AN IMPROVED IN SITU BIO-OPTICAL DATA SET FOR OCEAN COLOR ALGORITHM DEVELOPMENT AND SATELLITE DATA PRODUCT VALIDATION

P. Jeremy Werdell[1*] and Sean W. Bailey[2]

[1] Science Systems and Applications, Inc.
NASA Goddard Space Flight Center
Greenbelt, Maryland, USA
jeremy.werdell@gsfc.nasa.gov

[2] Futuretech Corporation
NASA Goddard Space Flight Center
Greenbelt, Maryland, USA
sean.bailey@gsfc.nasa.gov

* to whom correspondence should be addressed

Werdell, P.J. and Bailey, S.W. (2005). An improved in situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, 98(1), 122-140.

0. ABSTRACT

Global satellite ocean color instruments provide the scientific community a high-resolution means of studying the marine biosphere. Satellite data product validation and algorithm development activities both require the substantial accumulation of high-quality *in situ* observations. The NASA Ocean Biology Processing Group maintains a local repository of *in situ* marine bio-optical data, the SeaWiFS Bio-optical Archive and Storage System (SeaBASS), to facilitate their ocean color satellite validation analyses. Data were acquired from SeaBASS and used to compile a large set of coincident radiometric observations and phytoplankton pigment concentrations for use in bio-optical algorithm development. This new data set, the NASA bio-Optical Marine Algorithm Data set (NOMAD), includes over 3,400 stations of spectral water-leaving radiances, surface irradiances, and diffuse downwelling attenuation coefficients, encompassing chlorophyll *a* concentrations ranging from 0.012 to 72.12 mg m^{-3}. Metadata, such as the date, time, and location of data collection, and ancillary data, including sea surface temperatures and water depths, accompany each record. This paper describes the assembly and evaluation of NOMAD, and further illustrates the broad geophysical range of stations incorporated into NOMAD.

1. INTRODUCTION

The oceans contribute substantially to biological processes that help regulate the Earth's climate, yet, the geographic and temporal extent of such contributions is only partially understood. Coastal and inland waters support a diverse assortment of ecosystems, many of which possess commercial and ecological value. Still, a scientific understanding of the biological responses of these ecosystems to perturbations (e.g., climatic disturbances, erosion, and industrial pollution) has not been fully realized. The improved scientific understanding of such processes requires surmounting a considerable obstacle – the routine acquisition of high-quality, globally distributed scientific observations – a feat near impossible using only ships and other marine platforms. Fortunately, satellite-borne ocean color instruments provide a regular, synoptic view of the productivity and variability of the Earth's oceans. A nadir-viewing, polar orbiting instrument with a 1-km^2 Earth footprint and 55º half-scan width, for example, observes an average of 15% of the ocean each day, and up to 50% over 4-days, after accounting for the effects of cloud coverage and contamination by excessive sun glint (Gregg et al. 1998). At such spatial and temporal scales, satellite ocean color instruments provide the scientific community a high-resolution means of studying the marine biosphere.

The color of seawater relies on the relative concentrations of optically active water-column constituents, including phytoplankton pigments, non-algal particulate and dissolved organic carbon, and water molecules themselves (Morel and Prieur 1977). Chlorophyll *a*, the primary photosynthetic pigment in phytoplankton, absorbs relatively more blue and red light than green, and as its concentration increases, the spectrum of backscattered sunlight progressively shifts from blue to green (Yentsch 1960). Satellite-borne ocean color instruments measure the spectral radiant flux emanating upward from the top of the Earth's atmosphere at discrete visible and near-infrared wavelengths. Algorithms are applied to these data to remove the contribution of the atmosphere from the signal (e.g., Gordon and Wang 1994), thereby producing an estimate of the upwelling spectral radiant flux at the sea surface. The resulting water-leaving radiances, $L_w(\lambda)$, are in turn used to estimate a number of geophysical data parameters, such as the concentration of chlorophyll *a*, C_a, via the application of additional bio-optical algorithms (e.g., O'Reilly et al. 1998, Carder et al. 1999, and Maritorena et al. 2002). The radiances and derived parameters, C_a in particular, are subsequently used to assess and monitor temporal changes in the marine ecosystem (e.g., Denman and Abbott, 1994, Siegel et al. 2002, and Subramaniam et al. 2002, Tomlinson et al. 2004) and to investigate the role of marine photosynthesis and net primary productivity in the Earth's carbon budget (e.g., Platt et al. 1991, Longhurst et al. 1995, Antoine et al. 1996, Behrenfeld et al. 2001, and Sarmiento et al. 2004).

Clark and co-authors (1970) successfully used data collected onboard high-altitude aircraft to relate the color of the ocean to its coincident chlorophyll *a* concentration. Less than a decade later, the Coastal Zone Color Scanner, launched onboard the National Aeronautics and Space Administration (NASA) Nimbus-7 spacecraft, provided the first (i.e., proof-of-concept) satellite ocean color data set (CZCS; Hovis et al. 1980). The extensive utility of this data for both coastal and open ocean research prompted the launch of a series of subsequent missions, including the Ocean Color and Temperature Sensor (OCTS; Tanii et al. 1991), the Sea-viewing Wide Field-of-view Sensor (SeaWiFS; McClain et al. 1998), the Moderate Resolution Imaging Spectroradiometer (MODIS; Esaias et al. 1998), and the Medium Resolution Imaging

Spectrometer (MERIS; Rast and Bézy 1995). All employ empirical algorithms to remotely estimate marine chlorophyll *a* concentrations.

The empirical relationship between satellite-derived ocean color and chlorophyll *a* concentrations has been studied for several decades (e.g., Gordon et al. 1980, Clark 1981, and Gordon et al. 1983), culminating recently in the SeaWiFS Bio-optical Algorithm Mini-workshop (SeaBAM; Firestone and Hooker 1998), an international collaboration whose primary goal was the identification of an operational chlorophyll *a* algorithm for SeaWiFS. A by-product of the SeaBAM effort was the compilation of a global, high-quality *in situ* data set of coincident radiometric and chlorophyll *a* concentrations (O'Reilly et al. 1998). O'Reilly and co-authors (2000) expanded this SeaBAM data set (SBDS) prior to using it to define the current ocean chlorophyll *a* algorithms for OCTS, SeaWiFS, and MODIS. Today, the SBDS remains one of the largest *in situ* data sets assembled and realizes continued value as a resource for the refinement and verification of bio-optical reflectance models (Carder et al. 1999, Maritorena et al. 2002, Yan et al. 2002, Tanaka et al. 2004). To our knowledge, it prevails as the most widely available public source of global, high-quality reflectance and chlorophyll *a* data for bio-optical algorithm development.

This notwithstanding, the observations in SBDS suffer from several deficiencies, such as a lack of associated metadata (e.g., date and time of collection and station latitude and longitude), making regional and temporal analyses impossible. Further, a mechanism for adding new observations or updating existing data has yet to be developed. Although more contemporary empirical algorithms have been successfully developed using regional data (Kahru and Mitchell

1999, D'Ortenzio et al. 2002, Gohin et al. 2002, Darecki and Stramski 2004, Garcia et al. 2005), they have rarely been verified on global scales, and generally consider only a local range of geophysical conditions. In principal, globally applied chlorophyll *a* algorithms produce the most widely acceptable results when developed using a cohesive global data set (e.g., Maritorena et al. 2002). The ongoing satellite missions (e.g., SeaWiFS, MODIS, and MERIS), and those scheduled to launch in the next decade (e.g., the Visible/Infrared Imager/Radiometer Suite (VIIRS) mission, part of the National Polar-orbiting Operational Enviromental Satellite System (NPOESS) Preparatory Project), require current *in situ* data for their respective calibration and validation activities (Hooker and McClain 2000, Werdell et al. 2003). Following, these missions, and the merger of their respective data products, also benefit from the use of bio-optical algorithms consistently developed with contemporary field data (Barnes et al. 2003). Further, as ocean color satellite data products (McClain et al. 2004) and processing utilities (Baith et al. 2001) have become widely available to an international community, the utility of and need for regionally tuned algorithms has increased significantly.

As such, we propose that a modern data set, with detailed metadata and an inherent mechanism for adding and updating data records, be considered prerequisite to adequately support present and future satellite ocean color missions. In this article, we describe the compilation and evaluation of such a data set, the NASA bio-Optical Marine Algorithm Data set (NOMAD), which is co-located with the NASA Ocean Biology Processing Group (OBPG) at Goddard Space Flight Center, in Greenbelt, Maryland, U.S.A. The data products included in NOMAD (Table 1) are used simultaneously for OBPG calibration and validation activities. Following the legacy of the SBDS, for remote sensing studies, these pertinent geophysical data products include spectral water-leaving radiances and surface irradiances (the ratio of which provides remote sensing reflectance), spectral column-averaged diffuse attenuation coefficients, and chlorophyll *a* concentrations. We also recorded weekly-averaged sea surface temperature (optimum interpolation sea surface temperature; Reynolds et al. 2002) and station water depths (National Geophysical Data Center ETOPO2; Smith and Sandwell 1997, Jakobsson et al. 2000) with each data record, but do not discuss either in this article. NOMAD was designed for the future inclusion of aerosol optical depths and inherent optical properties (specifically, spectral absorption and scattering coefficients) to support validation of atmospheric correction products (Knobelspiesse et al. 2004, Wang et al. 2005) and marine semi-analytical algorithms (Roesler and Perry 1995, Hoge and Lyon 1996, Garver and Siegel 1997, Carder et al. 1999, Lee et al. 2002), respectively; a detailed discussion of either, however, is beyond the scope of this article.

2. METHODS

2. 1 data acquisition

The NASA SeaWiFS Bio-optical Archive and Storage System (SeaBASS) serves as the local repository for *in situ* radiometric and phytoplankton pigment data used by the NASA OBPG in their satellite validation activities (Werdell and Bailey 2002, Werdell et al. 2003). SeaBASS was originally populated with data used in the NASA SeaWiFS Project’s algorithm development activities (Hooker et al. 1994, O’Reilly et al. 1998). The archive was expanded, both spatially and temporally, as part of the NASA Sensor Intercomparison and Merger of Biological and Interdisciplinary Ocean Studies (SIMBIOS) Program (McClain et al. 2002, Fargion et al. 2003, Barnes et al. 2003). Currently, SeaBASS consists of oceanographic and atmospheric field data from over 1350 field campaigns contributed by researchers from 48 institutions in 14 countries.

As of February 2005, participants in the SIMBIOS Program contributed just over 90% of the data archived in SeaBASS (Fargion and McClain 2003). Other NASA-funded researchers and many U.S. and international voluntary contributors provided the remainder (Table 2). To facilitate consistency across these many data contributors, the SeaWiFS and SIMBIOS Project Offices (SSPO) specified *a priori* a series of *in situ* data requirements and sampling strategies that, when followed, ensure the observations are acceptable for algorithm development and ocean color sensor validation (Mueller et al. 2003a, 2003b.).

We acquired approximately 15,000 radiometric and 32,000 pigment observations from SeaBASS (Table 3). The radiometric data sources include both in-water profiling instruments and handheld or platform mounted above-water instrumentation. We considered both multi- and hyperspectral resolution instruments. The phytoplankton pigment concentrations were calculated via laboratory analysis of discrete water samples, fluorometric instruments onboard profiling packages, and fluorometric instruments continuously sampling shipboard flow-through systems. The former included both laboratory fluorometry and high performance liquid chromatography (HPLC) measurements. The volume of data considered and the range of sources prohibit comprehensive description of data sampling in this article (although a brief effort is made to do so in each subsequent section). As such, and also for consistency, we limited data to those collected following SSPO protocols, or compatible methods when appropriate documentation was provided by the data contributor. Data are fully processed to depth-registered, calibrated geophysical values by the data contributor prior to inclusion in SeaBASS, thus eliminating the need for any additional calibration or normalization efforts by the SSPO. Data contributors were queried when outliers and questionable measurements were identified, or when data processing methods or instrument calibrations were uncertain. Otherwise, the data were considered accurate *as is* after acquisition from SeaBASS.

To facilitate the post-processing evaluation of uncertainties resulting from varying observation types and measurement resolutions (such as, in-water versus above-water radiometry, or analysis of water samples collected via profiling rosettes versus shipboard sea chests), the OBPG established a series of binary flags to record collection and processing details for each measurement (Table 4). As such, a *flag* field accompanies every measurement in the final compiled data set. In addition, intermediate processing minutiae were also logged for reference, as described in each of the subsequent sections. Briefly, the latter includes statistics generated from each processing method and contributor- and processor-provided comments.

2.2 radiometry

2.2.1 radiometric profiles

Radiometric profiles were limited to those with coincident observations of upwelling radiance, $L_u(\lambda,z)$ and downwelling irradiance, $E_d(\lambda,z)$. When available, measurements of surface irradiance, $E_s(\lambda,0^+)$, usually collected near-simultaneously either on the deck of the vessel or nearby buoy, were also acquired. The term z is used to indicate the observations' dependence on depth, while 0^+ denotes an above water observation. Spectral dependence is hereafter implied, but, for clarity, will no longer be explicitly denoted. Each observation was viewed and processed using visualization software developed by the OBPG (Werdell and Bailey 2002). Measurements without near-surface profiler data (less than 5 meters) or without significant overall stability in the reference surface irradiance were excluded. All remaining measurements of radiance and irradiance were corrected for variations in solar irradiance using the surface reference cell, if available and not performed prior to data submission. Contaminated observations (e.g., those resulting from wave focusing near the surface or high tilt) were removed from the profiles. Data collected under non-ideal, or cloudy, sky conditions were not excluded. The comments provided by the data contributor in each data file were also considered and recorded.

Near-surface diffuse attenuation coefficients were calculated from the radiance and irradiance profiles via a linear exponential fit to the corrected data. These coefficients were used to propagate the radiances and irradiances to just beneath the surface, $L_u(0^-)$ and $E_d(0^-)$, respectively. For consistency, upwelling radiances were not corrected for instrument self-shading (Gordon and Ding 1992, Zibordi and Ferrari 1995), as the required supporting data were often inadequate (i.e., the absorption coefficient of the water mass and the ratio between diffuse and direct Sun irradiance). The uncertainty introduced by omitting this correction varies geographically and temporally and by instrument, however, it may be selectively applied at a later time as supporting data become available or as deemed appropriate. Water leaving radiances were then determined as $L_w = t_u L_u(0^-) n^{-2}$, where t_u is the upward Fresnel transmittance of the air-sea interface (≅ 0.975) and n is the refractive index of seawater (Austin 1974). Similarly, extrapolated surface irradiances were computed as $E_s = t_d^{-1} E_d(0^-)$, where t_d is the downward Fresnel irradiance transmittance across the air-sea interface (≅ 0.96; Mueller et al. 2003a). Observations were considered questionable and discarded if extrapolated surface irradiances could not be reconciled with reference surface irradiances (extrapolated $E_s \cong$ reference $E_s \pm 25\%$, Figure 1). Intermediate processing details were logged, including the extrapolation depth intervals, which were occasionally variable as a function of wavelength, extrapolation statistics, cast direction, and processor-defined comments.

