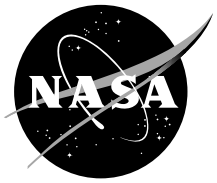
NASA/TM—2003-212249



MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003

G.S. Fargion, C.R. McClain

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

September 2003

The NASA STI Program Office … in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA's counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and mission, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at http://www.sti.nasa.gov/STI-homepage.html
- E-mail your question via the Internet to help@sti.nasa.gov
- Fax your question to the NASA Access Help Desk at (301) 621-0134
- Telephone the NASA Access Help Desk at (301) 621-0390
- Write to:
  NASA Access Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

NASA/TM—2003–212249

MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003

Giulietta S. Fargion, SIMBIOS Project, Science Applications International Corp., Greenbelt, MD
Charles R. McClain, Goddard Space Flight Center, Greenbelt, MD

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

September 2003

Available from:

NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320
Price Code: A17

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Price Code: A10

Preface

The purpose of this technical report is to provide current documentation of the Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Project activities, satellite data processing, and data product validation. This documentation is necessary to ensure that critical information is related to the scientific community and NASA management. This critical information includes the technical difficulties and challenges of validating and combining ocean color data from an array of independent satellite systems to form consistent and accurate global bio-optical time series products. This technical report focuses on the SIMBIOS Project's efforts in support of the Moderate-Resolution Imaging Spectroradiometer (MODIS) on the Earth Observing System (EOS) Terra platform (similar evaluations of MODIS/Aqua are underway). This technical report is not meant as a substitute for scientific literature. Instead, it will provide a ready and responsive vehicle for the multitude of technical reports issued by an operational project.

TABLE OF CONTENTS

Chapter 1

An Overview of MODIS Support and Accomplishments by SIMBIOS and SeaWiFS Projects

Giulietta S. Fargion
Science Applications International Corporation (SAIC), Beltsville, Maryland

Charles R. McClain and Gene C. Feldman
NASA Goddard Space Flight Center, Greenbelt, Maryland

Over the course of the SeaWiFS and SIMBIOS Projects, there has been a substantial level of collaboration with the MODIS program, e.g., the development of the Marine Optical Buoy (MOBY) and the prototyping of the MODIS Adaptive Processing System (MODAPS). Over the past year, in particular, both Projects have provided substantial support to the MODIS Ocean Team as guidance from NASA Headquarters was for SIMBIOS to focus on MODIS. The following list is a summary of specific types of assistance with names of the main people involved. The following chapters detail some of these contributions as of May 2003. More recent analyses associated with the pending MODIS (Terra) reprocessing are posted on various project websites and will be documented in more detail once the reprocessing is underway. Also, as part of the data system prototyping activities in preparation for the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP) Visible and Infrared Imaging Radiometer Suite (VIIRS) data, the Project is using MODIS data to explore new data formats, data access and distribution approaches, and processing system architectures. If deemed beneficial to the ocean color community, developments derived from this prototyping may eventually be integrated into the operational MODIS data processing system.

- MODIS-sensor characterization and on-board calibration analyses (Bob Barnes and Gerhard Meister)
- MODAPS (MODIS data processing system) prototype development (John Wilding)
- Distribution of MODIS diagnostic data (Sean Bailey and John Wilding) and Web Support (Norman Kuring)
- SeaDAS support of MODIS data display and analysis (Mark Rubens and Xiao-Long)
- MODIS data geolocation (Fred Patt)
- MODIS ancillary data (Wayne Robinson and Bryan Franz)
- MODIS over flight predictions in support of field campaigns (Sean Bailey and Jeremy Werdell)
- Bio-optical and atmospheric data from SeaBASS (Sean Bailey, Jeremy Werdell, and SIMBIOS Science Team members)
- MODIS L2 ocean and atmosphere measurement match-up analysis (Sean Bailey)
- MODIS comparison analysis (over time and by region) with SeaWiFS (Bryan Franz and Ewa Kwiatkowska)
- MODIS data merger with SeaWiFS and data distribution (Joel Gales and Bryan Franz)
- SIMBIOS laboratory radiometric intercomparisons with MODIS Team members (Gerhard Meister)
- SIMBIOS instrument pool support (Christophe Pietras and Kirk Knobelspiesse)

