Chapter 6

Merging Ocean Color Data From Multiple Missions

Watson W. Gregg

NASA Goddard Space Flight Center, Greenbelt, Maryland

6.1 INTRODUCTION

Oceanic phytoplankton may play an important role in the cycling of carbon on the Earth, through the uptake of carbon dioxide in the process of photosynthesis. Although they are ubiquitous in the global oceans, their abundances and dynamics are difficult to estimate, primarily due to the vast spatial extent of the oceans and the short time scales over which their abundances can change. Consequently, the effects of oceanic phytoplankton on biogeochemical cycling, climate change, and fisheries are not well known.

In response to the potential importance of phytoplankton in the global carbon cycle and the lack of comprehensive data, NASA and the international community have established high priority satellite missions designed to acquire and produce high quality ocean color data (Table 6.1). Ten of the missions are routine global observational missions: the Ocean Color and Temperature Sensor (OCTS), the Polarization and Directionality of the Earth's Reflectances sensor (POLDER), Sea-viewing Wide Field-of-view Sensor (SeaWiFS), Moderate Resolution Imaging Spectrometer-AM (MODIS-AM), Medium Resolution Imaging Spectrometer (MERIS), Global Imager (GLI), MODIS-PM, Super-GLI (S-GLI), and the Visible/Infrared Imager and Radiometer Suite (VIIRS) on the NPOESS Preparatory Project (NPP) and the National Polar-orbiting Operational Environmental Satellite System (NPOESS). In addition, there are several other missions capable of providing ocean color data on smaller scales. Most of these missions contain the spectral band complement considered necessary to derive oceanic chlorophyll concentrations and other related parameters. Many contain additional bands that can provide important ancillary information about the optical and biological state of the oceans.

In previous efforts, we have established that better ocean coverage can be obtained in less time if the data from several missions are combined (Gregg et al., 1998; Gregg and Woodward, 1998). In addition to improved coverage, data can be taken from different local times of day if the missions are placed in different orbits, which they are. This can potentially lead to information on diel variability of phytoplankton abundances. Since phytoplankton populations can increase their biomasses by more than double in a single day under favorable circumstances (Eppley, 1972; Doney at el., 1995), observations of their abundances at different times within a single day would be useful.

We proposed to investigate, develop, and test algorithms for merging ocean color data from multiple missions. We seek general algorithms that are applicable to any retrieved Level-3 (derived geophysical products mapped to an Earth grid) ocean color data products, and that maximize the amount of information available in the combination of data from multiple missions. Most importantly, we will investigate merging methods that produce the most complete coverage in the smallest amount of time, nominally, global daily coverage. We will emphasize 3 primary methods: 1) averaging, 2) blending, and 3) statistical interpolation.

6.2. RESEARCH ACTIVITIES

We investigated a set of 3 merging algorithms utilizing Level-3 data products. None of the candidate algorithms were limited to any Level-3 grid size or temporal frequency. The choice of grid size and frequency issue depends on how sparse the final fields are and the acceptance level for data gaps. We leave this choice to the SIMBIOS Project. For our analyses, however, we used 25-km equiangular spatial, and daily time fields.

Candidate merger algorithms under investigation in this proposed effort were: averaging, blending, and statistical (optimal) interpolation.

Averaging

This method is a simple, straightforward application of weighting data from each sensor equally. At grid points where only data from one satellite are available, it enters the merged field unadjusted.

$$C_{ij} \frac{\sum_s C_{ijs}}{\sum_s n_{ijs}}$$

where C indicates chlorophyll from sensor s, n is the number of observations from sensor s,ij represents the Level-3 grid point in question, and the summations are over the sensors. Although we use chlorophyll to represent the equation, any Level-3 data product can be used. This method has the advantage of simplicity and total objectivity, i.e., no sensor data are preferred over others. It can potentially suffer from this same objectivity in the case of relatively poorer performance. If Level-3 grid locations are common among the different sensor products, the application of the method is straightforward. If they are not, then interpolation may be required.

Blended analysis

The blended analysis has traditionally been applied to merging satellite and in situ data (Reynolds, 1988). Also known as the Conditional Relaxation Analysis Method (CRAM; Oort, 1983), this analysis assumes that in situ data are valid and uses these data directly in the final product. The satellite chlorophyll data are inserted into the final field using Poisson's equation

$$\nabla^2 C^b = \psi$$

where C^b is the final blended field of chlorophyll, and Ψ is a forcing term, which is defined to be the Laplacian of the gridded satellite chlorophyll data ($\nabla^2 S$). In situ data serve as internal boundary conditions, and are inserted directly into the solution field C^b

$$C_{ibc} = I$$

where the subscript $_{ibc}$ indicates internal boundary condition (IBC) and I is the in situ value of chlorophyll. Thus in situ data appear un-adjusted in the final blended product. In its application to multiple ocean color data sets, in situ data would be replaced by a determination of superior performance by one of the sensors data, and utilized as the IBC. This could occur across the domain for an individual sensor, if its calibration was considered superior, for example. Or it could occur by location as the environmental conditions provide for better performance of one sensor over the others (e.g., location of sun glint, individual scan problems, etc.). Where one sensor data could be established as superior, it would serve as the IBC. If no distinction could be provided, the data could simply be merged using one or more of the other methods. Then the final merged product would be blended, so that the internal boundary conditions are upheld, and the rest of the Level-3 field is adjusted according to the spatial variability of the other sensors. This can provide a bias correction to the non-IBC points, while retaining their spatial structure, and potentially produce an overall enhanced data set.