2.2.2 the remote sensing diffuse attenuation coefficient

Gordon and McCluney (1975) demonstrated that 90% of remotely-sensed radiance originates in the upper layer, defined by depth z_{90}, corresponding to the first optical attenuation length as defined by Beer's Law. As such, existing ocean color algorithms are designed to return an average diffuse attenuation coefficient over the depth range from just beneath the surface ($z = 0^-$) to depth z_{90} (Austin and Petzold 1981, Mueller 2000). Measurements of $E_d(\lambda,z)$ were smoothed using a weighted least-square polynomial fit. Using the smoothed data and the previously calculated subsurface irradiance, values for $z_{90}(\lambda)$ were identified as the depth which satisfied the condition $E_d(\lambda,z_{90}) = E_d(\lambda,0^-)\ e^{-1}$. Remote sensing diffuse attenuation coefficients, $K_{rs}(\lambda)$, were calculated from the original irradiance profiles by applying a linear exponential fit over the depth range from $z = 0^-$ to $z_{90}(\lambda)$. Radiometric profiles with retrieved $K_{rs}(\lambda)$ values less than the value for pure water ($K_w(490) = 0.016\ m^{-1}$; Mueller 2000) were considered questionable and discarded. Otherwise, both $K_{rs}(\lambda)$ and $z_{90}(\lambda)$ were recorded (Figure 2).

2.2.3 above-water radiometry

Above-water measurements of surface and sky radiance under known geometric conditions may be used to derive water-leaving radiances (Mobley 1999, Hooker et al. 2002, Deschamps et al. 2004). Contributors of above-water radiometric observations performed this derivation prior to submission to SeaBASS, thus eliminating the need for additional data preparation (Arnone et al. 1994, Mueller et al. 2003a). For 12% of these field campaigns (relating to 39% of all above-water observations), the data contributor provided remote sensing reflectances in lieu of water-leaving radiances. Under this circumstance, water-leaving radiances were estimated from remote sensing reflectances via $L_w = R_{rs}\ E_s$. If surface irradiances were not explicitly provided, they were derived using a clear sky model based on Frouin et al. (1989), an operation requiring an assumption of ideal sky conditions at the time of data collection. The uncertainty introduced by such an assumption will be minimal when developing algorithms using radiance ratios, as the modeled surface irradiances are, in general, spectrally flat. That is, the ratio of any two discrete modeled irradiance values is approximately unity, and, following, the errors associated with the magnitude of the modeled irradiances are mathematically cancelled. Conceptually, the uncertainty will also be negligible for satellite validation activities, where only the clearest days are considered (Werdell et al. 2003). While the demonstrated uncertainties associated with above-water radiometry are significant even under ideal conditions (Toole et al. 2000, Hooker et al. 2002), data collected onboard underway research vessels were considered in this analysis when located in an otherwise under-sampled bio-regime. Underway observations were reduced as described in section 2.3.3.

2.2.4 hyperspectral radiometry

We maintained the inherent spectral resolution of all data collected with multispectral instruments. While a hyperspectral data are desirable for remote-sensing analyses whose foci are marine component identification, for example, seafloor type (Dierssen et al. 2003, Werdell and Roesler 2003), phytoplankton and microbial pigments (e.g., Richardson et al. 1994 and Andréfouët et al. 2003), and water-column constituents (e.g., Lee et al. 1999 and Lee and Carder 2003), such studies are beyond the intended purpose of this data set. To expedite the merger of the hyperspectral data with the multispectral data, and therefore increase the overall geographic coverage by including such data sets, we reduced all high-resolution data to discrete values at predefined wavebands, λ_c, the selection of which is described in section 2.7. To mimic the approximately rectangular 10-nm full width at half-maximum band pass inherent to most commercially available multispectral radiometers, data collected with hyperspectral instruments were degraded to 11-nm averages centered on λ_c, as defined by:

$$X(\lambda_c) = \frac{\sum_{i=\lambda_c-5}^{\lambda_c+5} X(\lambda_i)}{n} \quad (1)$$

where X is some radiometric quantity, such as remote-sensing reflectance, and n is the number of wavelengths considered (= 11).

2.2.5 radiometric data reduction

While on station, replicate radiometric observations are often acquired (e.g., multiple up and down profiles during a single cast) to increase the statistical reliability of the measurement. Replicate observations of both in-water and above-water radiometry were identified via a combination of spatial and temporal thresholds, defined independently for each field campaign based on water type and data collection method. Often, minimal interference was required as the recorded latitude and longitude coordinates remained fixed for a given station while the time stamp varied. In the presence of coordinate drift, spatial thresholds of less than 0.01 degree were typically assigned (roughly equating to < 1.1 km at the Equator, also equivalent to a nadir-viewing MODIS-Aqua or SeaWiFS Local Area Coverage footprint). Typically, temporal thresholds of one hour or less were assigned, as coincident changes in station location often preempted the need for longer periods.

All replicate radiometric measurements were individually viewed and reduced via analysis of coincident water-leaving radiances, surface irradiances, and remote sensing reflectances, the combined evaluation of which provides simultaneous insight into processing artifacts, changing sky conditions, and water-mass variability that results from erroneous replicate identification.

For example, for a given station with multiple measurements, comparable surface irradiance spectra indicated stable sky conditions and similar remote sensing reflectances implied a consistent water mass. Under ideal circumstances, when the statistical variance of all three products was low, the geometric mean was calculated. If only the remote sensing reflectances were stable, we retained the single observation with the highest surface irradiance, an indicator of the clearest sky conditions. Stable surface irradiances and highly variable water-leaving radiances suggested errors in data processing or replicate evaluation (e.g., insufficiently small spatial thresholds for frontal regions). Data with these symptoms were reevaluated, and eventually, discarded upon unsuccessful reconciliation. For all of the above, the average observation time, latitude, and longitude were recorded.

2.3 phytoplankton pigments

2.3.1 phytoplankton pigment methods

Both fluorometric- and HPLC-determined concentrations of the phytoplankton pigment chlorophyll *a* were acquired (Table 3). For HPLC, following SSPO protocols, only total chlorophyll *a* was considered, and calculated as the sum of chlorophyllide *a*, chlorophyll *a* epimer, chlorophyll *a* allomer, monovinyl chlorophyll *a*, and divinyl chlorophyll *a*, where the latter two were physically separated (Mueller et al. 2003b). Laboratory analysis of discrete water samples often yielded coincident fluorometric- and HPLC-determined chlorophyll *a* concentrations, although, overall, fluorometric determinations far outnumber HPLC determinations in SeaBASS. Additional continuous depth profiles and underway observations were collected via calibrated *in situ* fluorometers, either mounted to CTD packages or coupled to shipboard sea chests. For both, only calibrated data (concentrations, not voltages) were considered to ensure first-order quality assurance by the data contributor and to eliminate the need for additional OBPG data preparation. Discrete pigment measurements made only at the sea surface were also acquired, and replicate measurements were averaged.

2.3.2 profiles

Following the logic of section 2.2.2, weighted remote sensing estimates of chlorophyll *a* were calculated from profiles of concentrations using the method described in Gordon and Clark (1980), a computation that requires *a priori* knowledge of the local attenuation conditions. The remote sensing diffuse attenuation coefficient, $K_{rs}(490)$, was used for consistency and computational simplicity. Coincident radiometric observations provided the most reliable source of $K_{rs}(490)$. When unavailable, the diffuse attenuation coefficient was estimated using one, or more, of the following supplemental data sources: (a) *in situ* observations of the absorption and scattering coefficients (Kirk 1984); (b) above-water measurements of normalized water-leaving radiance (Mueller 2002); and (c) the chlorophyll *a* profile itself (Morel and Maritorena 2001). No preference was given to any of the above, although relative closure amongst the methods was required when multiple supplemental data sources were available. The required degree of agreement varied based on the sampling resolution of the phytoplankton pigment profile and the reported structure of the water mass. For example, vertically homogeneous observations where the depth of the chlorophyll *a* maxima exceeded $z_{90}(490)$ required less agreement (~ ± 25%) than stations with shallow, vertically stratified measurements (~ ± 10%). Data were discarded upon unsuccessful reconciliation.

2.3.3 underway observations

Continuous underway, or flow-through, fluorometric measurements of chlorophyll *a* at fixed depths were systematically reduced to observations collected at local 10:30, 12:00, and 13:30, coinciding with typical MODIS-Terra, SeaWiFS, and MODIS-Aqua overpass times, respectively. Software developed by the OBPG was used to display the underway data series for each sampling day, remove statistical outliers around each overpass time, and calculate a 15-minute average, t_c, of measurements centered on each overpass time, t_o:

$$C_a(t_c) = \frac{\sum_{i=t_o-450}^{t_o+450} C_a(t_i)}{n} \quad (2)$$

where t_o has units of seconds and n is the number of observations considered. The average chlorophyll *a* concentrations were retained when the coefficient of variation for the 15-minute sampling interval was less than 0.2.

2.4 exceptions

Naturally, the volume of data considered and the wide range of sources prompted several exceptions to be made in how specific data were treated (Table 2). SeaBASS data contributors occasionally provided water-leaving radiances and diffuse attenuation coefficients derived from in-water measurements without providing the radiance and irradiance profiles. In such cases, the contributor commonly estimated diffuse attenuation coefficients over the irradiance extrapolation interval, $K_d(\lambda, z_1 \text{ to } z_2)$, where z_1 and z_2 indicate the minimum and maximum depths in the interval. Such values differ from the remote sensing diffuse attenuation coefficient, $K_{rs}(\lambda, 0^- \text{ to } z_{90})$, in the presence of a stratified water column where the water mass is heterogeneous at depths less than z_{90}. Similarly, SeaBASS data contributors occasionally provided water-leaving radiances that included a correction for instrument self-shading (Zibordi and Ferrari 1995), either without providing the radiance profiles, or with a documented perspective in favor of the correction for their particular field campaign. As discussed in section 2.2.1, the associated uncertainty for including such data varies irregularly, and future work might consider the routine application of the correction as the required supporting data become available. Data from both of the above examples were not excluded from this analysis. We excluded, however, all radiometric data collected solely on tethered buoys (e.g., the Satlantic, Inc. Tethered Spectral Radiometry Buoy) and moorings, as these data are predominantly scarce in SeaBASS and, when available, rarely included supporting radiometric information for use in the extrapolation of $L_u(\lambda,z)$ to $L_u(\lambda,0^-)$.

2.5 radiometry and pigment data merger

The operational definition of a coincident observation follows the approach described in section 2.2.5, albeit with reverse logic. As before, concurrent measurements were identified via a combination of temporal and spatial thresholds. We defined each threshold independently for every field campaign based on marine and atmospheric conditions, sampling rates, water mass, and geography. Subjectivity was rarely required, as the recorded latitude and longitude coordinates and time stamp often remained fixed for a given sampling station. In the presence of temporal drift, usually the result of a time lag between radiometric sampling and discrete water collection, a threshold of one hour was typically assigned. A spatial threshold of 0.1 degree was used, as coincident changes in sampling time and logistical spatial improbabilities preempted the need for larger distances. Turbid and ecologically patchy locations necessitated more stringent spatial and temporal thresholds, while horizontally homogeneous regions (e.g., subtropical gyres) permitted relaxed thresholds. Relaxed thresholds were occasionally applied to accommodate data merger in under-sampled geographic regions (Table 5).

Although replicate radiometric and pigment measurements were previously consolidated, differences in thresholds at those stages occasionally created nonsymmetrical stations for the two data types, thus permitting multiple radiometric observations to be associated with multiple chlorophyll *a* measurements. The time and location stamp of every chlorophyll *a* measurement was iteratively compared with those of the radiometric observations to identify the closest match in time and space. Next, this process was reversed (radiometry iteratively compared with chlorophyll *a*) to verify the association. Once associated, observations were removed from the pool of available data. Such logic ensured that coincident observations were accurately identified and that measurements were only used once. The merged data set at this point included 3,720 coincident radiometric and phytoplankton pigment observations.

2.6 data storage

The OBPG developed a relational database management solution (RDBMS; SQL Server, Sybase, Inc.) to organize and catalog the radiometric and phytoplankton pigment measurements. A similar approach is employed to distribute SeaBASS data (Werdell and Bailey 2002). Once processed as described in section 2, the geophysical data and metadata, including binary flags, were ingested into a series of database tables, simultaneously making this information available to the OBPG validation system (Werdell et al. 2003). Briefly, station metadata (e.g., date and time, and latitude and longitude) reside in one table, while the geophysical radiometric and pigment data occupy two others. A quaternary database table associates coincident radiometric and pigment observations and contains merged flags from the independent measurements (Figure 3). Although not discussed in this article, the RDBMS design also supports the ability to catalog coincident aerosol optical depths and marine inherent optical properties. Once cataloged, the merged radiometric and chlorophyll *a* data become available for export into a variety of external data storage formats, and via a World Wide Web search engine that interfaces with the RDBMS. For simplicity, we hereafter refer to the merged data set as NOMAD, the NASA bio-Optical

Marine Algorithm Data set. We discuss the public availability of NOMAD and data acquisition methods in section 5.

2.7 wavelength generalization

Water-leaving radiances, surface irradiances, and diffuse attenuation coefficients retain their native instrument-resolution in the RDBMS, for example, $L_w(411.8)$ is not rounded to $L_w(412)$, yielding approximately 250 uniquely cataloged wavelengths. In general, such exact radiometric precision is not required for algorithm development (O'Reilly et al. 1998), so to simplify the data for generalized and efficient use, wavelengths are rounded in each data export process. We predefined a series of 21 nominal wavelengths after both reviewing the spectral resolution of past, present, and future ocean color satellites and considering the frequency of occurrence of center wavelengths in the merged data set (Figure 4, Table 6). When exported from the database, radiometric data are assigned the predefined wavelength, λ_{pd}, that satisfies the condition $\{\lambda_{pd} - 2\text{-nm}\} \leq \lambda_n \leq \{\lambda_{pd} + 2\text{-nm}\}$, where λ_n is the native instrument wavelength.

3. RESULTS

3.1 quality assurance

We adopted a rigorous approach to applying secondary quality assurance metrics to NOMAD to ensure that observations fell within expected ranges and did not clearly exhibit characteristics of measurement or calibration problems (preliminary metrics were inherently imbedded into the data processing and reduction steps of section 2). Rejection criteria, however, were cautiously defined in order to remove only the most extreme stations. Our intent was to eliminate anomalous and spurious data, yet retain a diverse range of marine geophysical conditions ranging from oligotrophic to eutrophic and Case-1 to Case-2 conditions. Following, the first quality measures focus on the internal consistency of the radiometric and chlorophyll *a* measurements. Data were initially evaluated on a cruise-level, where only observations collected on a given field campaign were considered. Subsequent tests compared these data against the full suite of observations. While both steps incorporated identical analyses, as described below, the regional comparisons revealed discrete measurement and processing errors that would have otherwise been disguised in the global comparisons. In addition, the low data volume of a single field campaign facilitated the discrimination of unique geophysical conditions versus outliers (i.e., the identification of false-positives). Conversely, comparing data from one cruise against the full data set identified systematic calibration and processing errors for that field campaign. It is conceivable, for example, that statistical relations between water-leaving radiances and chlorophyll *a* will be conserved when comparing observations from various oligotrophic regions. Differences, highlighted by the latter global analyses, indicate data with possible errors.

