TESTS OF A SEMI-ANALYTICAL CASE 1 AND GELBSTOFF CASE 2 SEAWIFS ALGORITHM WITH A GLOBAL DATA SET

Kendall L. Carder, Steve K. Hawes, and Zhongping Lee

Marine Science Department
University of South Florida
140 Seventh Avenue South
St. Petersburg, Florida 33701
kcarder@monty.marine.usf.edu

1.0 Introduction

A number of empirical and semi-analytical optical models have been developed to simulate the behavior of the underwater light field for Case 1 waters (Morel and Prieur, 1977; Smith and Baker, 1982; Baker and Smith, 1982; Gordon et al., 1988; Morel, 1988; Mitchell and Holm-Hansen, 1991). Case 1 waters are dominated by the optical properties of phytoplankton and covarying detrital byproducts of production. Such models have been used as the basis for classifying water types and/or for developing remote sensing algorithms.

However, the accuracies of these models decrease when environmental conditions depart from those represented in the data set used to empirically derive the covariance relationships. For instance, gelbstoff is produced when grazing, photolysis, and other mechanisms degrade the viable plant matter at and downstream from phytoplankton blooms. The gelbstoff-to-chlorophyll ratio will change dramatically for a parcel of upwelled water over a relatively short time, from chlorophyll-rich and gelbstoff-poor to gelbstoff-rich and chlorophyll-poor. Solid evidence for the occurrence of this scenario can be found in two separate studies. Peacock et al. (1988) found that absorption attributed to gelbstoff at 440 nm was at least 16 fold that due to phytoplankton pigments within an offshore jet from an upwelling region, whereas pigments were the dominant absorption agents at the upwelling center near the coast. Similarly, Carder et al. (1989) found that particulate absorption at 440 nm decreased 13 fold while gelbstoff absorption at 440 nm increased by 60% in ten days for a phytoplankton bloom tracked from the Mississippi River plume to Cape San Blas. This widely varying gelbstoff-to-chlorophyll ratio has a profound effect on upwelled radiance in the blue 443 nm band of the CZCS, and a smaller but still significant effect in the green 520 nm band. The correspondence in absorption at 443 nm and 520 nm between gelbstoff and chlorophyll creates erroneously high estimates of pigment concentration in those models which rely solely upon either of these spectral bands to indicate absorption due to phytoplankton.

Carder et al. (1991) proposed that a short wavelength channel at around 410 nm could be used to distinguish gelbstoff (and other degradation products) from chlorophyll. A channel at 412 nm will be available not only on the Sea-Viewing-Wide-Field-Sensor (SeaWiFS), but also on the Ocean Color and Temperature Scanner (OCTS), and on the Moderate Resolution Imaging Spectrometer (MODIS). A semi-analytical chlorophyll *a* algorithm for Case 1 and gelbstoff-rich Case 2 waters has been developed (Carder et al., 1996; Carder et al,. 1997) and will be thoroughly tested during the SeaWiFS project. Only a brief synopsis of the algorithm and recent upgrades will be reported here.

Extensive field data sets are needed to evaluate model performance with time and space. Performance should be best where parameterization has captured the natural variations in pigment packaging for the dominant plankton groups present, largely a function of nutrient availability and recent light history. Acquiring such data sets on a global scale is a major community goal during the next few years, and SeaBAM provides just the beginning of such tests. We have developed a scenario that can both guide the parameterization process and provide an initial implementation of the algorithm for much of the ocean. We will test the model against a well-calibrated global data set contributed by investigators using four different sensors, two with measurements collected just below the sea surface and two collected just above.

2.0 Algorithm Description

After light enters the ocean, some of it is eventually scattered back up through the surface. This light is called the water-leaving radiance, $L_w(\lambda)$, and it can be detected from space. The magnitude, spectral variation, and angular distribution of this radiance depend on the following: the absorption and backscattering coefficients of the seawater, $a(\lambda)$ and $b_b(\lambda)$, respectively (known as the inherent optical properties); the downwelling irradiance incident on the sea surface, $E_d(\lambda,0^+)$; and the angular distribution of the light within the ocean. To make things easier, we divide seawater into three components, each one having distinct optical properties of its own. These components are the seawater itself (water and salts), the particle fraction, and the dissolved fraction. Fortunately, $a(\lambda)$ is simply equal to the sum of the absorption coefficients for each component, and, to first order, $b_b(\lambda)$ is equal to the sum of the backscattering coefficients. If we can accurately describe or model each spectrally distinct component of the absorption and backscattering coefficients, then we can determine the magnitude of each one from measurements of $L_w(\lambda)$ and $E_d(0^+,\lambda)$, given some assumptions about the angular distribution of light in the water. The key here is to accurately model the spectral behavior of $a(\lambda)$ for each component. The spectral behavior of $b_b(\lambda)$ is less important.

The R_{rs} model is given by the following general equation, which is adapted from Lee et al. (1994):

$$R_{rs}(\lambda) = \frac{f\ t^2}{Q(\lambda)\ n^2} \frac{b_b(\lambda)}{[a(\lambda) + b_b(\lambda)]} \qquad (1)$$

where f is an empirical factor averaging about 0.29–0.33 (Gordon et al., 1975; Morel and Prieur, 1977; Jerome et al., 1988; Kirk, 1991; Morel and Gentili 1996), t is the transmittance of the air-sea interface, $Q(\lambda)$ is the upwelling irradiance-to-radiance ratio, $E_u(\lambda)/L_u(\lambda)$, and n is the real part of the index of refraction of seawater. By making three approximations, Eq. 1 can be greatly simplified.

1) In general, f is a function of the solar zenith angle, θ_0 (Kirk, 1984; Jerome et al., 1988; Morel and Gentili, 1991). However, Morel and Gentili (1993) have shown that the ratio f/Q is relatively independent of θ_0 for sun and satellite viewing angles expected for the SeaWiFS orbit. They estimate that f/Q = 0.0936, 0.0944, and 0.0929 (standard deviation ± 0.005), for λ = 440, 500, and 565, respectively. Also, Gordon et al. (1988) estimate that f/Q = 0.0949, at least for

$\theta_0 \geq 20°$. Thus, we assume that f/Q is independent of λ and θ_0 for all SeaWiFS wavebands of interest, except perhaps for the band centered at 670 nm.

2) t^2/n^2 is approximately equal to 0.54, and although it can change with sea-state (Austin, 1974), it is relatively independent of wavelength.

3) Many studies have confirmed that $b_b(\lambda)$ is usually much smaller than $a(\lambda)$ and can thus be safely removed from the denominator of Eq. 1 (Morel and Prieur, 1977; references cited in Gordon and Morel, 1983), except for highly turbid waters.

These three approximations lead to a simplified version of Eq. 1,

$$R_{rs}(\lambda) \approx constant \frac{b_b(\lambda)}{a(\lambda)} \tag{2}$$

where the "constant" is unchanging with respect to λ and θ_0. The value of the constant is not relevant to the algorithm since, as will be shown later, the algorithm uses spectral ratios of $R_{rs}(\lambda)$ and the constant term factors out.

