Methodology for Evaluation of Ocean Color Products

Bryan A. Franz, Ewa Kwiatkowska, Sean W. Bailey
NASA Ocean Biology Processing Group

Introduction

The Ocean Biology Processing Group (OBPG) at NASA/GSFC is currently tasked with the production of ocean color and related products from the SeaWiFS, MODIS/Aqua, OCTS, and CZCS spaceborne sensors. This poster describes some of the standard methods used by the OBPG to evaluate the oceanic optical properties derived from these instruments. The analyses described here are performed routinely for standard production products, to assess and track data quality. The same procedures are also used to evaluate proposed changes in processing algorithms or calibration, prior to any full-mission reprocessing. In this capacity, the analyses serve to verify the implementation of proposed changes, and to provide quantitative feedback as to the impact of those changes on field-data comparisons, sensor-to-sensor agreement, and temporal and spatial stability in derived product retrievals.

When evaluating multiple proposed changes to standard processing, the OBPG philosophy is to make incremental changes and perform a systematic and comprehensive set of analyses at each step in the progression toward a final reprocessing configuration. In this way, the effect of each change in processing algorithm or calibration can be predicted and verified, thus minimizing chances for implementation error and enhancing understanding of the effect of each change on the global data set.

Any evaluation of instrument calibration or processing algorithm changes is normally preceded by a reevaluation of the vicarious calibration (Eplee, 2003). This effectively removes any bias on the mission-mean normalized water-leaving radiance retrievals at the Marine Optical Buoy (MOBY) vicarious calibration site (Clark et al., 2001). When comparing products from different sensors, any algorithm changes that are applicable to both sensors are applied equally, and both sensors are vicariously recalibrated at MOBY.

For each proposed change in processing algorithm or calibration:

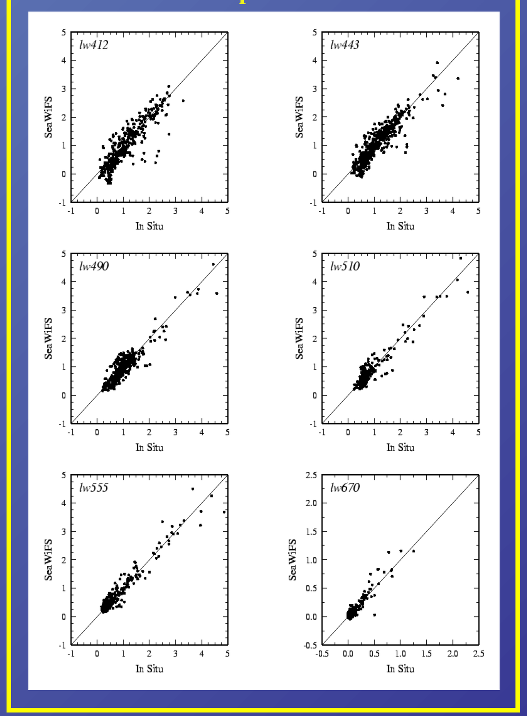
1) Recompute vicarious calibration to MOBY
2) Reprocess entire data set through Level-3 global products, or a mission-long temporal subset
3) Evaluate consistency in global data set over time
4) Compare global and regional trends relative to other missions
5) Quantify systematic artifacts (e.g., striping)

In the sections that follow, some details are provided on the analysis methods used for these evaluations. Some sample results are provided to illustrate the concepts, but these results do not represent the current operational quality of the OBPG products. Current results can be found at http://oceancolor.gsfc.nasa.gov/.

I. Comparison with *In Situ* Measurements

The primary mechanism for assessing the quality of retrieved ocean color properties is through comparison with ground-truth measurements. A full description of the *in situ* match-up process and current operational results is available at http://seabass.gsfc.nasa.gov/matchup_results.html. Figure 1 provides an example of water-leaving radiance comparisons between in situ measurements and a set of satellite retrievals (SeaWiFS in this case). The comparison to field data provides a measure of the quality of satellite retrievals for a single, instantaneous observation at Level-2. It is most useful as a tool for identifying systematic bias in satellite retrievals. The scatter in these comparisons may be due to a variety of sources, including the satellite calibration and retrieval algorithms, the quality of the *in situ* data set, the possibility that the field measurement is not representative of the surrounding region encompassed by the satellite observational footprint, and the possibility that the optical properties of the water have changed between the time of the field measurement and the time of the satellite observation (typically within 2-hours). It should also be recognized that the temporal and geographic distribution of the *in situ* data set (Werdell et al., 2005) is limited. The *in situ* match-ups are generally not sufficient for assessing the quality of satellite remote sensed ocean color data over the full range of geometries through which the spaceborne sensor views the earth, or over the full temporal and geographic distribution of the Level-3 products, nor do they account for the effects of temporal and spatial averaging or systematic errors associated with Level-3 masking decisions.