The requirement of superior data field insertion unaltered into the merged field can be relaxed. For example, the IBC weight could be 0.25 for sensor 1 and 0.75 for sensor 2 at grid point ij. This can be a useful modification of several sensor data sets are superior to others but not necessarily from one another, or if clear superiority is difficult to quantify.

Statistical Interpolation

This method is often referred to as optimal interpolation (e.g., Reynolds and Smith, 1994), but is technically only optimal when all of the error correlations are known (Daley, 1991), which is rare. In this method the weights W are chosen to minimize the expected error variance of the analyzed field (Daley,

1991). It differs from the spatial analysis method by allowing error correlations to determine the effective separation distance, and from the blended analysis by use of a statistical approach for defining the weights. A weight matrix **W** represents the error correlations, and is referred to as the error covariance matrix.

$$C_{ij} = C_{sij} + \sum W_{ijkm}(C_{S+1,km} - C_{s,km})$$

This method has the advantage of widespread use in data assimilation problems, and objectivity in selection of the weights. The disadvantage is the statistical interpretation of the merged data set, as opposed to a scientific evaluation.

It is possible that the best merging method will be one that utilizes combinations of these algorithms. For example, some level of subjective analysis will be used in data masking, and then averaged. Reynolds and Smith (1994) combine the use of blending for bias correction followed by statistical interpolation to recover the grid resolution. We may easily envision a combination of approaches.

Table 6.1. Mission characteristics of proposed and present global ocean color sensors. For node, D indicates descending, A indicates ascending. Incl. Indicate inclination (degrees). ECT means local equator crossing time on the node. GIFOV means ground instantaneous field of view at nadir. Advanced Earth Observing Satellite (ADEOS); Earth Observing System (EOS); Environmental Satellite (Envisat); NPOESS Preparatory Project (NPP) and National Polar-orbiting Operational Environmental Satellite System (NPOESS).

Sensor	Launch	Spacecraft	Altitude	Incl.	ECT	Node	Swath	Tilt	GIFOV
SeaWiFS	1997	OrbView-2	705 km	98.2	noon	D	45°	±20°	1 km
MODIS-AM	1999	EOS-Terra	705 km	98.2	10:30 AM	D	55°	none	1 km
MERIS	2002	Envisat	780 km	98.5	10:00 AM	D	41°	none	1 km
GLI	2003	ADEOS-II	803 km	98.6	10:30 AM	D	45°	±20°	1km
POLDER-II	2003	ADEOS-II	803 km	98.6	10:30 AM	D	51°	±20°	7 km
MODIS-PM	2002	EOS-Aqua	705 km	98.2	1:30 PM	A	55°	none	1 km
S-GLI	2005	ADEOS-III	803 km	98.6	10:30 AM	D	45°	±20°	1
VIIRS	2006	NPP	TBD	TBD	TBD	TBD	TBD	TBD	1 km
VIIRS	2009	NPP	TBD	TBD	TBD	TBD	TBD	TBD	1 km

6.3. RESEARCH RESULTS

The three candidate merger algorithms have been tested using SeaWiFS and MODIS data. The SeaWiFS data is Version 4 and the MODIS is Collection 4. Results indicate promising behavior from all three candidate algorithms (Figure 6.1). However, there are some problems remaining, most associated with data quality of the sensors and our ability to understand and correct for them prior to application of the algorithms. Overall the averaging method is best for data with no biases, because it is simple, objective, and computationally fast. If there are biases in either or both data sets that are uncorrected or unrecognized, this method will propagate these errors into the merged field, and produce a poor quality data set. Knowledge of biases in the new versions of each sensor is presently lacking, and requires substantial effort. The blended method is effective at eliminating biases if a "truth field" can be identified. In the analyses done so far, we assumed SeaWiFS to be a truth field unilaterally, and MODIS was the data blended to produce the final merged product. The effectiveness of the bias-correction capability of the blended analysis is quite well known in in situ-satellite data merging, but not in satellite-satellite merging. Our results indicate that significant differences in satellite data quality coupled with the very large coverage of both sensors, results in over-correction by the blended method.

The statistical (optimal) interpolation (OI) method has many of the advantages of the blended method in bias-correction. However, the over-correction behavior of the blended method is reduced to the point that it is not readily apparent in the resulting merged field. The method suffers from computational complexity and is very slow.
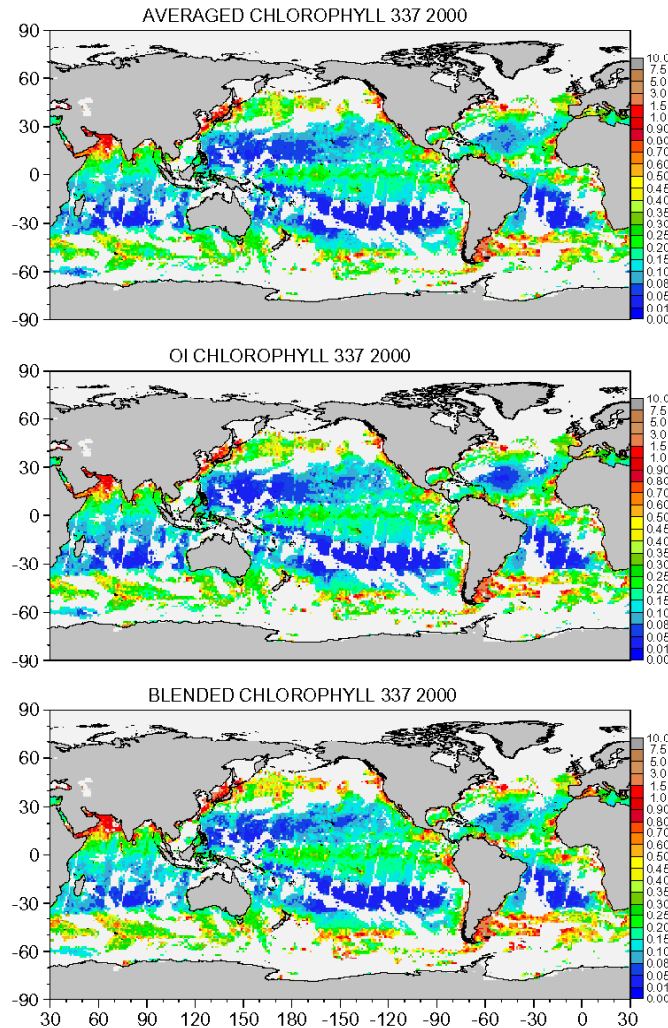
Figure 6.1: Comparison of 3 different merging methodologies for SeaWiFS and MODIS on Dec. 2, 2000

