standard deviation equal to the cv value, and the resulting matrix of error factors is multiplied by the original ratio matrix.

The second step in the data generation creates random compositions for the desired number of samples. The function runif generates random numbers drawn from a uniform distribution between 0 and 1, after which a sort and subtraction method is

employed to force the composition amounts to be proportional (equal to 1) within each sample.

Artificial data for input into CHEMTAX and BCE was created by multiplying the adjusted ratio matrix (incorporating error) with the composition matrix to create a series of pigment weights per sample. For all artificial data generation, the largest ratio matrix, comprising six algal classes, was used. The true composition of the samples was not viewed prior to any data runs.

This method did have one possible drawback in that algal class amounts of zero were not well represented: almost every generated sample had some amount of each of the algal classes specified in the ratio matrix. In actual Everglades periphyton samples it is not uncommon for at least one, and often two to three, of the six algal classes measured here to be absent. To check for any large effect of this on the model performance, a small set of artificial samples (4 sets of 5-10 samples each) was calculated by hand using simple proportions which included zeroes. Initial tests did not show any obvious difference in handling by the models, so all further runs utilized the R-generated data. This testing was only employed in order to check the usability of the artificial data. More sophisticated evaluations of the effect of zeroes in the sample compositions were undertaken using the culture mixes.

Artificial Mixes

Three unialgal cultures were procured and cultured for two weeks under a medium light regime in the FAU Environmental Biogeochemistry Laboratory by Panne Mongkronsri, who then created known mixes of the cultures and determined the pigment

content of each mix. This data was used to test the different models' ability to derive the correct algal class composition of the samples from the pigment measurements.

Pure cultures of a coccoidal cyanobacteria (*Anacystis* sp., syn. *Synechoccus*), a chlorophyte (*Scenedesmus* sp.), and a diatom (*Cyclotella* sp.), were obtained from either Carolina Biological Supply (Burlington, North Carolina) or the University of Texas (UTEX) collection (Austin, Texas) in 2009. After two weeks of growth, the cultures were mixed in known ratios by cell number (Coulter Counter) to create five different mixes with three replicates each. The cell ratios in the mixes were 1:1:1, 1:1:3, 1:3:1, 3:1:1, and 1:3:3 (*Anacystis* : *Scenedesmus* : *Cyclotella*), respectively. Pigment content of the pure cultures and of each mix replicate was determined using the same methodology and instrumentation as that used for the field samples.

Initial analysis of the data with the SLE method by Ms. Mongkronsri indicated that the cell-count ratios were not directly comparable to the chemotaxonomic results due to a much larger difference in chlorophyll-*a* content per cell per species than previously suspected. As a result of this finding, the cell count ratios were corrected by a factor of 23:4:1, reflecting the relative amount of chlorophyll-*a* per cell (*Anacystis* : *Scenedesmus* : *Cyclotella*), and this taxon-specific chlorophyll-*a* estimate was used for comparison with the chemotaxonomic results. The mixes were assigned random identification numbers prior to analysis with chemotaxonomic methods in order to reduce any unconscious bias when setting up the data runs.

Models

Simultaneous Linear Equations

Simultaneous linear equations (SLE) is a chemotaxonomic method that calculates algal class composition of one sample at a time using a set of biomarker coefficients in which each biomarker is unique to one algal class. Unlike numerical methods, in which the coefficients are allowed to vary within certain parameters, in SLE the coefficients cannot be altered during the calculations Therefore, SLE is sometimes called the fixed-coefficient method. Each coefficient is the estimated ratio of a biomarker pigment to chlorophyll-*a* which is considered typical in a given algal class. The amount of each biomarker pigment in the sample is multiplied by its respective coefficient to determine the estimated amount of chlorophyll-*a* contributed by each algal class (the taxon-specific chlorophyll-*a*). Each of the estimated class contribution amounts is divided by the sum of all estimated contributions to arrive at an algal class composition in percentage form.

SLE was used to determine algal class composition on all field and artificial mix samples using MS Excel (v. 2007, running on a Dell PC with a 1.6 GHz Pentium-4 processor and 256MB RAM under WindowsXP sp3). Although these equations were calculated for the field samples on each of the original pigment calculation spreadsheets (see Louda 2008 and references therein) prior to data compilation (see the Field Sample section above, and Appendix A), calculations were set up separately and redone on the compiled data for two reasons. The initial calculations used moles rather than weights, which, although more sound from a biochemistry standpoint, did not allow direct comparison to most of the literature, which uses pigment weights. In actual practice, algal

class compositions derived from molar percentages of pigments will seldom differ substantially from those derived from weight percentages of pigments. Setting up the data to allow easy recalculation also allowed more flexibility in using different sets of coefficients or choosing to handle specific coefficients differently.

The SLE method requires relatively few decisions as to input. Choices to be made prior to running this model are: which algal classes should be included, which pigments will be chosen for biomarkers of specific algal classes, and what ratios of pigment to chlorophyll-*a* should be used as the calculation coefficients.

Algal Classes. A set of six algal classes comprising the periphyton groups typically found in the Everglades was used in this model. These were: chlorophytes (including both desmid and non-desmid green algae), cryptophytes, coccoidal cyanobacteria, filamentous (including some calcium-precipitating) cyanobacteria, diatoms, and dinoflagellates. It is possible to break this down into subsets, since there are often predictable differences in the relative abundance and occurrence of these groups throughout the Everglades, especially between the northern and southern Everglades. The SLE model, however, works best when applied to the widest set of possible algal classes. If a biomarker for an algal class is absent in the sample, that term in the equation is simply set to zero and a corresponding zero is returned for the class composition. Since there is a one-to-one correspondence in SLE between biomarkers and algal classes there is no ambiguity in this process. In contrast, numerical methods can attempt to compensate for a zero value, particularly when it occurs in combination with some amount of a shared biomarker, and can return a compositional value for an algal class which is actually absent. For these

methods, as will be discussed below, a truncated group of algal classes works best, where each model run is only applied to the subset of classes which are estimated to be present based on biomarker pigment presence.

Biomarker Pigments. The biomarkers chosen follow those that are standard in the literature and are the biomarkers currently in use by the FAU Environmental Biogeochemistry Laboratory (see Wright and Jeffrey 2006 and Louda 2008). Table 3 shows a comparison of biomarkers and ratios used in recent freshwater and estuarine chemotaxonomic studies. The biomarkers used in this study for SLE can be found in Table 4.

Since SLE has no provision for shared pigments, only unique or nearly unique biomarkers can be used. Non-unique biomarkers are those where the pigment occurs in much larger amounts, or more consistently, in one class of algae (for examples, see Jeffrey et al. 1997, Table 2.3). In these cases, the pigment ratio for the primary group was the only one considered. In other words, all of this pigment present in the sample was assumed to belong to the one primary algal class. Some biomarkers are best considered as the sum of two or more individual, though metabolically related, pigments. For example, the biomarker for diatoms, fucoxanthin, was considered to be the sum of three pigments: fucoxanthin, fucoxanthinol (desterified fucoxanthin), and cis-fucoxanthin (isomerized fucoxanthin). Similarly, chlorophyll-*b,* the biomarker for chlorophytes, was considered to be the sum of chlorophyll-*b* and chlorophyllide-*b* (chlorophyll-*b* having lost the phytol ester).

Occasionally, a shared pigment may be encountered which can be used by means of an adjusted value. In the case of the six algal classes considered here, the pigment zeaxanthin necessitated this type of adjustment in order to consider two distinct types of cyanobacteria. Zeaxanthin is a major pigment in both coccoidal and filamentous cyanobacteria, while the pigment echinenone is found primarily in filamentous cyanobacteria. Since SLE can only consider zeaxanthin in one algal subset, a subtraction method was used which made a working assumption that filamentous cyanobacteria would contain an equivalent amount of echinenone and zeaxanthin. This method took the total zeaxanthin in each sample, subtracted the amount equivalent to echinenone in the same sample, and assigned the remaining zeaxanthin as the biomarker for coccoidal cyanobacteria. Although this method is not without precedent in the literature, there are shortcomings. For example, the assumption of equivalent amounts is not always valid and, in particular, negative numbers must be avoided. For all model analyses detailed here, any subtraction which generated a negative number was forced to zero. Some other ramifications of use of this method are noted in the CHEMTAX and BCE sections below.

Ratios. The initial set of coefficients for SLE was originally developed in a study of phytoplankton samples from Florida Bay (Louda 2008). In the Florida Bay study, estimates of biomarker to chlorophyll-*a* ratios were derived from unialgal laboratory cultures and refined during analysis of field samples. The final ratios (coefficients) performed well (estimated / measured total chlorophyll-*a* ≈ 0.9-1.1) when applied to a set of simultaneous linear equations to determine algal class composition of Florida Bay phytoplankton. These coefficients were considered a valid starting point for the

periphyton ratios since, although they were developed for coastal phytoplankton, they include families typically found in the periphyton and have returned reasonable composition values in initial studies in the laboratory based on contributed chlorophyll-*a*. These coefficients were used for the first run of SLE (Table 4, top).

These initial coefficients were refined in a series of light-response studies performed by Ms. Cidya Grant in the FAU Environmental Biogeochemistry Laboratory (Grant 2006). Grant cultured 10 algae species at different light levels and measured pigments at 2- to 5-day intervals over the course of 2 weeks of growth. Pigments were extracted and measured using the same procedures as those described in the field sample analysis section above. Algae species and light levels were chosen specifically to gain insight on the Everglades ecosystem. Therefore, the light levels chosen matched those actually measured in the Everglades, from a low-light range of 150-180 μmoles photons m^{-2} s^{-1} (representative of surface light on a cloudy day in winter) up to 1800 μmoles photons m^{-2} s^{-1} for the high-light level (surface light with no attenuation on a very hot summer day). Grant's work resulted in revised pigment : chlorophyll-*a* ratios for chlorophyll-*b* in chlorophytes, fucoxanthin in diatoms, and peridinin in dinoflagellates (Grant 2006). This set of coefficients was used for a separate second run of SLE (Table 4, bottom).

CHEMTAX

CHEMTAX is a factor analysis program developed in 1996 (M. Mackey et al. 1996) and licensed through CSIRO Marine Laboratories. It is written to run inside MATLAB (The MathWorks, Inc. 2008). Although revised versions of CHEMTAX (v.

1.95, v. 2) have been referenced by the program's developers (Wright and Jeffrey 2006, Wright et al. 2009), they were not available at the time of this study and the original 1997 release (v. 1.0) was used. Since the major revisions in the program are reported to be in the interface, leaving the algorithms unchanged (Wright et al. 2009), it is expected that the results obtained by the original version should not be substantially different than those that would be produced by the newer version. The data preprocessing program, PrePro, that ships with CHEMTAX would not run correctly on any of the computers tested and so was not used. CHEMTAX was run inside MATLAB version R2008b, using a Dell PC with a 1.6 GHz Pentium4 processor and 256MB RAM under WindowsXP sp3.

CHEMTAX works by evaluating groups of samples, with pigment data arranged in matrix form. Biomarker ratios to chlorophyll-*a* are also arranged in a matrix. Using the (unknown) algal class composition of the samples as a third matrix, this forms a linear inverse problem which is solved by matrix factorization, using a straightforward algorithm to provide the least-squares solution (M. Mackey et al. 1996). Unlike SLE, which takes only sample data and biomarker ratios as input, CHEMTAX allows input as to how much and which ratios are allowed to vary and how the data are weighted. Several other input options control more specifically how the calculations are run. Results from CHEMTAX include not only the algal class composition of each sample, but also a revised ratio matrix, residuals (from the least-squares calculations), and, if requested, breakdowns of pigments assigned to each algal class within each sample and more information regarding the iterative calculation process.

Choices which must be made prior to running CHEMTAX are: which algal classes should be included, which pigments will be chosen for biomarkers of specific

algal classes, what ratios of pigment to chlorophyll-*a* should be used for the calculation coefficients, how should sample groups be constructed, and how to set the program parameters.

Setting Up the Program. The CHEMTAX program is a series of individual MATLAB (.m) files which include the main CHEMTAX routine as well as separate files for the various algorithms, or functions, used in the operations. These files are all called from the main chemtax.m file. General instructions for setup are contained within the CHEMTAX user's manual (M. Mackey et al. 1997). Prior to setup, a working version of MATLAB must be installed on the computer. For the present study, a bare-bones student version (v. R2008b) proved to be more than adequate. The algorithms which CHEMTAX uses are, with one exception, functions which are built in to MATLAB. However, the CHEMTAX-supplied functions will be used if the MATLAB path is set to the directory containing the CHEMTAX files. The one exception, nnls (nonnegative least squares), has been replaced in newer versions of MATLAB by the lsqnonneg function, and the authors recommend editing the files to call this newer function (D. Mackey and Higgins 2001). For this study, the path was set to use the CHEMTAX-supplied functions and two minor alterations were made in the .m files to avoid MATLAB error statements: these were limited to correcting case to avoid a "case-insensitive" error when calling files, and commenting out a request for number of flops which is no longer supported.

Although MATLAB will automatically call the lsqnonneg function when it encounters a call for the older nnls function, this method does not allow the user to set any function options, and the default lsqnonneg options often prevented the program

from converging on an answer. However, it was found that convergence was only assured when setting the function tolerance very high, which was unsettling in that it allowed (albeit very small) negative values to be returned in the composition. Since CHEMTAX does supply a version of nnls, the program was edited to call this function and results from artificial data using both methods were compared. Other than the difference accounted for by the increase in tolerance in lsqnonneg, there was no discernable difference in the results, and therefore the nnls function was retained.

Data and parameters must be furnished to the program in a separate .m file. To facilitate construction of this file, CHEMTAX authors include a stand-alone program, PrePro, which builds a .m file for the user based on a simple interactive interface, running inside an MS-DOS window. However, this study found PrePro to be incompatible with newer PC architecture, freezing at 100% of the CPU resources when attempting to process anything but the default ratio matrix. Accordingly, .m files were written as needed using the instructions supplied in the documentation (Appendix A in M. Mackey et al. 1997).

Algal Classes. The maximum algal classes considered were the same six as for SLE: chlorophytes, cryptophytes, coccoidal cyanobacteria, filamentous cyanobacteria, diatoms, and dinoflagellates (see Table 5). Unlike SLE however, CHEMTAX does not work well when algal classes are included which are actually absent in the samples (M. Mackey et al. 1997). Runs made with artificial data and artificial mixes confirmed that CHEMTAX often returns a compositional value for a class which is absent, if that class was included in the initial ratio matrix. Ratio matrices were therefore constructed for certain subsets of

the six algal classes. The exact subsets were determined from evaluation of sample pigment signatures. Evaluation by pigment signatures is done by examining samples for presence or absence of biomarkers rather than by ecological considerations. Absence of a unique biomarker is taken as absence of that algal class. For example, a sample without any alloxanthin measured was assumed to contain no cryptophytes. Where no clear pigment signature was apparent, the maximum algal class configuration was used as a default. See Table 6 for examples of algal class subsets used.

Biomarker Pigments. CHEMTAX recommends that more than one biomarker pigment be used per algal class if possible. The authors also caution that the number of pigments must be greater than the number of algal classes in order to allow the mathematical problem to be effectively solved (M. Mackey et al. 1996; M. Mackey et al. 1997). However, when considering secondary biomarker pigments for inclusion, a tradeoff becomes apparent. The increase in the number of biomarkers often necessitates the use of less stable ratios. For example, lutein, although a major pigment in chlorophytes, is highly sensitive to varying light regimes, making it harder find a consistent, or stable, value for the pigment to chlorophyll-*a* ratio (Grant 2006). Thus, while secondary pigments enhance the ability of the program to effectively calculate compositions and possibly adjust ratios, they also increase the original uncertainty. Many researchers working in estuarine and freshwater systems appear to have found it difficult to strictly follow the CHEMTAX recommendations for two or more biomarkers per algal class (see Table 3). For our system as well, it was not always possible to locate stable secondary pigments with well-known pigment to chlorophyll-*a* ratios.