In the following sections, both $b_b(\lambda)$ and $a(\lambda)$ will be divided into several separate terms. Each term will be described empirically. The equations are written in a general fashion — i.e., the empirically derived parameters that describe each term are written as variables — and the actual values of the parameters that are used in the algorithm are shown in Table 1.

2.1 Backscattering term

The total backscattering coefficient, $b_b(\lambda)$, can be expanded as

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda) \tag{3}$$

where the subscripts "w" and "p" refer to water and particles, respectively. $b_{bw}(\lambda)$ is constant and well known (Smith and Baker, 1981). $b_{bp}(\lambda)$ is modeled as

$$b_{bp}(\lambda) = X\left[\frac{555}{\lambda}\right]^Y \tag{4}$$

The magnitude of particle backscattering is indicated by X, which is approximately equal to $b_{bp}(555)$, while Y describes the spectral shape of the particle backscattering.

We now need expressions for X and Y. Lee et al. (1994) use a quasi single-scattering form of the R_{rs} model, summarized by the following three equations:

$$R_{rs}(\lambda) = \frac{0.176\ b_b(\lambda)}{Q(\lambda)\ a(\lambda)} \tag{5}$$

$$\frac{b_b(\lambda)}{Q(\lambda)} \approx \frac{b_{bw}(\lambda)}{Q_w(\lambda)} + \frac{b_{bp}(\lambda)}{Q_p(\lambda)} \tag{6}$$

$$\frac{b_{bp}(\lambda)}{Q_p(\lambda)} = X' \left[\frac{400}{\lambda}\right]^{Y'} \tag{7}$$

The main differences here are that b_b/Q is modeled explicitly rather than just b_b (compare Eqs. 3 and 6), and that 400 nm is used rather than 555 nm as the normalizing point in the particle backscattering term (compare Eqs. 4 and 7). Eq. 6 is an approximation derived from single and quasi-single scattering theory (Lee et al., 1994).

They developed a method to determine X' and Y' empirically for a given optical station by model inversion. The method uses measured values of $R_{rs}(\lambda)$ and $a(\lambda)$ at ≈ 200 wavelengths. The best-fit values for X' and Y' are determined using Eqs. 5–7 on a station-by-station basis. Using this method Carder et al. (1996; 1997) determined X' and Y' for a number of optical stations taken from 4 separate cruises to the Gulf of Mexico. We then converted the X' and Y' values to our X and Y via

$$X = X' Q_p \left[\frac{400}{555}\right]^{Y'}$$
$$Y = Y' \tag{8}$$

using a value of 3.55 for Q_p. We then compared these values of X and Y to $R_{rs}(\lambda)$ values measured at the corresponding station providing empirical relationships for both X and Y as a function of $R_{rs}(\lambda)$ (Carder et al., 1996; 1997).

The general expressions for X and Y are

$$X = X_0 + X_1 R_{rs}(555) \tag{9}$$

$$Y = Y_0 + Y_1 \frac{R_{rs}(443)}{R_{rs}(490)} \quad (10)$$

where X_0, X_1, Y_0, and Y_1 are the empirically derived constants shown in Table 1 (Carder et al., 1996; 1997).

Accurate measurements of $a_g(\lambda)$ and accurate removal of reflected skylight from the R_{rs} measurements are critical in determining Y by model inversion. Only data from the GOMEX and COLOR cruises are used here because the $a_g(\lambda)$ values were determined with a long-path (> 0.5 m) spectrophotometer (Peacock et al., 1994).

2.2 Absorption term

The total absorption coefficient can be expanded as

$$a(\lambda) = a_w(\lambda) + a_\phi(\lambda) + a_d(\lambda) + a_g(\lambda) \quad (11)$$

where the subscripts "w", "ϕ," "d," and "g" refer to water, phytoplankton, detritus, and gelbstoff ("g" stands for gelbstoff). $a_w(\lambda)$ is taken from Pope and Fry (1997). Expressions for $a_\phi(\lambda)$, $a_d(\lambda)$, and $a_g(\lambda)$ were developed as (Carder et al., 1996; 1997)

The shape of the $a_\phi(\lambda)$ spectrum for a given water-mass changes due to the pigment-package effect (i.e., the flattening of absorption peaks with increasing intracellular pigment concentration due to self-shading; Morel and Bricaud, 1981) and due to changes in pigment composition. A hyperbolic tangent function was chosen to model this relationship in order to ensure that the value of $a_\phi(\lambda)/a_\phi(675)$ approaches an asymptote at very high or very low values of $a_\phi(675)$ (Carder et al., 1991). Using logarithmic scaling for both axes results in the following model equation for $a_\phi(\lambda)$ as a function of $a_\phi(675)$,

$$a_\phi(\lambda) = a_0(\lambda)\exp\left[a_1(\lambda)\tanh\left[a_2(\lambda)\ln\left(a_\phi(675)/a_3(\lambda)\right)\right]\right] \bullet a_\phi(675) \quad (12)$$

where the parameters $a_0(\lambda)$–$a_3(\lambda)$ are empirically determined for each SeaWiFS wavelength of interest. The measured data and the modeled curves for $a_\phi(\lambda)$ measurements were developed by Carder et al. (1996; 1997) from GOMEX, COLOR, and TN048 cruise data, and the parameters $a_0(\lambda)$–$a_3(\lambda)$ are listed in Table 1.

The method used to determine absorption coefficients for particles and for detritus involves filtering as much as 4 liters of water through a 25 mm diameter, Gelman glass-fiber filter (GFF). This large amount of water is used to concentrate the sample enough for accurate measurements of the pad optical density (OD) to be determined (Shibata, 1958; Mitchell, 1990; Nelson and Robertson, 1993; Moore et al., 1995). In order to estimate absorption coefficients from the OD measurements, an optical path elongation factor, called β, which is dependent upon OD, is employed. Recently however, it has been shown that β varies with the particle size prevalent to a region (Moore et al., 1995). This happens because smaller particles get more

deeply imbedded into the pad, providing a greater absorption cross-section for photons scattered numerous times than for the large particles remaining at the surface of the pad. Carder et al. (1996; 1997) chose a β factor appropriate for small, subtropical particles by averaging two published β factors, one developed for detritus (Nelson and Robertson, 1993) and one for Synechococcus (Moore et al., 1995). Their β factor was

$$\beta = 1.0 + 0.6\,OD^{-0.5} \tag{13}$$

$a_d(\lambda)$ and $a_g(\lambda)$ can both be fit to a curve of the form $a_x(\lambda) = a_x(400)\exp[-S_x(\lambda-400)]$ where the subscript "x" refers to either "d" or "g" (Bricaud et al., 1981; Roesler et al., 1989; Carder et al., 1991). Due to this similarity in spectral shape, the $a_d(\lambda)$ term can be eliminated, allowing both detrital and gelbstoff absorption to be represented by $a_g(\lambda)$. The combined gelbstoff and detritus absorption term is thus written

$$a_g(\lambda) = a_g(400)\exp^{-S(\lambda-400)} \tag{14}$$

where S is empirically determined.