It is clear that proper selection of merging algorithms depends critically upon knowledge of the error characteristics of the data sets being merged. Consequently, we invested significant effort into analyzing the SeaWiFS and MODIS data behavior as compared against the SeaBASS data archive.

SeaWiFS Reprocessing 3 vs. Reprocessing 4

Comparisons were made between the newly reprocessed SeaWiFS Level-3 chlorophyll product (Reprocessing 4 or R4) and the previous version (Reprocessing 3 or R3) using in situ measurements. There were 2,470 SeaWiFS/SeaBASS match-up measurements of fluorometrically/spectrophotometrically-derived chlorophyll-a (mg m^{-3}) at depths of 0.0 to 10.0 meters available for the SeaWiFS mission period of September 15, 1997 through June 1, 2002 (Figure 6.2). The results showed that the newly reprocessed SeaWiFS data matched up better with the surface measurements than the previous version did (Figure 6.3). Globally, the slope of the match-ups improves to 0.85 from 0.78 in log-log scale. A significant trend that

contributed to this improvement was the overall decrease in SeaWiFS chlorophyll levels less than 1.0 mg m^{-3}. Regional analyses reveal that the match-ups improve in every oceanic basin, except the Antarctic. However, SeaWiFS continues to exhibit poor correspondence with in situ data in the North Atlantic where the match-ups have a slope of 0.54.
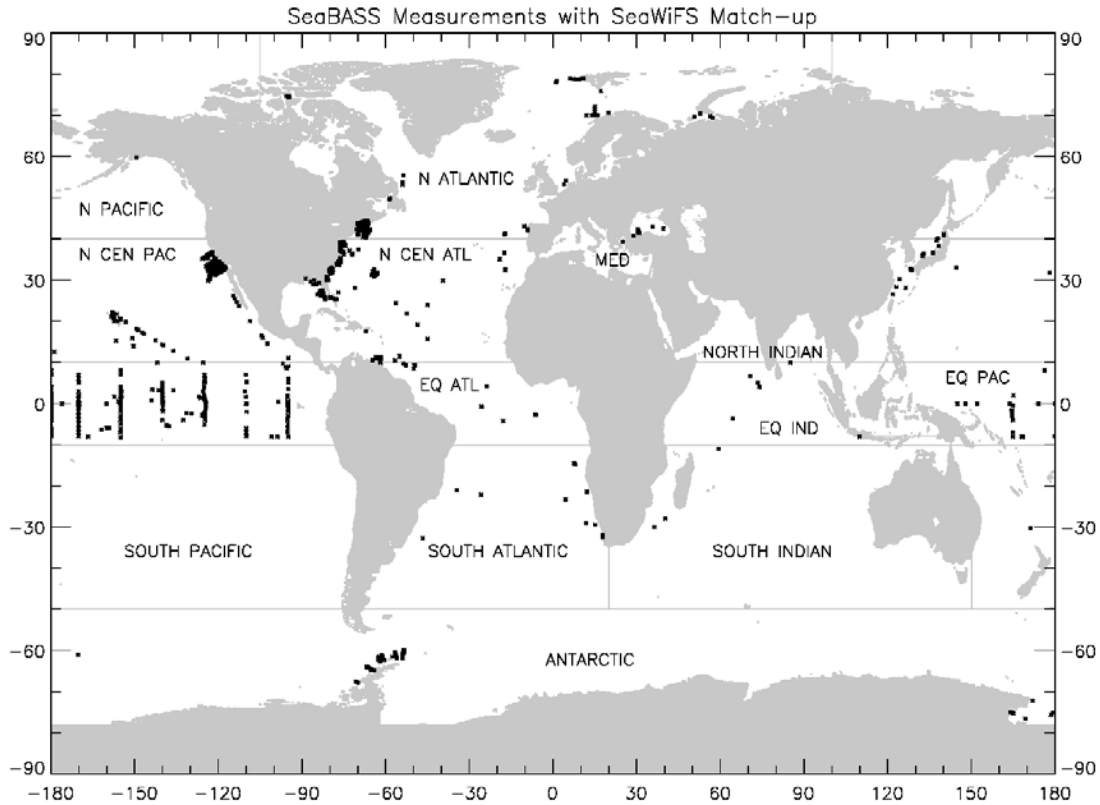

Figure 6.2: Global distribution of the SeaBASS measurements with a co-located SeaWiFS pixel for both R3 and R4 (N=2,470). The oceanic regions used for statistical comparisons are shown.