For the CHEMTAX biomarkers, the original set of SLE biomarkers was utilized, along with three additional pigment groups and an additional allocation for one of the original pigments (Table 5). Along with peridinin from the original coefficient set, dinoxanthin was used as a biomarker for dinoflagellates. Diatoxanthin and diadinoxanthin were used as their sum due to rapid cycling between the two pigments and were allocated in differing amounts to dinoflagellates and diatoms, supplementing the original biomarker pigments of fucoxanthin for diatoms and peridinin for dinoflagellates. Lutein was added as a secondary biomarker for chlorophytes. Zeaxanthin was allowed a small allocation to chlorophytes, in addition to its use as a primary biomarker for coccoidal cyanobacteria.

Although β,ε-carotene (also known as α-carotene) was initially considered as a secondary biomarker for cryptophytes, inspection of the field sample data showed that this pigment occurs in very small amounts relative to alloxanthin, the primary biomarker for cryptophytes, and was in fact absent (or at least not measurable) in over half the field samples where alloxanthin was measured. The authors of CHEMTAX caution against this scenario (D. Mackey and Higgins 2001, M. Mackey et al. 1997). Violaxanthin was initially considered as well, as a biomarker for chlorophytes. However, it was dropped from consideration due to its extreme sensitivity to varying light levels. It was felt that it was not possible to arrive at one overall violaxanthin : chlorophyll-*a* ratio that would have accounted for all varying light levels under which the field samples might have been collected, even allowing for the flexibility in ratios which are built in to the numerical routines. A different approach to this problem would have been to include violaxanthin only in combination with its xanthophyll-cycle (see Hager 1980; Demmig-Adams 1990)

counterparts (neoxanthin and antheraxanthin), similar to the approach that was chosen for diadinoxanthin and diatoxanthin. However, since chlorophytes were already assigned three pigments (chlorophyll-*b*, lutein, and zeaxanthin), this was not considered necessary. Further, violaxanthin is present in various *Chromophyta* groups (Jeffrey et al. 1997) and this would have been an added complication in the ratio matrix construction.

Zeaxanthin was somewhat problematical. The subtraction approach was retained in order to facilitate comparison with results from SLE. However, this method does have some shortcomings. Notably, it assigns a set value of zeaxanthin to filamentous cyanobacteria prior to any assignation to other algal classes (namely, coccoidal cyanobacteria and chlorophytes). This poses a danger that when the echinenone : zeaxanthin ratio is more than 1:1 in filamentous cyanobacteria, zeaxanthin which truly belongs to another algal class will not be included in the data. However, when both the subtraction method and an alternate approach utilizing the original zeaxanthin values were employed in BCE (see following sections), the final composition results were very similar.

The end result of the pigments and pigment groups which were chosen as biomarkers, shown in Table 5, was a ratio matrix with single biomarkers for cryptophytes, coccoidal cyanobacteria, and filamentous cyanobacteria, two biomarkers for diatoms, and three biomarkers each for chlorophytes and dinoflagellates. Only two of these biomarkers (‘diadinoxanthin plus diatoxanthin’ and zeaxanthin) were allocated to more than one algal class.

Ratios. Unlike SLE, CHEMTAX requires some consideration of the biomarkers as a group and not just individually. Since the ratios are now a matrix rather than individual coefficients, care must be taken to include a value for each biomarker in every class in which it appears. For example, since zeaxanthin, the biomarker for cyanobacteria, also occurs in small amounts in chlorophytes, a ratio of zeaxanthin : chlorophyll-*a* in chlorophytes must now be supplied.

Primary biomarker : chlorophyll-*a* ratios used in CHEMTAX were identical to the second, or revised, set used for the SLE model as described above. Grant's 2006 study, which was used to provide this revised coefficient set for the SLE model, was also used to provide ratios for the additional and shared biomarker pigments (Grant 2006 and Grant unpublished data). A review of the literature determined that the algal species used by Grant were, not surprisingly, closest to those expected to be encountered in the periphyton field samples. Also, her values were determined using the same extraction and analysis protocol as that used for analysis of the samples used in this study. Particularly, this extraction protocol is considered more efficacious for the algal groups typically encountered in periphyton (see Hagerthey et al. 2006).

These additional biomarker ratios were derived from the medium-light level (300 μmoles photons $m^{-2}s^{-1}$) results from Grant's study of pigment responses to varying light levels in unialgal cultures (see Grant 2006). Medium-light values were chosen for two reasons. First, the bulk of the field samples were taken from periphytometers with three months of growth, which is typically a much thinner growth than the natural mats, which can be several centimeters thick. The medium light value would likely represent an integrated value over the mat, from the senescent and low-light adapted algae

(selfshading of the mid- to lower-light areas) to the medium- and high-light adapted cells of the upper layer. A medium-light value also compensates for seasonal and daily light variations since field sampling was performed throughout the year and not necessarily only on sunny days, or even at a certain time day.

Cosmarium and *Closterium* cultures (3 replicates each) were used to determine the lutein : chlorophyll-*a* and zeaxanthin : chlorophyll-*a* ratios for chlorophytes. Cultures of *Navicula* and *Phaeodactylum* (3 replicates each) were used to find a 'diatoxanthin plus diadinoxanthin' : chlorophyll-*a* ratio for diatoms. For dinoflagellates, the 'diatoxanthin plus diadinoxanthin' : chlorophyll-*a* ratio and the dinoxanthin : chlorophyll-*a* ratio were determined using *Amphidinium* cultures (3 replicates). In each case, the mean of all measurements within each algal class was used. All initial measurements were taken from Grant 2006 and Grant unpublished data). With the exception of dinoxanthin, coefficients derived from these values were all within the range of values reported in the literature (see Table 3). Dinoxanthin typically occurs in very small amounts and appears to be seldom used for chemotaxonomy: only one of the surveyed references (M. Mackey 1996) reported a value. Initial testing of both values using artificial data (for which the true composition of a sample is known) showed that Grant's ratio returned results closer to the true dinoflagellate percentage.

Initially the ratio matrices were constructed without a chlorophyll-*a* term. However, artificial data runs showed that CHEMTAX returned much more accurate results when this term was included. In fact, accuracy appeared to depend quite heavily on how close the row sums of the ratio matrix were to each other. This was a somewhat surprising development and is discussed in more detail in the Results and Discussion

section. Accordingly, a chlorophyll-*a* term of 1 (since all biomarker ratios are calculated against a unit of chlorophyll-*a*) was added to ratio matrices, and the actual amount of total chlorophyll-*a* in each sample was included with the data. Chlorophyll-*a* in the samples was calculated as the sum of chlorophyll-*a* and all chlorophyll-*a* derivatives (see Table 2).

Sample Groups. Samples were first divided into the following categories: artificial mixes (all), grab samples (all), and all periphytometer samples from each geographical area (Water Conservation Areas 1A / Loxahatchee Wildlife, 2A, and 3A, Shark River Slough, and Taylor Slough). These groups were kept distinct throughout the analyses. Further subdivision based on transects within the larger areas failed to produce good results in initial tests, primarily because similarity of pigment profiles could not be assured based solely on geographical origin of the samples.

The authors of CHEMTAX caution that the data must be carefully inspected for outliers, as the method is not robust in respect to this (M. Mackey et al. 1997; Wright 2005). Samples were inspected for total chlorophyll-*a* and biomarker outliers. There was a large amount of natural variation in the samples and only two extreme outliers were removed. These were instances where a biomarker varied by nearly an order of magnitude from the other measurements in the sample group. One outlier from the Shark River Slough samples which likely represented a measurement error (see Figure 1) and one outlier from the WCA-1A (Loxahatchee Wildlife Refuge) samples were removed prior to any further analyses. Figure 1 also shows examples of natural chlorophyll-*a* variation which were not considered outliers.

Samples in which pheopigments constituted over 10% of the total chlorophyll-*a* and derivatives were considered problematical. Previous work has shown that pigment ratios tend to be increasingly unpredictable as the percentage of pheopigments increases (Deydier-Stephan et al. 2003; Louda 2008). Aside from contributing to an incorrect answer for the algal class composition in both SLE and numerical methods, unreliable pigment ratios can sometimes prevent convergence in numerical methods. SLE runs and one run of CHEMTAX were performed on samples with high pheopigments, for comparison purposes, and after this these samples were excluded from the analyses. In total, this excluded 137 periphytometer samples and 1 grab sample (see Table 1 for a breakdown by area).

CHEMTAX can process fairly large sample groups. For example, 40-sample groups are used in the synthetic datasets in Mackey et al. (1996). The main danger for this method lies in specifying sample groups which are too small. A baseline recommendation is that the number of samples be at least greater than the number of pigments (D. Mackey and Higgins 2001). This can present a challenge when attempting to construct what the authors call homogenous groups (i.e., groups with similar pigment signatures). In very heterogeneous communities such as periphyton, the homogenous groups can be quite small even when a large initial sample group is available. Accordingly, groups of samples were constructed using common biomarker pigment signatures while maintaining a minimum number of samples. For example, samples which contained no alloxanthin, the biomarker for cryptophytes, were grouped separately from samples which did contain this pigment. To continue the example, the first group would be run in CHEMTAX using with a ratio matrix which did not include

cryptophytes, but the second group would be run with a ratio matrix which did include this algal class. Other than the default configuration of all classes, the most common sample groups in this study were those which did not include the biomarkers for cryptophytes and/or dinoflagellates. (See Table 6 for a detailed breakdown of sample groups).

Parameters. Parameters for CHEMTAX are included in the same .m file as the data and ratio matrices (see example in Appendix C). Parameters were changed rather significantly from the defaults in order to take advantage of newer computing capabilities, with the best parameters found to be very close to those used by Latasa (2007). Maximum iterations were set at 10000, with a cutoff value for the residual of 0.0001; initial step size (divisor) = 25, step ratio = 2, and the step limit = 30000. All pigments were allowed to vary at each iteration and subiterations were set at 1. Ratio limits were set to 500, allowing all elements of the ratio matrix to vary somewhat freely (from 1/5 to 5 times the original ratio).

Setting the weighting was more problematical. The exact way in which the PrePro program determines pigment weights was unclear, and parallel weighting calculations could not be reconstructed manually. (Attempts to recreate the weighting calculations from the documentation produced numbers far different from those found in the CHEMTAX example files.) These weights are designed to assign a relative error to sample data for each biomarker pigment, and (if the "bounded" option is chosen in PrePro) to set an absolute lower limit for error in minor pigment amounts so they are not swamped by the major pigments. For this study, a weighting scheme was constructed

using the upper and lower limits from the CHEMTAX example files and choosing values which corresponded to the relative weight of the pigments. These weights are shown along with the ratio matrices used for CHEMTAX in Table 5.

Bayesian Compositional Estimator

The Bayesian Compositional Estimator (BCE) is a chemotaxonomic program which was developed in 2007 by researchers at the Netherlands Institute of Ecology (Van den Meersche et al. 2008). BCE is implemented as a package (Van den Meersche and Soetaert 2009) in the open source software R (R Development Core Team 2009). Periodic updates of the package are available by downloading the updated package from the R website; this study used BCE version 1.4 and R version 2.9.1. BCE was run on a Compaq laptop with a 2.0 GHz AMD Athlon dual processor and 2GB RAM under a 32-bit version of Windows Vista.

Similarly to CHEMTAX, BCE uses a ratio matrix, a data matrix, and an unknown sample composition matrix to compose a linear inverse problem. BCE, however, uses Bayesian methods to fit a probability distribution to the data and find a maximum likelihood solution for the problem. BCE first finds a least-squares solution (although this methodology can be altered by specifying different parameters when starting the program) and uses it a starting point for a Markov Chain Monte Carlo (MCMC) simulation. It should be noted that this initial solution is not identical to the CHEMTAX solution. The program provides a number of diagnostic outputs in order to check the performance of the simulation, and this output must be inspected prior to acceptance of any results. This output includes the number of runs accepted as well as plots which

indicate the extent and randomness of the sampling of the solution space (i.e., mixing in the markov chain). Final results of the program include the algal class composition of each sample, a revised ratio matrix, and the standard deviations and covariance matrices for the ratios and class compositions.

Choices to be made prior to running BCE are similar to choices in CHEMTAX: which algal classes should be included, which pigments will be chosen for biomarkers of specific algal classes, what ratios of pigment to chlorophyll-*a* should be used for the calculation coefficients, how should sample groups be constructed, and how should the program parameters be set. Choices should also be made as to when the results of the random walk will be considered acceptable. For example, this includes the choice of a minimum value for the number of runs accepted, and a determination of what will constitute acceptable mixing.

Setting Up the Program. A working version of R (R Development Core Team 2009) must first be installed on the computer in order to run BCE. BCE is most easily installed using the package menu in R, which downloads the files from the R website and installs them into the program. Although BCE is installed as a package, the documentation files can be opened separately if desired. The user's manual (BCE-manual.pdf), which contains detailed examples, is particularly useful. As with CHEMTAX, some minor discrepancies between documentation and actual syntax must be resolved, the most common being case inconsistencies (R is case-sensitive).

Prior to running BCE, the package must be loaded in and at least the ratio and data matrices must be specified. If desired, matrices can also be specified for standard

deviations and jump lengths. For this study, the data and ratio matrices, and sometimes a standard deviation matrix, were stored as text files and read into R using the read.table function (see examples in Appendix D). The BCE () function is then used to perform the basic operation. Other functions are also available in the package, such as graphical views of the MCMC random walks and an alternate least-squares regression function. Although a summary function is available to view BCE output, the use of the export option in the main BCE function proved more tractable.

Algal Classes. The maximum algal classes used for BCE runs were identical to those used in SLE and CHEMTAX: chlorophytes, cryptophytes, coccoidal cyanobacteria, filamentous cyanobacteria, diatoms, and dinoflagellates. BCE, in contrast to CHEMTAX, does not rescale the original data and ratio matrices. The authors assert that avoiding the dependence of data columns which results from rescaling should allow the program to work with more variation in the biomarker pigments of the samples than CHEMTAX allows (Van den Meersche et al. 2008). However, in practice, runs of sample groups which exhibited large differences in pigment signatures created more problems with mixing and convergence than did runs with more homogenous groups. In addition, initial tests with the artificial mixes showed that, similarly to CHEMTAX, use of a ratio matrix which included algal classes not actually present in the sample led to highly inaccurate results. For these reasons, as well as for ease of comparison, samples for BCE input were grouped, within area, by pigment signatures and ratio matrices were adjusted accordingly.

Biomarker Pigments. The set of biomarker pigments used in BCE was identical to that used in CHEMTAX. BCE has the same general requirements as CHEMTAX in that more biomarkers than algal classes must be supplied in order to keep the linear inverse problem overdetermined (and thus allow a solution), and the set developed for use in CHEMTAX was thought to be the best set for this data (see discussion in the CHEMTAX Biomarker Pigments section above). In addition, use of identical biomarkers allowed a more direct comparison between methods.

Ratios. Biomarker pigment to chlorophyll-*a* ratios used in BCE were identical to those used for CHEMTAX, with the exception of one set of BCE runs which handled the allocation of zeaxanthin slightly differently. Originally, the subtraction method for zeaxanthin utilized in SLE and CHEMTAX was retained. In this method, an amount of zeaxanthin equivalent to the amount of echinenone in the samples is subtracted from total zeaxanthin. The remaining zeaxanthin is assigned to coccoidal cyanobacteria and, in methods allowing shared biomarkers, also to chlorophytes. However, due to the model's recurring difficulty in correctly identifying coccoidal vs. filamentous cyanobacteria, a different method was employed for a second set of BCE runs. The original zeaxanthin amount from the sample was retained in the data matrix and a ratio matrix specifying a 1:1 ratio for echinenone : zeaxanthin for filamentous cyanobacteria was used. This allowed the program to vary the echinenone : chlorophyll-*a* and zeaxanthin : chorophyll-*a* ratios separately for filamentous samples, rather than forcing them to vary identically, which is the effect of the subtraction method. Since the subtraction method occasionally returns zeaxanthin values of less than zero, which are then forced to zero so as to avoid

negative numbers, inclusion of the actual zeaxanthin amount should allow BCE to compensate more directly for the not uncommon case where the echinenone : zeaxanthin ratio was more than 1:1 in the sample. It was also hoped that this change would help the BCE handling of chlorophytes, which occasionally was very weak, by allowing a more flexible allocation of the zeaxanthin (included as a minor pigment in chlorophytes).