Many researchers have reported that $S_d = 0.011$ nm^{-1}, on average (Roesler et al., 1989). For the GOMEX and COLOR cruises, an average value of 0.017 nm^{-1} was measured for S_g. Values reported by F. Hoge and R. Bidigare (personal communication) for the Sargasso Sea were somewhat higher as are those found near swampy regions of the Gulf of Mexico. Also, a higher value is needed to compensate for gelbstoff fluorescence, which was not included in the model. The algorithm performance was optimized by varying S_g, with the value 0.0225 nm^{-1} providing the smallest residual error compared to field measurements.

2.3 Inverting the semi-analytical model

Using spectral ratios of R_{rs} eliminates the "constant" term in Eq. 2, since it is largely independent of wavelength. In principle, two spectral ratio equations can be used to solve for the two remaining unknowns, $a_\phi(675)$ and $a_g(400)$. Based on the shape of the absorption curve for phytoplankton versus those for gelbstoff and detritus, equations using spectral ratios of 412:443 and 443:555 for $R_{rs}(\lambda)$ provide a good separation of the two absorption contributions. The two equations are

$$\frac{R_{rs}(412)}{R_{rs}(443)} = \frac{b_b(412)}{b_b(443)}\,\frac{a(443)}{a(412)}$$

$$\frac{R_{rs}(443)}{R_{rs}(555)} = \frac{b_b(443)}{b_b(555)}\,\frac{a(555)}{a(443)} \tag{15}$$

The right-hand side of each equation is a function of $a_\phi(675)$, $a_g(400)$, $R_{rs}(443)$, $R_{rs}(490)$ and $R_{rs}(555)$. Since the R_{rs} values are provided on input, we now have two equations in two unknowns. The equations can be solved algebraically to provide values for $a_\phi(675)$ and $a_g(400)$. The computational method of solving these equations is described in Section 2.7.

For waters with high concentrations of gelbstoff and chlorophyll, $R_{rs}(412)$ and $R_{rs}(443)$ values are small, and the semi-analytical algorithm cannot perform properly. It is thus designed to return values only when modeled $a_\phi(675)$ is less than 0.06 m^{-1}, which is equivalent to chl *a* of about 3–4 mg m^{-3}. Otherwise, an empirical algorithm for chl *a* is used, which is described in Section 2.5. There is presently no output for $a_\phi(675)$ and $a_g(400)$ when the empirical chl *a* algorithm is employed, but empirical algorithms for these variables are under development.

2.4 Pigment algorithm for semi-analytical case

When the semi-analytical algorithm returns a value for $a_\phi(675)$, chl *a* is determined via a direct relationship to this value. This step requires precise knowledge of the chlorophyll-specific absorption coefficient for phytoplankton at 675 nm, $a_\phi^*(675)$. Quadratic regression of log(chl *a*) vs. log($a_\phi(675)$) yields an equation of the form

$$[chl\ a] = p_0 \left[a_\phi(675) \right]^{p_1} \tag{16}$$

For a global data set of 95 points, an $r^2 = 0.97$ coefficient of regression on the log-transformed values was found (Carder et al., 1996; 1997), and the coefficients are displayed in Table 1. Note that these data were determined in laboratories aboard ships and in no way were reliant upon field measurements of R_{rs}.

2.5 Pigment algorithm for the default case

When the semi-analytical algorithm does not return a value for $a_\phi(675)$, usually due to low Rrs values in high-pigment waters, we provide an empirical, two-wavelength algorithm for chl *a* to use by default. Aiken et al. (1995) found that the $L_w(490)/L_w(555)$ ratio is best for empirical chl *a* determination due to its low response to gelbstoff and high saturation levels. We use an equation of the form

$$[chl\ a]_{emp} = 10^{c_0 + c_1 R + c_2 R^2 + c_3 R^3} \tag{17}$$

where

$$R = \log \left[\frac{R_{rs}(490)}{R_{rs}(555)} \right] \tag{18}$$

chl a_{emp} is called the "empirically-derived" or "default" chl *a* concentration, and c_0, c_1, c_2, and c_3 are empirically derived constants (see Table 1).

A data set consisting of subtropical, temperate summer, and high-latitude summer stations was created from the Carder subtropical and high-scattering data sets, NABE, and the EqPac above- and below-water data sets (n = 378; see Section 3.2). It includes both open-ocean and riverine-influenced stations. Third-order regression of log(chl *a*) against $\log(r_{35})$ for measured chl *a* and $R_{rs}(\lambda)$ in this data set resulted in values of $c_0 = 0.2818$, $c_1 = -2.783$, $c_2 = 1.863$, and $c_3 = -2.387$. The root-mean-square (RMS) error of 0.327 for three orders of magnitude variation in chl a, including Case 2 river-plume data near the Mississippi.

2.6 Weighted pigment algorithm

Another consideration is that there should be a smooth transition in chl *a* values when the algorithm switches from the semi-analytical to the empirical method. This is achieved by using a weighted average of the chl *a* values returned by the two algorithms when near the transition border. When the semi-analytical algorithm returns an $a_\phi(675)$ value between 0.03 and 0.06 m^{-1}, chl *a* is calculated as

$$[chl\ a] = w\,[chl\ a]_{sa} + (1 - w)\,[chl\ a]_{emp} \tag{19}$$

where chl a_{sa} is the semi-analytically derived value and chl a_{emp} is the empirically derived value, and the weighting factor is $w = [0.06 - a_\phi(675)]/0.03$. For lower absorption data, the semi-analytical algorithm is used, while for higher absorption data the default algorithm is used.

2.7 Numerical computation

$a_\phi(675)$ and $a_g(400)$ are determined from Eqs. 15 by inverting one of the equations to isolate $a_g(400)$, substituting into the other equation, and moving all terms to one side, yielding a function that depends only on $a_\phi(675)$ (given values for R_{rs} and Table 1 for the algorithm parameters). The value of $a_\phi(675)$ at which the function crosses zero is the solution we seek. This solution is determined computationally via the bisection method. A 33-element array of $a_\phi(675)$ values, scaled logarithmically from 0.0001 to 0.06 m^{-1} is created, and the function is evaluated at the two extrema. If the function changes sign between the two outermost values, a solution exists on the $a_\phi(675)$ interval. The function is then evaluated at the mid-point of the array, and the half in which the function changes sign becomes the new search interval. In this manner, the solution interval, which will be two adjacent points on the $a_\phi(675)$ array, is determined in 5 iterations. Linear interpolation across the interval yields the semi-analytical $a_\phi(675)$ value, and $a_g(400)$ is determined via either one of the R_{rs}-ratio equations using the modeled value of $a_\phi(675)$. If the function does not change sign across the two outermost values, a switch is made to the empirical, two-wavelength default algorithm.

When compared to an older lookup-table-based method (Carder et al., 1991), the bisection method gave identical solutions to within 5 significant digits for $a_\phi(675)$ and $a_g(400)$, and the

code ran in 75% of the time that the lookup-table-based version of the code took.