Figure 1
SeaWiFS Comparison to In Situ



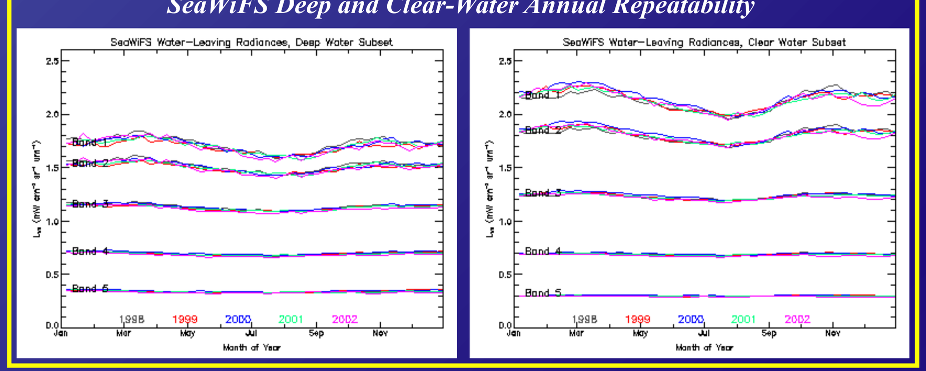
II. Temporal Trending and Annual Repeatability

The temporal analysis looks at long-term trends in Level-3 products and the consistency of those trends from year to year. It provides a standard mechanism for evaluating derived product and sensor stability, and it quantifies the relative impact of calibration and algorithm changes on global spatial scales and life-of-mission time-scales. The approach begins with standard Level-3 products. These products are global binned, multi-day averages at 4.6 or 9-km resolution, with bins distributed in an equal-area, sinusoidal projection (Campbell et al., 1995). The typical compositing period is 8-days, but for quick turn-around test processing the OBPG uses a temporal subset of the mission lifespan consisting of 4-day composites generated from the start of each consecutive 32-day period (i.e., 12.5% of the mission data set). The temporal subset is generated at 9-km resolution, and it can be processed within 1-day. The 4-day compositing period generally provides sufficient opportunity to observe most of the daylit side of the earth, including coverage in orbit and glint gaps.

From these global, multi-day composites, various geographic subsets of the filled bins are selected, and standard ocean products are spatially averaged and then trended with time. The analysis is focused on the trends in normalized water-leaving radiances (nLw), but trends in chlorophyll and atmospheric products are also evaluated. For bin selection and averaging, three global subsets are defined, corresponding to clear water, deep water, and coastal water. The deep water subset consists of all bins where water depth is greater than 1000 meters. Clear water is defined as deep water where the retrieved chlorophyll is less than 0.15 mg/m^3. Coastal water is defined as all bins where water depth is between 30 and 1000 meters, as defined by a shallow water mask and the deep water mask.

An example of this analysis is the SeaWiFS 5-year annual cycle for nLw shown in Figure 2. In the absence of any major geophysical events, we expect the trend in global deep-water or global clear-water nLw to repeat from year to year. Low-level differences may be due to geographic sampling biases or real geophysical changes, but on the large-scale these plots tell us that SeaWiFS products are self-consistent over time (i.e., there is no long-term drift).

Figure 2
SeaWiFS Deep and Clear-Water Annual Repeatability



III. Sensor-to-Sensor Temporal Comparison

This analysis looks at average values of sensor-to-sensor coincident Level-3 retrievals on global and regional spatial scales, and presents the results as a comparative time-series over the common mission lifespan. The goal is to produce a statistically rigorous comparison of equivalent ocean color products, to provide a quantitative assessment to the users of these data sets as to their relative agreement, and to evaluate the impact of calibration and algorithm changes on sensor-to-sensor consistency at global spatial scales and life-of-mission time scales. The ocean color products compared are the standard chlorophyll products derived from each mission data set, as well as the normalized water-leaving radiances in the four closest visible wavelengths. The equivalent wavelengths are listed in Table 1. The chlorophyll products compared are the standard "chlor_a" products produced for each sensor, which are all empirical max-band-ratio algorithms developed by O'Reilly (O'Reilly et al., 2000) from a common *in situ* archive. Additional insight into the atmospheric correction performance is gained by comparing aerosol optical thickness (AOT) retrievals and single-scattering epsilon of the near IR band pair (Gordon & Wang, 1994).