Sample Groups. BCE cautions against using large sample group sizes (in their example, n=40) as they may cause overfitting (Van den Meersche et al. 2008), but the literature otherwise says very little about what sample sizes are recommended. In strong contrast to CHEMTAX, single samples are allowed and will return correct results. In practice, it was also found that larger samples sizes (n> 50) tended to be unwieldy in terms of runtime and size of export files. A test with the artificial mixes used one parameter set to run the samples as one 15-sample group and as three 5-sample groups: each run returned the same results for algal class compositions. A similar test was run with artificial data, comparing results from a group of 30 samples with results from subgroups of 10, and in this test as well virtually identical algal class compositions were returned.

Since parameters had to be set individually for each run, in order to obtain the best combination of good mixing and acceptance rates, very small sample sizes were not the best option for analysis of this amount of field data. Accordingly, for most BCE runs, sample groups were kept identical to those run in CHEMTAX if n<30, and groups where n>30 were split into smaller groups. When problems with mixing or convergence were experienced, runs with smaller, more homogenous groups were attempted. Due to the

difficulty in obtaining results in general with BCE, samples for which pheopigments comprised over 10% of chlorophyll-*a* and derivatives were excluded from all runs.

Parameters. In contrast to CHEMTAX, where many of the default parameters were clearly outdated, the BCE methodology included default parameters representing the current best option for running these types of simulations. In the absence of any compelling evidence that our datasets necessitated special parameters, these defaults were retained. The most important of these were the use of the least-squares solution as the composition matrix starting point in the markov chain and the use of a gamma distribution for the prior probabilities of the ratio and data matrices. In a few cases a uniform prior probability distribution for the ratio matrix was specified in order to more directly compare BCE results with CHEMTAX results. (See Van den Meersche et al. 2008 for a full discussion of these parameters.)

The primary parameters altered were: the relative or absolute standard deviations for the ratio and data matrices, the jump lengths for the ratio and composition matrices, the number of iterations, and the burn-in length (see Appendix D for a typical example of BCE parameters). These parameters had to be tuned in each run by inspecting the diagnostic output of the previous runs. For example, there is no guarantee that a parameter set which returns excellent mixing or good acceptance rates in one group of samples will do the same for the next group of samples. In addition, all runs were performed several times to ensure that results were consistent, as occasionally parameter sets were used which would return widely varying results on successive runs. Consistency issues in BCE are discussed in more detail in the Results section following.

Typically, one standard deviation value was used for the data and a matrix of values was used for the ratios. The ratio standard deviations were set up to allow wider deviations in photoprotective pigments than in photoaccessory pigments, with suspected dual-action pigments accorded values in between these (see Table 5). In theory, the standard deviations should not have had to be altered as a result of program output. It was found, however, that some flexibility on this issue, particularly increasing or decreasing data standard deviations relative to ratio standard deviations, improved mixing and occasionally even convergence. In general, relative standard deviations were preferred over absolute ones.

The suggestion of the authors (Van den Meersche and Soetaert 2009, user's manual) that ratio mixing is often improved by making the ratio jump length linear to the ratio standard deviation was found to be true and therefore this method was employed consistently (see Appendix D for an example). In most cases, the jmpX value (the jump length for the composition matrix) had to be set quite small, usually between 0.001 and 0.005, to attain a reasonable acceptance rate.

Acceptance of Results. For each run of BCE, the random walk results must be evaluated and a decision made as to whether or not to accept the result. (See Van den Meersche and Soetaert 2009, user's manual, for a full discussion of this.) Three main criteria were used for this study: the number of points accepted, the mixing plots, and whether or not the solution was trivial or inconsistent. This was done in a stepwise procedure. That is, mixing was only evaluated if the accepted points were high enough and the composition was only evaluated if mixing was acceptable.

Each BCE run returns a value at the command line for number and percentage of accepted points. The percentage will typically lower as the amount of iterations increases. For 10,000-iteration runs in this study, mixing was evaluated if the acceptance was over 10%. In these cases, the plot function built into BCE was used to view the random walk summaries for the ratio pigments and the sample compositions. Figure 2 shows examples of what, for this study, was considered excellent, acceptable, and unacceptable mixing. Adjustment of parameters to achieve both good acceptance and good mixing in all categories was often not possible. Excellent mixing could often be achieved in the ratio pigments but was rare in the sample compositions. Even very good overall composition mixing often showed poor mixing in certain algal classes in some samples. However, testing with artificial data showed that the program could find a close approximation of the sample composition even with the occasional poor sample mixing.

For runs where acceptance of points and mixing were both considered acceptable, the run was repeated at least three times and the algal class composition results and revised ratios were inspected. Although a trivial solution (nearly 100% of the algae in all samples assigned to a class which not actually present) was encountered in a very few instances, the larger issue was consistency of results. Occasionally runs which returned widely varying ratios and similarly varying compositions were encountered. In these cases, parameters could not always be located which would give a consistent result.

Analysis

Bases of Comparison and Statistical Methods

The primary comparison between models was that of accuracy of results. Statistical comparisons were made by finding the Pearson's product-moment coefficient, using the R function cor.test (R Development Core Team 2009) for various pairs of algal class compositions and comparing the results for one class from one sample from two methods. This method of comparison, showing the linear correlation of class determinations resulting from various methodologies such as comparison between CHEMTAX and microscopy results, is standard in the chemotaxonomic literature. Comparisons were made not only between models but within as well, to test results from various parameterizations.

Although it is tempting to analyze broader patterns between model results, there is no theoretical basis upon which to make any rigorous statistical comparison other than linear correlation (Zar 1999). However, any general trends which might be the basis for further experimentation and testing were noted. In addition, model results were summarized and shown graphically for quick comparison.

Other comparisons were made between the models as a whole. For example, the relative efficiency, initial cost, expandability, and overall reliability of results for different models were assessed. This type of comparison follows Millie et al. (1993).

Choice of Sample Groups for Comparison

Due to the amount of actual data available, results from the computer-generated artificial data were used primarily to parameterize the numerical methods (CHEMTAX and BCE) and were not extensively analyzed. The artificial data runs did pinpoint potential weak spots in these methods which are discussed in the following Results section. Results for various sets of parameters and the correlations for the best parameters were compared between the two methods.

Algal class composition results from runs on the artificial mix data (algal cultures) were compared between the known composition, SLE using both sets of coefficients, both BCE and CHEMTAX using a ratio matrix which included only the algal classes present (note that this can be determined from the pigment signatures of the biomarkers and did not require previous knowledge of the class composition), and CHEMTAX using a ratio matrix which included all six algal classes.

The comparison of microscopy to chemotaxonomy results was performed using a subset of the grab samples for which biovolume data was available. Originally 122 grab samples were analyzed at FAU for pigment weights and algal class composition (HPLC/PDA and PigCalc). However, all but 29 were analyzed with a slightly different protocol and pigment weight calculation that precluded easy conversion for comparison here. The n=29 subsample included sites from all grab sampling areas (Loxahatchee, Shark River Slough, Taylor Slough, Water Conservation Areas 2 and 3, and PAL plus PEN) and spanned a nearly identical date range for the subsample (9/27/06-11/28/06) as compared to the entire sample group (9/25/06-11/28/06). Comparisons were made

between the sample compositions as determined by biovolume and the results of all chemotaxonomic methods.

Sample compositions were determined for all 475 periphytometer samples using both SLE configurations, 2-5 CHEMTAX configurations, and, for a few groups of samples, 1-2 BCE configurations. Correlation analysis, however, was limited to a few representative groups. This analysis was also confined to the 338 samples with less than 10% pheopigments. Correlations between compositions as determined by different models were assessed for one sample group each in WCA-1, WCA-3A, and Shark River Slough, and two sample groups each in WCA-2A and Taylor Slough. Since the true composition of these samples was unknown, correlations were assessed with the expectation that weaker areas in the compositions would correspond to lower correlation values. For example, it was hypothesized that the compositions for groups with more heterogeneous communities, which might be less well-represented by one set of ratios, would correlate less well than compositions from more homogenous groups.

For all field sample data, results from samples for which pheopigments comprised over 10% of chlorophyll-*a* and derivatives were compared to determine the influence of high pheopigments on reliability of results. This comparison was made within SLE using two sets of coefficients. It was hypothesized that groups including samples having high pheopigments would have lower correlation coefficients.

Table 1. Sites and sampling data. Sampling summaries for periphytometers and grab samples from Shark River Slough, Taylor Slough, Water Conservation Areas 1 (Loxahatchee National Wildlife Refuge), 2, and 3.

PERIPHYTOMETERS

area	transects	total dates and sites	n	n with <10% pheopigments
SRS	S12A, S12C, S345B, S345C, S355A, S355B	17 retrieval dates, 16 sites	146	111
TS	T1 T2, T3, T5, T6, T7	14 retrieval dates, 12 sites	84	78
WCA-1	Meso, X, Y, Z	12 retrieval dates, 10 sites	91	49
WCA-2A	E, F, U	12 retrieval dates, 13 sites	113	82
WCA-3A	Meso E, W	12 retrieval dates for 2 sites, 4 retrieval dates for another 8 sites	41	18
total			**475**	**338**

GRAB SAMPLES

area	total n	subsample n
LOX	14	4
PAL	7	3
PEN	2	1
SRS	43	7
TS	5	2
WCA-2	9	3
WCA-3	42	9
total	**122**	**29**

Table 2. Biomarkers. Photosynthetic pigments used in the chemotaxonomic calculations are shown in the left column. The middle column contains biomarkers, which are sometimes the sum of various pigments, and their abbreviations. The right column shows the algal classes associated with these biomarkers for this study (for a comprehensive list of biomarker pigment occurrence in specific algal classes, see Jeffrey et al. 1997).

Measured Pigment	**Biomarker**	**Algal Class**
chlorophyll-*a* chlorophyllide-*a* pyrochlorophyllide-*a* pheophorbide-*a** pyropheophorbide-*a** chlorophyll-*a* allomer pheophytin-*a* allomer* pheophytin-*a** pheophytin-*a'* * pyropheophytin-*a**	total chlorophyll-*a* (cha)	all
chlorophyll-*b* pheophytin-*b*	total chlorophyll-*b* (chb)	chlorophytes
lutein	lutein (lut)	chlorophytes
alloxanthin	alloxanthin (alx)	cryptophytes
zeaxanthin	zeaxanthin (zea)	coccoidal cyanobacteria, chlorophytes
echinenone	echinenone (ech)	filamentous cyanobacteria
fucoxanthin *cis*-fucoxanthin fucoxanthinol	total fucoxanthin (fuc)	diatoms
diadinoxanthin diatoxanthin	diadinoxanthin plus diatoxanthin (dia)	diatoms, dinoflagellates
dinoxanthin	dinoxanthin (din)	dinoflagellates
peridinin	peridinin (per)	dinoflagellates

* indicates pheopigment

Table 3-a. Comparison of biomarkers and ratios, showing biomarkers chosen for use in selected freshwater and estuarine studies. Note that only the biomarkers used for the 6 classes of algae considered here are shown and the original references must be consulted for the entire suite of biomarkers used in each study.

biomarker pigments used in recent freshwater and estuarine studies

source	system	chloro	crypto	cyano total	cyano type 1	cyano type 2	diatom	dinofl
Descy et al. 2000	northern Wisconsin lakes	chb lut neo vio zea	alx β, ε car	-----	zea	zea ech	fuc did	per did
Greisburger and Teubner 2007	alpine lakes	chb lut vio	alx mon	zea ech myx osc	-----	-----	fuc did	per did
Guisande et al. 2008	freshwater (various)	chb lut neo vio zea	alx chc1+2	zea	-----	-----	fuc did chc1+2 vio zea	per did chc1+2
Marinho and Rodrigues 2003	tropical eutrophic reservoirs	chb lut	alx	zea	-----	-----	fuc	per
Schlüter et al. 2006	northern European eutrophic lakes	chb lut	alx	zea ech	-----	-----	fuc	per
Lewitus et al. 2005	southeastern US estuaries	chb lut neo vio zea	alx chc1+2	zea	zea	zea	fuc did dit chc1+2	per did dit
Lionard et al. 2008	European estuary	chb lut zea	alx	zea ech	-----	-----	fuc did+dit	per did+dit

Pigment abbreviations: alx (alloxanthin), β, ε car (β, ε carotene), chb (chlorophyll-*b*), chc1+2 (sum of chlorophyll-*c*1 and chlorophyll-*c*2), did (diadinoxanthin), din (dinoxanthin), dit (diatoxanthin), ech (echinenone), fuc (fucoxanthin), lut (lutein), mon (monadoxanthin; not measured in hypolimnion), myx (myxoxanthophyll; not measured in hypolimnion), neo (neoxanthin), osc (oscillaxanthin), per (peridinin), vio (vioxanthin), zea (zeaxanthin).

Table 3-b. Comparison of biomarkers and ratios. This shows secondary biomarker ratios used in selected freshwater and estuarine studies, as well as Grant's (2006) cultures and M. Mackey et al.'s (1996) original ratios.

diadinoxanthin and diatoxanthin (sum) in diatoms

0.126 Lionard et al. 2008
0.129 Sarmento and Descy 2008
0.149 Lewitus et al. 2005
0.278 Grant unpublished data
0.717 M. Mackey et al. 1996 (max value; min value = 0.03)

diadinoxanthin and diatoxanthin (sum) in dinoflagellates

0.250 Sarmento and Descy 2008
0.275 Lewitus et al. 2005
0.385 Grant unpublished data
0.400 Lionard et al. 2008
0.442 M. Mackey et al. 1996 (max value; min value = 0)

dinoxanthin in dinoflagellates

0.005 M. Mackey et al. 1996 (max value; min value = 0)
0.042 Grant unpublished data

lutein in chlorophytes

0.117 Sarmento and Descy 2008
0.126 Descy et al. 2000 (max avg value)
0.149 Marinho and Rodrigues 2003
0.165 Grant 2006
0.221 Lewitus et al. 2005
0.251 Lionard et al. 2008
0.283 M. Mackey et al. 1996 (max value; min value = 0)
0.362 Guisande et al. 2008
0.654 Greisberger and Teubner 2007

zeaxanthin in chlorophytes

0.002 Lewitus et al. 2005
0.008 Grant 2006
0.009 Marinho and Rodrigues 2003
0.015 Descy et al. 2000 (max avg value)
0.018 Sarmento and Descy 2008
0.030 Lionard et al. 2008
0.035 Guisande et al. 2008
0.118 M. Mackey et al. 1996 (max value; min value = 0)

Table 4. Ratios used in simultaneous linear equations. Note that although the coefficients are reproduced here in matrix format for ease of reading, in SLE they are used as separate linear equations.