The algorithm code is written in C. The program file contains about 300 lines of code and comments. It was developed and tested on a DEC Alpha machine which uses the DEC OSF/1 C Compiler. All of the algorithm parameters listed in Table 1 are read in from a file, so different parameter tables can easily be constructed for different applications. The code is available via anonymous ftp at:

ftp montypython.marine.usf.edu
/pub/swf_alg/

3.0 Algorithm Evaluation

3.1 Statistical criteria

To evaluate algorithm performance we generated the same statistics described in the Algorithm Evaluation chapter (O'Reilly and Maritorena, this volume) using O'Reilly's stats2.pro IDL program. These statistics are determined on the log-transformed variables, and the slope and intercept are from Type II RMA regression. The RMS statistic they describe will be referred to here as RMS1. We also generated values for r^2 and root-mean-square error on the non-log-transformed (linear) data. Our RMS statistic will be referred to as RMS2 and is described by

$$RMS2 = \sqrt{\frac{\sum_{i=1}^{n}\left[\frac{X_{mod,i} - X_{obs,i}}{X_{obs,i}}\right]^2}{n-2}} \qquad (20)$$

where $x_{mod,i}$ is the modeled value of the *i*th element, $x_{obs,i}$ is the observed (or *in situ* or measured) value of the *i*th element, and n is the number of elements.

We used two graphical means of evaluating algorithm performance: scatter plots of modeled versus observed values and quantile-quantile plots (see Algorithm Evaluation chapter, O'Reilly and Maritorena, this volume).

3.2 Tests with USF data (Carder data set)

We initially tested our algorithm with our own data set, called the Carder data set in the Evaluation Data Set chapter (Maritorena et al., this volume). However, the data set we present here differs from the Carder data used in the global evaluation data set in two ways. First, we include observed values of $a_\phi(675)$, and $a_g(400)$ wherever possible to go along with observed $R_{rs}(\lambda)$ and chl *a*. Second, 17 points of high-chlorophyll, high-scattering stations, mostly from the Mississippi River Plume region, are included. The data sources are listed in Table 2.

$R_{rs}(412)$, $R_{rs}(443)$, $R_{rs}(490)$, $R_{rs}(510)$, and $R_{rs}(555)$ were derived from hyperspectral $R_{rs}(\lambda)$ measurements collected just above the sea surface (for measurement protocols, see Lee et al., 1996) by weighting to simulate the SeaWiFS band responses (Barnes et al., 1994). All chl *a* values were determined fluorometrically (Holm-Hansen and Riemann, 1978). $a_\phi(675)$ was determined as described in section 2.2. $a_g(400)$ was determined by measuring 0.2 uM filtered

seawater in a spectrophotometer.

Algorithm performance was evaluated on both the n=87 subset of stations which correspond to the data available in the global evaluation data set and on the full n=104 set. The algorithm parameters used are shown in Table 1. For the n=87 subset, all but one of the points were determined via the semi-analytical portion of the algorithm. chl *a*, $a_\phi(675)$, and $a_g(400)$ were predicted with RMS1 errors of 0.122, 0.131, and 0.252, respectively, and RMS2 errors of 0.289, 0.302, and 0.405, respectively. All of the statistics for this and for all evaluations are shown in Table 3. The results are shown as scatter (Figure 1a) and quantile (Figure 1b) plots. The crosses on the plots are the points determined with the semi-analytical blended algorithm, and all but 4 of these points are from the n=87 data set. The chl *a* and $a_\phi(675)$ data appear to be quite evenly clustered about the one-to-one line on both scatter and quantile plots with no tails at either end. The $a_g(400)$ points are predominantly below the one-to-one line and show a very low bias. There are only 26 points in this plot because measured values of $a_g(400)$ are infrequently available for comparison.

4 of the 17 additional high-chlorophyll points are determined by either the semi-analytical or blended portion of the algorithm. chl *a* values for the other 13 points are thus determined by the default empirical algorithm. However, since the default portion of the algorithm does not yet return values for $a_\phi(675)$ and $a_g(400)$, these high-chlorophyll points add little to the tests for those variables. The RMS1 and RMS2 errors for chl *a* for this composite data set were 0.132 and 0.300, respectively. The results are also shown in Figure 1a and 1b (diamonds). The additional high-chlorophyll points extend nicely along the one-to-one line on both the scatter and quantile plots.

3.3 Partitioning the global evaluation data set

A large (n=919) global evaluation data set consisting of measured R_{rs} at the SeaWiFS wavelengths and pigment measurements was collected by the SeaWiFS Project for the SeaBAM exercise (see the Evaluation Data Set chapter, Maritorena et al., this volume). These data came from various researchers around the U.S. and Europe. There are no observed (*in situ*) values of $a_\phi(675)$ or $a_g(400)$ provided in this data set. In addition to these data, we received 36 data points from the equatorial Pacific, which consisted of R_{rs} measurements made above the surface (EqPac, courtesy of C. Davis).

Since many different locations and sensors were involved with the data collection, and as many as four separate sensor channels must be well calibrated to provide accurate spectral ratios of R_{rs}, an attempt was made to select an initial core set of data consistent with Case 1 waters and with each other. Also, an attempt was made to partition the data sets into ones for regions where little pigment packaging is to be expected (e.g., high-light, non-upwelling locations in warm, tropical and subtropical waters), and one where more packaging might be expected (e.g., western boundary upwelling, non-summer, high latitude, etc.). To help in this task, the data were examined with the help of two numerical filters.

The first numerical filter developed was to compare the data sets with the CZCS chlorophyll pigment algorithm ($C = 1.14\ [r_{25}]^{-1.71}$, $r_{25} = R_{rs}(443)/R_{rs}(555)$) to check for consistency with this classical determinant of Case 1 waters. Figures 2c, 3c, 4c, and 5c show scatter-plots of observed chl *a* versus r_{25} for different groups of data with the CZCS algorithm illustrated by the dotted line. The warm-water, subtropical and tropical data sets (Figure 2c) were mostly

consistent with the CZCS algorithm for pigment values less than about 1 mg m^{-3}. When data from eastern boundary and upwelling locations (Figure 3c) were applied to the CZCS algorithm, however, they provided chlorophyll *a* values typically 50% to 90% lower than measured, suggesting that perhaps regional algorithms are needed to obtain best results for such waters. This helped separate the data into two water types which we will call "unpackaged" pigment waters and "packaged" pigment waters. Since this "packaging" filter is not applicable using only spacecraft-derived data, a second type of packaging filter was sought.

A second numerical filter was developed using the ratios r_{12} (= $R_{rs}(412)/R_{rs}(443)$) and r_{25} (Figures 2d, 3d). For waters with unpackaged pigments, the line $r_{12} = 0.95\ [r_{25}]^{0.16}$ was used to separate high-gelbstoff data points (those below the line in Figures 2d, 3d) from the Case 1 data. The gelbstoff-rich Case 2 data shown in Figure 2c and 2d had $a_g(400)$ values typically in excess of the relationship 0.12 $[chl\ a]^{0.7}$ (Figure 2e), where 0.12 has the units m^2 (mg chl)$^{-1}$. Since this data set contained both gelbstoff and chlorophyll a measurements and had been acquired by making R_{rs} measurements against a reflectance standard, minimizing calibration uncertainties (see Carder and Steward 1985), it was used to evaluate tropical and subtropical waters for gelbstoff-rich conditions and to flag data sets with sensor-calibration uncertainties.