Radiance band ratios and diffuse attenuation coefficient band ratios were plotted versus themselves, such as $R_{rs}(443)$ / $R_{rs}(510)$ versus $R_{rs}(490)$ / $R_{rs}(555)$, and versus chlorophyll *a* to identify outliers. Single radiance bands were also plotted versus chlorophyll *a*. All were effective in revealing data with errors and in determining which data were correctable and recoverable, versus those to be removed from the data set. Additional measures used two widely reviewed four-band empirical chlorophyll algorithms (OC4 and OC4O; O'Reilly et al. 2000) and the operational SeaWiFS diffuse attenuation coefficient algorithm (K490; Mueller 2000) to identify anomalous radiance measurements (Figure 5). Following O'Reilly and co-authors (1998), an observation was considered an outlier when the ratio of the modeled chlorophyll *a* concentration or $K_{rs}(490)$ to the *in situ* measurement exceeded 5:1 or was less than 1:5. As the performance of these models degrades in the most eutrophic waters, the latter outlier rejection criteria were only applied to stations with C_a less than 3.0 mg m^{-3} to facilitate the inclusion of the highest chlorophyll observations. Still, the maximum ratio of the modeled to *in situ* values was 1:7.6, at a station with an observed concentration of 17.7 mg m^{-3}. When both were available, fluorometrically-derived chlorophyll *a* concentrations were compared with coincident HPLC total chlorophyll *a* measurements (Figures 6). Some differences were anticipated, as, for example, fluorometric accuracy degrades in the presence of accessory pigments (Mantoura et al. 1997, Trees et al. 2000). We, again, applied a five to one threshold on the ratio of fluorometric to HPLC values (and vice-versa) to reveal outliers. Quality measures applied to the SIMBIOS-era pigment data prior to submission to SeaBASS significantly minimized the volume anomalous stations identified in this analysis (Fargion and McClain 2003). Observed spectral surface irradiances were compared with modeled clear sky values, based on the algorithm of Frouin and co-authors (1989). Stations were considered questionable and discarded when the *in situ* value exceeded the modeled value by more than 33%. As such algorithms require date and location inputs, the latter analysis also proved effective in locating erroneous station metadata. Finally, those data collected as part of a multi-year (i.e., time series) experiment, such as the Bermuda Atlantic Time Series (Siegel et al. 2001), were plotted as a function of both time and season. These plots were useful in revealing both discrete spurious stations and possible long-term instrument calibration biases. After elimination of 245 questionable stations, 3,475 stations remained in the final NOMAD data set (Figure 7).

3.2 geophysical distribution

The final NOMAD data consists of fluorometrically-derived chlorophyll *a* concentrations ranging from 0.012 to 72.12 mg m^{-3}, and HPLC-derived values ranging from 0.021 to 48.99 mg

m^{-3}. The geometric means were 0.82 and 0.38 mg m^{-3}, respectively, both significant increases from the 0.27 mg m^{-3} reported in O'Reilly and co-authors (1998) and far greater than the global ocean mean of 0.19 mg m^{-3} reported in Antoine and co-authors (1996). The fluorometric data (n = 2,974) consist of approximately 7, 48, and 45% of oligotrophic, mesotrophic, and eutrophic stations, respectively, if 0.1 and 1 mg m^{-3} are taken as approximate limits between oligotrophic and mesotrophic waters and between mesotrophic and eutrophic waters. Likewise, the HPLC data (n = 986) consist of 19, 49, and 32%, respectively (Figure 8). As such, when considering the approximate proportions for the world ocean, 56% oligotrophic, 42% mesotrophic, and 2% eutrophic (Antoine et al. 1996), eutrophic waters are severely over-represented in NOMAD, at least to concentrations up 10 mg m^{-3}. Concentrations greater than 10 mg m^{-3} comprise only 6.7% of the combined fluorometric and HPLC data set. That aside, the relative representation of oceanic regions with chlorophyll *a* concentrations between 1 and 10 mg m^{-3} in NOMAD is equivalent or improved in comparison with its predecessors (O'Reilly et al. 1998, 2000). Of the 245 stations excluded from NOMAD, 7.0, 40.8, and 52.2% were oligotrophic, mesotrophic, and eutrophic stations, respectively.

Not surprisingly, the evaluation of R_{rs} ratios as a function of chlorophyll *a* concentration suggests that $R_{rs}(443) / R_{rs}(555)$ dominates where $C_a < 0.3$ mg m^{-3}, $R_{rs}(490) / R_{rs}(555)$ where $0.3 < C_a < 2$ mg m^{-3}, and $R_{rs}(510) / R_{rs}(555)$ where $C_a > 2$ mg m^{-3} (Figure 5A). Overall, the R_{rs} band ratios versus chlorophyll *a* concentration demonstrate a conservative sigmoid relationship, as indicated repeatedly in the past (e.g., Gordon et al. 1988 and Morel and Maritorena 2001). Our results, however, find this relationship to be more symmetric than previously suggested (O'Reilly et al. 1998), as indicated the more asymptotic correlation at chlorophyll *a* concentrations greater than

10 mg m^{-3}. Past analyses, however, suffered from a more significant paucity of observations in high chlorophyll waters.

Plots of the NOMAD visible remote-sensing reflectance spectra clearly exhibit this blue-to-green shift in their maximum value with increasing chlorophyll *a* concentration (Figures 9 and 10). The spectral shapes are, in general, fairly conserved for $C_a < 0.2$ mg m^{-3}, although the dispersion of the blue wavelengths increases with increasing C_a in this data range. Comparisons of blue-to-green R_{rs} band ratios further highlight such dispersion (Figure 11). As C_a approaches concentrations of 1 mg m^{-3} and greater, variations in spectral magnitude increase significantly, particularly at green wavelengths (Figure 10), with an overall flattening of the spectral blue-to-green shape (Figure 11).

Geophysical and experimental conditions are both potential contributors to such variation. While the marine optical backscattering efficiency of the water column tends to decrease with increasing C_a (Gordon et al. 1988, Twardowski et al. 2001, Morel and Maritorena 2001), other particulate and dissolved non-algal constituents regularly exist at the most biologically productive NOMAD stations (evident to a first order in the predominantly coastal distribution in Figure 7). Small, particulate, non-algal constituents, such as mineralic silt, effectively scatter photons backwards, resulting in elevated reflectance spectra (Stramski et al. 2004, Woźniak and Stramski 2004). Following, stations with high C_a, but otherwise observing strict Case-1 conditions, often yield reflectance spectra depressed relative to those with additional optically relevant constituents. At the most biologically productive stations, larger biomasses elevate the role of absorption relative to scattering, thus again reducing the magnitude of the reflectance spectra (e.g., Carder et al. 1989 and Magnuson et al. 2004)

Note that the chlorophyll *a* ranges with the most obvious spectral R_{rs} variability are also the most highly sampled ranges (Figures 8 and 9), and, hence, geographic diversity likely contributes to the spectral dispersal in these ranges. Following, the general reduction in variability in the most productive waters (C_a > 20 mg m^{-3}), may result from a lack of geographic, and thus geophysical, variability in NOMAD. In addition, data processing errors (resulting from, for example, incorrect depth registration) also contribute to errors in spectral magnitude. Comparisons of extrapolated surface irradiances with averaged values from the above-water reference instruments, however, suggest that the resulting differences are not systematically correlated with water column turbidity (Figure 12).

4. DISCUSSION

To our knowledge, NOMAD is the largest public *in situ* data set ever assembled for bio-optical algorithm development and ocean color satellite validation activities. The quality of such analyses, however, relies on the quality of the *in situ* data set itself. Quality assurance metrics were designed to ensure the removal of extreme outliers and spurious data without the use of overly restrictive rejection criteria. Hence, stations departing slightly from established bio-optical trends deliberately remain. We reduced the variability associated with data acquisition methodologies and post-processing techniques by accepting only data collected following SSPO collection protocols (Mueller et al. 2003a and 2003b) and by using consistent processing techniques to derive water-leaving radiances and optically-weighted chlorophyll *a* concentrations (Werdell and Bailey 2002). Variability associated with instrument design and calibration, and environmental factors, such as sea and sky state, remain. For the most part, however, the latter uncertainties reflect the inherent variability of the waters sampled, and alternatively, may be desirable for global bio-optical algorithm development activities. Users of NOMAD might consider the additional application of instrument self-shading (Zibordi and Ferrari 1995) or bi-directional reflectance distribution function (Morel et al. 2002) corrections, contingent on their particular analyses.

Unfortunately, several major provinces of the world ocean are either poorly or not represented in NOMAD, for example, the southern Pacific and Indian Oceans south of –30ºS latitude. (Figure 7). Others are statistically over-sampled, such as the east and west coasts of North America. Such geographic biases are not surprising given the operational difficulty of visiting the remote open ocean. The resulting geophysical biases, however, such as the prevalence of eutrophic stations over oligotrophic stations, are undesirable for the development of general bio-optical algorithms to be applied to the world ocean (O'Reilly et al. 1998). We advise users of NOMAD to be cautious when performing subsequent statistical analyses of the data set, as particular water types are over-represented. NOMAD includes a larger volume of high chlorophyll *a* concentrations (> 20 mg m^{-3}) than its predecessors (O'Reilly et al. 1998), but like them, continues to suffer from a paucity of measurements below 0.05 mg m^{-3}. Data from the clearest ocean waters are prerequisite to determine if existing global ocean color chlorophyll *a* algorithms accurately extrapolate to the lowest concentrations (O'Reilly et al. 2000).

In general, however, NOMAD includes a broad range of chlorophyll *a* concentrations (0.012 to 72.12 mg m^{-3}) and encompasses a significant variety of mesotrophic and eutrophic water types (Figure 9). The comparison of calculated $z_{90}(\lambda)$ with its theoretical value (= $K_{rs}(\lambda)^{-1}$) reveals a number of stratified, or heterogeneous, stations (Figure 2), information relevant towards the development of Case-1 and Case-2 prediction algorithms. This variability in water mass is also evident when evaluating the function form of the SeaWiFS operation K490 algorithm (Mueller 2000; Figure 5B), when populations of different water masses become statistically obvious (e.g., $K_{rs}(490) - K_{w}(490) > 0.3$ m^{-1}). Naturally, analyses of such marine variability will improve appreciably with the addition of inherent optical properties to the data set (specifically, spectral absorption and backscattering coefficients). But, overall, the radiometric and phytoplankton pigment data of NOMAD adhere to established bio-optical relationships (Figures 2 and 5A), and maintain significant internal consistency both optically and biologically (Figure 13). For the full data set, ratios of the diffuse attenuation coefficient compare sensibly with both ratios of remote-sensing reflectance and chlorophyll *a* concentration. $K_{rs}(\lambda)$ encompasses the effects of all absorbing material in the water column, including that of dissolved organics. The ratio of blue (≈ 443 nm) to green (≈ 555 nm) reflectances provides a reasonable index of the biological properties of the water column (Gordon et al. 1983). While such results (Figure 13) are rarely unanticipated, the relative agreement of the various parameters suggests a significant level of internal consistency amongst the wide variety of observations accumulated in NOMAD.

Unlike its predecessors (e.g., O'Reilly et al. 1998 and 2000) NOMAD benefits from systematic and consistent data processing and evaluation, the inclusion of fundamental metadata, such as date and time of collection, and latitude and longitude coordinates, and the addition of accessory data products, including sea surface temperatures (OISST; Reynolds et al. 2002) and water depths (ETOPO2; Smith and Sandwell 1997, Jakobsson et al. 2000). These features, in combination with the large volume of coincident radiometry and fluorometric- and HPLC-derived chlorophyll *a* concentrations, both foster the improvement of existing ocean color chlorophyll algorithms and facilitate the development of regionally-tuned algorithms. Further, the binary flags that accompany each station (Table 4) provide a unique resource for evaluating, to a first degree, variability in data collection and preparation on a large scale. For example, the NOMAD flags permit the identification of radiometric data collected without an above-water reference irradiance sensor. Further, following recent efforts that compare the estimation of chlorophyll *a* concentrations by both fluorometric and HPLC techniques (e.g., Bianchi et al. 1995), and the derivation of water-leaving radiances by both in-water and above-water methods (e.g., Hooker et al. 2002), the NOMAD flags permit the selection of geophysical values collected by any combination of the latter methods. OBPG-sponsored satellite validation analyses include iterative comparisons of satellite data products with *in situ* values collected via the various radiometric and pigment methods (Werdell et al. 2003, Bailey et al. 2005). Ongoing OBPG validation analyses utilize the integration flags (INT_CHL and INT_HPLC) to compare satellite-derived chlorophyll *a* concentrations with both optically weighted *in situ* concentrations and discrete *in situ* concentrations, in order to quantify uncertainties associated with the use of the latter (Gordon et al. 1983, Mueller et al. 2003b).

5. CONCLUSIONS

NOMAD is a publicly available, global, high quality *in situ* bio-optical data set for use in ocean color algorithm development and satellite data product validation activities. Current data products include coincident observations of water-leaving radiances and chlorophyll *a* concentrations, along with relevant metadata, such as date, time, and coordinates of data collection, and ancillary data, including sea surface temperatures and water depths. This existing suite of data products both encourages the improvement of existing global ocean color chlorophyll algorithms (O'Reilly et al. 2000), and facilitates the development of regional empirical algorithms. We plan on including inherent optical properties (e.g., spectral absorption and backscattering coefficients) in the near future to support the analysis and evaluation of semi-analytic algorithms (Roesler and Perry 1995, Hoge and Lyon 1996, Garver and Siegel 1997, Carder et al. 1999, Lee et al. 2002). The radiometric and pigment profiles used in the development of NOMAD are publicly available in SeaBASS (Werdell and Bailey 2002). The NOMAD data set is currently available online via two mechanisms, a digital text file, adhering to ASCII format, which includes the full merged bio-optical data set, and an Internet search engine that provides a means of limiting the data to specific data products, field campaigns, or date and location ranges. All are available via the SeaBASS Web site (http://seabass.gsfc.nasa.gov).

6. ACKNOWLEDGEMENTS

This work was funded by the NASA Earth Observing System (EOS) / MODIS and NASA Ocean Biogeochemistry Programs. The authors are grateful to C. McClain, G. Feldman, G. Fargion, and all OBPG Staff members for their regular support and assistance in this effort. We thank J. O'Reilly for valuable advice and encouragement. We also thank W. Balch, F. Chavez, L. Harding, S. Hooker, G. Mitchell, R. Morrison, F. Muller-Karger, N. Nelson, D. Siegel, A. Subramaniam, R. Stumpf, and all of their co-investigators for releasing their SeaBASS data contributions to the public ahead of schedule. The utility and success of SeaBASS relies on the tremendous efforts put forth by these, and the numerous other, data contributors.