SLE original coefficients

	alx	chb	ech	fuc	per	zea
chlorophytes		0.317				
cryptophytes	0.166					
cocc. cyanobacteria						0.579
fil. cyanobacteria			0.056			
diatoms				0.615		
dinoflagellates					0.471	

SLE revised coefficients

	alx	chb	ech	fuc	per	zea
chlorophytes		0.406				
cryptophytes	0.166					
cocc. cyanobacteria						0.579
fil. cyanobacteria			0.056			
diatoms				0.738		
dinoflagellates					0.543	

Table 5. Ratios used in CHEMTAX and BCE. At the top is the ratio matrix for the maximum 6 algal classes. Ratio matrices used for fewer classes were subsets of this matrix. Also shown are CHEMTAX weighting and examples of ratio relative standard deviation matrices used in BCE.

ratio matrix used in CHEMTAX and BCE (maximum algal classes)

	alx	chb	dia	din	ech	fuc	lut	per	zea	chla
chlorophytes	0	0.406	0	0	0	0	0.165	0	0.008	1
cryptophytes	0.166	0	0	0	0	0	0	0	0	1
cocc. cyanobacteria	0	0	0	0	0	0	0	0	0.579	1
fil. cyanobacteria	0	0	0	0	0.056	0	0	0	0	1
diatoms	0	0	0.278	0	0	0.738	0	0	0	1
dinoflagellates	0	0	0.385	0.042	0	0	0	0.543	0	1

CHEMTAX weighting

alx	chb	dia	din	ech	fuc	lut	per	zea	chla
30	5	10	30	10	3	10	20	20	1

2 standard deviation matrices used in BCE

	alx	chb	dia	din	ech	fuc	lut	per	zea	chla
chlorophytes	0	0.02	0	0	0	0	0.05	0	0.1	0.01
cryptophytes	0.1	0	0	0	0	0	0	0	0	0.01
cocc. cyanobacteria	0	0	0	0	0	0	0	0	0.1	0.01
fil. cyanobacteria	0	0	0	0	0.05	0	0	0	0	0.01
diatoms	0	0	0.1	0	0	0.02	0	0	0	0.01
dinoflagellates	0	0	0.1	0.1	0	0	0	0.02	0	0.01

	chb	dia	ech	fuc	lut	zea	chla
chlorophytes	0.05	0	0	0	0.1	0.2	0.01
cocc. cyanobacteria	0	0	0	0	0	0.2	0.01
fil. cyanobacteria	0	0	0.1	0	0	0	0.01
diatoms	0	0.1	0	0.05	0	0	0.01

Table 6. Pigment signature divisions of sample groups. This shows the numbers of samples within areas which were run in CHEMTAX using configurations which differed from the default configuration of all 6 algal classes. The "pigment profile" column shows which biomarker pigments were absent in these groups of samples.

area	pigment profile	algal classes	sample n
SRS	no alx	all except cryptophytes	16
SRS	no din or per	all except dinoflagellates	9
SRS	no alx, din, or per	all except cryptophytes and dinoflagellates	48
TS	no alx, din, or per	all except cryptophytes and dinoflagellates	53
TS	no alx, chb, din, lut, or per	no chlorophytes, cryptophytes, or dinoflagellates	14
WCA-1	no alx, din, or per	all except cryptophytes and dinoflagellates	15
WCA-2A	no alx	all except cryptophytes	13
WCA-2A	no din or per	all except dinoflagellates	23
WCA-2A	no alx, din, or per	all except cryptophytes and dinoflagellates	24
WCA-3A	no alx	all except cryptophytes	9
WCA-3A	no din or per	all except dinoflagellates	12

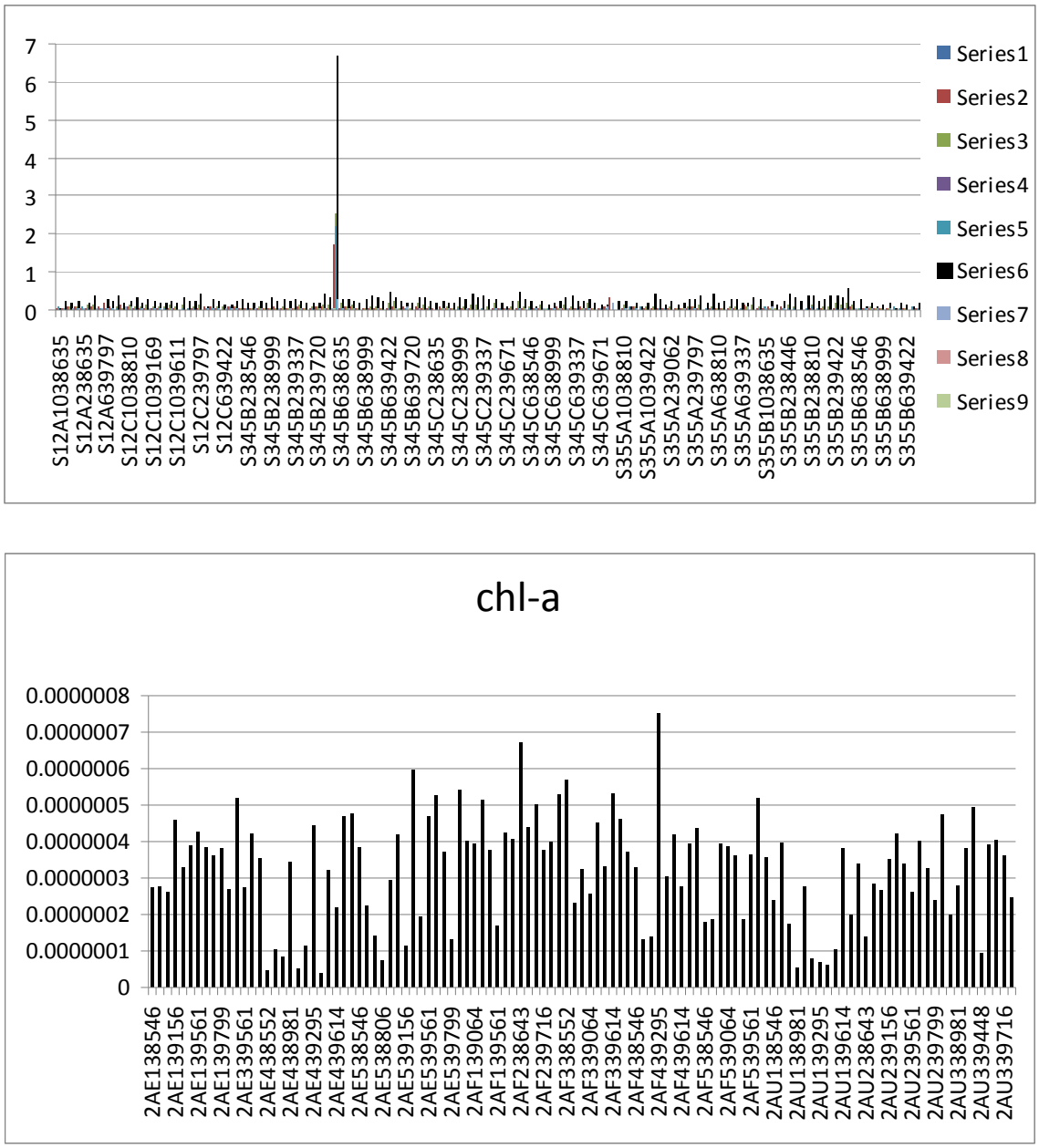
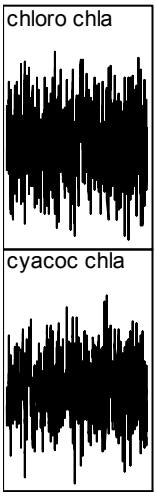
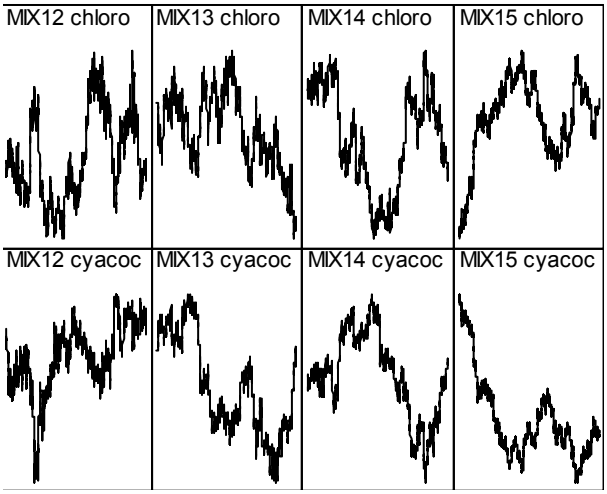


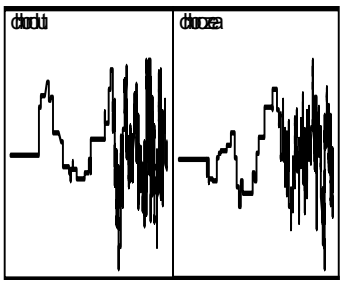
Figure 1. Outlier determination. The top chart shows an outlier which is probably a result of measurement error, where a decimal point has shifted. The lower chart shows variation in absolute chlorophyll-*a*. For this study none of these chlorophyll-*a* values were considered outliers.



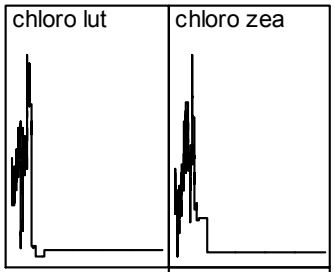
a) Excellent mixing.



b) Acceptable mixing.



c) Acceptable with burn-in period.



b) Unacceptable mixing.

Figure 2. Examples of mixing in the Bayesian Compositional Estimator. Mixing plots which for this study were considered excellent (a), acceptable (b), acceptable after an initial burn-in period (c), and unacceptable (d).

RESULTS AND DISCUSSION

This chapter first presents results for the artificial mixes and grab samples, discussing how well each of the three models was able to match the true algal composition (the artificial mixes) or the microscopically-determined algal composition (the grab samples). Following this is a presentation of results from the periphytometer samples, comparing them across methods, and discussing the possible effects of high pheopigment percentages. The influence of different biomarker pigment : chlorophyll-*a* ratios is discussed and some of the possible consequences of the ratio sets chosen are assessed. Finally, the models are compared along a number of qualitative and quantitative factors to assess the relative strengths and weaknesses of each model, particularly in relationship to the analysis of periphyton.

Artificial Mixes: Comparison to Known Compositions

The artificial mixes tested the ability of the models to determine algal compositions for very constrained samples. Since the mixes only contained chlorophytes, coccoidal cyanobacteria, and diatoms, the number of available biomarker pigments was limited. Only 3 algal classes and 5 biomarker pigments along with chlorophyll-*a* were used. As discussed previously in the Methods section, each class contained only one species of algae. The chlorophyll-*a*/cell ratio was determined for each species and the cell count ratios were corrected by this factor to return a taxon-specific chlorophyll-*a* ratio.

Correlations for all analyses were determined using the R function cor.test, with an alternative hypothesis that the correlation was greater than zero. CHEMTAX and BCE were both run using the SLE revised coefficient set.

All three models performed very well (see Figure 3 and Table 7). For the diatom, CHEMTAX showed the weakest correlation, while for the chlorophyte and coccoidal cyanobacteria the weakest correlation was found in BCE. The two different coefficient sets used in separate runs of SLE contained different ratios for 2 of the 5 biomarker pigments : chlorophyll-*b* for the chlorophyte class and fucoxanthin for the diatom class. SLE performed very well under these single-species per class conditions, with both sets of coefficients returning the highest correlations for all 3 classes.

Grab Samples: Comparison to Microscopically-Determined Compositions

Work with the artificial mixes made it clear that the microscopy comparison must be made with biovolume-based compositions rather than cell-count based compositions. Accordingly, the number of samples available for comparison was limited. Due to both the minimum group size requirement for CHEMTAX and the fact that the microscopy analysis did not measure cryptophytes, the single sample which contained alloxanthin was not used. Also not used was the one sample for which pheopigments constituted more than 10% of the total chlorophyll-*a* and derivatives. The remaining 27 samples were constrained in time but varied widely in geographical origin.

Two CHEMTAX runs were used for the comparison. One run contained all 27 samples with a pigment profile which excluded cryptophytes and dinoflagellates. Alloxanthin, dinoxanthin, and peridinin were absent from both the samples and the ratio

matrix. An additional CHEMTAX run was performed excluding 5 samples which contained no chlorophyll-*b*. Results were also returned from BCE and from SLE using both sets of coefficients. Coccoidal cyanobacteria and filamentous cyanobacteria were estimated separately in the models but were totaled for comparison to total cyanobacteria in the biovolume results.

Composition results from each model were plotted against the biovolume compositions (see Figure 4). Although most plots appeared to show a normal distribution, chlorophytes were a possible exception. Correlations were initially run using both the Pearson correlation, which assumes a normal distribution, and the Spearman rank correlation, which is a nonparametric method which does not make the assumption of normal distribution of data. Both sets of coefficients were very similar, showing the same order of correlations. Since Spearman coefficients are based on rank and exact p-values are difficult to compute when there are ties, which are abundant in this data, only the Pearson coefficients are reproduced here.

Correlations in general were very poor (Table 8), with the highest correlation coefficients being 0.47 for diatoms and 0.57 for total cyanobacteria ($df=25$, $p < 0.01$). For chlorophytes, the highest correlation coefficient was only 0.36 ($df=25$, $p < 0.05$). A second CHEMTAX configuration which excluded the 5 samples without chlorophyll-*b* showed similarly poor correlations. BCE performed far more poorly than any other method. For this dataset, CHEMTAX appeared to estimate chlorophytes and cyanobacteria best, while SLE performed best on diatoms.

This poor correlation might have occurred for several reasons. Chemotaxonomic calculations determine the amount of chlorophyll-*a* contributed by each taxon. While this is more closely approximated by a biovolume based composition than a cell-count composition, there is still quite a bit of potential variability (see, for example, Wright and Jeffrey 2006). Although biovolume calculations are performed on a finer resolution than chemotaxonomic calculations, they are still based on groups of algae rather than individual species. Additionally, any algae which are very small or poorly preserved may be underestimated (Millie et al. 1993, Wright and Jeffrey 2006). Another important consideration is that chlorophyll-*a*/cell ratios vary not only between types of algae, but also along many scales of environmental gradients (see, for example, Grant 2006). Wright and Jeffrey (2006) report that some studies (for example, Havskum 2004) obtained better results in this area after removing large diatoms from consideration. Although periphyton may be too complex for such a relatively simple correction, the wide ratio of 23:1 in chlorophyll-*a*/cell for cyanobacteria vs. diatoms indicates this may be worth an attempt.

The most likely reason for the poor correlation may be the variability of the sample group. Although collected during one 3-month time span, the samples originated from throughout the Everglades. This area comprises a number of gradients as well as differences in the overall heterogeneity of the algal community. It is quite likely that a single ratio set cannot adequately encompass this variability.

A final reason for the poor correlation may lie in the fact that these samples were taken from existing periphyton mats rather than from periphytometers. Periphytometers are invaluable for providing samples of relatively young and thin periphyton layers. They

exclude the possible confounding influence of an existing, natural substrate or the self-shading and competition, such as movement to higher- or lower-light areas by more mobile algal species, that can occur in a thicker mat. Periphytometers thus allow use of a much simpler chemotaxonomic model. In the grab sample analysis, samples of the actual mats without regard to age were essentially forced to the simpler chemotaxonomic model. This is based on the assumption that medium-light pigment ratios can represent an average of true pigment ratios throughout the mat, from high light to low light. The inability of any of the chemotaxonomic methods considered in this paper to match microscope results from the grab samples may indicate this particular assumption cannot be made.

Periphytometer Sample Results

Due to the poor correlation of any of the methods with the biovolume compositions, it was determined that no model could return results which could be used as a baseline best answer for the periphytometer data. Analysis of the periphytometer compositions therefore had to be structured on the assumption that the true composition was unknown. However, even given this constraint, certain comparisons could be made. Properly configured CHEMTAX runs and SLE runs using the same coefficients had returned nearly identical answers for artificial data groups, as well as good correlations for the artificial mixes. Periphytometer results from these same methods were compared with the expectation that weaker parts of the analysis would reveal themselves through lower correlations. Just as important, low correlations might indicate areas where the secondary biomarker pigments included in CHEMTAX may be having a stronger effect.

There is no way to know, however, if the effect is tending towards or away from the correct answer.

Care must be taken not to overinterpret these particular results as the sample groups contain a high level of variability. As discussed in the Methods section, due to the constraints imposed by CHEMTAX on the minimum group size and homogenous pigment profiles, groups could only be subdivided geographically at the coarsest level. Subdivision by time of collection, such as collection in wet season vs. dry season, was not possible for the same reasons. Light regime at the time of collection is an additional unknown.

Compositional correlations were determined between SLE and the best configuration(s) in CHEMTAX, using identical coefficients, for representative sample groups in all 5 geographical areas. These included a freshwater-driven area with total phosphorous (TP) and total dissolved solids (TDS) gradients (WCA-1, transects X and Z) and an area with a predominance of calcareous (filamentous) cyanobacteria of marine origin (WCA-2A, transect U). WCA-3A, Shark River Slough, and two sample groups from Taylor Slough were also included. Community heterogeneity moves from very high in the north (WCA-1) to much lower in the south (Shark River and Taylor Sloughs).