An examination of monthly images for May 1999 revealed that the number and magnitude of high-value chlorophyll pixels had increased in the high latitude open ocean of the South Pacific. There were more high-value outliers in the R4 image as compared to the R3 in the open ocean of the South Pacific, an area generally characterized by relatively homogeneous levels of low-chlorophyll. On a monthly timescale and in a largely homogenous area, increased numbers and magnitudes of high outliers in the new R4 data set as compared to the previous version is some cause for concern. The incidence of outliers should tend to be relatively low in monthly images in general due to the smoothing effects of averaging the dailies. In addition, the lower latitude areas such as the South Pacific may have fewer valid daily measurements than in other areas, but an examination of the daily images for this region exhibited the same general pattern shown in the monthly values. These results were published as a NASA Technical Memorandum (Casey and Gregg, 2003).

Global and Regional SeaWiFS Chlorophyll Data Evaluation

The SeaWiFS chlorophyll data set was compared to comprehensive archives of in situ chlorophyll data from NASA and NOAA, involving 4168 point-to-point daily matchups. The global comparison indicated

an RMS log error of 31%, with a coefficient of determination (r^2) of 0.76 (Figure 6.5). RMS log error for open ocean (defined as bottom depth > 200m) was 5% lower than for coastal regions, indicating a small deterioration of quality of the SeaWiFS data set in coastal regions. All of the Pacific oceanographic basins generally showed very good agreement with SeaWiFS, as did the South Atlantic basin. However, poorer agreement was found in the Mediterranean/Black Seas, North Central and Equatorial Atlantic basins, and the Antarctic. Optical complexity arising from riverine inputs, Saharan dust, and anomalous oceanic constituents contributed to the differences observed in the Atlantic, where a trend of overestimation by SeaWiFS occurred. The Antarctic indicated a pronounced negative bias, indicating an underestimation, especially for chlorophyll concentrations > 0.15 mg m^{-3}. The results provide a comprehensive global and geographic analysis of the SeaWiFS data set, which will assist data users and policy makers in assessing the uncertainty of estimates of global and regional ocean chlorophyll and primary production. The results have been submitted to Remote Sensing of Environment (Gregg and Casey 2003).
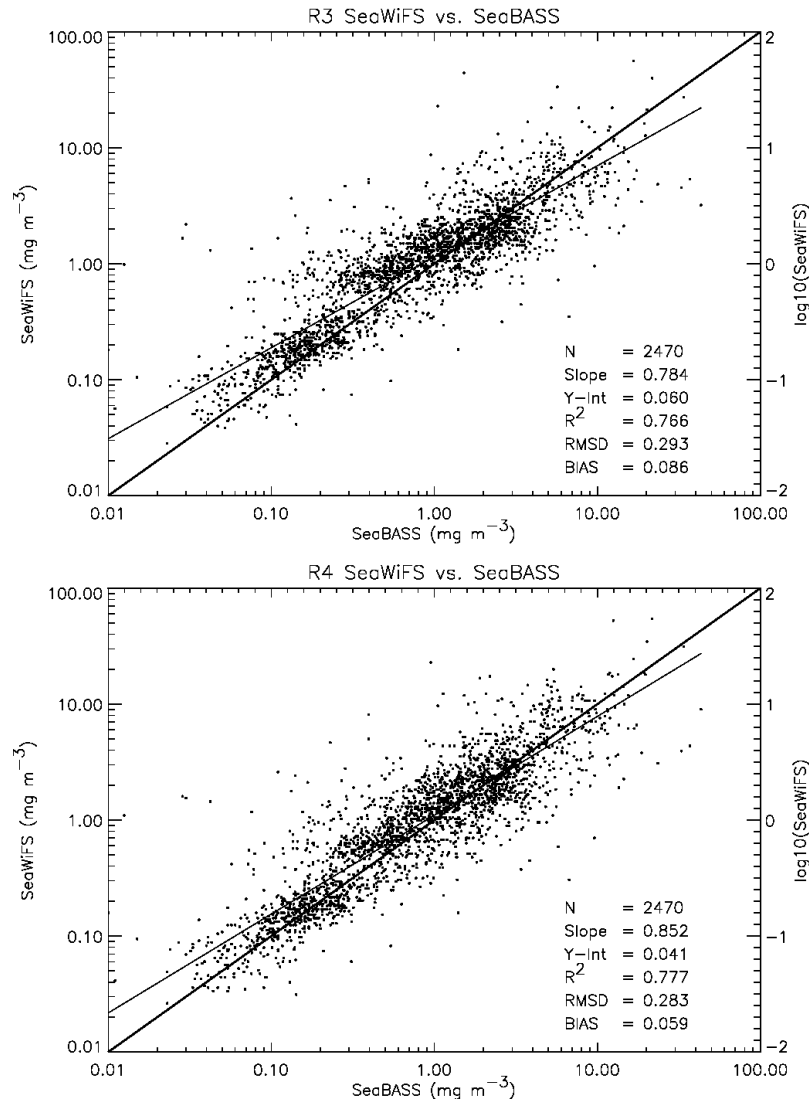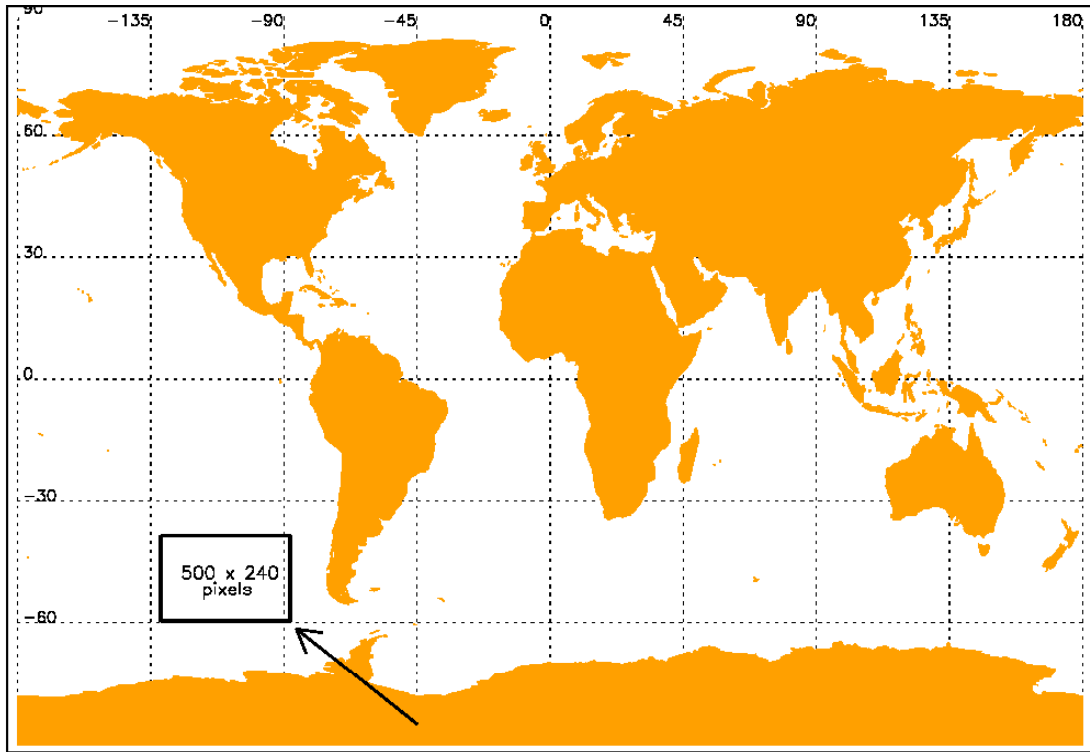