7. REFERENCES

Andréfouët, S.A., Payri, C., Hochberg, E.J., Che, L.M, and Atkinson, M.J. (2003), Airborne hyperspectral detection of microbial mat pigmentation in Rangiroa atoll (French Polynesia). *Limnology and Oceanography*, 48, 426-430.

Antoine, D, André, J.M., and Morel, A. (1996), Oceanic primary production, 2, Estimation at global scale from satellite (CZCS) chlorophyll. *Global Biogeochemical Cycles*, 10, 57-70.

Arnone, R.A., Gould, R.W., Oriol, R.A., and Terrie, G. (1994), Effects of vertical chlorophyll structure and solar irradiance on remote sensing ocean color spectrum. In J.S. Jaffe (Ed.), *Ocean Optics XII*. *Proceedings SPIE* (Vol. 2258, pp. 322-331). Bellingham: The Society of Photo-Optical Instrumentation Engineers.

Austin, R.W. (1974), The remote sensing of spectral radiance from below the ocean surface. In N.G. Jerlov and E.S. Nielson (Eds.), *Optical Aspects of Oceanography* (pp. 317-344). London: Academic Press.

Austin, R.W. and Petzold, T. (1981), The determination of the diffuse attenuation coefficient of sea water using the Coastal Zone Color Scanner. In J.F.R. Gower (Ed.), *Oceanography from Space* (pp. 239-356). New York: Plenum Press.

Baith, K., Lindsay, R., Fu, G., and McClain, C.R. (2001), Data analysis system developed for ocean color satellite sensors. *EOS Transactions AGU*, 82, 202.

Bailey, S.W. and Werdell, P.J. (2005), The operational NASA Ocean Biology Processing Group satellite data product validation results. http://seabass.gsfc.nasa.gov/matchup_results.html.

Barnes, R.A., Clark, D.K., Esaias, W.E., Fargion, G.S., Feldman, G.C., and McClain, C.R. (2003), Development of a consistent multi-sensor global ocean colour time series. *International Journal of Remote Sensing*, 24, 4047-4064.

Behrenfeld, M.J., Randerson, J.T., McClain, C.R., Feldman, G.C., Los, S.O., Tucker, C.J., Falkowski, P.G., Field, C.B., Frouin, R., Esaias, W.E., Kolber, D.D., and Pollack, N.H. (2001), Biospheric primary production during an ENSO transition. *Science*, 291, 2594-2597.

Bianchi, T.S., Lambert, C., and Biggs, D.C. (1995), Distribution of chlorophyll *a* and phaeopigments in the northwester Gulf of Mexico: a comparison between fluorometric and high-performance liquid chromatography measurements. *Bulletin of Marine Science*, 56, 25-32.

Carder, K.L., Steward, R.G., Harvey, G.R., and Ortner, P.B. (1989), Marine humic and fulvic acids: Their effects on remote sensing of ocean chlorophyll. *Limnology and Oceanography*, 34, 68-81.

Carder, K.L., Lee, Z.P., Hawes, S.K., and Kamykowski, D. (1999), Semianalytic Moderate-Resolution Imaging Spectrometer algorithms for chlorophyll *a* and absorption with bio-optical domains based on nitrate-depletion temperatures. *Journal of Geophysical Research*, 104, 5403-5421.

Clark, G.L., Ewing, G.C., and Lorenzen, C.J. (1970), Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration. *Science*, 167, 1119-1121.

Clark, D.K. (1981), Phytoplankton pigment algorithms for the Nimbus-7 CZCS. In J.F.R. Gower (Ed.), *Oceanography from Space* (pp. 227-237). New York: Plenum Press.

Darecki, M. and Stramski, D. (2004), An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sensing of Environment*, 89, 326-350.

Denman, K.L. and Abbott, M.R. (1994), Timescales of pattern evolution from cross-spectrum analysis of advanced very high resolution radiometer and coastal zone color scanner imagery. *Journal of Geophysical Research*, 99, 7433-7442.

Deschamps, P., Fougnie, B., Frouin, R., Lecomte, P., and Verwaerde, C. (2004), SIMBAD : a field radiometer for satellite ocean-color validation. *Applied Optics*, 43, 4055-4096.

Dierssen, H.M., Zimmerman, R.C., Leathers, R.A., Downes, T.V., and Davis, C.O. (2003), Ocean color remote sensing of seagrass and bathymetry in the Bahamas Banks by high-resolution airborne imagery. *Limnology and Oceanography*, 48, 444-455.

D'Ortenzio, F., Marullo, S., Ragni, M., d'Alcal·, M.R., and Santoleri, R. (2002), Validation of empirical SeaWiFS algorithms for chlorophyll-*a* retrieval in the Mediterranean Sea: A case study for oligotrophic seas. *Remote Sensing of Environment*, 82, 79-94.

Esaias, W.E., Abbott, M.R., Barton, I., Brown, O.B., Campbell, J.W., Carder, K.L., Clark, D.K., Evans, R.H., Hoge, F.E., Gordon, H.R., Balch, W.M., Letelier, R., and Minnett, P.J. (1998), An overview of MODIS capabilities for ocean science observations. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1250-1265.

Fargion, G.S. and McClain, C.R. (2003), *SIMBIOS Project 2003 Annual Report*. NASA Tech. Memo. 2003-212251. Greenbelt: NASA Goddard Space Flight Center, pp. 202.

Fargion, G.S., Franz, B.A., Kwiatkowska, E.J., Pietras, C.M., Bailey, S.W., Gales, J., Meister, G., Knobelspiesse, K.D., Werdell, J., and McClain, C.R. (2003), SIMBIOS program in support of ocean color missions: 1997-2003. In R.J. Frouin, G.D. Gilbert, and D. Pan (Eds.), *Ocean Remote Sensing and Imaging II*. *Proceedings SPIE* (Vol. 5155, pp. 49-60). Bellingham: The Society of Photo-Optical Instrumentation Engineers.

Firestone, E.R. and Hooker, S.B. (1998), *SeaWiFS Prelaunch Technical Report Series Final Cumulative Index, Vol. 43.* NASA Tech. Memo. 1998-104566. Greenbelt: NASA Goddard Space Flight Center, pp 69.

Frouin, R., Longner, D.W., Gautier, C., Baker, K.S., and Smith, R.C. (1989), A simple analytical formula to compute clear sky total and photosynthetically available solar irradiance at the ocean surface. *Journal of Geophysical Research*, 94, 9731-9742.

Garcia, C.A.E., Garcia, V.M.T., and McClain, C.R. (2005), Evaluation of SeaWiFS chlorophyll algorithms in the Southwestern Atlantic and Southern Oceans, *Remote Sensing of Environment*, 95, 125-137.

Garver, S.A. and Siegel, D.A. (1997), Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation .1. Time series from the Sargasso Sea. *Journal of Geophysical Research*, 102, 18607-18625.

Gohin, F., Druon, J.N., and Lampert, L. (2002), A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters. *International Journal of Remote Sensing*, 23, 1639-1661.

Gordon, H.R. and McCluney, W.R. (1975), Estimation of depth of sunlight penetration in sea for remote-sensing. *Applied Optics*, 14, 413-416.

Gordon, H.R. and Clark, D.K. (1980), Remote sensing optical properties of a stratified ocean: and improved interpretation. *Applied Optics*, 19, 3428-3430.

Gordon, H.R., Clark, D.K., Mueller, J.L., and Hovis, W.A. (1980), Phytoplankton pigments from the Nimbus-7 Coastal Zone Color Scanner: Comparisons with surface measurements. *Science*, 210, 63-66.

Gordon, H.R., Clark, D.K., Brown, J.W., Brown, O.B., Evans, R.H., and Broenkow, W.W. (1983), Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparison of ship determinationos and CZCS estimates. *Applied Optics*, 22, 20-36.

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J.W., Smith, R.C., Baker, K.S., and Clark, D.K. (1988), A semianalytic radiance model of ocean color. *Journal of Geophysical Research*, 93, 10909-10924.

Gordon, H.R. and Ding, K. (1992), Self shading of in-water optical instruments. *Limnology and Oceanography*, 37, 491-500.

Gordon, H.R., and Wang, M.H. (1994), Retrieval of water-leaving radiance and aerosol optical-thickness over the oceans with SeaWiFS – a preliminary algorithm. *Applied Optics*, 33, 443-452.

Gregg, W.W., Esaias, W.E., Feldman, G.C., Frouin, R., Hooker, S.B., McClain, C.R., and Woodward, R.H. (1998), Coverage opportunities for global ocean color in a multimission era. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1620-1627.

Hoepffner, N. and Sathyendranath, S. (1992), Bio-optical characteristics of coastal waters: Absorption spectra of phytoplankton and pigment distribution in the western North Atlantic. *Limnology and Oceanography*, 37, 1660-1679.

Hoge, F.E. and Lyon, P.E. (1996), Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: An analysis of model and radiance measurement errors. *Journal of Geophysical Research*, 101, 16631-16648.

Hooker, S.B., McClain, C.R., Firestone, J.K., Westphal, T.L., Yeh, E.N., and Geo Y. (1994), *The SeaWiFS Bio-optical Archive and Storage System (SeaBASS), Part 1*. NASA Tech. Memo. 104566, Vol. 20. Greenbelt: NASA Goddard Space Flight Center, pp. 37.

Hooker, S.B. and McClain, C.R. (2000), The calibration and validation of SeaWiFS data. *Progress in Oceanography*, 45, 427-465.

Hooker, S.B., Lazin, G., Zibordi, G., and McLean, S. (2002), An evaluation of above- and in-water methods for determining water-leaving radiances. *Journal of Atmospheric and Oceanic Technology*, 19, 486-515.

Hovis, W.A., Clark, D.K., Anderson, F., Austin, R.W., Wilson, W.H., Baker, E.T., Ball, D., Gordon, H.R., Mueller, J.L., El Sayed, S.Y., Sturm, B., Wrigley, R.C., and Yentsch, C.S. (1980), Nimbus-7 Coastal Zone Color Scanner: System description and initial imagery. *Science*, 210, 60-63.

Kahru, M. and Mitchell, B.G. (1999), Empirical chlorophyll algorithm and preliminary SeaWiFS validation for the California Current. *International Journal of Remote Sensing*, 20, 3423-3429.

Jakobsson, M., Cherkis, N.Z., Woodward, J., Macnab, R., and Coakley, B. (2000), New grid of Arctic bathymetry aids scientists and mapmakers. *EOS Transactions AGU*, 81, 89, 93, 96.

Kirk, J.T.O (1984), Dependence of relationship between inherent and apparent optical properties of water on solar altitude. *Limnology and Oceanography*, 29, 350-356.

Knobelspiesse, K.D., Pietras, C., Fargion, G.S., Wang, M.H., Frouin, R., Miller, M.A., Subramaniam, A., and Balch, W.M. (2004), Maritime aerosol optical thickness measured by handheld sun photometers. *Remote Sensing of Environment*, 93, 87-106.

Lee, Z.P., Carder, K.L., and Arnone, R.A. (2002), Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Applied Optics*, 41, 5755-5772.

Lee, Z.P. and Carder, K.L. (2003), Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance. *Remote Sensing of Environment*, 89, 361-368.

Lee, Z.P., Carder, K.L., Mobley, C.D., Steward, R.G., and Patch, J.S. (1999), Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization. *Applied Optics*, 38, 3831-3843.

Longhurst, A., Sathyendranath, S., Platt, T., and Caverhill, C. (1995), An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research*, 17, 1245-1271.

Magnuson, A., Harding, L.W., Mallonee, M.E., and Adolf, J.E. (2004), Bio-optical model for Chesapeake Bay and the Middle Atlantic Bight. *Estuarine, Coastal, and Shelf Science*, 61, 403-424.

Mantoura, R.F.C, Jeffrey, S.W., Llewellyn, C.A., Claustre, H., and Morales, C.E. (1997), Comparison between spectrophotometric, fluorometric, and HPLC methods for chlorophyll analysis. In S.W. Jeffrey, R.F.C. Mantoura, and S.W. Wright (Eds.), *Phytoplankton Pigments in Oceanography* (pp. 361-380). Paris: UNESCO Publishing.

Maritorena, S., Siegel, D.A., and Peterson, A.R. (2002), Optimization of a semianalytical ocean color model for global-scale applications. *Applied Optics*, 41, 2705-2714.

McClain, C.R., Cleave, M.L., Feldman, G.C., Gregg, W.W., and Hooker, S.B. (1998), Science quality SeaWiFS data for global biosphere research. *Sea Technology*, 39, 10-15.

McClain, C.R., Esaias, W., Feldman, G., Frouin, R., Gregg, W., and Hooker, S. (2002), *The Proposal for the NASA Sensor Intercalibration and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Program, 1995*. NASA Tech. Memo. 2002-210008, Greenbelt: NASA Goddard Space Flight Center, pp. 62.

McClain, C.R., Feldman, G.C., and Hooker, S.B. (2004), An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. *Deep Sea Research II*, 51, 5-42.

Mobley, C.D. (1999), Estimation of the remote-sensing reflectance from above-surface measurements. *Applied Optics*, 38, 7442-7455.

Morel, A. and Prieur, L. (1977), Analysis of variations in ocean color. *Limnology and Oceanography*, 22, 709-722.

Morel, A. and Maritorena, S. (2001), Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, 106, 7163-7180.

Morel, A., Antoine, D., and Gentili, B. (2002) Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function. *Applied Optics*, 41, 6289-6306.

Mueller, J.L., 2000: SeaWiFS algorithm for the diffuse attenuation coefficient, K(490), using water-leaving radiances at 490 and 555 nm. In S.B. Hooker and E.R. Firestone (Eds.), *SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3* (pp. 24-27). NASA Tech. Memo. 2000-206892 Vol. 11, Greenbelt: NASA Goddard Space Flight Center.

Mueller, J.L., Morel, A., Frouin, R., Davis, C., Arnone, R., Carder, K., Lee, Z.P., Steward, R.G., Hooker, S.B., Mobley, C.D., McLean, S., Holben, B., Miller, M., Pietras, C., Knobelspiesse, K.D., Fargion, G.S., Porter, J., and Voss, K. (2003a), *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4, Volume III: Radiometric Measurements and Data Analysis Protocols*. NASA Tech. Memo. 2003-211621/Rev4-Vol.III, Greenbelt: NASA Goddard Space Flight Center, pp. 78.

Mueller, J.L., Bidigare, R.R., Trees, C., Balch, W.M., Dore, J., Drapeau, D.T., Karl, D., Van Heukelem, L., and Perl, J. (2003b): *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 5, Volume V: Biogeochemical and Bio-optical Measurements and Data Analysis Protocols*. NASA Tech. Memo. 2003-211621, Rev. 5, Vol. V, Greenbelt: NASA Goddard Space Flight Center, pp. 36.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., and McClain, C. (1998), Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, 103, 24937-24953.