All 7 groups contained estimates for chlorophytes, coccoidal cyanobacteria, filamentous cyanobacteria, and diatoms. Three groups contained estimates for cryptophytes and 2 of those groups also included dinoflagellates. The percentages of cryptophytes, coccoidal cyanobacteria, and dinoflagellates were universally low (<5% of the total sample) within the groups. As a whole, correlations were quite high (see Figure 5 and Table 9). Lower correlations were seen in the dinoflagellates and diatoms.

Surprisingly, one of the more homogenous areas, Taylor Slough, exhibited some of the lowest correlations. Hydroperiod is a main driver of variability in this area, however, and since the samples were analyzed without regard to time the low correlations may reflect a level of seasonal variability. Differences between the actual composition values returned by the two methods can be seen in Table 10. Minimum and maximum values of sample compositions shown in this table illustrate the level of variability between samples within the same group.

Influence of Pheopigments

High percentages of pheopigments in a sample indicate that processes are occurring which may make compositional determination much more difficult. Pigment ratios tend to be increasingly unpredictable, in part because different types of algae form pheopigments at very different rates (Deydier-Stephan et al. 2003; Louda 2008). The composition in particular is less predictable, and by logical extension, one overall set of ratios may no longer apply. The periphytometer field samples included 137 samples with >10% pheopigments, calculated as a percentage of total chlorophyll-*a* and derivatives. Since the true composition of the samples was unknown, a test was devised using two methods which should return compositions which are very close to each other. These results were assessed to see if incorporating samples with high pheopigments would cause the them to diverge more widely.

For this analysis, one sample group from a highly variable freshwater area, WCA-1, and one from a less variable freshwater area, Shark River Slough, were used. The Shark River Slough (SRS) sample was limited to transect 12C and contained 22 samples,

6 of which had >10% pheopigments. The WCA-1 group was limited to transect Z and contained 34 samples, of which 16 had >10% pheopigments. The most straightforward comparison to be made was between the two SLE models, using two sets of coefficients. All 6 algal classes were considered, with 1 biomarker pigment per class. For the SRS samples, correlations were measured for a group which included all samples and a group which excluded samples with high pheopigments. For WCA-1 samples, correlations were measured for a group which included all samples, a group which excluded samples with high pheopigments, and a group which only included samples with high pheopigments.

Somewhat surprisingly, all correlations were extremely close, with $r => 0.99$ ($p < 0.0001$) for all but one group (Table 11). This group, dinoflagellates in the WCA-1 Z transect measuring the sample with high pheopigments included, had a correlation coefficient of 0.95 ($p < 0.0001$). The most likely reason for the lack of effect of high pheopigments on the correlations is the similarity of the models. Only 3 biomarkers vary between these coefficient sets and this may not have been enough for an effect. Also, the classes for which ratios differed may not be the strongest influence in these samples.

Ratio Considerations

Initial testing of CHEMTAX on artificially generated datasets showed a potential bias. CHEMTAX results were very poor when using a ratio matrix with severely unequal row-sums, and were excellent when using a ratio matrix where row-sums were equal. This is somewhat surprising since the CHEMTAX program rescales the matrices prior to other operations. This bias was evaluated using several different ratio matrices form the literature to generate artificial datasets. Although no particular weakness could be found

in the artificial data generator, it is possible that one existed. The same bias occurred when using varied ratio matrices on the artificial mixes but was not nearly as pronounced (see Figure 6). It is possible that using a rescaled ratio matrix may give the model an easier problem to solve, increasing the likelihood that it will locate the global best solution rather than a local best solution.

If this is a true bias it may not be apparent unless the program is run using a ratio matrix with unequal row-sums against a dataset where the answers are known. The original CHEMTAX paper (M. Mackey et al. 1996) mentions use of artificial datasets to calibrate the performance of model to ecosystems other than the two that it was originally tested in. There is, however, no indication in the literature that this has been done. Mikel Latasa's (2007) use of artificial datasets, for example, was not designed to expand the analysis to other algal sets but rather to test a revised method of implementing CHEMTAX, and, as such, uses the values from the original Southern Ocean system as starting points. Based on the results above, it would seem that testing revised ratio matrices on samples with known compositions might be a wise precaution.

The treatment of zeaxanthin is another important ratio consideration. As outlined in the Methods section, the subtraction method of zeaxanthin allocation was retained for most model runs in order to facilitate comparison. An alternative method which included the actual zeaxanthin allocation, partitioning it between both types of cyanobacteria as well as chlorophytes, was utilized in a few sets of BCE runs. However, this change had very little consistent effect, particularly on chlorophytes. Quite probably this was due to the fact that the zeaxanthin : chlorophyll-*a* ratios involved were small (0.056 for

filamentous cyanobacteria and 0.008 for chlorophytes) relative to the allowed standard deviation of the ratio matrix, which was typically 0.2 (see Van den Meersche et al. 2008). Since the allocation of coccoidal versus filamentous cyanobacteria was a recurring weak spot in the analyses, further work with the zeaxanthin ratios may be needed.

Model Comparison

These models can be compared in a number of ways, both qualitatively and quantitatively. See Table 12 for a summary comparison. In terms of accuracy, SLE and CHEMTAX performed equivalently throughout, and both outperformed BCE on actual laboratory and field samples. No model, however, performed well on the field sample microscopy comparison.

CHEMTAX was far more difficult to set up, due to the age of the program and outdated parameters. Once set up, however, CHEMTAX was easier to utilize than BCE. A great strength of BCE is that it can handle smaller sample groups. However, since each group has to be separately parameterized and adjusted through several runs, sometimes by looking at extensive output, the small group size does render it ultimately less efficient than CHEMTAX. Based on this, SLE might be the current method of choice for very small samples.

CHEMTAX is relatively insensitive to small parameter changes. Small changes in ratios as well as large changes in weights and ratio limits often had no effect on the final composition. Even with ratio limits set at 500, the CHEMTAX revised ratio matrices were seldom very different from the original for our samples. BCE, by contrast, is

extremely sensitive to small changes. Often a change in the BCE parameters will prevent the program from converging on an answer.

Consistency of BCE was an issue. Some BCE runs using the same sets of parameters produced widely varying revised ratio matrices. It was found that if subsequent runs produced varying or trivial answers, often there was a mixing problem. Although mixing is better in version 1.4 than in version 1.3, this is still a shortfall. Often a BCE best answer was still not ideal in terms of mixing or covariance.

A drawback for BCE is that no literature yet exists on implementing it other than the authors' original paper (Van den Meersche et al. 2008). CHEMTAX could not have been implemented in this study had it not been for the extensive literature describing its use in the field (see especially Latasa 2007; Wright 2006; D. Mackey and Higgins 2001).

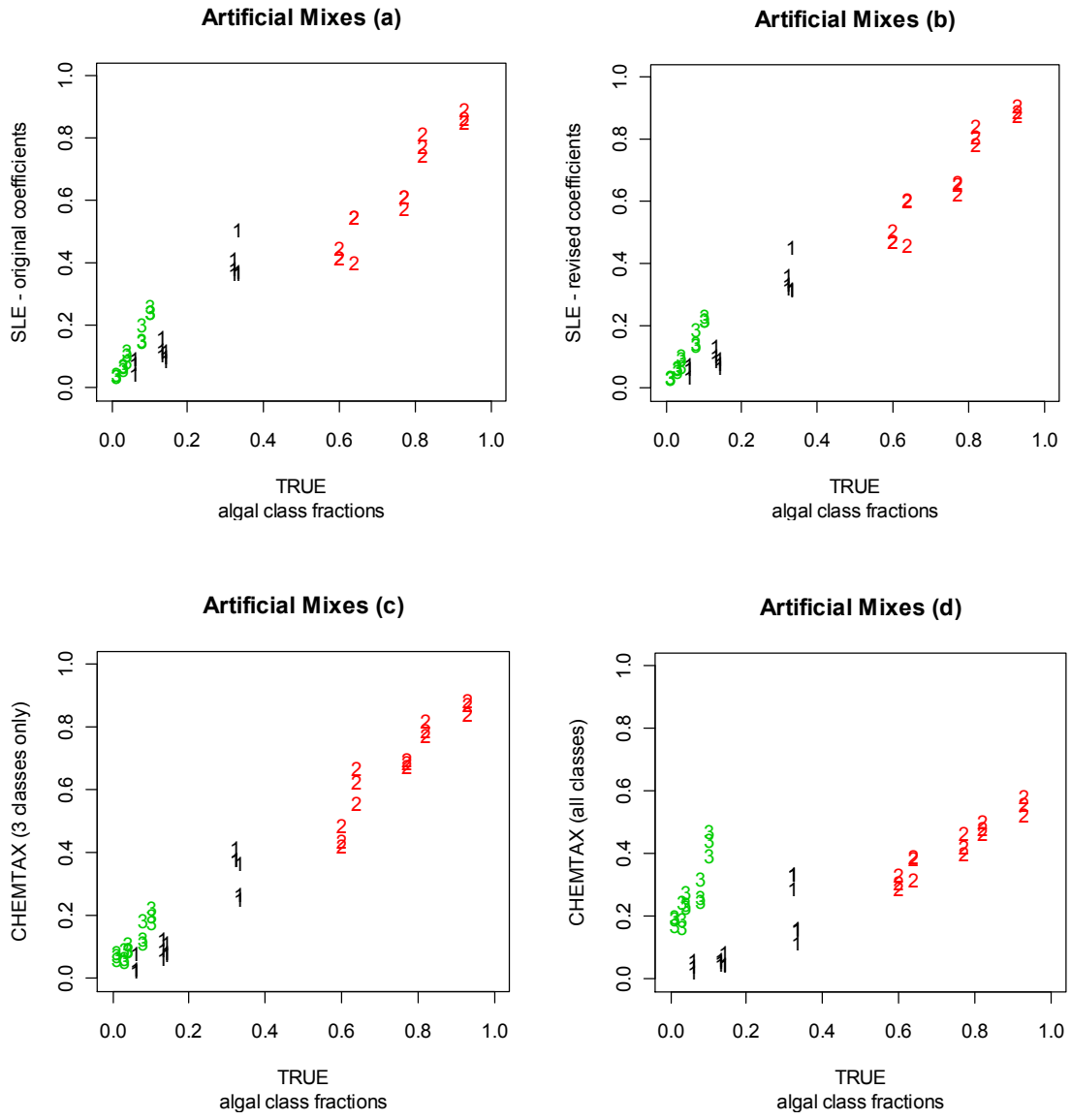


Figure 3. Summary plots of algal composition fractions for artificial mixes, as determined by various methods. (**1** = chlorophytes, **2** = coccoidal cyanobacteria, **3** = diatoms.)

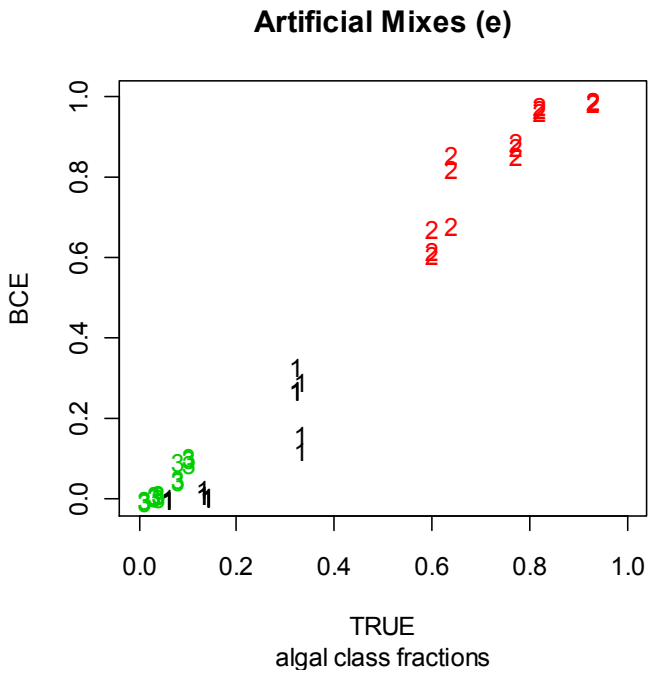


Figure 3, cont. Summary plots of algal composition fractions for artificial mixes, as determined by various methods. (**1** = chlorophytes, **2** = coccoidal cyanobacteria, **3** = diatoms.)

Table 7. Correlations for artificial mixes. This table shows the Pearson product-moment correlations between known compositions and compositions returned by the different methods.

true composition vs:	chloro	cya-cocc	diatom
a) SLE - original coefficients	0.960	0.956	0.970
b) SLE - revised coefficients	0.957	0.953	0.977
c) CHEMTAX (3 classes)	0.936	0.938	0.905
d) CHEMTAX (6 classes)	0.808	0.956	0.871
e) BCE	0.892	0.898	0.946

(df=13, $p < 0.0001$ for all)

Table 8. Correlations for grab samples. This table shows the Pearson product-moment correlations between microscopically-determined biovolume compositions and compositions returned by the different methods. The second group excluded 5 samples where chlorophyll-*b* = 0.

biovolume vs:	df	chloro	tot cyano	diatom
a) SLE - revised coeff	25	0.311 *	0.445 **	0.475 **
b) SLE - original coeff	25	0.309 *	0.444 **	0.453 **
c) CHEMTAX	25	0.365 **	0.568 **	0.448 **
d) BCE	25	0.252	0.191	-0.003
e) SLE - revised coeff	20	0.272	0.260	0.345 *
f) SLE - original coeff	20	0.269	0.257	0.319 *
g) CHEMTAX	20	0.317 *	0.353 *	0.286 *
h) BCE	20	0.216	0.072	0.023

(significant p=values are indicated by * ($0.05 < p < 0.1$) and ** ($p < 0.05$))

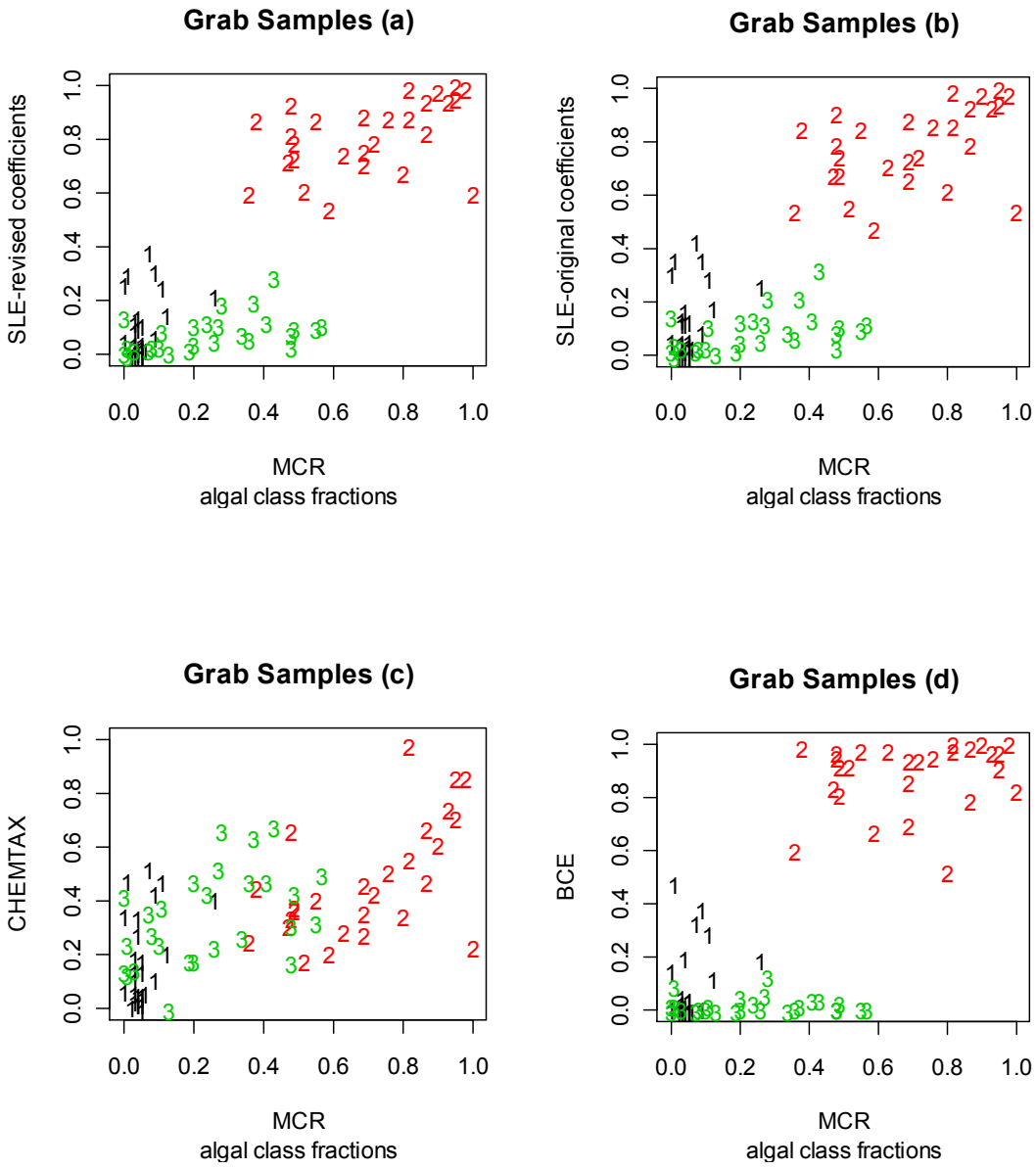


Figure 4. Grab Samples: Correlations between compositions determined by microscopy (MCR) and other methods. (**1** = chlorophytes, **2** = total cyanobacteria, **3** = diatoms.)