Figure 6.3: Scatterplots comparing SeaBASS chlorophyll measurements (mg m-3) with co-located SeaWiFS values for versions R3 (top) and R4 (bottom). The 1-to-1 line (thick) and the least-squares regression line (thin) are shown, as well as regression coefficients, root mean squared differences (RMSDs), and biases.
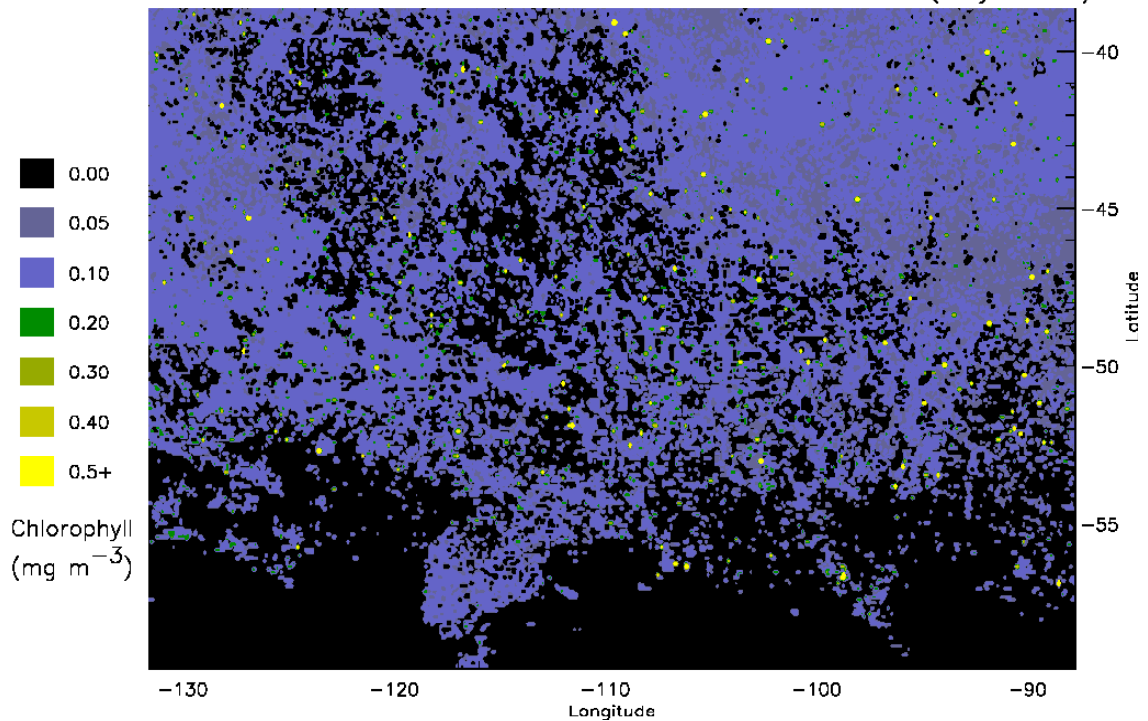
Figure 6. 4: (Top) Global map showing the location of the South Pacific region between -131.6 and -87.8 degrees longitude and -59.6 and -38.6 degrees latitude (500 x 240 pixels). (Bottom) Contour plot of the monthly SeaWiFS R4 chlorophyll values (mg m-3) which have a co-located R3 value for the South Pacific region in May 1999.