To identify waters with more packaged pigments using remotely sensed data, Case 1 data from a traditional upwelling region (e.g., CalCOFI) were examined. These data are included in Figure 3c for comparison to the unpackaged data of Figure 2c. Since pigment packaging reduces the absorption for a given concentration of pigments far more at 443 nm than at 555 nm, and somewhat more at 443 nm than at 412 nm, packaging significantly reduces r_{25} while increasing the r_{12} ratio somewhat. This, then places packaged data points below the $r_{12} = 0.95\ [r_{25}\ ]^{0.16}$ line even without excessive gelbstoff concentrations (see Figures 3d and 3e), at least for points with r_{25} values in excess of a value of about 3.0.

For the numerical filter approach to work consistently at separating even more heavily packaged data sets from unpackaged ones, more data sets need to be evaluated. Measurements of particulate and detrital absorption would be useful. There are regions with pigments packaged even more extensively than those represented in this study (Section 4.2), and algorithm parameterization for those environmental situations is being pursued. A nascent outline of an approach to vary algorithm parameters using measurements from space is suggested by our work with the r_{12} vs. r_{25} numerical filter. In future research, this approach will be expanded to other band ratios and data sets, and supplemented with a temperature-anomaly approach based upon estimating nutrient-replete conditions (Kamykowski 1987). This should improve our facility and accuracy in modulating the pigment-absorption parameters for future ocean-color algorithms.

3.4 Algorithm evaluation with the "unpackaged" data set

Those data sets generally found consistent with the CZCS algorithm line as well as occurring above the line $r_{12} = 0.95\ [r_{25}]^{0.16}$ for points where $r_{25} > 3.0$, were classified as "unpackaged", in reference to the pigment effects on the optics prevalent at those locations at the time of data collection. There are 378 data points in this ensemble "unpackaged" data set: 104 USF data points and 37 EqPac equatorial Pacific points, all measured above-surface using the Lee et al. (1996) protocols, and 126 EqPac points and 112 North Atlantic Bloom Experiment (NABE) points, all measured below-surface using the Mueller and Austin (1995) protocols. Of these points, 339 (90%) were processed by the semi-analytical portion of the algorithm yielding RMS1

and RMS2 errors of 0.103 and 0.240, respectively. The scatter (Figure 2a) and quantile (Figure 2b) plots overlay the one-to-one line at the ends as well as in the middle. For the log-transformed variables, the Type II RMA slope and intercept were 1.003 and 0.001, the bias was 0.000, and r^2 was 0.943. When all 378 data points were considered using the semi-analytical algorithm plus the blended and empirical algorithms RMS1 and RMS2 errors were 0.107 and 0.251, respectively. The Type II RMA slope and intercept were 1.001 and 0.002, the bias was 0.001, and r^2 was 0.962. Table 3 has a a complete summary of these statistics.

3.5 Algorithm evaluation with the "packaged" data set

Three data sets within the global evaluation set were numerically diagnosed as coming from waters where the pigments were more "packaged" or at least different from those of the unpackaged, largely tropical and subtropical data sets. Simulations of the optical properties for these regions required some minor alterations of the phytoplankton absorption characteristics, based upon decreased specific absorption observed in the CalCOFI study area. The new parameters, shown in Table 4, are used to define a slightly different, "packaged" algorithm. The forms of the algorithm equations are the same except for the chl *a*–$a_\phi(675)$ relationship, which is

$$chl\ a = 10^{\left[p_0 + p_1 \log_{10}(a_\phi(675)) + p_2 [\log_{10}(a_\phi(675))]^2\right]} \quad (21)$$

There are 355 points in this ensemble "packaged" data set, consisting of CalCOFI (n=303), AMT (n=42), and North Sea (n=10). 341 (96%) points from this ensemble "packaged" data set passed the semi-analytical portion of the new algorithm, yielding RMS1 and RMS2 errors for chl *a* retrieval of 0.118 and 0.289, respectively. The Type II RMA slope and intercept were 1.003 and 0.000, the bias was 0.002, and r^2 was 0.931. The scatter plot (Figure 3a) overlays the one-to-one line and the quantile plot (Figure 3b) is linear, overlies the line, but has a slight discountinuity near a chlorophyll value of 3. With all 355 data points the statistics are about the same (Table 3).

3.6 Algorithm evaluation with the combined data set

Combining the results for the unpackaged and packaged data sets provides an estimate of how the algorithm might perform if the appropriate algorithm parameters (Table 4) can be smoothly (or unsmoothly) varied from unpackaged to packaged regimes. For the combined data set of 733 points, using the appropriate parameters for each set, 675 (92%) of the points passed the semi-analytical portion of the algorithm, yielding RMS1 and RMS2 errors in algorithm-derived chl *a* of 0.112 and 0.280, respectively. The Type II RMA slope and intercept were 1.009 and 0.004, the bias was –0.001, and r^2 was 0.936. Statistics for the entire n=733 set were similar. The scatter and quantile plots overlaid the one-to-one line closely (Figures 4a and 4b).

3.7 Algorithm evaluation with a modified global data set

We developed a modified global data set which differs from the evaluation data set (Maritorena et al., this volume) in two ways. First, the Cota and U. Maryland points were excluded pending further study of hyper-packaging possibilities (Cota data) and possible suspended sediments (U. Md. data). Second, the 17 high chl *a* points in the Carder data set and the above-water data from EqpPac were included. This data set has 955 data points. We then developed a set of compromise parameters for our algorithm, shown in Table 4, for use at times and places where "packaging" is unknown. For this data set and these "average" parameters, 870 points (91%) of the points passed the semi-analytical portion of the algorithm, yielding RMS1 and RMS2 errors in algorithm-derived chl *a* of 0.173 and 0.441, respectively. The Type II RMA slope and intercept were 0.999 and 0.003, the bias was 0.004, and r^2 was 0863. Statistics for the entire n=955 set were similar except r^2 was higher (0.915). The scatter plot (Figure 5a) looks evenly clustered about the one-to-one line and the quantile plot (Figure 5b), though wiggly, overlays the one-to-one line for the most part.

4.0 Discussion

The biggest limitation is the lack of bio-optical field data from around the globe that are complete with ancillary particle and gelbstoff absorption spectra. These data are needed in order to assess the spatial and temporal variation in the key algorithm parameters X, Y, S, $a_g(400)$, and most importantly, $a_0(\lambda)$ and $a_1(\lambda)$. In order to derive chl *a*, it is vitally important to be able to predict how $a_\phi^*(\lambda)$ will vary. Thus, we must study the effect of light history, which is related to season, cloudiness, latitude, and nutrient history, which is influenced by mixed-layer depth, upwelling, river plumes, and offshore/onshore proximity.