O'Reilly, J.E., Maritorena, S., O'Brien, M.C., Siegel, D.A., Toole, D., Menzies, D., Smith, R.C., Mueller, J.L., Mitchell, B.G., Kahru, M., Chavez, F.P., Strutton, P., Cota, G.F., Hooker, S.B., McClain, S.B., Carder, K.L., Muller-Karger, F., Harding, L., Magnuson, A., Phinney, D., Moore, G.F., Aiken, J., Arrigo, K.R., Letelier, R., and Culver, M. (2000), Ocean color chlorophyll *a* algorithms for SeaWiFS, OC2, and OC4: Version 4. In S.B. Hooker and E.R. Firestone (Eds.), *SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3* (pp. 9-23). NASA Tech. Memo. 2000-206892, Vol. 11, Greenbelt: NASA Goddard Space Flight Center.

Platt, T., Caverhill, C., and Sathyendranath, S. (1991), Basin-scale estimates of oceanic primary production by remote sensing: The North Atlantic. *Journal of Geophysical Research*, 96, 15147-15159.

Rast, M. and Bézy, J.L. (1995), The ESA medium resolution imaging spectrometer (MERIS): requirements to its mission and performance of its system. In P.J. Curran and Y.C. Robertson (Eds.), *RSS95, Remote Sensing in Action, Proceedings of the 21st Annual Conference of the Remote Sensing Society, University of Southampton, UK, 11-14 September 1995* (pp. 125-132). Oxfordshire: Taylor and Francis.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., and Wang, W. (2002), An improved *in situ* and satellite SST analysis for climate. *Journal of Climate*, 15, 1609-1625.

Richardson, L.L, Buisson, D., Liu, C.J., and Ambrosia, V. (1994), The detection of algal photosynthetic accessory pigments using airborne visible-infrared imaging spectrometer (AVIRIS) spectral data. *Marine Technology Society Journal*, 28, 10-21.

Roesler, C.S. and Perry, M.J. (1995), In-situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance. *Journal of Geophysical Research*, 100, 13279-13294.

Sarmiento, J.L., Slater, R., Barber, R., Bopp, L., Doney, S.C., Hirst, A.C., Kleypas, J., Matear, R., Mikolajewicz, U., Soldatov, V., Spall, S.A., and Stouffer R. (2004), Response of ocean ecosystems to climate warming. *Global Biogeochemistry Cycles*, 18, GP3003, doi:10.1029/2003GB002134.

Siegel, D.A., Westberry, T.K., O'Brien, M.C., Nelson, N.B., Michaels, A.F., Morrison, J.R., Scott, A., Caporelli, E.A., Sorensen, J.C., Maritorena, S., Garver, S.A., Brody, E.A., Ubante, J., and Hammer, M.A. (2001), Bio-optical modeling of primary production on regional scales: the Bermuda BioOptics project. *Deep Sea Research II*, 48, 1865-1896.

Siegel, D.A., Maritorena, S., Nelson, N.B., Hansell, D.A., and Lorenzi-Kayser, M. (2002), Global distribution and synamics of colored dissolved and detrital organic materials. *Journal of Geophysical Research*, 107, 3228, doi:10.1029/2001JC000965.

Smith, W.H.F. and Sandwell, D.T. (1997), Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, 277, 5334, 1956-1962.

Stramski, D. Boss, E., Bogucki, D., and Voss, K.J. (2004), The role of seawater consituents in light backscattering in the ocean. *Progress in Oceanography*, 61, 27-56.

Subramaniam, A., Brown, C.W., Hood, R.R., Carpenter, E.J., and Capone, D.G. (2002), Detecting *Trichodesmium* blooms in SeaWiFS imagery. *Deep Sea Research II*, 49, 107-121.

Tanaka, A., Kishino, M., Doerffer, R., Schiller, H., Oishi, T., and Kubota, T. (2004), Development of a neural network algorithm for retrieving concentrations of chlorophyll, suspended matter and yellow substance from radiance data of the ocean color and temperature scanner. *Journal of Oceanography*, 60, 519-530.

Tanii, J., Machida, T., Ayada, H., Katsuyama, Y., Ishida, J., Iwasaki, N., Tange, Y., Miyachi, Y., and Sato, R. (1991), Ocean Color and Temperature Scanner (OCTS) for ADEOS. In P.N. Slater (Ed.), *Future European and Japanese Remote-Sensing Sensors and Programs: Proceedings of the Meeting, Orlando, Fla, April 1, 2, 1991. Proceedings SPIE* (Vol.1490, pp. 200-206). Bellingham: The Society of Photo-Optical Instrumentation Engineers.

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

Tomlinson, M.C., Stumpf, R.P., Ransibrahmanakul, V., Truby, E.W., Kirkpatrick, G.J., Pederson, B.A., Vargo, G.A., and Heil, C.A. (2004), Evaluation of the use of SeaWiFS imagery for detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico. *Remote Sensing of Environment*, 91, 293-303.

Toole, D.A., Siegel, D.A., Menzies, D.W., Neumann, M.J., and Smith, R.C. (2000), Remote-sensing reflectance determinations in the coastal ocean environment: impact of instrumental characteristics and environmental variability. *Applied Optics*, 39, 456-469.

Trees, C.C., Clark, D.K., Bidigare, R.R., Ondrusek, M.E., and Mueller, J.L. (2000), Accessory pigments versus chlorophyll *a* concentrations within the euphotic zone: A ubiquitous relationship. *Limnology and Oceanography*, 45, 1130-1143.

Twardowski, M.S., Boss, E., Macdonald, J.B., Pegau, W.S., Barnard, A.H., and Zaneveld, J.R.V. (2001), A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. *Journal of Geophysical Research*, 106, 14129-14142.

Wang, M.H., Knobelspiesse, K.D., and McClain, C.R. (2005), Study of the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) aerosol optical property data over ocean in combination with the ocean color products. *Journal of Geophysical Research*, 110, D10S06, doi:10.1029/2004JD004950.

Werdell, P.J. and Bailey, S.W. (2002), *The SeaWiFS Bio-optical Archive and Storage System (SeaBASS): Current Architecture and Implementation*. NASA Tech. Memo. 2002-211617, Greenbelt: NASA Goddard Space Flight Center, pp. 45.

Werdell, P.J. and Roesler, C.S. (2003), Remote assessment of benthic substrate composition in shallow waters using multispectral reflectance. *Limnology and Oceanography*, 48, 547-556.

Werdell, P.J., Bailey, S., Fargion, G., Pietras, C., Knobelspiesse, K., Feldman, G., and McClain, C. (2003), Unique data repository facilitates ocean color satellite validation. *EOS Transactions AGU*, 84, 38, 379.

Woźniak, S.B., and Stramski, D. (2004), Modeling the optical properties of mineral particles suspended in seawater and their influence on ocean reflectance and chlorophyll estimation from remote sensing algorithms. *Applied Optics*, 43, 3489-3503.

Yan, B., Stamnes, K., Toratani, M., Li, W., and Stamnes, J.J. (2002), Evaluation of a reflectance model used in the SeaWiFS ocean color algorithm: implications for chlorophyll concentration retrievals. *Applied Optics*, 41, 6243-6259.

Yentsch, C.S. (1960), The influence of phytoplankton pigments on the colour of seawater. *Deep Sea Research*, 7, 1-9.

Zibordi, G. and Ferrari, G.M. (1995), Instrument self-shading in underwater optical measurements: experimental data. *Applied Optics*, 34, 2750-2754.

8. TABLES

Table 1. Notations and descriptions of relevant geophysical data products.

abbreviation	units	description
$C_a(z)$	mg m^{-3}	concentration of chlorophyll *a* at depth *z*
$E_d(\lambda,z)$	$\mu W\ cm^{-2}\ nm^{-2}$	spectral downwelling irradiance at depth *z*
$E_s(\lambda)$	$\mu W\ cm^{-2}\ nm^{-2}$	spectral downwelling surface irradiance
$K_{rs}(\lambda)$	m^{-1}	spectral remote sensing diffuse attenuation coefficient
$L_u(\lambda,z)$	$\mu W\ cm^{-2}\ nm^{-2}\ sr^{-1}$	spectral upwelling radiance at depth *z*
$L_w(\lambda)$	$\mu W\ cm^{-2}\ nm^{-2}\ sr^{-1}$	spectral surface water-leaving radiance
$R_{rs}(\lambda)$	sr^{-1}	spectral surface remote-sensing reflectance

Table 2. Data sources and providers. For brevity, only the principal contributing investigator is listed. In the *AOP* column, *A* indicates above-water radiometry and *B* in-water radiometry. In the *CHL* column, *F* indicates fluorometric methods and *H* high-performance liquid chromatography. In the *exceptions* column, *K* indicates that the data contributor provided the downwelling diffuse attenuation coefficients and *S* indicates that the data contributor applied an instrument self-shading correction.

experiment	investigator	location	dates	N	AOP	CHL	exceptions
ACE-Asia	G. Mitchell	Japan & Western Pacific	Mar-Apr 01	36	B	FH	
Aerosols / INDOEX	G. Mitchell	Atlantic & Indian Oceans	Jan-Mar 99	50	B	F	
Akwanavt	O. Kopelevich	Black Sea	Oct 97	5	A	F	
AMLR	G. Mitchell	Weddell Sea	Feb 00-02	68	B	FH	
AMT	S. Hooker	Meridional Atlantic Ocean	variable 95-99	279	B	H	K
BBOP	D. Siegel	Sargasso Sea	monthly 94-03	104	B	F	
BENCAL	S. Hooker	West South Africa	Oct 02	47	B	H	K
Biocomplexity	L. Harding	Chesapeake Bay	Apr,Jul,Oct 01-03	73	B	F	S
BOUSSOLE	S. Hooker	Mediterranean Sea	Jul 01	3	B	H	K
CalCOFI	G. Mitchell	California Current	quarterly 93-03	223	B	FH	
CARIACO	F. Muller-Karger	Cariaco Basin, Venezuela	monthly 98-04	48	B	F	
CaTS	J. Corredor	Caribbean Sea	monthly 98-01	12	B	F	
Chesapeake Light Tower	G. Cota	Chesapeake Bay	Mar 00	5	A	F	
COASTAL NASA	S. Hooker	Caribbean Sea	Feb-Mar 00-01	40	B	H	K
COJET	R. Arnone, R. Stumpf	Mississippi Sound	variable 01-02	124	A	FH	
EcoHAB	K. Carder	West Florida Shelf	monthly 99-01	218	A	F	
FRONT	R. Morrison	East Long Island Sound	Dec 00, Oct 02	3	B	F	
GasEx II	F. Chavez	East Equatorial Pacific	Feb-Mar 01	11	B	F	K
Gulf of Mexico	F. Muller-Karger	Gulf of Mexico	Jun-Jul 97	4	B	F	
HIVE	D. Eslinger	Prince William Sound	Aug-Sep 98	11	B	F	
Horn Island	R. Arnone, R. Stumpf	Mississippi Sound	Sep 03	12	A	FH	
IOFFE	O. Kopelevich	Meridional Atlantic Ocean	Oct 01 - Apr 02	138	A	FH	
JGOFS Arabian Sea	J. Mueller	Arabian Sea	Mar,Nov,Dec 95	69	B	F	
JGOFS Sea of Japan	G. Mitchell	Japan & Western Pacific	Jun-Jul 99	25	B	FH	
JGOFS Southern Ocean	G. Mitchell	Ross Sea	Nov 97, Jan-Feb 98	41	B	F	
Lake Bourne NRL	R. Arnone, R. Stumpf	Mississippi Sound	Apr-Sep 01	10	A	FH	
LEO	R. Arnone, R. Stumpf	Northeastern Atlantic	Jul-Aug 01	10	A	FH	
LMER-TIES Chesapeake	L. Harding	Chesapeake Bay	Apr,Jul,Oct 95-00	172	B	F	S
LTER PAL	R. Smith	West Antarctic Peninsula	variable 91-99	1005	B	F	
LTER SBC	D. Siegel	Santa Barbara, California	variable 01-02	13	B	FH	
MANTRA PIRANA	A. Subramaniam	West Equatorial Atantic	Feb,Jun 01	20	B	FH	
MASS BAY	A. Subramaniam	Massachussetts Bay	Jul 02	7	B	H	
MOCE	D. Clark	Hawaii	Jan-Feb 98	20	B	H	
Monterey Bay NRL	R. Arnone, R. Stumpf	Monterey Bay	Apr 03	52	A	FH	
NOAA CSC	M. Culver	Northeastern Atlantic	variable 96-99	75	B	F	
OceanLIDAR	M. Lewis	Equatorial Pacific	variable 97-99	31	B	FH	
OMEXII	T. Smyth	Northwestern Atlantic	Jun-Jul 98	9	B	F	
ONR Chesapeake	L. Harding	Chesapeake Bay	variable 96-98	78	B	F	S
ORCA	G. Cota	Canadian Arctic	variable 97-00	22	B	F	
ORINOCO	F. Muller-Karger	Orinoco River, Venezuela	variable 98-00	17	B	F	
Pamlico	R. Arnone, R. Stumpf	Pamlico Sound	May,Jul,Oct 01	10	A	FH	
Plumes and Blooms	D. Siegel	Santa Barbara, California	monthly 00-03	159	B	FH	
PROSOPE	S. Hooker	Mediterranean Sea	Sep-Oct 99	12	B	H	
ROAVERRS	K. Arrigo	Ross Sea	Dec 97 - Jan 98	8	B	F	
Scotia Prince Ferry	W. Balch	Gulf of Maine	May-Oct 99-03	51	A	F	

TAO MBARI	F. Chavez	Equatorial Pacific	variable 97-01	19	B	F
WHOI Active Fluorescence	R. Morrison	Northeastern Atlantic	Jun 01	8	B	FH
WHOI Photochemistry	R. Morrison	Northeastern Atlantic	Jul 02	18	B	F

Table 3. Distribution of radiometric and phytoplankton pigment data archived in SeaBASS as a function of instrument and sampling method. For this analysis, as described in section 2.3.3, the volume of flow-through phytoplankton pigment concentrations was significantly reduced via the limitation of considered observations to satellite overpass times.

radiometry		phytoplankton pigments	
type	percentage	type	percentage
in-water	53.3	discrete HPLC	13.3
above-water	46.7	discrete fluorometry	27.0
		profiled fluorometry	14.3
		flow-through fluorometry	45.4
total observations	15,400	total observations	32,094

Table 4. Descriptions of the NOMAD binary flagging system. While only radiometric and phytoplankton pigment observations are described in this article, flags for the expanded version of the system (to include aerosol optical depths and inherent optical properties) are also listed. The *usage* field describes the purpose of the flag: *D* for those describing available data; *I* for those describing instruments; and *P* for those describing data processing. An example flag for a data record accumulated in this article would be 19141 {= 2^0 (AOP) + 2^2 (HPLC) + 2^6 (KD) + 2^7 (CAST) + 2^9 (VSB) + 2^{11}(INT_HPLC) + 2^{14}(ES)}, suggesting that water-leaving radiances were acquired from a profile processed internally, and are accompanied by coincident reference surface irradiances and depth-integrated high performance pigment chromatography chlorophyll measurements.

bit	abbreviaton	usage	description
0	AOP	D	radiometry, $L_w(\lambda)$ or $R_{rs}(\lambda)$, available (*always set*)
1	CHL	D	fluorometrically-derived C_a available
2	HPLC	D	HPLC-derived C_a available
3	AOT	D	aerosol optical depths, $\tau_a(\lambda)$, available
4	A	D	absorption coefficients, $a(\lambda)$, available
5	BB	D	backscattering coefficients, $b_b(\lambda)$, available
6	KD	D	diffuse downwelling attenuation coefficient, $K_{rs}(\lambda)$, available
7	CAST	I	data from radiometric or pigment depth profile
8	SPEC	I	data from laboratory spectrophotometry
9	VSB	P	$L_w(\lambda)$ processed using OBPG software
10	INT_CHL	P	depth-integrated (optically weighted) fluorometric C_a
11	INT_HPLC	P	depth-integrated (optically weighted) HPLC-derived C_a
12	SHADE	P	instrument self-shading correction applied to $L_u(\lambda,0^-)$
13	FQ	P	*f*/*Q* correction applied to $L_w(\lambda)$
14	ES	I	$E_s(\lambda)$ available from reference instrument
15	RRS	I	$L_w(\lambda)$ estimated from $R_{rs}(\lambda)$
16	HYPER	I	hyperspectral observation of $L_w(\lambda)$ or $R_{rs}(\lambda)$

Table 5. The percentage use of the temporal thresholds applied in radiometric and pigment data merger.