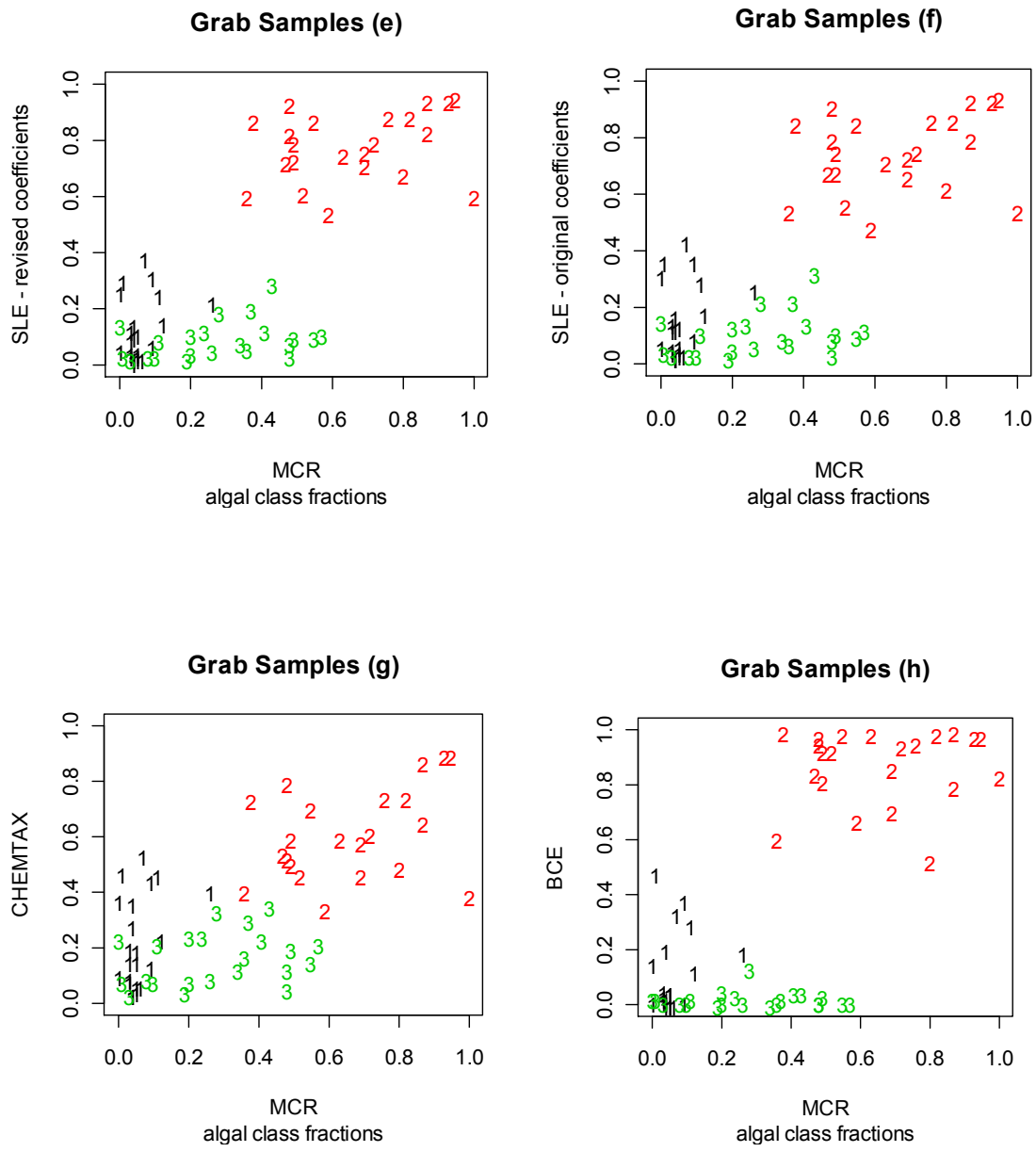


Figure 4, cont. Grab Samples: Correlations between compositions determined by microscopy (MCR) and other methods. (**1** = chlorophytes, **2** = total cyanobacteria, **3** = diatoms.)

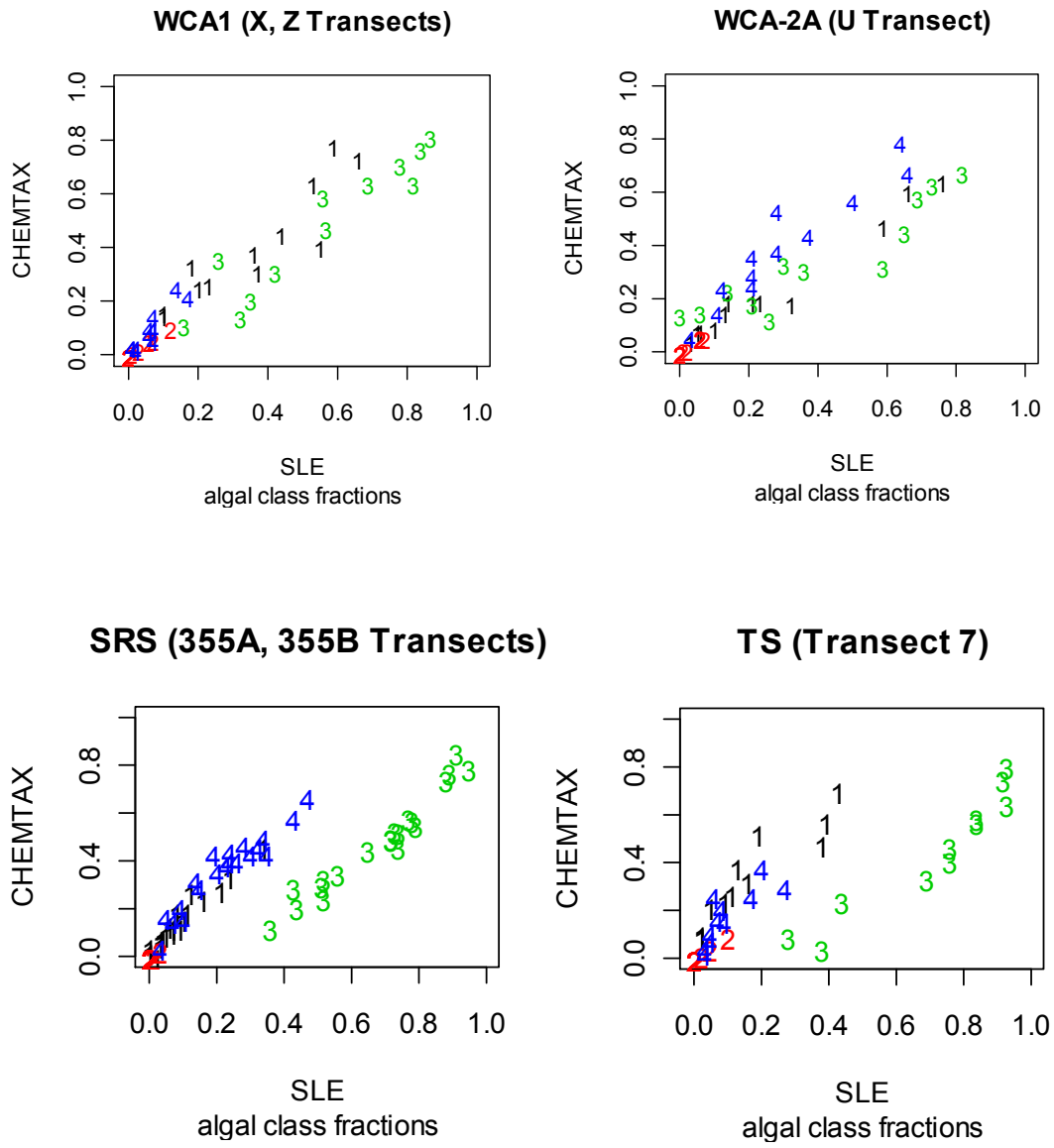


Figure 5-a. Periphytometers: Correlations between compositions determined by various methods for 4 algal classes. (**1** = chlorophytes, **2** = coccoidal cyanobacteria, **3** = filamentous cyanobacteria, **4** = diatoms.)

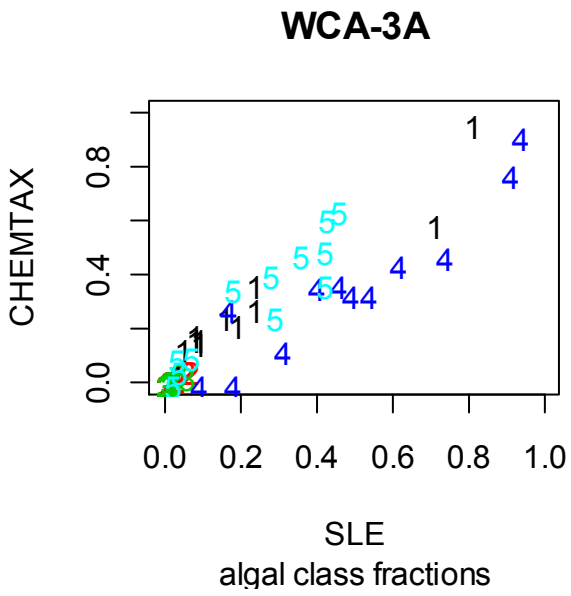


Figure 5-b. Periphytometers: Correlations between compositions determined by various methods for 5 algal classes. (**1** = chlorophytes, **2** = cryptophytes, **3** = coccoidal cyanobacteria, **4** = filamentous cyanobacteria, **5** = diatoms.)

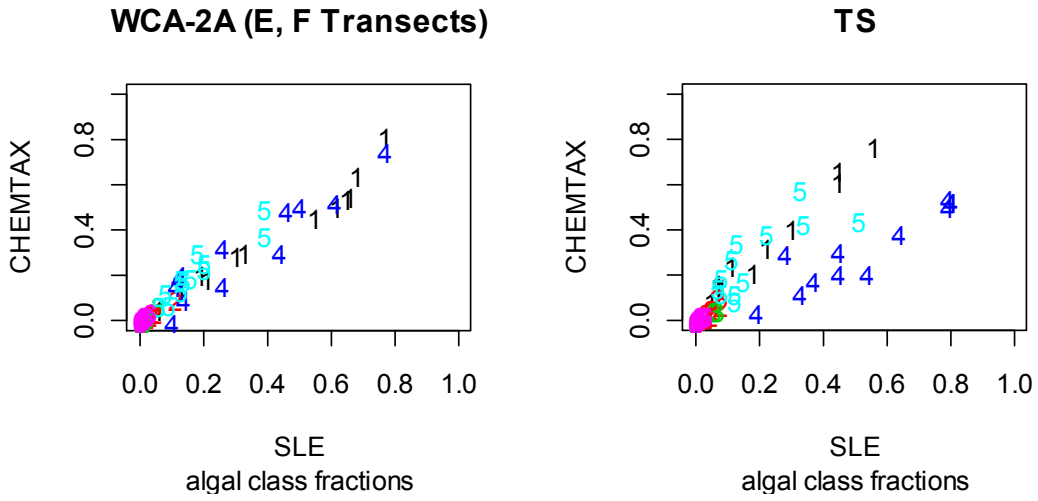


Figure 5-c. Periphytometers: Correlations between compositions determined by various methods for 6 algal classes. (**1** = chlorophytes, **2** = cryptophytes, **3** = coccoidal cyanobacteria, **4** = filamentous cyanobacteria, **5** = diatoms, **6** = dinoflagellates.)

Table 9. Correlations for periphytometer samples. This shows the Pearson product-moment correlations for results obtained by SLE and CHEMTAX.

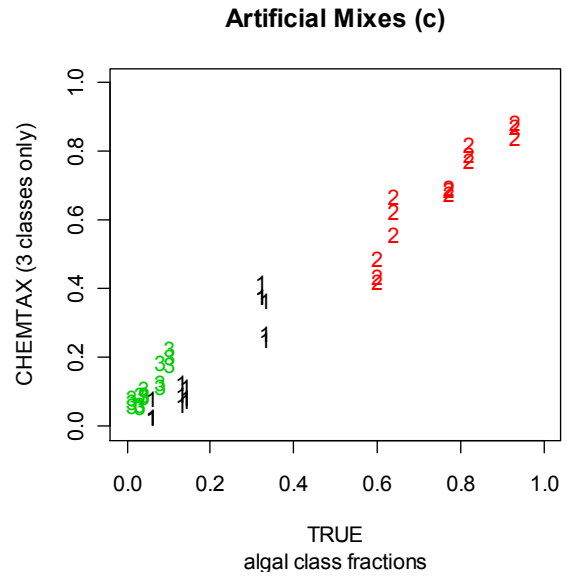
SLE vs CT	**chloro**	**crypto**	**cya-cocc**	**cya-fil**	**diatom**	**dinofl**
WCA-1, transects X&Z						
n=12, df=10, p<.001	0.906	NE	0.999	0.947	0.938	NE
WCA-2A, transect U						
n=12, df=10, p<.001	0.982	NE	0.984	0.922	0.950	NE
WCA-2A, transects E&F						
n=12, df=10, p<.001	0.976	0.994	0.955	0.943	0.942	0.873
WCA-3A, all transects						
n=12, df=10, p<.001	0.963	0.987	0.898	0.932	0.934	NE
SRS, transects 355A&B						
n=20, df=18, p<.001	0.968	NE	0.947	0.973	0.953	NE
TS, all transects						
n=12, df=10, p<.01	0.990	0.854	0.996	0.922	0.787	0.861
TS, transect 7						
n=11, df=9, p<.01	0.923	NE	0.996	0.955	0.814	NE

Table 10. Periphytometer sample compositions, as determined by SLE and CHEMTAX.