Table 1: *Comparable Sensor Products*

Band	SeaWiFS	MODIS	OCTS
nLw 1	412	412	412
nLw 2	443	443	443
nLw 3	490	488	490
nLw 4	510	531	520
nLw 5	555	551	565
nLw 6	670	667 & 678	670
Chlor_a	OC4v4	OC3M	OC4O
AOT	865	870	865
Epsilon	765/865	750/870	765/865

The comparative analysis begins with standard Level-3 products composited over a common time period (usually 4 or 8 days). All OBPG Level-3 ocean color products use the same, equal area binning approach (Cambell et al., 1995), but standard MODIS products are distributed at 4.6-km resolution while SeaWiFS is distributed at 9-km resolution. To allow for a direct, bin-for-bin comparison, the MODIS products are rebinned to the SeaWiFS 9-km resolution using standard binning algorithms. For quick turn-around test processing, the OBPG processes a temporal subset of the common mission lifespan. Again, the temporal subset consists of 4-day composites generated from the start of each consecutive 32-day period, generated at 9-km resolution. With Level-3 composited data products in an equivalent form, the data sets are further reduced to a set of common bins. This means that only those bins for which a retrieval exists for both sensors are included in subsequent averaging and trending. This is critical to the statistics, as some sensors show systematic geographic gaps in coverage, even after 8-days of compositing, and this can result in sampling bias if both sensors are not equivalently masked. Finally, with the products in common bin form, the data are divided into geographic subsets for averaging and trending. The process flow for Level-3 comparison of MODIS to SeaWiFS is illustrated in Figure 3. The standard geographic subsets include the global deep, clear, and coastal-water subsets described in Section II, as well as a set of standard regions and a set of latitudinally distributed zones. Figure 4 presents a sample pair of MODIS and SeaWiFS deep-water subsetted chlorophyll images for one 8-day period in May of 2003, after mapping to a platte carre projection. The images show the geographic extent of the common-binned, deep-water subset. Such images are routinely generated to provide some insight into the qualitative agreement between the two sensors, but quantitative analysis is performed using the original, equal-area bin data.

Figure 3
Sensor-to-Sensor Level-3 Comparison Process

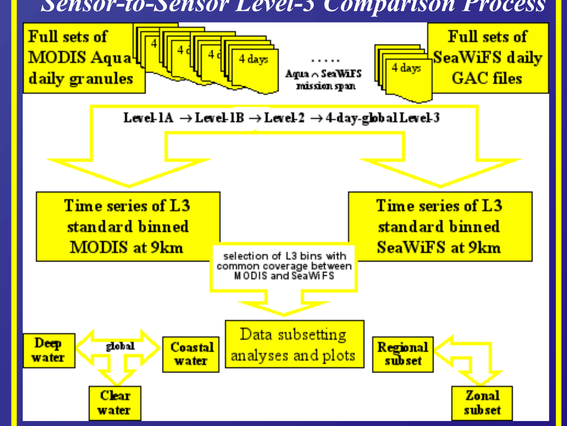
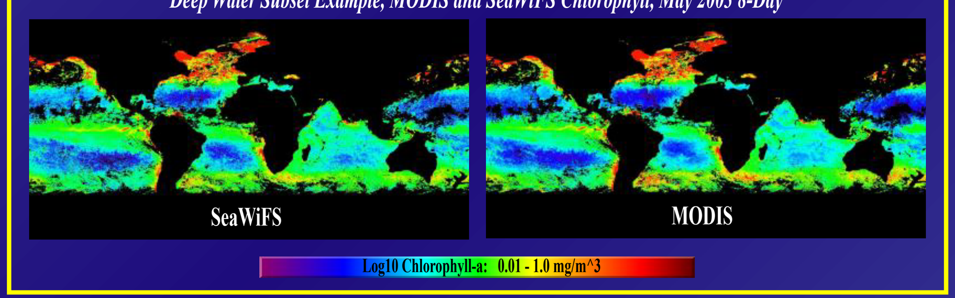
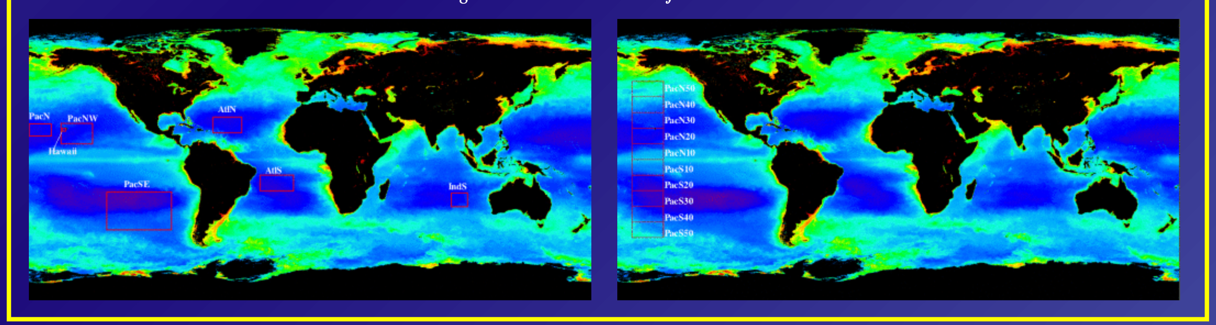