4.1 High b_b pixels

Since the R_{rs} model does not specifically account for absorption and backscattering from suspended sediments or coccolithophores or for reflection from the bottom, a method is needed to determine which pixels are influenced by any of these. Such waters will be referred to as "high-b_b Case 2" waters, as opposed to high-gelbstoff Case 2 waters, which the model explicitly accounts for. Although not yet implemented, a possible means of identifying high-b_b Case 2 stations is to examine the $R_{rs}(670)$:$R_{rs}(555)$ ratio. Retaining $b_b(\lambda)$ in the denominator of Eq. 1 is required, and the site-specific behavior of sediment absorption characteristics must be known.

4.2 a_ϕ in other environments

We have learned from trends in the data observed so far that the semi-analytical algorithm performs as well with temperate summer data (TT010 north of 45^0 and MLML 2 north of 50^0) as it does with subtropical data for all seasons. How, then, might temperate data from other seasons and/or data from upwelling and high-latitude areas differ from the temperate summer, non-upwelling data?

To address this question we compare $a_\phi(\lambda)$ data from MLML 1 (May, 50^0–60^0 N), MLML 2 (August, 50^0–60^0 N), TT010 (July, north of 45^0), Monterey Bay (fall, upwelling region), and 2 coastal upwelling stations from the Arabian Sea. Although these two data points were collected from a subtropical summer environment, the water was about 4 ^{0}C cooler than offshore waters,

indicating a lower-light, nutrient-rich, upwelling source, conditioning the water for highly packaged, fast-growing species such as diatoms. This is manifest in Figure 6, where these data fall among the more packaged points. Here, the ratios of the blue peak to the red peak, $a_\phi(443):a_\phi(675)$, are plotted as a function of the height of the red peak itself, $a_\phi(675)$, which can be thought of as an indicator of pigment concentration. The subtropical algorithm values (solid line) and trend lines for the high and low out-lying points for the entire data set (dashed lines) are also shown. The dotted line represents a median trend for the entire data set, and it approximates the mean line for two years of data from the Southern California Bight (SCB) (B. G. Mitchell, personal communication). The SCB data also ranged widely between the top and bottom dashed curves.

The first thing to note in Figure 6 is how well the subtropical line is followed by the high-latitude summer data. In fact, two of the summer TT010 points along the Washington coast fall among the highest of the subtropical data. The phaeocystis-rich, spring-bloom, MLML 1 data, however, represent data with the lowest specific absorption coefficients of the entire study. Similarly, upwelling data from the Arabian Sea and Monterey Bay fall below the median line for the data set. These data trends suggest that there is less packaging in summer temperate data than at other times. Maximal packaging appears associated with high-latitude, low-light, spring bloom stations (MLML 1) and with upwelling sites. The data also suggest that a single global algorithm will lack the accuracy needed to address data sets that include subtropical, high-latitude, and upwelling areas. For the non-subtropical areas, some of the parameters in Table 1 need to be functions of region and season.

In addition to the numerical filter approach mentioned above, one trend in the data that will be exploited to condition a smooth transition between a subtropical algorithm and upwelling sites or between temperate versions for different seasons is that sites with heavily packaged pigments have relatively low stability in the upper water-column. For several stations, we found that the temperature difference between the sea surface and the top of the permanent thermocline was minimal when packaging was highest. The MLML2 temperatures were 4–5 ^{0}C warmer than for MLML 1 along the same transect line, while both share essentially the same permanent thermocline. Also, the Arabian Sea upwelling stations had water 3–4 ^{0}C cooler than found offshore, while again sharing a common permanent thermocline.

Low-temperature anomalies have been used extensively to predict availability of major nutrients. Kamykowski and Zentara (1986) and Kamykowski (1987) used the anomalies relative to historical monthly mean temperatures for a given location to predict nutrient availability, while Gong et al. (1995) used anomalies relative to the temperature at the top of the permanent thermocline to predict nitrogen levels. Significant injections of nitrogen into surface waters are typically followed by blooms of larger-celled phytoplankton such as diatoms or phaeocystis, resulting in high packaging. It is consistent with these trends, then, to explore ways of conditioning changes in the algorithm parameters for $a_\phi(\lambda)$ based on sea-surface temperature measurements from satellites.

While we can observe subtle hints of a strategy to develop a truly global algorithm, it would be premature to presently attempt to seamlessly adjust the subtropical algorithm to address all high-latitude, upwelling or other less stratified environments. Much more data are needed before attempting such a task. Researchers can, however, develop $a_\phi(\lambda)/a_\phi(675)$ vs. $a_\phi(675)$ and chl *a* vs. $a_\phi(675)$ relationships specific to their favorite study region, noting seasonal changes. These relationships can be used to modify the subtropical algorithm to improve its performance

on a regional basis. A method to transition between regions with packaging differences similar to those expressed in Tables 1 and 4 appears feasible now, however, using numerical filters and space-based data.

5.0 Conclusions

A semi-analytical algorithm was tested with a total of 733 points of either unpackaged- or packaged-pigment data, with corresponding algorithm parameters for each data type. The "unpackaged" type consisted of data sets that were generally consistent with the Case 1 CZCS algorithm and other well calibrated data sets. The "packaged" type consisted of data sets apparently containing somewhat more packaged pigments, requiring modification of the absorption parameters of the model consistent with the CalCOFI study area. This resulted in two equally divided data sets. A more thourough scrutiny of these and other data sets using a semi-analytical model requires improved knowledge of the phytoplankton and gelbstoff of the specific environment studied. Since the semi-analytical algorithm is dependent upon 4 spectral channels including the 412 nm channel, while most other algorithms are not, a means of testing data sets for consistency was sought. A numerical filter was developed to classify data sets into the above classes. The filter uses reflectance ratios, which can be determined from space. The sensitivity of such numerical filters to measurement resulting from atmospheric correction and sensor noise errors requires further study.

The semi-analytical algorithm performed superbly on each of the data sets after classification, resulting in RMS1 errors of 0.107 and 0.121, respectively, for the unpackaged and packaged data-set classes, with little bias and slopes near 1.0. In combination, the RMS1 performance was 0.114.

While these numbers appear rather sterling, one must bear in mind what mis-classification does to the results. Using an average or compromise parameterization on the modified global data set yielded an RMS1 error of 0.171, while using the unpackaged parameterization on the global evaluation data set (Maritorena et al., this volume) yielded an RMS1 error of 0.284 (O'Reilly and Maritorena, this volume). So, without classification, the algorithm performs better globally using the average parameters than it does using the unpackaged parameters.

Finally, the effects of even more extreme pigment packaging (Figure 6) must be examined in order to improve algorithm performance at high latitudes. Note, however, that the North SEa and Mississippi River plume studies contributed data to the packaged and unpackaged classess, respectively, with little effect on algorithm performance. This suggests that gelbstoff-rich Case 2 waters do not seriously degrade performance of the semi-analytical algorithm.

6.0 References

Aiken, J., G. F. Moore, C. C. Trees, S. B. Hooker, and D. K. Clark, The SeaWiFS CZCS-type pigment algorithm, in *SeaWiFS Technical Report Series*, S. B. Hooker and E. R. Firestone, eds., Vol. 29, 1995.

Austin, R. W., Inherent spectral radiance signals of the ocean surface, in *Ocean Color Analysis, SIO ref. 74–10*, pp. 2.1–2.20, Scripps Inst. of Oceanogr., La Jolla, Calif, 1974.