SLE	MEAN	MIN	MAX	***CHEMTAX***	MEAN	MIN	MAX
WCA-1 (X, Z)				WCA-1 (X, Z)			
chloro	0.36	0.10	0.66	chloro	0.40	0.15	0.78
cya-coc	0.02	0.00	0.12	cya-coc	0.02	0.00	0.10
cya-fil	0.55	0.16	0.87	cya-fil	0.48	0.11	0.81
diatom	0.06	0.01	0.17	diatom	0.10	0.02	0.25
WCA-2A (U)				WCA-2A (U)			
chloro	0.27	0.03	0.76	chloro	0.24	0.04	0.64
cya-coc	0.02	0.00	0.07	cya-coc	0.02	0.00	0.06
cya-fil	0.40	0.00	0.82	cya-fil	0.34	0.12	0.67
diatom	0.30	0.03	0.66	diatom	0.39	0.06	0.79
WCA-2A (E, F)				WCA-2A (E, F)			
chloro	0.43	0.06	0.77	chloro	0.39	0.06	0.82
crypto	0.04	0.01	0.11	crypto	0.04	0.01	0.10
cya-coc	0.02	0.00	0.06	cya-coc	0.01	0.00	0.07
cya-fil	0.33	0.10	0.77	cya-fil	0.31	0.00	0.75
diatom	0.18	0.06	0.39	diatom	0.22	0.07	0.50
dinofl	0.01	0.00	0.03	dinofl	0.02	0.00	0.04
WCA-3A				WCA-3A			
chloro	0.23	0.03	0.81	chloro	0.29	0.04	0.96
crypto	0.02	0.00	0.07	crypto	0.01	0.00	0.05
cya-coc	0.01	0.00	0.06	cya-coc	0.00	0.00	0.04
cya-fil	0.49	0.09	0.94	cya-fil	0.37	0.00	0.91
diatom	0.25	0.02	0.46	diatom	0.32	0.00	0.64
SRS (355 A, B)				SRS (355 A, B)			
chloro	0.09	0.00	0.34	chloro	0.15	0.00	0.46
cya-coc	0.00	0.00	0.03	cya-coc	0.00	0.00	0.03
cya-fil	0.68	0.36	0.95	cya-fil	0.48	0.12	0.85
diatom	0.22	0.03	0.47	diatom	0.37	0.04	0.67
TS				TS			
chloro	0.22	0.05	0.56	chloro	0.33	0.09	0.77
crypto	0.04	0.01	0.08	crypto	0.04	0.01	0.11
cya-coc	0.01	0.00	0.07	cya-coc	0.00	0.00	0.05
cya-fil	0.54	0.19	0.80	cya-fil	0.32	0.04	0.54
diatom	0.19	0.07	0.51	diatom	0.28	0.09	0.58
dinofl	0.01	0.00	0.03	dinofl	0.02	0.00	0.03
TS (7)				TS (7)			
chloro	0.18	0.02	0.43	chloro	0.35	0.09	0.70
cya-coc	0.02	0.00	0.10	cya-coc	0.01	0.00	0.09
cya-fil	0.71	0.28	0.93	cya-fil	0.45	0.04	0.80
diatom	0.10	0.03	0.27	diatom	0.18	0.03	0.38

Table 11. Pheopigment correlations. This shows the Pearson product-moment correlations for results obtained by SLE using original and revised coefficients, comparing samples with <10% and >10% pheopigments.

SLE (original) vs SLE (revised)	**chloro**	**crypto**	**cya-coc**	**cya-fil**	**diatom**	**dinofl**
SRS (12C) all samples n=22; df=20; p<.0001	0.998	0.990	0.983	0.999	0.997	1.000
SRS (12C) low pheo samples only n=16; df=14; p<.0001	0.998	0.988	0.991	0.999	0.997	1.000
WCA-1 (Z) all samples n=34; df=32; p<.0001	0.998	0.987	0.995	0.998	0.999	0.948
WCA-1 (Z) low pheo samples only n=18; df=16; p<.0001	0.998	0.995	0.995	0.999	0.999	1.000
WCA-1 (Z) high pheo samples only n=16; df=14; p<.0001	0.998	0.986	0.995	0.999	1.000	1.000



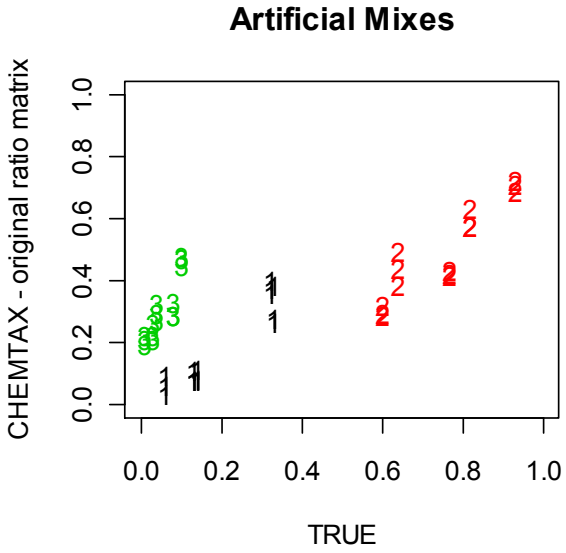


Figure 6. Effect of different ratio matrices on CHEMTAX correlations. These correlation plots for the artificial mixes show the effect of removing the chlorophyll-*a* term from the CHEMTAX ratio matrix. The second plot shows results when using a ratio matrix without chlorophyll-*a,* which increases the inequality in the ratio matrix row-sums.

Table 12. Model comparison. A qualitative comparison of the three models used in this study.

method	initial cost	updates	setup	implementation
SLE	none	n/a	easy	easy
CHEMTAX	none, but requires MatLab purchase	by request (1 since 1996)	difficult due to outdated interface	relatively easy
BCE	none (open source software)	available for download as they are written	easy	relatively difficult (especially MCMC behavior)

method	weak areas	strengths
SLE	each class must have a unique biomarker	no ambiguity in results; sum of chlorophyll-*a* can be used to verify
CHEMTAX	often finds a local rather than a global best solution; groups must be a minimum size and have a very similar biomarker makeup	handles groups of samples; can resolve classes with shared biomarkers; revises the ratios
BCE	can be difficult to get consistent results; very sensitive to initial input parameters; can be hard to tell if the MCMC has actually converged	similar to CHEMTAX, but can handle smaller groups with more variability; result is a probability distribution

CONCLUSIONS

None of the models evaluated were able to adequately describe the periphyton community, based on the poor correlations with microscopy data. All three models performed well on artificial datasets and very constrained artificial mixes. As samples became more complex, BCE correlations progressively worsened. However, all BCE runs with poor correlations showed weaknesses in mixing, acceptance, or covariance. This indicates that if parameters can be located which will return good acceptance and mixing and low covariance, results may be able to be trusted.

In general, SLE and CHEMTAX performed equivalently. There were indications that the different models might handle certain algal classes best, but more testing would need to be done to be certain.

Several recommendations can be made to improve the performance of the models and to isolate and potentially correct weak spots. I recommend running BCE first against more complex samples with known compositions, since parameterization of BCE currently is problematical. For CHEMTAX, the issue may lie more in the correct ratios than in parameterization, especially since it usually performed equivalently to SLE. Ratios might be able to be tested more effectively on a more constrained periphytometer field sample set. Knowledge of the light regime at the time of sampling may be essential for good results from periphyton. Ideally, the results from both these models could be tested against a larger biovolume comparison, preferably from a constrained sample area.

The echinenone : zeaxanthin ratio in filamentous cyanobacteria may need to be reevaluated. The subtraction method used in the periphytometer samples returned a substantial amount of negative values. Although these were forced to zero for the current analysis, this indicates that the 1:1 ratio now in use may not be valid at least for all light regimes. Although there is little precedent in the literature, I would also recommend including a secondary biomarker pigment for coccoidal cyanobacteria. Some of the composition problems in cyanobacteria determinations may have arisen because zeaxanthin is shared between half the algal classes considered, but is the only biomarker for coccoidal cyanobacteria. Inclusion of a secondary pigment for this class might help the numerical routines adjust the zeaxanthin allocation better, leading to a more accurate estimation of the class compositions.

Some specific improvements in the way BCE and CHEMTAX are run may need to be made. CHEMTAX has been reported to more easily converge on a global solution when several runs are performed using different ratios as starting points (Latasa 2007; Wright and Jeffrey 2006). BCE showed a substantial improvement when a relative standard deviation matrix was used which specified a very small standard deviation for chlorophyll-*a* and small vs. large standard deviations for photoaccessory vs. photoprotective pigments. There are several other parameters for which matrices rather than one overall value can be used. In particular, a matrix for the jump length in the compositions (jmpX) may improve mixing and convergence. A method which increases the chances of global convergence in MCMC simulations by using multiple runs with very different starting points (Gelman 1996) may also be worth implementing.

APPENDIXES

Appendix A. Field Sample Data Preparation

Written records of all samples and analyses are maintained in the Florida Atlantic University Environmental Biogeochemistry Laboratory. Data from the field sample pigment analyses were also stored digitally at the time of analysis in Microsoft Excel (v. 2000 and 2003) as one pigment calculation ("PigCalc") record per spreadsheet file. A record is defined as the analysis of one sample, i.e., the initial, hplc, and calculated results for one site on one date. The PigCalc spreadsheets were grouped into folders by "pickup date," which is the date each group of samples was picked up to be brought to the lab and analyzed. For each pickup date, there is also an Excel summary sheet which contains more information about all the samples. This includes a transcription of the data from the sample labels such as deployment and retrieval dates, the condition of the samples, and any technical issues encountered in the analysis. For the entire sample set, a running summary sheet was maintained which included pigment concentrations for each sample, with area name, site name, deploy date, retrieve date, pickup date, and analysis date also recorded.

Originally, the protocol for recording the South Florida Water Management Division sample analyses was set up with the lab's current chemotaxonomic method, simultaneous equations (SLE), in mind. In contrast to factor analysis methods such as

CHEMTAX, which use a large set of samples in a data matrix, SLE is effectively only applied to one sample at a time. Although batch calculations can be set up which provide an answer for many samples at once, the actual calculation is performed independently on each sample. The PigCalc spreadsheet is constructed to allow several analyses in addition to the algal class breakdown, and consequently each record contains a significant amount of data, calculations, and graphics. Thus, initially there were more reasons to keep the samples separate than to combine them into one file. However, this arrangement did present a challenge when attempting to compile the data for matrix input.

The original spreadsheet records had been constructed over a span of three years and the PigCalc spreadsheet itself had evolved over time, with, for example, 26 pigments measured in 2005 as opposed to 29 pigments in 2008. Thus, data fields were not necessarily in identical locations between spreadsheets. Also, retrieval dates were not transcribed onto the individual spreadsheets. These factors made use of any type of macro or Visual Basic code to quickly compile the data impractical. It was determined that each spreadsheet would need to be individually handled to some extent, and thus the data was compiled by hand, using cut-and-paste where possible, in a several-step process.

The retrieval date was determined from cross-referencing the pickup date on the sample sheet's folder to the summary sheet that existed for each pickup date. Area names were determined by matching site names to a master list of sampling sites. For each spreadsheet, a second worksheet, or page, was created containing just the pertinent data, consisting of pigments weights and the sum of chlorophyll-*a* and derivative weights, in a set cell range. A standardized area name, site name, and retrieval date were input by

hand. Where an exact retrieval date was not supplied an estimate was recorded and the criteria for the estimation were noted.

These pages were collated into a master spreadsheet and a unique identifier was generated for each record based on area, site number, and retrieval date. A randomly chosen subset of records, approximately 5% of the total, was proofread against the original PigCalc sheets. This spreadsheet thus identified each sample uniquely and contained the following data for each: area name, site name, identifier, deploy date if known, retrieval date, and pickup, processing, and analysis dates, measured weight of all pigments individually, sum of chlorophyll-*a* and derivatives, percentage of pheopigments, and any calculated sums of pigments necessary for the chemotaxonomic equations. This sheet was used as the source to pull out data necessary for input files, which were the identifier, biomarker pigment weights including total chlorophyll-*a*, and percent pheopigments.

Appendix B. R Code for Artificial Data Generation

```
#input ratio matrix, desired number of samples, and CV
build.x<-function(rat,n.samples,cv){

#read in the ratio matrix
species.pigs=rat
n.species=length(species.pigs[,1])
n.pigs=length(species.pigs[1,])
c(n.species,n.pigs)
species.pigs.matrix=matrix(nrow=n.species,ncol=n.pigs)
```

```
for(i in 1:n.species){
        for(j in 1:n.pigs){
                species.pigs.matrix[i,j]=species.pigs[i,j]
        }
}

#create an empty composition matrix for n samples (from input) and n species (from ratio matrix)
comp.matrix=matrix(nrow=n.samples,ncol=n.species)

#create an empty data matrix for n samples (from input) and n pigments (from ratio matrix)
data.matrix=matrix(nrow=n.samples,ncol=n.pigs)

#create one sample per loop up to n samples
for(s in 1:n.samples){

#create a matrix of absolute variations using the CV value you input at beginning
random.masses=matrix(rnorm(n.species*n.pigs,mean=1,sd=cv),nrow=n.species,ncol=n.pigs)

#create the species composition for each sample
species.comp=array(0,n.species)
propor=runif(n.species-1)
propor=sort(propor)
species.comp[1]=propor[1]
        for(i in 2:(n.species-1))species.comp[i]=propor[i]-propor[i-1]
species.comp[n.species]=1-propor[n.species-1]
comp.matrix[s,]=species.comp

#create the data
sample.matrix=species.pigs.matrix*random.masses
sample.masses=species.comp%*%sample.matrix
data.matrix[s,]=sample.masses
                }
#write out final matrices to files
#note: this function writes output to a directory specified in the function; change as needed
write.table(comp.matrix,file="c:/tempR/comp.txt")
write.table(data.matrix,file="c:/tempR/data.txt")
}
```

Appendix C. Sample .m File for CHEMTAX

This is an example of a .m file used for CHEMTAX input which includes all 6 classes of algae. The data has been truncated to 5 samples for presentation here. Normally, however, a sample size at least one greater than the number of biomarker pigments would be required.

```
%SAMPLE ID'S for reference
%3AE2038443  3AE2038981    3AE2039063    3AE4038981    3AE403906

% Data matrix
S = [0   6.62368E-09    1.49192E-08    0        5.39607E-09    4.71988E-08    4.60092E-09
         0        1.40308E-09    1.60553E-07
0        2.3728E-08     1.5957E-09     3.8E-09  2.60234E-08    7.46386E-09    1.5934E-08
         0        0        5.50311E-07
5.92E-10         8.2415E-09     6.05023E-09    0        2.17916E-08    1.05361E-08
         6.65274E-09    0        0        3.02767E-07
0        6.00234E-09    1.87527E-09    0        3.01542E-08    1.65482E-08    2.96699E-09
         0        0        3.63508E-07
0        4.64244E-09    9.35914E-09    0        5.86922E-09    1.91386E-08    4.06185E-09
         0        0        1.31714E-07
];

% Ratio matrix
F0 = [0  0.406   0        0        0        0        0.165    0        0.008    1
0.166    0        0        0        0        0        0        0        0        1
0        0        0        0        0        0        0        0        0.579    1
0        0        0        0        0.056    0        0        0        0        1
0        0        0.278    0        0        0.738    0        0        0        1
0        0        0.385    0.042    0        0        0        0.543    0        1
];
```

```
% Ratio limit matrix
ratiolimit = [500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500
500 500 500 500 500 500 500 500 500 500
];

% Names of the pigments
pignames = ['alx  '
'chb  '
'dia  '
'din  '
'ech  '
'fuc  '
'lut  '
'per  '
'zea  '
'cha  '
];

% Names of the classes
speciesnames = ['chloro      '
'crypto      '
'cyacoc      '
'cyafil      '
'diatom      '
'dinofl      '
];


% Diary file name
diaryname = 'samplediary.txt';

% Output file name
outname = 'sample';
```

```
% Data weights
weights = [30    5        10       30       10       3        10       20       20       1];

% Maximum number of iterations
maxiters = 10000;

% Stop calculation when residual is below:
errlimit = 1E-0004;

% Initial step size (inverse)
divisor = 25;

% Increase step size by a factor of:
stepratio = 2;

% Halt calculation when step is larger than:
steplimit = 30000;

% Verbosity
verbose = 3;

% Vary how many ratios at each step? default to number of pig's
numvaried = 10;

% Interval between redetermination of what ratios to vary; default to 1
numsubiter = 1;

% Vector indicating what files to output
outputtypes = [1 1 3 1 9 1 10 0 10 1];
```

Appendix D. BCE Input Examples

Reading files (data matrix, ratio matrix, and standard deviation matrix) into R:

```
> ratios<-read.table("RatT1.txt")
> data<-read.table("TNPT1.txt")
> sd<-read.table("sdT1.txt")
```

Running BCE using those same files:

```
> X<-BCE(ratios,data,relsdRat=sd,relsdDat=.1,iter=20000,outputlength=5000,burninlength=
1000, jmpX=.01,jmpRat=.2*(.2*ratios),export=TRUE)
```

REFERENCES

Browder, Joan A., Sally Black, Peter Schroeder, Melvin Brown, Mark Newman, Dan Cottrell, David Black, Robert Pope, and Peck Pope. 1981. Perspective on the ecological causes and effects of the variable algal composition of southern Everglades periphyton. South Florida Research Center Report T-643. Homestead, FL: South Florida Research Center, Everglades National Park.