Figure 4
Deep Water Subset Example, MODIS and SeaWiFS Chlorophyll, May 2003 8-Day



In addition to the global analysis, the temporal trends are generated over regional and zonal subsets. The regions were chosen for homogeneity (Fougnie, et al., 2002), and they are all in relatively clear water. A region is also included at Hawaii, to verify performance at the point of vicarious calibration (Eplee et al., 2003), where calibration biases should be minimal. The regions are shown graphically in Figure 5. The zonal subsets were added to provide a systematic means for investigating latitudinally-dependent differences between the two sensors. These are also shown in Figure 5.

Figure 5
Regional and Zonal Subset Definitions



For each sensor Level-3 product, the common, filled bins associated with a particular subset are identified and used to compute the mean, standard deviation, and average observation time. Figure 6 shows an example of a typical trend plot derived from this analysis. For the plot on the left, the common MODIS and SeaWiFS bins for the deep-water subset were spatially averaged for each 8-day-binned water-leaving radiance product, and the resulting means were then plotted as a function of time. MODIS is shown as dashed lines. The colors indicate different bands, as summarized in Table 1. The plot on the right shows the same data as a ratio, with MODIS means normalized by SeaWiFS means.

Taken alone, the comparative temporal analysis can not be used to determine absolute error, since relative differences may be due to errors in either sensor data set, or real geophysical effects related disparity in observation times. However, when taken in concert with the self consistency analyses described in section II and the *in situ* comparisons of Section I, the sensor-to-sensor comparisons can serve to identify and isolate the likely cause for differences. An example of this is Figure 7, which shows water-leaving radiance trend results for the PacN50 zonal subset (40N-50N, 150W-170W) for two test cases. The plot on the left is before a correction was made to the MODIS/Aqua polarization sensitivity, while the plot on the right is after the correction.