Baker, K. S., and R. C. Smith, Bio-optical classification and model of natural waters 2, *Limnol. Oceanogr.*, 27, 500–509, 1982.

Barnes, R. A., W. L. Barnes, W. E. Esaias, and C. R. McClain, Prelaunch acceptance report for the SeaWiFS radiometer, in *SeaWiFS Technical Report Series*, S. B. Hooker, E. R. Firestone, and J. G. Acker, eds., Vol. 22, 1994.

Bricaud, A., A. Morel, and L. Prieur, Absorption by dissolved organic matter in the sea (yellow substance) in the UV and visible domains, *Limnol. Oceanogr.*, 26, 43–53, 1981.

Carder, K. L., and R. G. Steward, A remote-sensing reflectance model of a red tide dinoflagellate off West Florida, *Limnol. Oceanogr.*, 30, 286–298, 1985.

Carder, K. L., R. G. Steward, G. R. Harvey, and P. B. Ortner, Marine humic and fulvic acids: Their effects on remote sensing of ocean chlorophyll, *Limnol. Oceanogr.*, 34, 68–81, 1989.

Carder, K. L., S. K. Hawes, K. A. Baker, R. C. Smith, R. G. Steward, and B. G. Mitchell, Reflectance model for quantifying chlorophyll *a* in the presence of productivity degradation products, *J. Geophys. Res.*, 96(C11), 20,599–20,611, 1991.

Carder, K. L., S. K. Hawes, and Z. Lee, MODIS Ocean Science Team ATBD, Case 2 Chlorophyll *a*, 1996.

Carder, K. L., S. K. Hawes, and Z. Lee, SeaWiFS algorithm for chlorophyll *a* and colored dissolved organic matter in subtropical environments, submitted to *JGR*, 1997.

Gong, G.-C., K.-K. Liu, and S.-U. Pai, Prediction of nitrate concentration from two end member mixing in the southern East China Sea, *Cont. Shelf Res.*, 15(7), 827–842, 1995

Gordon, H. R., O. B. Brown, and M. M. Jacobs, Computed relationships between the inherent and apparent optical properties of a flat homogeneous ocean, *Appl. Opt.*, 14, 417–427, 1975.

Gordon, H. R., O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark, A semi-analytic model of ocean color, *J. Geophys. Res.*, 93, 10,909–10,924, 1988.

Gordon, H. R., D. K. Clark, J. W. Brown, O. B. Brown, R. H. Evans, and W. W. Broenkow, Phytoplankton pigment concentrations in the Middle Atlantic Bight: Comparison of ship determinations and CZCS estimates, *Appl. Opt.*, 22, 20–36, 1983.

Gordon, H. R., and A. Y. Morel, *Remote assessment of ocean color for interpretation of satellite visible imagery: A review*, Springer, 1983.

Holm-Hansen, O., and B. Riemann, Chlorophyll *a* determination: Improvements in methodology, *Oikos*, 30, 438–437, 1978.

Jerome, J. H., R. P. Bukata, and J. E. Burton, Utilizing the components of vector irradiance to estimate the scalar irradiance in natural waters, *Appl. Opt.*, 27(19), 4012–4018, 1988.

Kamykowski, D., A preliminary biophysical model of the relationship between temperature and plant nutrients in the upper ocean, *Deep Sea Res.*, 34(7), 1067–1079, 1987.

Kamykowski, D., and S.-J. Zentara, Predicting plant nutrient concentrations from temperature and sigma-*t* in the upper kilometer of the world ocean, *Deep Sea Res.*, 33(1), 89–105, 1986.

Kirk, J. T. O., Dependence of relationship between inherent and apparent optical properties of water on solar altitude, *Limnol. Oceanogr.*, 29, 350–356, 1984.

Kirk, J. T. O., Volume scattering function, average cosines, and the underwater light field, *Limnol. Oceanogr.*, 36(3), 455–467, 1991.

Lee, Z., K. L. Carder, S. K. Hawes, R. G. Steward, T. G. Peacock, and C. O. Davis, A model for interpretation of hyperspectral remote-sensing reflectance, *Appl. Opt.*, 33, 5721–5732, 1994.

Lee, Z., K. L. Carder, R. G. Steward, T. G. Peacock, C. O. Davis, and J. L. Mueller, Remote-sensing reflectance and inherent optical properties of oceanic waters derived from above-water measurements, SPIE vol. 2963, 160–166, 1996.

Mitchell, B. G., Algorithms for determining the absorption coefficients for aquatic particulates using the Quantitative Filter Technique, in *Ocean Optics X*, Proc. SPIE, 1302, 137–148, 1990.

Mitchell, B. G., and O. Holm-Hansen, Bio-optical properties of Antarctic waters: differentiation from temperate ocean models, *Deep Sea Res.*, 38(8–9), 1009–1028, 1991.

Moore, L. R., R. Goericke, S. W. Chisolm, Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties, *Mar. Ecol. Prog. Ser.*, 116, 259–275, 1995.

Morel, A. Y., Optical modeling of the upper ocean in relation to its biogenous matter content (Case 1 waters), *J. Geophys. Res.*, 93, 10,749–10,768, 1988.

Morel, A. Y, and A. Bricaud, Theoretical results concerning light absorption in a discrete medium and application to the specific absorption of phytoplankton, *Deep Sea Res.*, 28, 1357–1393, 1981.

Morel, A. Y., and B. Gentili, Diffuse reflectance of oceanic waters: its dependence on Sun angle as influenced by the molecular scattering contribution, *Appl. Opt.*, 30(30), 4427–4438, 1991.

Morel, A. Y., and B. Gentili, Diffuse reflectance of oceanic waters II. Bi-directional aspects, *Appl. Opt.*, 32(33), 6864–6879, 1993.

Morel, A. Y., and B. Gentili, Diffuse reflectance of oceanic waters. III. Implication of bi-directionality for the remote sensing problem, *Appl. Opt.*, 35(24), 4850–4862, 1996.

Morel, A. Y., and L. Prieur, Analysis of variations in ocean color, *Limnol. Oceanogr.*, 22, 709–722, 1977.

Mueller, J. L., and R. W. Austin, Ocean optics protocols for SeaWiFS validation, revision 1, *NASA Tech. Memo. 104566*, vol. 25.

Nelson, J. R., and C. Y. Robertson, Detrital spectral absorption: Laboratory studies of visible light effects on phytodetritus absorption, bacterial spectral signal, and comparison to field measurements, *J. Marine Res.*, 51, 181–207, 1993.

Peacock, T. P., K. L. Carder, and R. G. Steward, Components of spectral attenuation for an offshore jet in the Coastal Transition Zone, *Eos Trans. AGU*, 69, 1125, 1988.

Peacock, T. G., K. L. Carder, P. G. Coble, Z. P. Lee, and S. K. Hawes, Long-path spectrophotometer for measuring gelbstoff absorption in clear waters, *EOS, Trans., AGU*, 75(3), pp. 22, 1994/supplement.