RMS

	Antarct	S Indian	S Pacific	S Atlantic	Eq Indian	Eq Pacific	Eq Atlantic	N Cen Pacific	N Cen Atlantic	Med/ Black Sea	N Pacific	N Atlantic	Global
RMS %	38.5	27.5	12.7	20.1	24.0	16.8	48.5	24.7	35.8	55.7	31.6	33.1	31.0

Bias

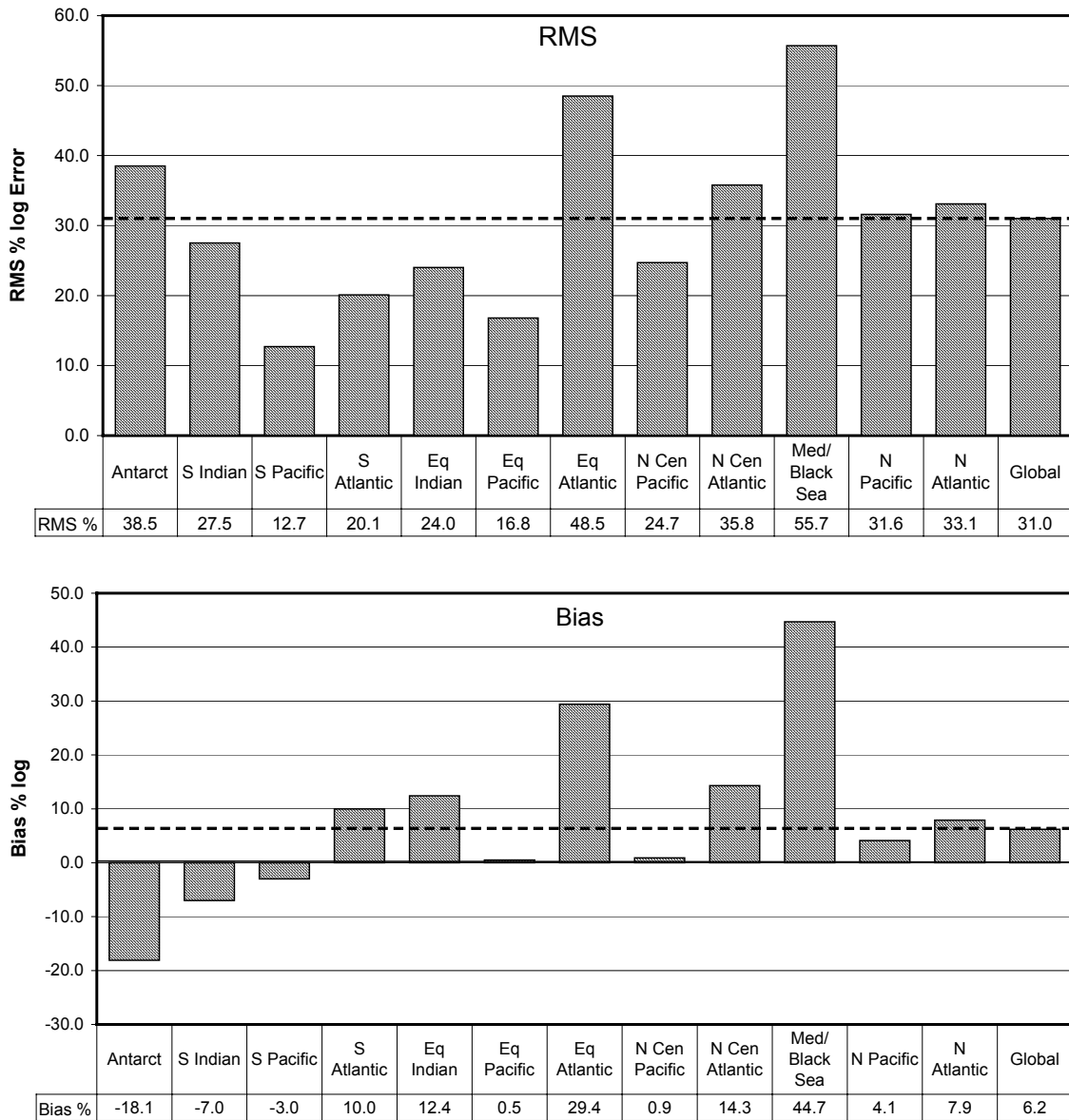|	Antarct	S Indian	S Pacific	S Atlantic	Eq Indian	Eq Pacific	Eq Atlantic	N Cen Pacific	N Cen Atlantic	Med/ Black Sea	N Pacific	N Atlantic	Global
Bias %	-18.1	-7.0	-3.0	10.0	12.4	0.5	29.4	0.9	14.3	44.7	4.1	7.9	6.2

Figure 6.5: Top: RMS error between in situ data and the SeaWiFS chlorophyll data set, separated into the 13 major oceanographic basins, and global. Bottom: average error or bias. Dashed lines indicate the global mean.

Global and Regional MODIS Chlorophyll Data Evaluation

The same methodology for evaluation of SeaWiFS was used to evaluate MODIS chlorophyll data. Only validated data from Collection 4 were used, spanning the period Nov. 2000 through Mar. 2002. In situ data were exclusively from the SeaBASS archive, as no NODC data were available.

There are 3 different bio-optical algorithms for chlorophyll from MODIS. Two (Chlor-MODIS and Chlor-a-2) are empirical expressions, while the third (Chlor-a-3) is a semi-analytic algorithm. Chlor-a-2 is the algorithm most similar to the SeaWiFS bio-optical algorithm.

All three bio-optical algorithms from MODIS produced log RMS error compared to in situ data comparable with SeaWiFS over the same time period (Figure 6.6). There were minor differences among the evaluations regionally as well. The North Atlantic basin was an exception, where the MODIS Chlor-MODIS and Chlor-a-2 algorithms performed much worse than SeaWiFS and the MODIS Chlor-a-3 algorithm. Overall these results suggest compatibility of the two chlorophyll data sets and supports initial merging analyses.