	1 hr	1.25 hr	1.5 hr	2 hr	5 hr	8 hr
percentage	85	7	3	3	1	1

Table 6. The discrete wavelengths (in nm) assigned to exported water-leaving radiances, surface irradiances, and downwelling diffuse attenuation coefficients. The *frequency* column provides the frequency of occurrence (in percent) of water-leaving radiance spectra in NOMAD that include each wavelength. The *heritage* column lists satellite instruments that possess each wavelength, within ± 2-nm (with the exception of three MOS channels, indicated by *, where the range was extended to ± 4-nm). Note that three additional operational satellite instruments, OCI, OCM, and OSMI, have configurations similar to that of SeaWiFS, with the exception of absent 411-nm channels for OCI and OSMI.

wavelength (nm)	frequency (%)	heritage
405	2.5	MOS*
411	99.5	OCTS, SeaWiFS, MODIS, MERIS, VIIRS
443	99.9	CZCS, OCTS, MOS, SeaWiFS, MODIS, MERIS, VIIRS
455	15.8	
465	3.8	
489	100.0	OCTS, MOS*, SeaWiFS, MODIS, MERIS, VIIRS
510	75.8	SeaWiFS, MERIS
520	42.6	CZCS, OCTS, MOS
530	34.9	MODIS
550	21.7	CZCS, MODIS
555	70.0	SeaWiFS, VIIRS
560	21.7	MERIS
565	46.2	OCTS
570	18.9	MOS
590	13.5	
619	17.2	MOS*, MERIS
625	43.5	
665	59.0	MODIS, MERIS
670	27.4	CZCS, OCTS, SeaWiFS, VIIRS
683	45.5	MOS, MERIS

9. FIGURES

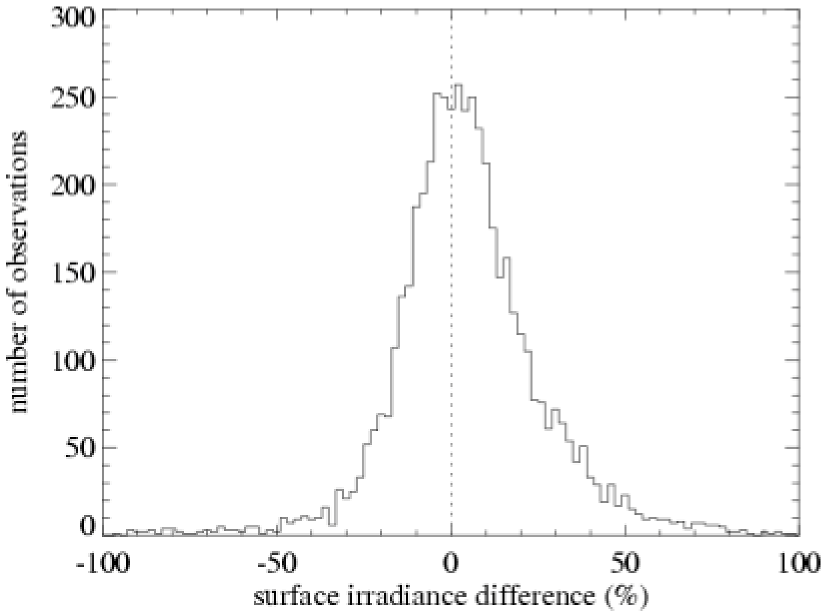

Figure 1. Frequency distribution of the differences between extrapolated surface irradiances, $E_d(0^+,490)$, and those measured using a reference deck cell, $E_s(490)$. Percent differences were calculated as $\{E_s(490) - E_d(0^+,490)\} / E_s(490) * 100\%$. As such, positive values indicate $E_s(490)$ values greater than $E_d(0^+,490)$.

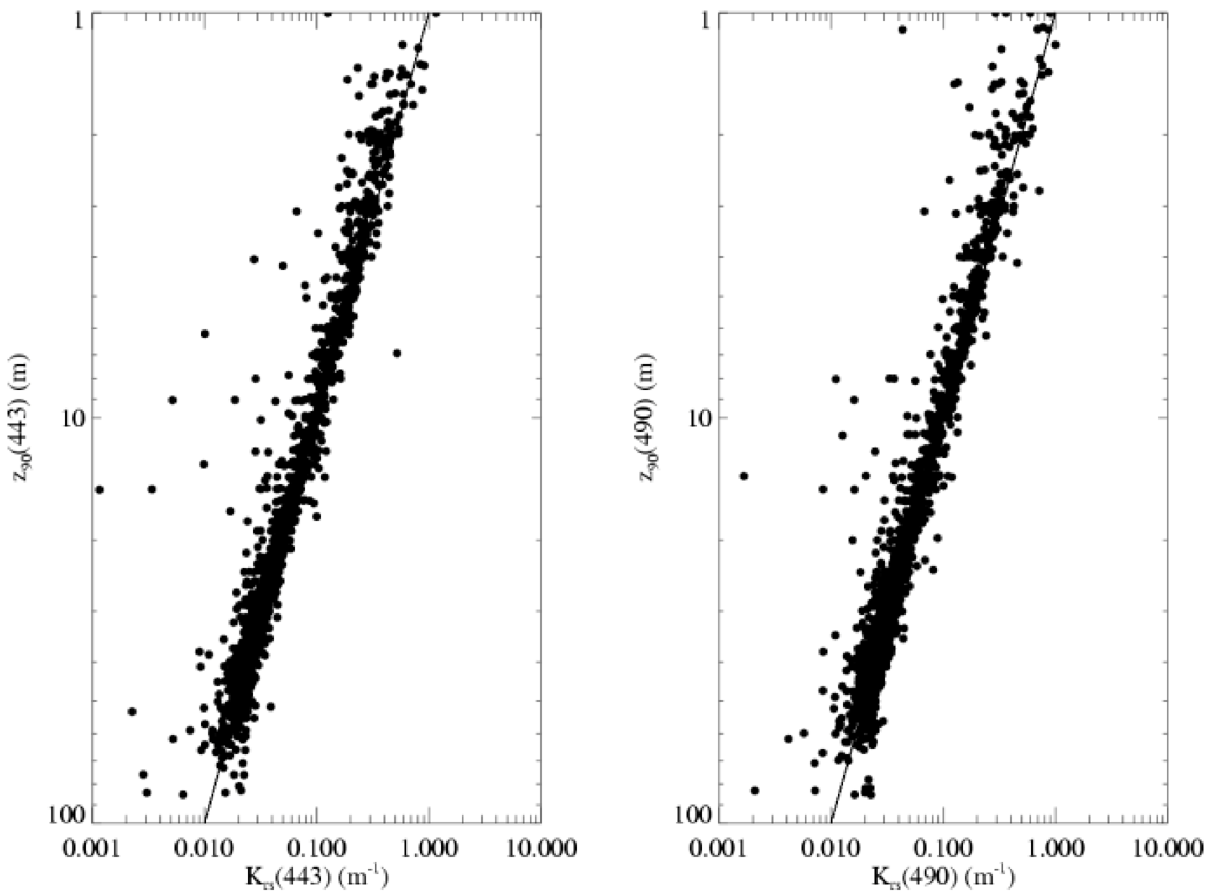

Figure 2. Remote-sensing diffuse attenuation coefficients, $K_{rs}(\lambda)$, as a function of depth of the first optical layer, $z_{90}(\lambda)$. For clarity, theoretical $z_{90}(\lambda)$ (= $K_{rs}(\lambda)^{-1}$) are shown as solid lines.

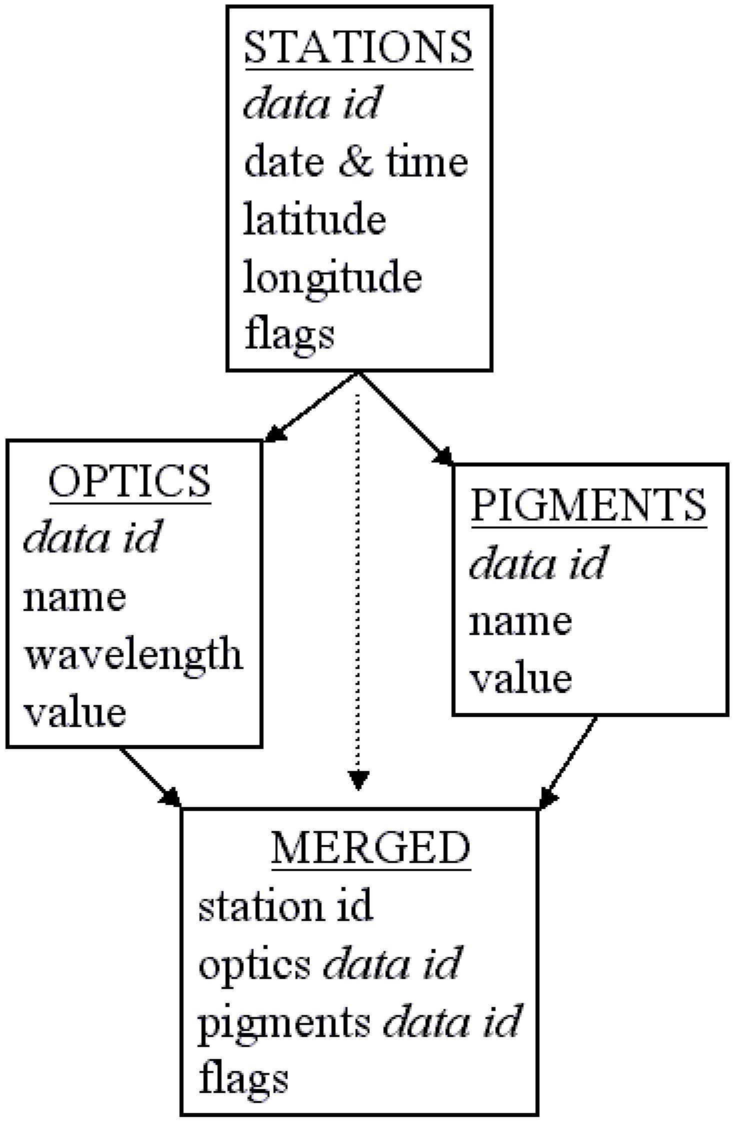

Figure 3. Generalized RDBMS table structure used to catalog coincident radiometric measurements and phytoplankton pigment concentrations. Metadata for each observation, such as time of data collection and station latitude and longitude, are stored in the *stations* table. The geophysical data values are stored in the *optics* and *pigments* tables, and are related back to the metadata via their assigned *data id*. The *data id*s of coincident observations are recorded in the *merged* database table. Export routines query the *merged* table to retrieve radiometric and pigment *data id*s, which are subsequently used to extract both geophysical data values and station metadata from the *optics*, *pigments*, and *stations* database tables.

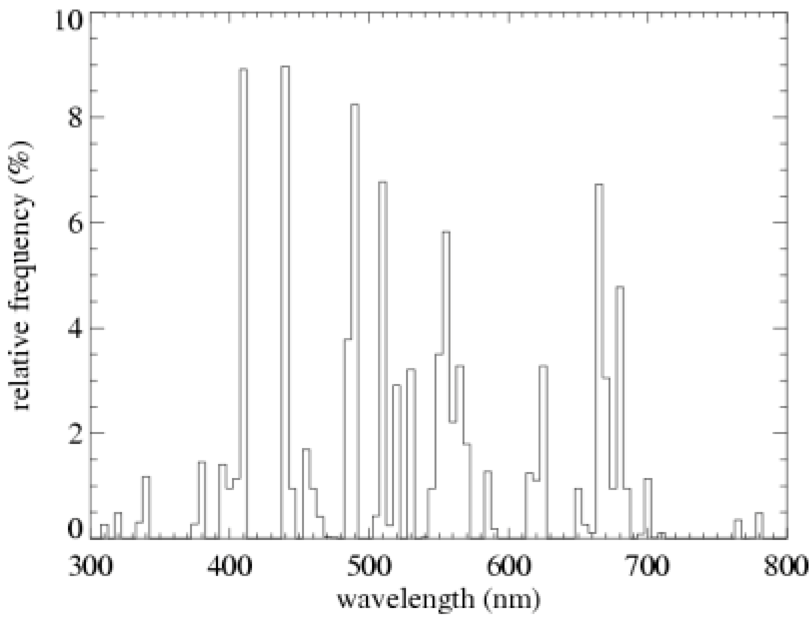

Figure 4. Relative spectral frequency of radiometry included in NOMAD, using 5-nm wide bars, defined as the ratio of the number of observations at a discrete wavelength to the total number of observations at all wavelengths, multiplied by 100 to generate units of percent.