Figure 6
SeaWiFS and MODIS Water-Leaving Radiances, Deep Water Subset

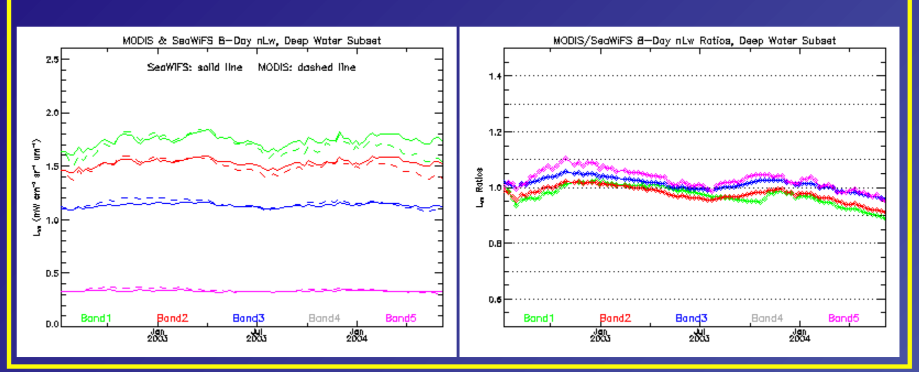
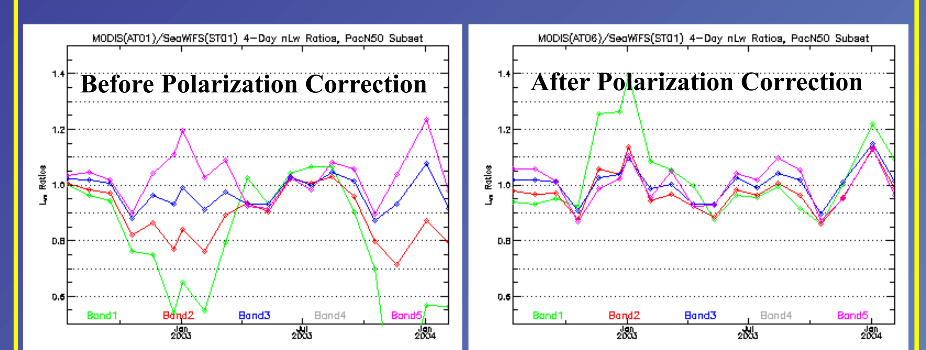


Figure 7
SeaWiFS and MODIS Water-Leaving Radiances, PacN50 Subset



V. Cross-Scan and Detector-Dependent Residuals

The final analysis seeks to quantify and track changes in residual cross-scan trends and, in the case of MODIS and OCTS, detector-to-detector relative differences (i.e., striping). The approach takes advantage of the fact that such Level-2 residual errors will tend to average-out over time and space. Software was developed to generate match-ups between Level-2 observations and Level-3 bins, where the Level-3 product is typically a 7-day mean at 9-km resolution, temporally centered on the date of the Level-2 granule. The software gathers all relevant information relating to the Level-2 to Level-3 match-up, including scan-pixel, detector number, and mirror side of the Level-2 observation. Match-ups for all granules collected over a complete day are screened, and those cases corresponding to deep, clear water (chlorophyll < 0.15 mg m-3) with minimal glint contamination are accepted. Standard binner masking is also employed, with the object being to obtain a large number of Level-2 to Level-3 match-ups from homogeneous, temporally stable waters, where the Level-3 retrieval is likely to be a good representation of what the Level-2 retrieval should be. The process is illustrated in Figure 8. The Level-2 to Level-3 ratios for each derived product can then be averaged within scan-pixel or detector number. Figure 9a shows an example of a cross-scan trend derived in this manner. The case shown is nLw at 443 for MODIS/Aqua. The dotted line in Figure 9a is the standard deviation within each scan-pixel mean. Figure 9b shows the same data as a function of detector number. These analyses are typically done for 6 dates covering the start and end of mission and a full set of solstice and equinox occurrences. In this way, the OBPG can quantify systematic artifacts in the Level-2 products, and track the consistency of such trends over time. The analysis is particularly useful for identifying problems with the instrument characterization, such as response-versus-scanangle (RVS) and detector relative calibration, and in testing the performance of correction algorithms which vary strongly with sensor view angle, such as bidirectional reflectance and polarization corrections.

Figure 8
Level-2 to Level-3 Match-up Process

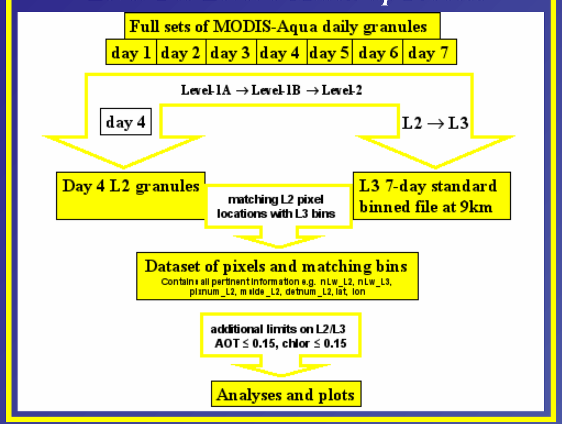
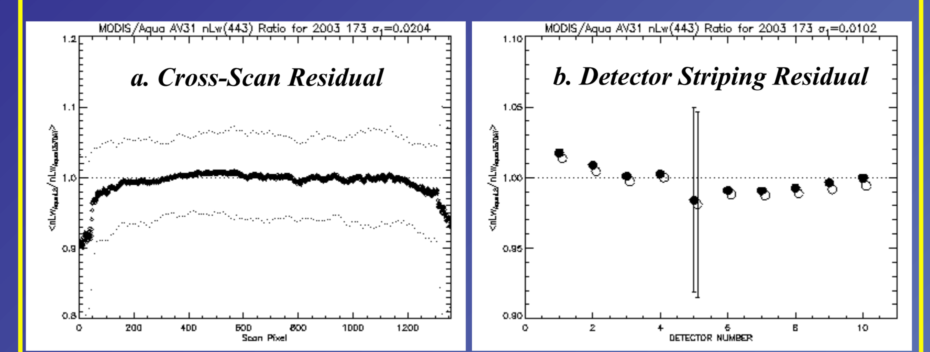