MODIS/SeaWiFS Chlorophyll Comparison vs. In Situ Data (11/2000-3/2002)

	Antarctic (n=40)	S Pacific (n=3)	S Atlantic (n=25)	Eq Pacific (n=100)	Eq Atlantic (n=10)	N Central Pacific (n=175)	N Central Atlantic (n=175)	N Atlantic (n=60)	Global (n=600)
chl_modis	26.6	7.8	24.6	18.2	13.8	33.2	38.9	35.9	32.5
chlor_a_2	26.6	9.9	21.8	17.9	14.7	30.4	32.7	27.8	28.0
chlor_a_3	27.7	16.7	25.5	17.0	19.2	26.8	38.9	19.1	29.1
SeaWiFS	27.7	12.7	22.4	15.0	19.7	28.7	34.7	17.1	27.0

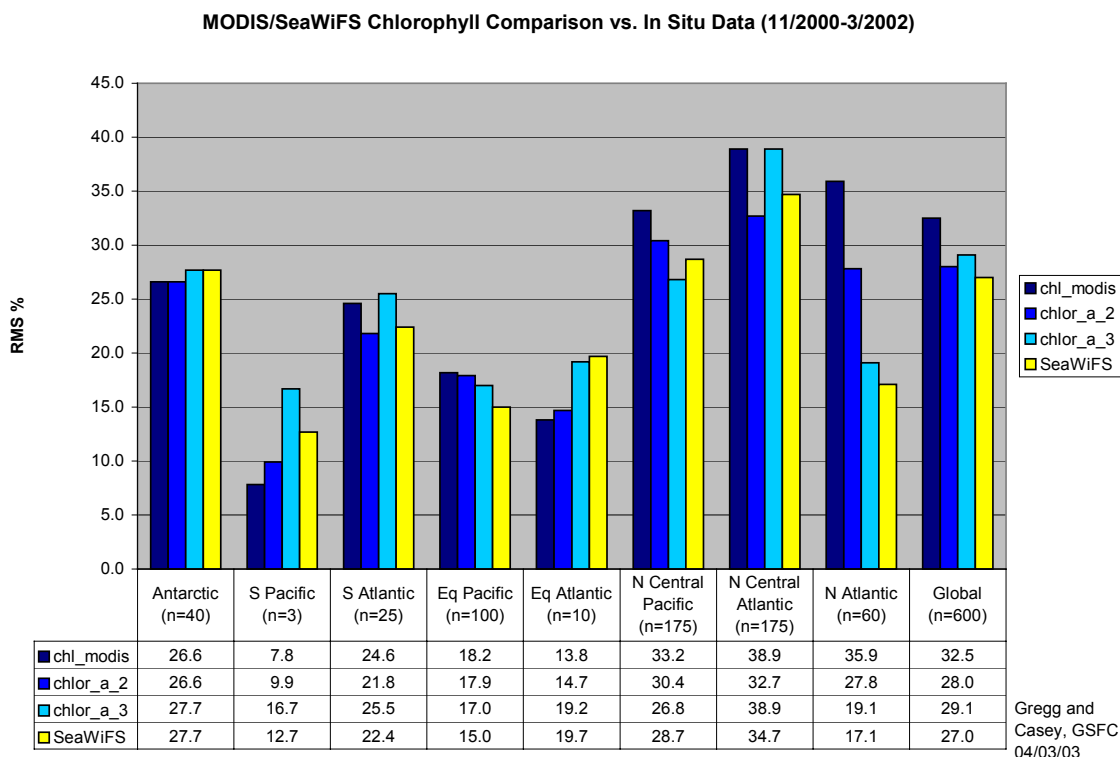Gregg and Casey, GSFC 04/03/03

Figure 6.6: Comparison of MODIS chlorophyll algorithms with SaeBASS in situ data globally and by region. SeaWiFS chlorophyll comparisons for the same time period are also shown.

Analyses of Decadal Changes in Ocean Primary Production

Although tangential to the main focus of our proposed effort, the SIMBIOS grant supported our work on analysis of decadal changes in global ocean primary production by providing in situ chlorophyll from the SeaBASS data set for blending and reduction of residual errors in the SeaWiFS data set. In this analysis, we found that satellite-in situ blended ocean chlorophyll records indicate that global ocean annual primary production has declined more than 6% since the early 1980's (Figure 6.7). Nearly 70% of the global decadal decline occurred in the high latitudes. In the northern high latitudes, these reductions in primary production corresponded with increases in sea surface temperature and decreases in atmospheric iron deposition to the oceans. In the Antarctic, the reductions were accompanied by increased wind stress. Three of four low latitude basins exhibited decadal increases in annual primary production. These results indicate that ocean photosynthetic uptake of carbon may be changing as a result of climatic changes and suggest major implications for the global carbon cycle. These results were published in Geophysical Research Letters (Gregg et al., 2003).

6.4 RECOMMENDATIONS

Comparisons between MODIS and in situ data from the SeaBASS archives for the period Nov. 2000 through Mar. 2002 indicated no significant difference from similar comparisons of SeaWiFS for the same time

period. This period is referred to as the validated MODIS time period. We therefore recommend the averaging methodology for merging MODIS and SeaWiFS chlorophyll. For time periods following the validated period, we have no knowledge of biases between the data sets, and also recommend the averaging methodology, which is the correct choice for these conditions.

For time periods preceding the MODIS validated period, biases are known to be present, and they vary in time and space. This averaging is not the proper choice for merging. However, the biases have not been fully characterized. Nevertheless, application of the blended analysis and statistical interpolation are appropriate for this time period. They both suffer from drawbacks, however, that limits their effective application. The blended analysis is subject to over-correction and produces artificially elevated chlorophyll values in places. The statistical interpolation, produces a more reasonable final merged analysis field, but is so computationally expensive as to limit its use. Further analysis is required to modify both methods prior to their implementation for satellite data merging, but both show promising results in the case where biases are known. In these circumstances one or both of these methods are superior to averaging because of the correction aspect of the methods. Averaging in the case of known biases results in degradation of the merged data set and is not recommended. If biases in the satellite data sets cannot be removed in later reprocessing, then additional effort should be expended on blending and statistical interpolation to correct for these errors in data merging.