Browder, Joan A., Patrick J. Gleason, and David R. Swift. 1994. Periphyton in the Everglades: spatial variation, environmental correlates, and ecological implications. In *Everglades: the ecosystem and its restoration,* ed. Steven M. Davis and John C. Ogden, 379-418. Delray Beach, FL: St. Lucie Press.

Browder, Joan A., and John C. Ogden. 1999. The natural South Florida system II: predrainage ecology. *Urban Ecosystems* 3: 245-77.

Carlin, Bradley P., and Thomas A. Louis. 2000. *Bayes and empirical Bayes methods for data analysis.* 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Clark, James S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8: 2-14.

Davis, Steven M., and John C. Ogden, eds. 1994. *Everglades: the ecosystem and its restoration.* Delray Beach, FL: St. Lucie Press.

Demmig-Adams, B. 1990. Carotenoids and photoprotection in plants: a role for the xanthophyll zeaxanthin. *Biochimica et Biophysica Acta* 1020: 1-24.

Descy, J.-P., M.-A. Hardy, S. Stenuite, S. Pirlot, B. Leporcq, I. Kimirei, B. Sekadende, S. R. Mwaitega, and D. Sinyenza. 2005. Phytoplankton pigments and community composition in Lake Tanganyika. *Freshwater Biology* 50: 668-84.

Descy, Jean-Pierre, Harry W. Higgins, Denis J. Mackey, James P. Hurley, and Thomas M. Frost. 2000. Pigment ratios and phytoplankton assessment in northern Wisconsin lakes. *Journal of Phycology* 36: 274-86.

Deydier-Stephan, Laurence, Georges Bertru, and Bertrand L. Rouzic. 2003. A chemotaxonomic method to quantify phytoplankton groups in freshwater lentic mesocosms: an approach including chlorophyll *a* breakdown products. *C. R. Biologies* 326: 95-105.

Ellison, Aaron M. 2004. Bayesian inference in ecology. *Ecology Letters* 7: 509-20.

Gameiro, C., P. Cartaxana, and V. Brotas. 2007. Environmental drivers of phytoplankton distribution and composition in Tagus Estuary, Portugal. *Estuarine, Coastal and Shelf Science* 75: 21-34.

Garibotti, Irene A., María Vernet, Wendy A. Kozlowski, and Martha E. Ferrario. 2003. Composition and biomass of phytoplankton assemblages in coastal Antarctic waters: a comparison of chemotaxonomic and microscopic analyses. *Marine Ecology Progress Series* 247: 27-42.

Gelman, A. 1996. Inference and monitoring convergence. In *Markov Chain Monte Carlo in practice,* ed. W. R. Gilkes, S. Richardson, and D. J. Spiegelhalter, 131-143, Boca Raton: Chapman & Hall / CRC.

Gieskes, Winfried W., and Gijsbert W. Kraay. 1983. Unknown chlorophyll *a* derivatives in the North Sea and the tropical Atlantic Ocean revealed by HPLC analysis. *Limnology and Oceanography* 28: 757-66.

Gilkes, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in practice,* ed. W. R. Gilkes, S. Richardson, and D. J. Spiegelhalter, 1-19. Boca Raton: Chapman & Hall / CRC.

Gilkes, W. R., and G. O. Roberts. 1996. Strategies for improving MCMC. In *Markov Chain Monte Carlo in practice,* ed. W. R. Gilkes, S. Richardson, and D. J. Spiegelhalter, 89-114. Boca Raton: Chapman & Hall / CRC.

Graham, Linda E., and Lee W. Wilcox. 2000. Algae. San Francisco: Benjamin Cummings.

Grant, Cidya S. 2006. Photosynthetic pigment ratios in relation to photic flux. M.S. thesis, Florida Atlantic University.

Greisberger, Sonja, and Katrin Teubner. 2007. Does pigment composition reflect phytoplankton community structure in differing temperature and light conditions in a deep alpine lake? An approach using HPLC and delayed fluorescence techniques. *Journal of Phycology* 43: 1108-19.

Guisande, C., A. Barreiro, A. Acuna, L. J. Marciales, E. Hernandez, A. M. Torres, N. Aranguren, W. Lopez, S. R. Duque,L. J. Gallo, N. Aguirre, M. Mogollon, J. Palacio, and G. Rueda-Delgado. 2008. Testing of the CHEMTAX program in contrasting Neotropical lakes, lagoons, and swamps. *Limnology and Oceanography: Methods* 6: 643-652.

Hager, A. 1980. The reversible, light-induced conversions of xanthophylls in the chloroplast. In *Pigments in Plants,* ed. F.-C. Czygan, 57-79. Stuttgart: Gustav Fischer.

Hagerthey, Scot E., J. William Louda, and Panne Mongkronsri. 2006. Evaluation of pigment extraction methods and a recommended protocol for periphyton chlorophyll *a* determination and chemotaxonomic assessment. *Journal of Phycology* 42, 1125-36.

Hammersley, J. M., and D. C. Hanscomb. 1964. *Monte Carlo methods.* London: Chapman & Hall.

Havskum, Harry, Louise Schlüter, Renate Scharek, Elisa Berdalet, and Stéphan Jacquet. 2004. Routine quantification of phytoplankton groups – microscopy or pigment analyses? *Marine Ecology Progress Series* 273: 31-42.

Higgins, H. W., and D. J. Mackey. 2000. Algal class abundances, estimated from chlorophyll and carotenoid pigments, in the western Equatorial Pacific under El Niño and non-El Niño conditions. *Deep-Sea Research I* 47: 1461-83.

Iwaniec, David M., Daniel L. Childers, Damon Rondeau, Christopher J. Madden, and Colin Saunders. 2006. Effects of hydrologic and water quality drivers on periphyton dynamics in the southern Everglades. *Hydrobiologia* 569: 223-35.

Jeffrey, S. W., R. F. C. Mantoura, and S. W. Wright, eds. 1997. *Phytoplankton pigments in oceanography: guidelines to modern methods.* Paris: United National Educational Scientific and Cultural Organization.

Kraay, Gijsbert W., Manuel Zapata, and Marcel J. W. Veldhuis. 1992. Separation of chlorophylls c_1, c_2, and c_3 of marine phytoplankton by reversed-phase-C18-high-performance liquid chromatography. *Journal of Phycology* 28: 708-12.

Latasa, Mikel. 2007. Improving estimations of phytoplankton class abundances using CHEMTAX. *Marine Ecology Progress Series* 329: 13-21.

Latasa, Mikel, and Robert R. Bidigare. 1998. A comparison of phytoplankton populations of the Arabian Sea during the Spring Intermonsoon and Southwest Monsoon of 1995 as described by HPLC-analyzed pigments. *Deep-Sea Research II* 45: 2133-70.

Lee, Jason H., Lisa Campbell, Tammi L. Richardson, James L. Pinckney, Rongjun Shen, and Norman L. Guisano Jr. 2005. Combining new technologies for determination of phytoplankton community structure in the northern Gulf of Mexico. *Journal of Phycology* 41: 305-10.

Letelier, R. M., R. R. Bidigare, D. V. Hebel, M. Ondrusek, C. D. Winn, and D. M. Karl. 1993. Temporal variability of phytoplankton community structure based on pigment analysis. *Limnology and Oceanography* 38: 1420-37.

Lewitus, Alan J., David L. White, Raphael G. Tymowski, Mark E. Geesey, Sabrina N. Hymel, and Peter A. Noble. 2005. Adapting the CHEMTAX method for assessing

phytoplankton taxonomic composition in southeastern U.S. estuaries. *Estuaries* 28: 160-72.

Lionard, Marie, Koenraad Muylaert, Michèle Tackx, and Wim Vyverman. 2008. Evaluation of the performance of HPLC-CHEMTAX analysis for determining phytoplankton biomass and composition in a turbid estuary (Schelde, Belgium). *Estuarine, Coastal and Shelf Science* 76: 809-17.

Louda, J. William. 2008. HPLC-based chemotaxonomy of Florida Bay phytoplankton: difficulties in coastal environments. *Journal of Liquid Chromatography & Related Technologies* 31: 295-323.

Louda, J. William, Jie Li, Lei Liu, M. Nancy Winfree, and Earl W. Baker. 1998. Chlorophyll-*a* degradation during cellular senescence and death. *Organic Geochemistry* 29: 1233-51.

Louda, J. William, Lei Liu, and Earl W. Baker. 2002. Senescence- and death-related alteration of chlorophylls and carotenoids in marine phytoplankton. *Organic Geochemistry* 33: 1635-53.

Mackey, Denis, and Harry Higgins. 2001. CHEMTAX – not just a 'black box.' Paper presented at the CSIC workshop "Pigments as a tool to estimate the biomass of different phytoplankton groups," Barcelona, Spain, April 29 – May 1, 2001.

Mackey, D. J., H. W. Higgins, M. D. Mackey, and D. Holdsworth. 1998. Algal class abundances in the western equatorial Pacific: estimation from HPLC measurements of chloroplast pigments using CHEMTAX. *Deep-Sea Research I* 45: 1441-68.

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

Mackey, M. D., D. J. Mackey, H. W. Higgins, and S. W. Wright. 1996. CHEMTAX a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Marine Ecology Progress Series* 144: 265-83.

Malinowski, Edmund R. 2002. *Factor analysis in chemistry.* 3rd ed. New York: John Wiley and Sons.

Mantoura, R. F. C., and C. A. Llewellyn. 1983. The rapid determination of algal chlorophyll and carotenoid pigments and their breakdown products in natural waters by reverse-phase high performance liquid chromatography. *Analyt. Chim. Acta* 151: 297-314.

Marinho, Marcelo Manzi, and Silvana Viana Rodrigues. 2003. Phytoplankton of an eutrophic tropical reservoir: comparison of biomass estimated from counts with chlorophyll-*a* biomass from HPLC measurements. *Hydrobiologia* 505: 77-88.

Marty, Jean-Claude, Jacques Chiavérini, Marie-Dominique Pizay, and Bernard Avril. 2002. Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991-1999). *Deep-Sea Research* II 49: 1965-85.

The MathWorks, Inc. 2008. MATLAB. Version R2008b. Natick, MA.

McCormick, Paul V., and R. Jan Stevenson. 1998. Periphyton as a tool for ecological assessment and management in the Florida Everglades. *Journal of Phycology* 34: 726-33.

Millie, David F., Hans W. Paerl, and James P. Hurley. 1993. Microalgal pigment assessments using high-performance liquid chromatography: a synopsis of organismal and ecological applications. *Canadian Journal of Fisheries and Aquatic Sciences* 50: 2513–27.

Muylaert, Koenraad, Rhia Gonzales, Melanie Franck, Marie Lionard, Claar Van der Zee, André Cattrijsse, Koen Sabbe, Lei Chou, and Wim Vyverman. 2006. Spatial variation in phytoplankton dynamics in the Belgian coastal zone of the North Sea studied by

microscopy, HPLC-CHEMTAX and underway fluorescence recordings. *Journal of Sea Research* 55: 253-65.

Obeysekeera, Jayantha, Joan Browder, Lewis Hornung, and Mark A. Harwell. 1999. The natural South Florida system I: climate, geology, and hydrology. *Urban Ecosystems* 3: 223-44.

O'Hagan, Anthony, and Jonathan Forster. 2004. *Kendall's advanced theory of statistics, vol. 2B: Bayesian inference.* 2nd ed. London: Arnold.

Omlin, Martin, and Peter Reichert. 1999. A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling* 115: 45-59.

Paerl, Hans W., Lexia M. Valdes, James L. Pinckney, Michael F. Piehler, Julianne Dyble, and Pia H. Moisander. 2003. Phytoplankton photopigments as indicators of coastal eutrophication. *BioScience* 53: 953-64.

Pinckney, James L., Tammi L. Richardson, David F. Millie, and Hans W. Paerl. 2001. Application of photopigment biomarkers for quantifying microalgal community composition and in situ growth rates. *Organic Geochemistry* 32: 585-95.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. 1992. *Numerical recipes in C: the art of scientific computing.* 2nd ed. Cambridge: Cambridge University Press.

R Development Core Team. 2009. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Raftery, A. E., and S. M. Lewis. 1996. Implementing MCMC. In *Markov Chain Monte Carlo in practice,* ed. W. R. Gilkes, S. Richardson, and D. J. Spiegelhalter, 115-130. Boca Raton: Chapman & Hall / CRC.

Rodriguez, Francisco, Manuel Varela, and Manuel Zapata. 2002. Phytoplankton assemblages in the Gerlache and Bransfield Straits (Antarctic Peninsula) determined by light microscopy and CHEMTAX analysis of HPLC pigment data. *Deep-Sea Research II* 49: 723-47.

Rufty, Alan. 2007. Comments on the reliability of Lawson and Hanson's linear distance programming algorithm: subroutine LDP. Cornell University e-print arXiv:0707.4651v1 [cs,MS]. http://arxiv.org/abs/0707.4651.

Sarmento, Hugo, and Jean-Pierre Descy. 2008. Use of marker pigments and functional groups for assessing the status of phytoplankotn assemblages in lakes. *Journal of Applied Phycology* 20: 1001-11.

Schlüter, L., T. L. Lauridsen, G. Krogh, and T. Jørgensen. 2006. Identification and quantification of phytoplankton groups in lakes using new pigment ratios – a comparison between pigment analysis by HPLC and microscopy. *Freshwater Biology* 51: 1474-85.

Schlüter, L., F. Møhlenberg, H. Havskum, and S. Larsen. 2000. The use of phytoplankton pigments for identifying and quantifying phytoplankton groups in coastal areas: testing the influence of light and nutrients on pigment/chlorophyll *a* ratios. *Marine Ecology Progress Series* 192: 49-63.

Stal, Lucas J. 1995. Physiological ecology of cyanobacteria in microbial mats and other communities. Tansley Review No. 84. *New Phytologist* 131: 1-32.

Van den Meersche, Karel, and Karline Soetaert. 2009. *BCE: Bayesian composition estimator: estimating sample (taxonomic) composition from biomarker data.* R package version 1.4. http://CRAN.R project.org/package=BCE.

Van den Meersche, Karel, Karline Soetaert, and Jack J. Middleburg. 2008. A Bayesian compositional estimator for microbial taxonomy based on biomarkers. *Limnology and Oceanography: Methods* 6: 190-99.

Van den Meersche, Karel, Karline Soetaert, and Dick Van Oevelen. 2009. xsample(): an R function for sampling linear inverse problems. *Journal of Statistical Software* 30 (Code Snippet 1): 1-15.

Vymazal, Jan, and Curtis J. Richardson. 1995. Species composition, biomass, and nutrient content of periphyton in the Florida Everglades. *Journal of Phycology* 31: 343-54.

Wright, Simon W. 2005. Analysis of phytoplankton populations using pigment markers. Course notes for a workshop "Pigment Analysis of Antarctic Microorganisms," University of Malaya, June 29 – July 1, 2005. http://www.aad.gov.au/MediaLibrary/asset/MediaItems/ml_38651523287037_Analysis of phytoplankton pigments.pdf.

Wright, Simon W., Akira Ishikawa, Harvey J. Marchant, Andrew T. Davidson, Rick L. van den Enden, and Geraldine V. Nash. 2009. Composition and significance of picophytoplankton in Antarctic waters. *Polar Biology* 32:797-808.

Wright, Simon W., and S. W. Jeffrey. 2006. Pigment markers for phytoplankton production. In *Handbook of environmental chemistry, vol. 2N: marine organic matter, biomarkers, isotopes and DNA,* ed. J. K. Volkamn, 71-104. Berlin: Springer.

Wright, S. W., S. W. Jeffrey, R. F. C. Mantoura, C. A. Llewellyn, T. Bjørnland, D. Repeta, and N. Welschmeyer. 1991. Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. *Marine Ecology Progress Series* 77: 183-96.

Wright, S. W., D. P. Thomas, H. J. Marchant, H. W. Higgins, M. D. Mackey, and D. J. Mackey. 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: 285-98.

Zar, Jerrold H. 1999. *Biostatistical Analysis.* 4th ed. Delhi: Dorling Kindersley / Pearson Education.