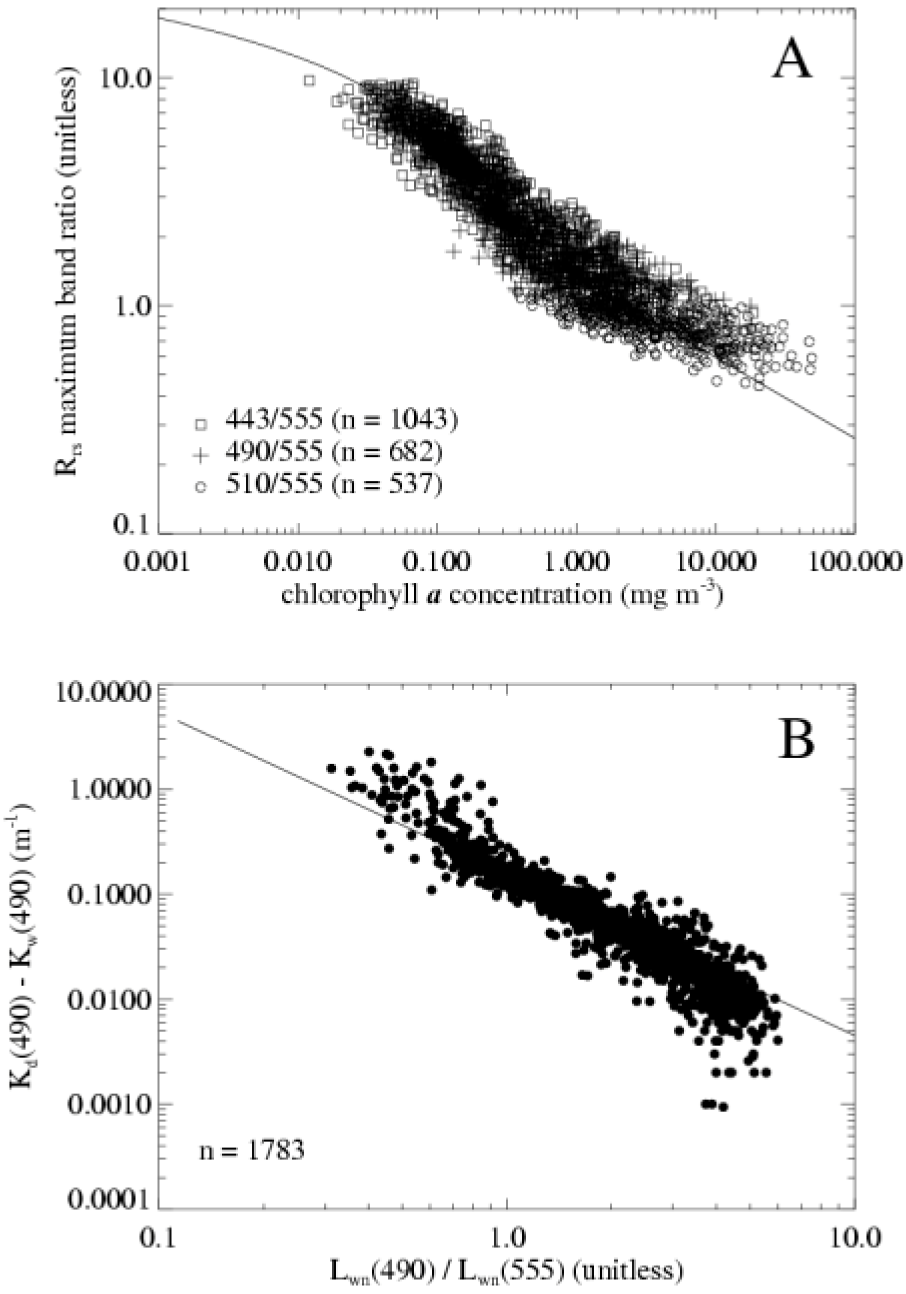

Figure 5. (a) An R_{rs} maximum band ratio, defined as the ratio of highest value of {$R_{rs}(443)$, $R_{rs}(490)$, and $R_{rs}(510)$} to $R_{rs}(555)$, as a function of chlorophyll *a* concentration. All chlorophyll *a* data were considered, but for a given station, HPLC data were selected if available. Different symbols indicate the different maximum bands. For reference, the solid line displays the ocean color chlorophyll algorithm OC4 version 4 (O'Reilly et al. 2000). (b) The ratio of $L_{wn}(490)$ to $L_{wn}(555)$ as a function of $K_d(490)$, less the pure-water diffuse attenuation coefficient at 490 nm. For reference, the solid line displays the operational SeaWiFS K490 algorithm (Mueller 2000).

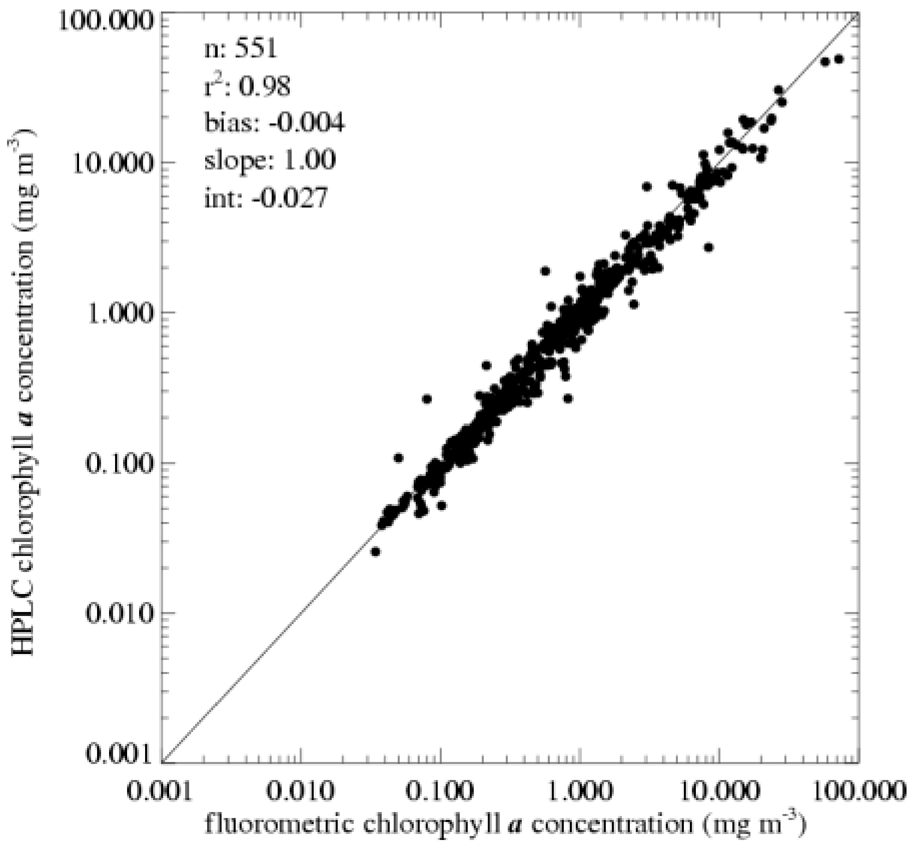

Figure 6. A comparison of coincidentally observed fluorometric- and HPLC-derived chlorophyll *a* concentrations. The data were transformed prior to regression analysis to account for their lognormal distribution. The comparison illustrates an overestimation of fluorometric C_a at very high concentrations ($C_a > 10$ mg m^{-3}), although only a few samples were available in this range. Recent studies suggest that the uncertainty in fluorometrically-derived C_a varies asymmetrically, as a function of geography, phytoplankton pigment population, and time of year (Hoepffner and Sathyendranath 1992, Bianchi et al. 1995, Tester et al. 1995).

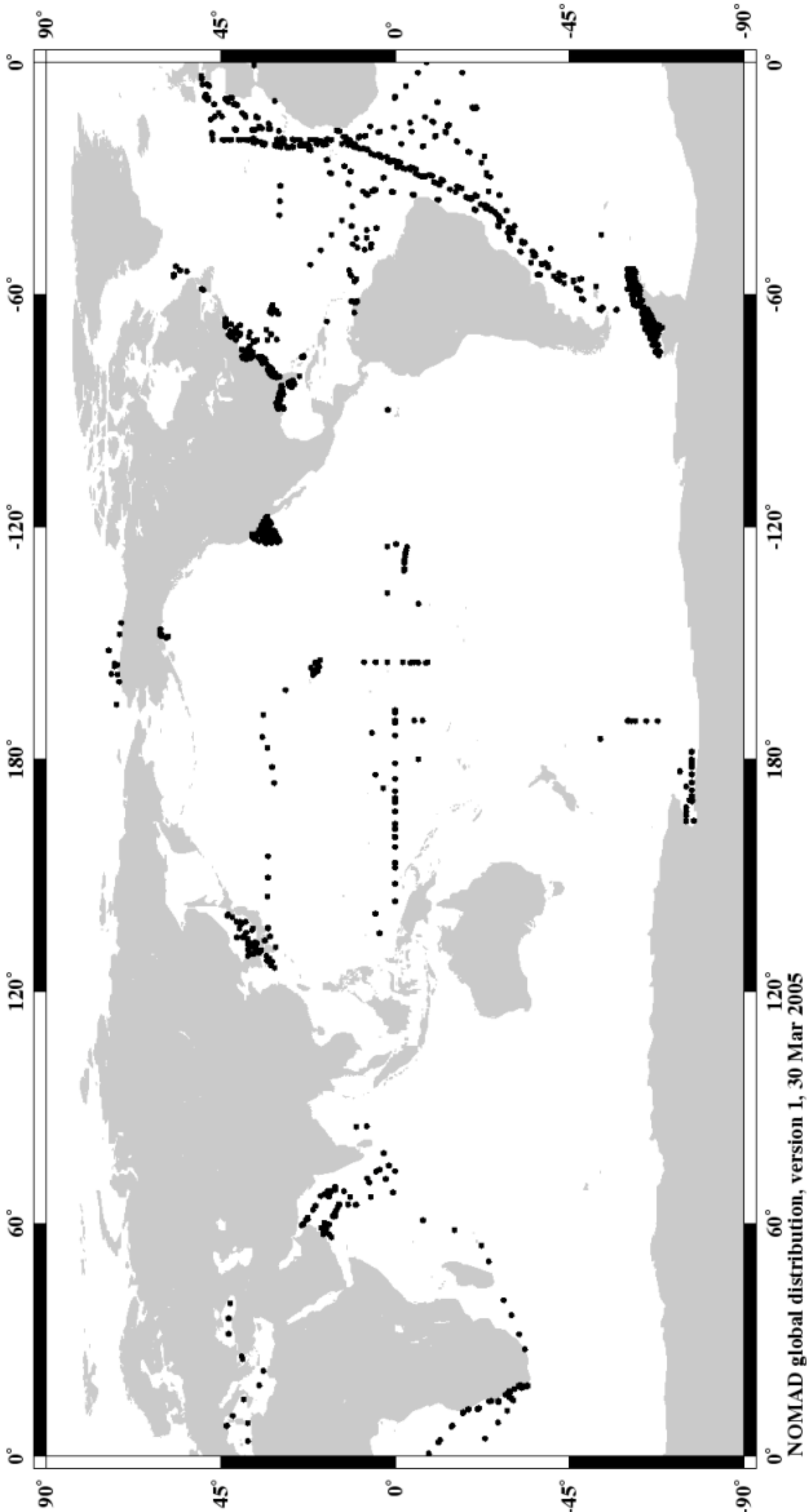

Figure 7. The global distribution of the NOMAD data set.

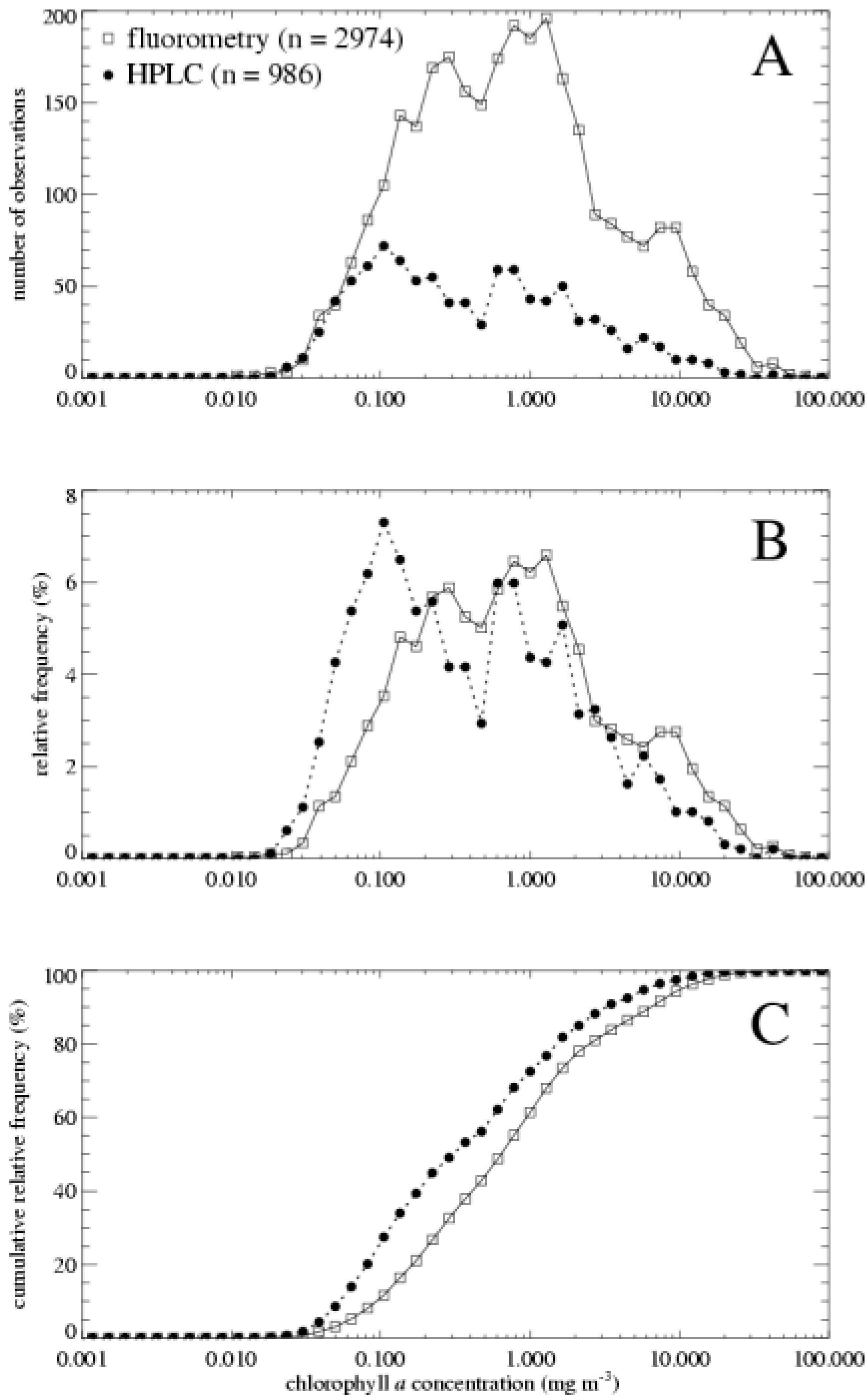

Figure 8. Frequency distributions of fluorometrically- and HPLC-derived chlorophyll *a* concentrations in NOMAD.