The SeaBASS chlorophyll data set has been an invaluable resource for evaluating data merging methodologies and it or a similar data set are required for further efforts. It has also been essential in analysis of long-term trends of global satellite chlorophyll as indicated by the identification of a 6% decline in global primary production, a finding that would not have been possible without the use of SeaBASS chlorophyll data for blending with SeaWiFS to reduce residual errors in the satellite data set. It is essential to continue this effort, or something similar, to improve data merger methodologies and to further our understanding of satellite chlorophyll trends.
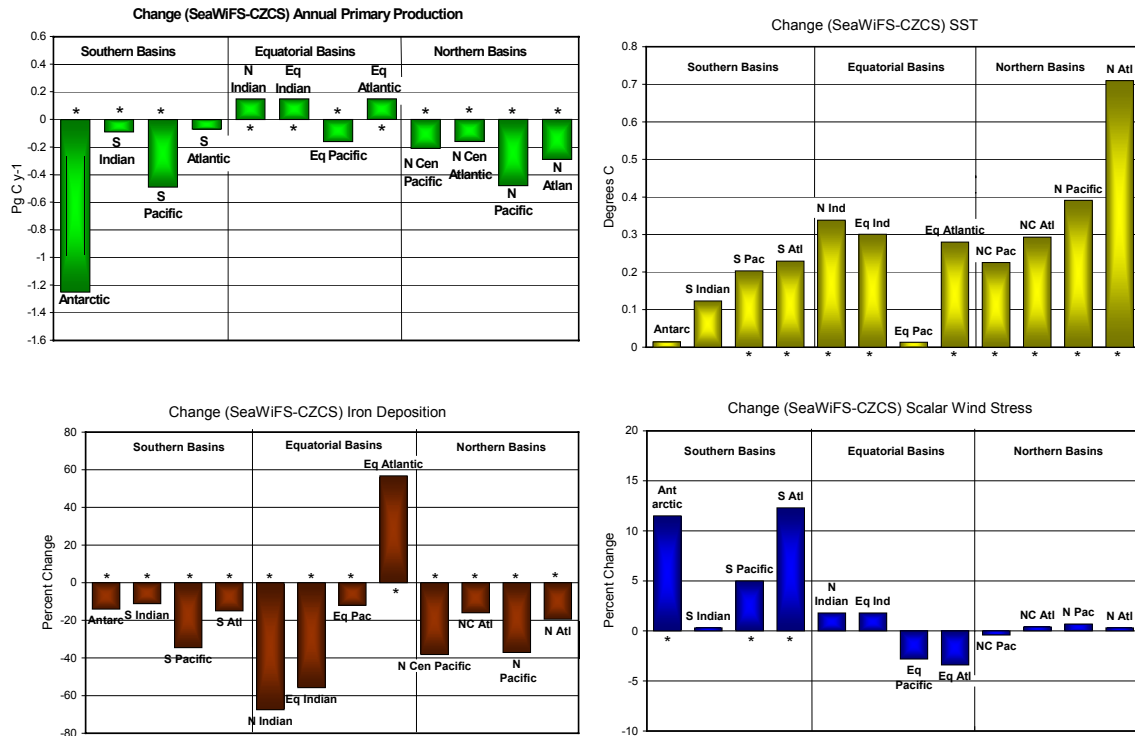

Figure 6.7: Differences between SeaWiFS (1997-2002) and CZCS (1979-1986) in the 12 major oceanographic basins. Differences are expressed as SeaWiFS-CZCS. Top left: Annual primary production (Pg C y-1). An asterisk indicates the difference is statistically significant at P < 0.05. Top right: SST (degrees C). Bottom left: iron deposition (%). Bottom right: mean scalar wind stress (%).

REFERENCES

Casey, N.W. and W.W. Gregg, 2003: Comparing Seawifs Reprocessing Versions (R3 vs. R4). *NASA Technical Memorandum 2003-212235.* 20 pp.

Daley, R., 1991: Atmospheric Data Analysis, Cambridge Univ. Press, Cambridge. 457 pp.

Doney, S.C., R.M. Najjar, and S.E. Stewart, 1995: Photochemistry, Mixing, and Diurnal Cycles in the Upper Ocean. *J. Mar. Res*. **53**, 341-369.

Eppley, R.W., 1972. Temperature And Phytoplankton Growth In The Sea. *Fish. Bull*., 70, 1063-1085.

Gregg, W.W. and N.W. Casey, 2003: Global And Regional Evaluation Of The Seawifs Chlorophyll Data Set. *Remote Sensing of Environment*, submitted.

Gregg, W.W., M.E. Conkright, P. Ginoux, J.E. O'Reilly, and N.W. Casey, 2003: Ocean Primary Production And Climate: Global Decadal Changes. *Geophysical Research Letters,* **30** (15) 1809, doi: 10.1029/2003GL016889.

Oort, A.H., 1983: Global Atmospheric Circulation Statistics, 1958-1973, NOAA Prof. Paper 14, 180 pp.

Reynolds, R.W., 1988: A Real-Time Global Sea Surface Temperature Analysis. *J. Clim*., **1**,75-86.

Reynolds, R.W. and T.M. Smith, 1994: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *J. Clim.,* **7**, 75-86.