Pope, R. M., and E. S. Fry, Absorption spectrum (380–700 nm) of pure water: II. Integrating cavity measurements, submitted to *Appl. Opt.* March 1997.

Roesler, C. S., M. J. Perry, and K. L. Carder, Modeling *in situ* phytoplankton absorption from total absorption spectra in productive inland marine waters, *Limnol. Oceanogr.*, 34, 1510–1523, 1989.

Shibata, K., Spectrophotometry of intact biological materials. Absolute and relative measurements of their transmission, reflection and absorption spectra, *J. Biochem.*, 45, 599–623, 1958.

Smith, R. C., and K. S. Baker, Optical properties of the clearest natural waters (200–800 nm), *Appl. Opt.*, 20, 177–184, 1981.

Smith, R. C., and K. S. Baker, Oceanic chlorophyll concentrations as determined by satellite (Nimbus-7 Coastal Zone Color Scanner), *Mar. Biol.*, 66, 269–279, 1982.

Table 1. Parameters for the Case 2 chlorophyll algorithm; see text for definitions.

wavelength dependent parameters					
λ	412	443	490	510	555
b_{bw} (m^{-1})	0.003341	0.002406	0.001563	0.001313	0.000929
a_w (m^{-1})	0.00480	0.00742	0.01632	0.03181	0.03181
a_0	2.20	3.59	2.27	1.40	0.42
a_1	0.75	0.80	0.59	0.35	–0.22
a_2	–0.5	–0.5	–0.5	–0.5	–0.5
a_3	0.010	0.010	0.010	0.010	0.010

wavelength independent parameters					
X_0	–0.00182	S	0.0225	c_0	0.2818
X_1	2.058	p_0	56.8	c_1	–2.783
Y_0	–1.13	p_1	1.03	c_2	1.863
Y_1	2.57			c_3	–2.387

Table 2. List of cruises with optical and bio-optical data collected by the University of South Florida (Carder data set). Numbers in parenthesis in the far left column indicate the number of stations included in the global evaluation data set.

cruise	start date	end date	region	# stations
MLML 2	13 Aug 91	29 Aug 91	North Atlantic, 42°N–60°N	7 (3)
TT010	20 Jul 92	02 Aug 92	North Pacific, 24°N–48°N	10 (10)
GOMEX	10 Apr 93	19 Apr 93	Northern Gulf of Mexico	21 (17)
COLOR	31 May 93	09 Jun 93	Northern Gulf of Mexico	13 (4)
TN042	29 Nov 94	18 Dec 94	Arabian Sea	12 (12)
TN048	21 Jun 95	13 Jul 95	Arabian Sea	41 (41)

total = 104 (87)

Table 3. Summary of regression statistics for each data set tested. The unpackaged data consists of the Carder, EqPac above-surface, EqPac below-surface, and NABE data sets. The packaged data consists of the CalCOFI, AMT, and North Sea data sets. The combined data consists of the unpackaged and packaged data, and uses appropriate algorithm parameters for each. The global data consists of the global evaluation data set, minus the Cota and U. Maryland data plus the high-chlorophyll Carder and EqPac above-surface data, and uses one set of average algorithm parameters for the whole data set. SA indicates that only the modeled values that passed the semi-analytical portion of the algorithm are used (including blended values). SA+EMP indicates that all modeled values—semi-analytical, blended, and empirical—are used. All statistics except RMS2 are calculated from $\log_{10}$-transformed variables.

data set	variable	n	intercept	slope	bias	R^2	RMS1	RMS2
Carder	chl SA	86	0.019	1.020	0.010	0.921	0.122	0.289
Carder	chl SA+EMP	104	–0.007	0.977	–0.002	0.963	0.132	0.300
Carder	$a_\phi(675)$ SA	82	0.098	1.052	–0.008	0.898	0.131	0.302
Carder	$a_g(400)$ SA	26	–0.278	0.905	–0.186	0.751	0.252	0.405
unpackaged	chl SA	339	0.001	1.003	0.000	0.943	0.103	0.240
unpackaged	chl SA+EMP	378	0.002	1.001	0.001	0.962	0.107	0.251
packaged	chl SA	341	0.000	1.003	0.002	0.931	0.118	0.289
packaged	chl SA+EMP	355	0.005	1.010	0.000	0.950	0.119	0.292
combined	chl SA	675	0.004	1.009	–0.001	0.936	0.112	0.280
combined	chl SA+EMP	733	0.004	1.008	0.001	0.958	0.114	0.285
global	chl SA	870	0.003	0.999	0.004	0.863	0.173	0.441
global	chl SA+EMP	955	0.005	1.000	0.005	0.915	0.171	0.436

Table 4. Algorithm parameters used with the "packaged" and modified global data sets. All algorithm parameters not listed here are the same as in Table 1. The values of $a_3(\lambda)$ shown apply to all of the SeaWiFS wavelengths. The equation to determine chl *a* from $a_\phi(675)$ for this data set is given by Equation 21.

parameter	packaged	global
$a_0(412)$	2.02	2.11
$a_0(443)$	3.16	3.38
$a_0(490)$	2.00	2.14
$a_3(\lambda)$	0.020	0.018
p_0	2.404	2.168
p_1	1.294	1.234
p_2	0.052	0.052
c_0	0.4818	0.3147
c_1	–2.783	–2.859
c_2	1.863	2.007
c_3	–2.387	–1.730

LIST OF FIGURE CAPTIONS

Figure 1. Algorithm performance for Carder data set. Top panels are observed vs. modeled chl *a*, middle panels are observed vs. modeled $a_{ph}(675)$, and bottom panels are observed vs. modeled $a_g(400)$. Left panels are scatter plots and right panels are quantile-quantile plots. The lines are the one-to-one lines in all panels. SA (cross) indicates points which are calculated semi-analytically or by a blend of semi-analytical and empirical values. EMP (diamond) indicatespoints that are calculated empirically.

Figure 2. Algorithm performance for and analysis of data sets passing the "unpackaged" numerical filter. Top left panel, a) scatter plot of observed vs. modeled chl *a* (mg m^{-3}). The dotted line is the one-to-one line. Top right panel, b) quantile-quantile plot of observed vs. modeled chl *a*. Middle left panel, c) observed chl *a* vs. r_{25}, with the CZCS algorithm line $C = 1.14[r_{25}]^{-1.71}$. Middle right panel, d) r_{12} vs. r_{25}, with the line, $r_{12} = 0.95[r_{25}]^{0.16}$, used to identify "unpackaged" Case 1 data (above line). Bottom left panel, e) modeled $a_g(400)$ (m^{-1}) vs. observed chl *a*.

Figure 3. As Figure 2 but for data sets not passing the "unpackaged" numerical filter.

Figure 4. As Figure 2 but for both unpackaged and packaged data sets.

Figure 5. As Figure 2, but for the modified global data set.

Figure 6. $a_\phi(443)/a_\phi(675)$ vs. $a_\phi(675)$ for stations from various non-subtropical environments. The solid line is the function used in the semi-analytical algorithm. The dashed lines represent the lower and upper bounds for all of the absorption ratio data that we have collected and the dotted line approximates the median trend.