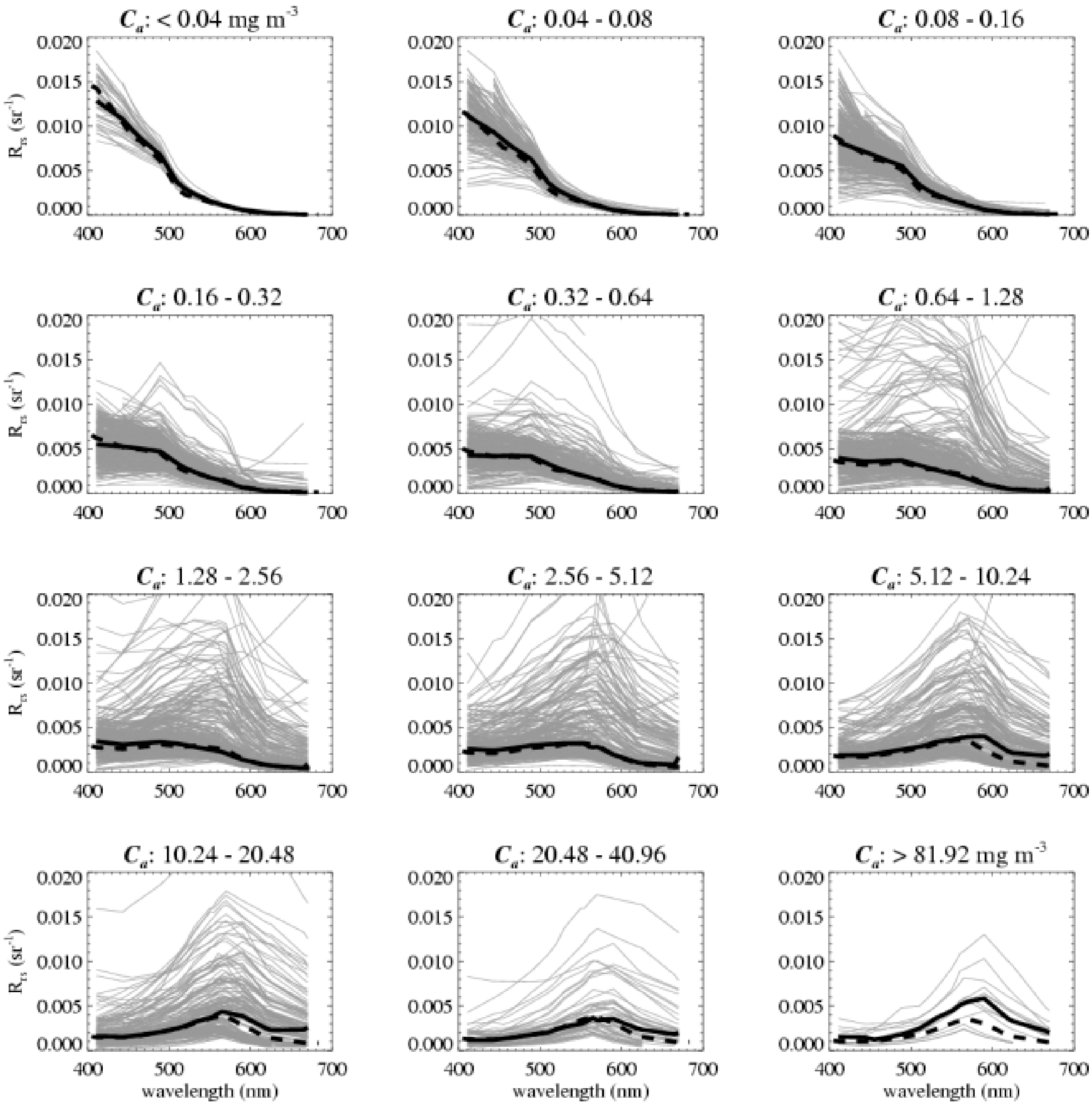

Figure 9. The NOMAD $R_{rs}(\lambda)$ spectra displayed for geometrically increasing chlorophyll *a* concentration ranges, as indicated by the title of each plot. Solid black lines indicate the median spectrum for each chlorophyll *a* range. Dashed black lines display a theoretical clear-water spectrum for the median chlorophyll *a* concentration for each range. The latter spectra were derived using the Case-1 approximations described in Morel and Maritorena (2001).

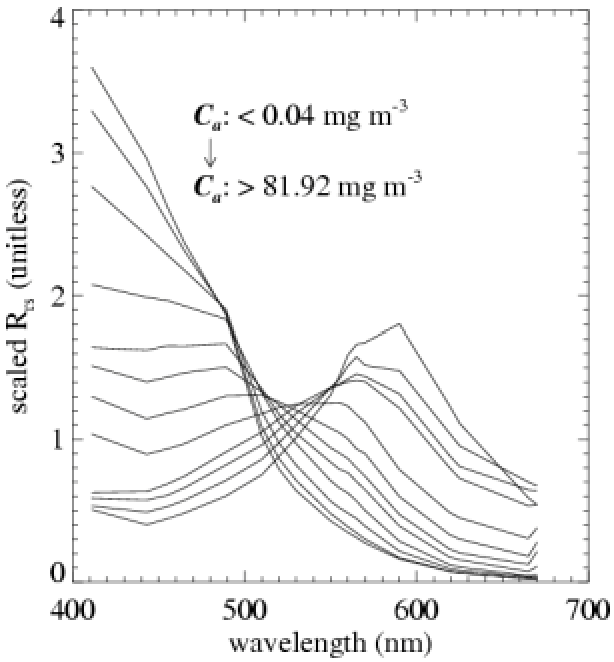

Figure 10. Average, scaled remote-sensing reflectance spectra for the 12 chlorophyll *a* concentration ranges described in Figure 9. Each spectra was first normalized to its mean, where *scaled* $R_{rs}(\lambda)$ = *observed* $R_{rs}(\lambda)$ / *average of observed* $R_{rs}(400)$ *to* $R_{rs}(700)$. All scaled $R_{rs}(\lambda)$ within a given chlorophyll *a* range were then averaged.

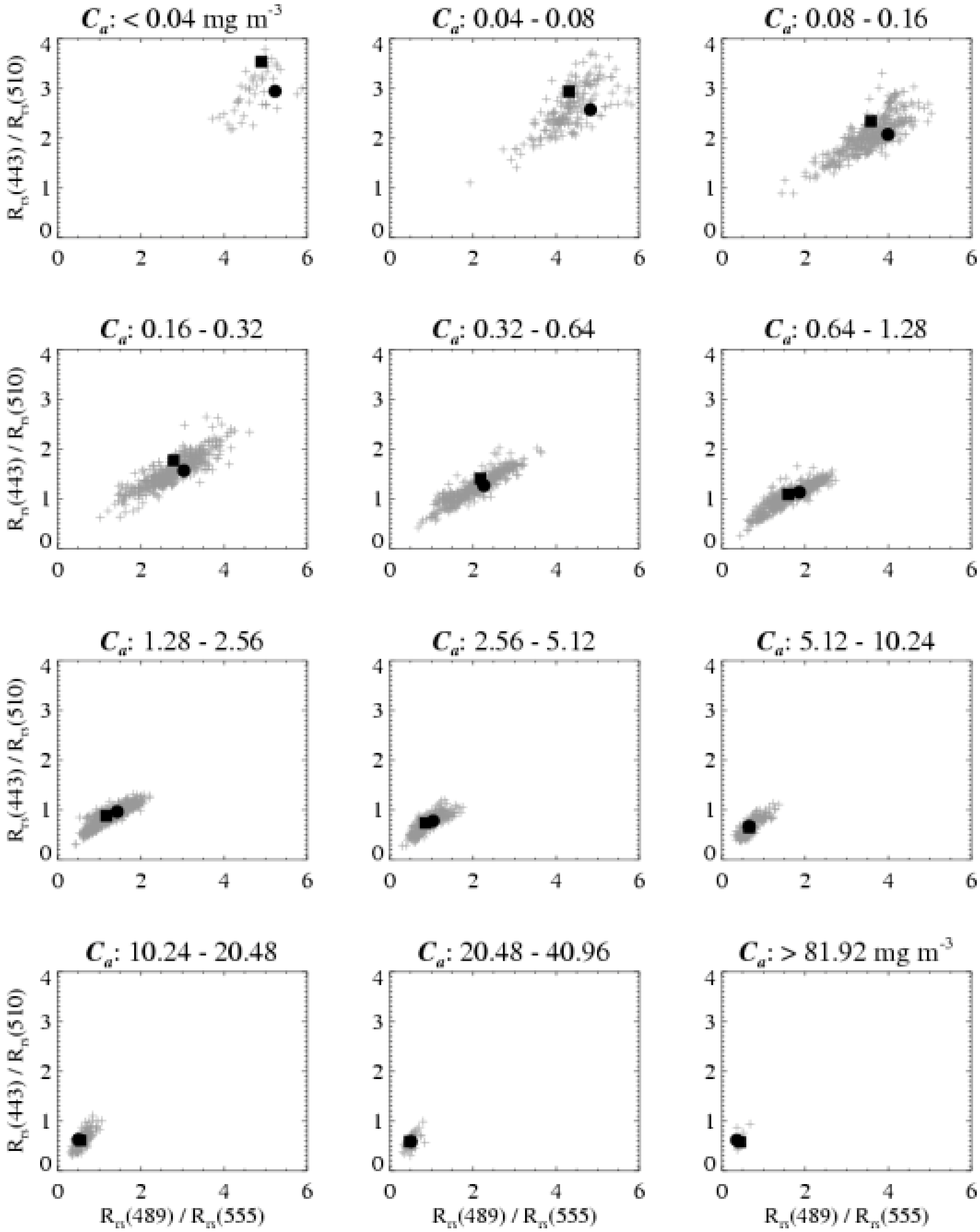

Figure 11. NOMAD $R_{rs}(\lambda)$ band ratios displayed for geometrically increasing chlorophyll *a* concentration ranges, as indicated by the title of each plot. Solid black circles indicate the median band ratios for each chlorophyll *a* range. Solid black squares display a theoretical clear-water band ratio for the median chlorophyll *a* concentration for each range. The latter ratios were derived using the Case-1 approximations described in Morel and Maritorena (2001). All ratios are unitless.

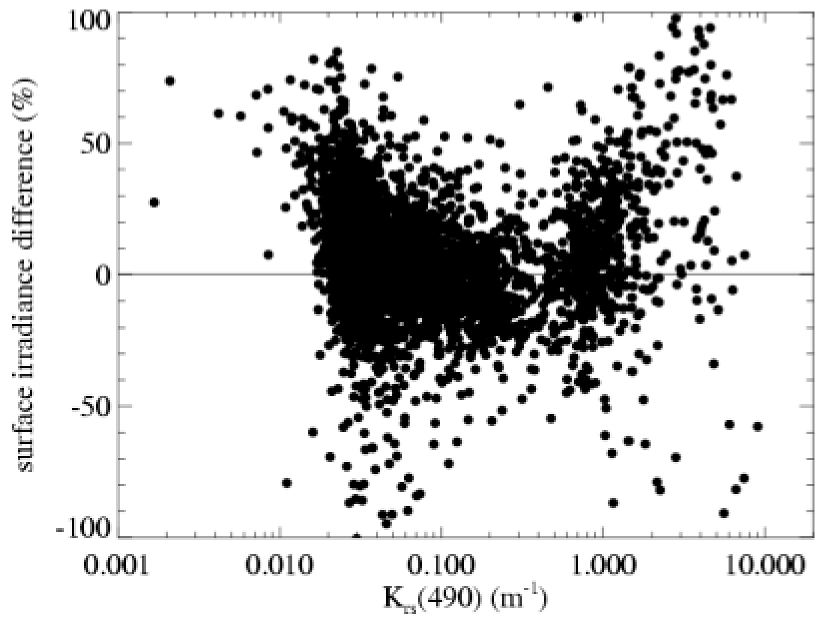

Figure 12. Differences between extrapolated surface irradiances, $E_d(0^+,490)$, and those measured using a reference deck cell, $E_s(490)$. Percent differences are displayed as a function of the remote-sensing diffuse attenuation coefficient, $K_{rs}(490)$. Percent differences were calculated as $\{E_s(490) - E_d(0^+,490)\} / E_s(490) * 100\%$. As such, positive values indicate $E_s(490)$ values greater than $E_d(0^+,490)$. Data from 4884 radiometric observations were considered. The simple linear correlation coefficient, r, is 0.008.

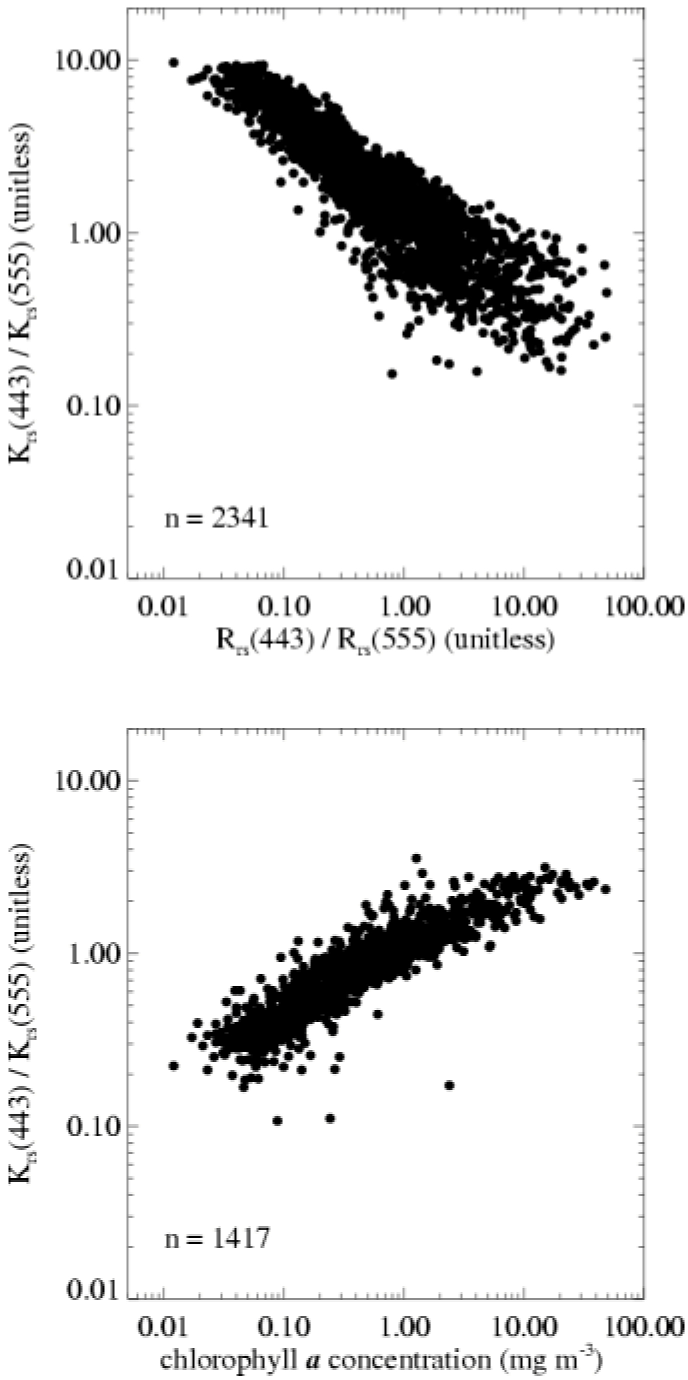

Figure 13. The ratio of $K_{rs}(443)$ to $K_{rs}(555)$ as a function of the ratio of $R_{rs}(443)$ to $R_{rs}(555)$ (top panel) and chlorophyll *a* concentration (bottom panel). For the latter, all chlorophyll *a* data were considered, but for a given station, HPLC data were selected if available.