Figure 9
MODIS Cross-Scan and Detector-Dependent Residuals, 443nm



Summary

The methods described in this poster are employed by the OBPG to provide a standard and systematic process for the evaluation of proposed algorithm and sensor calibration changes, and to quantify the quality of the resulting ocean products. The analyses are focused on assessment of the apparent optical properties (i.e., normalized water-leaving radiances), which are the basis for most of the derived ocean product algorithms. The evaluation process has been designed to provide a full picture of the data set, from the individual retrieval bias to assessment of large-scale geographic and temporal stability. The results derived from this process help to verify the implementation of algorithm updates, quantify the impact of processing changes, and provide feedback for the ongoing instrument calibration and characterization efforts. The use of Level-2 match-ups to Level-3 means can provide a quantitative measure of mean residual artifacts such as detector striping and scan-edge discontinuities. When making comparisons between sensors, the application of common analysis methods and the restriction to common coverage (common bins) ensures that relative assessments of quality and temporal stability are fair and accurate.

References

Campbell, J.W., J.M. Blaisdell, and M. Darzi, 1995: Level-3 SeaWiFS Data Products: Spatial and Temporal Binning Algorithms. NASA Tech. Memo. 104566, Vol. 32, S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland

Clark, D. K., M. E. Feinholz, M. A. Yarbrough, B. C. Johnson, S. W. Brown, Y. S. Kim, and R. A. Barnes, 2001: Overview of the radiometric calibration of MOBY, Proc. Spie, 4483, 64-76.

Gordon, H.R. & M. Wang, 1994: Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm, Appl. Opt., vol. 33, pp. 443-452.

Eplee, R.E., Jr., R.A. Barnes, S.W. Bailey, and P.J. Werdell, 2003: Changes to the vicarious calibration of SeaWiFS. In: Patt, F.S., R.A. Barnes, R.E. Eplee, Jr., B.A. Franz, W.D. Robinson, G.C. Feldman, S.W. Bailey, P.J. Werdell, R. Frouin, R.P. Stumpf, R.A. Arnone, R.W. Gould, Jr., P.M. Martinolich, and V. Ransibrahmanakul, Algorithm Updates for the Fourth SeaWiFS Data Reprocessing, NASA Tech. Memo. 2003--206892, Vol. 22, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland.

Fougnie, B., P. Henry, A. Morel, D. Antoine, and F. Montagner, 2002: Identification and Characterization of Stable Homogeneous Oceanic Zones: Climatology and Impact on In-Flight Calibration of Space Sensors over Rayleigh Scattering. Ocean Optics XVI, Santa Fe, NM, November 18-22, 2002.

O'Reilly, J., S. Maritorena, M. O'Brien, D. Siegel, D. Toole, D. Menzies, R. Smith, J. Mueller, B. Mitchell, M. Kahru, F. CHavez, P. Strutton, G. Cota, S. Hooker, C. McClain, K. Carder, F. Muller-Karger, L. Harding, A. Magnuson, D. Phinney, G. Moore, J. Aiken, K. Arrigo, R. Letelier and M. Culver (2000). SeaWiFS postlaunch technical report series, Volume 11, SeaWiFS postlaunch calibration and validation analyses, Part 3, NASA Technical Memorandum.

Werdell, P.J. & S.W. Bailey, 2005: An improved *in situ* bio-optical data set for ocean color algorithm development and satellite data product validation, to be submitted to Remote Sensing of Environment.

This information is also available at http://oceancolor.gsfc.nasa.gov*/DOCS/methods/*