Chapter 2

Comparisons of Daily Global Ocean Color Data Sets: MODIS-Terra/Aqua and SeaWiFS

Ewa Kwiatkowska

Science Applications International Corporation, Beltsville, MD

2.1 INTRODUCTION

To facilitate the SIMBIOS Project ocean-color data merger efforts, MODIS-Terra and MODIS-Aqua daily global ocean products were compared against SeaWiFS data. The analyses focused on assessing temporal biases in MODIS ocean data differences from SeaWiFS and the artifacts present in MODIS data. The artifacts were caused by the difficulties in accurately characterizing this complex sensor for features such as detector-to-detector variabilities, mirror-sidedness, response versus scan angle, and polarization sensitivity. The comparisons were vital for the ocean-color data merger because they enabled extraction of disparate trends and trend dependencies in data between the sensors. One of the goals of the data merger was then to eliminate these trends to produce integrated multi-instrument and multi-year products of a consistent spatial and temporal accuracy and uniform calibration and validation.

2.2 MATCHUP DATA, TIME SERIES, AND STATISTICS

The evaluations of MODIS global data products in comparison with SeaWiFS included qualitative analyses of chlorophyll-*a* and chlorophyll-*a* difference maps from both sensors as well as quantitative analyses. The subsequent results presented here are based on quantitative analyses obtained through sensor data comparisons, called matchups (Kilpatrick *et al*., 2002). Matchups used daily global overlapping level-3 (L3) bin coverage between MODIS and SeaWiFS at 9km resolution. L3 binned files were employed to facilitate comparisons between the sensors of data which corresponded to the same ground location. The 9km bins were used to extract statistically significant global trends in data discrepancies between the instruments averaged over a $9km^2$ coverage and over MODIS multiple detectors. MODIS and SeaWiFS data were processed using up-to-date algorithms, i.e. MODIS-Terra collection number 4, MODIS-Aqua collection number 3, and SeaWiFS reprocessing number 4. MODIS data used in the study were obtained from NASA GSFC Distributed Active Archive Center (GDAAC) and were rebinned from the native 4.6km resolution to 9km bins. SeaWiFS data were acquired at the standard 9km resolution from the SeaWiFS Data Processing System (SDPS).

A time series of daily water-leaving radiance and chlorophyll products, evenly spread over the three years of joint MODIS-Terra and SeaWiFS coverage, was used to study trends in discrepancies between MODIS-Terra and SeaWiFS data. The time series was composed of 74 days, roughly every 15^{th} day, of MODIS-Terra and SeaWiFS acquisitions from February 2000 to December 2002. A time series of MODIS-Aqua and SeaWiFS data was limited to three months of the overlapping sensor coverage. It was composed of 36 days of data, starting with daily and then in 4-day intervals, from the end of November 2002 to the beginning of March 2003. In all investigations only good quality data were applied, i.e. quality 0 MODIS data and standard SeaWiFS L3 quality data. Within the collection 4, MODIS-Terra best-calibrated data spanned the period from November 2000 to September 2001. All MODIS-Aqua collection 3 data were of provisional quality. All SeaWiFS data had a calibrated and validated quality.

Matchup data came from overlapping bin coverage between MODIS and SeaWiFS for each individual day and each common data product. Although both sensors operate using similar spectral bands for ocean applications, only two bands are identical, 412nm and 443nm. These two bands were used to quantitatively compare normalized water-leaving radiances (nLw) between the sensors. Chlorophyll-*a* concentration matchups were performed alongside the nLw comparisons. MODIS chlor_a_2 and SeaWiFS chlor_a products were used which were based on analogous algorithms between the sensors, OC3M and OC4v4 respectively. An example of daily common coverage bins between MODIS-Terra and SeaWiFS and a data scatter plot are displayed in Figure 2.1.

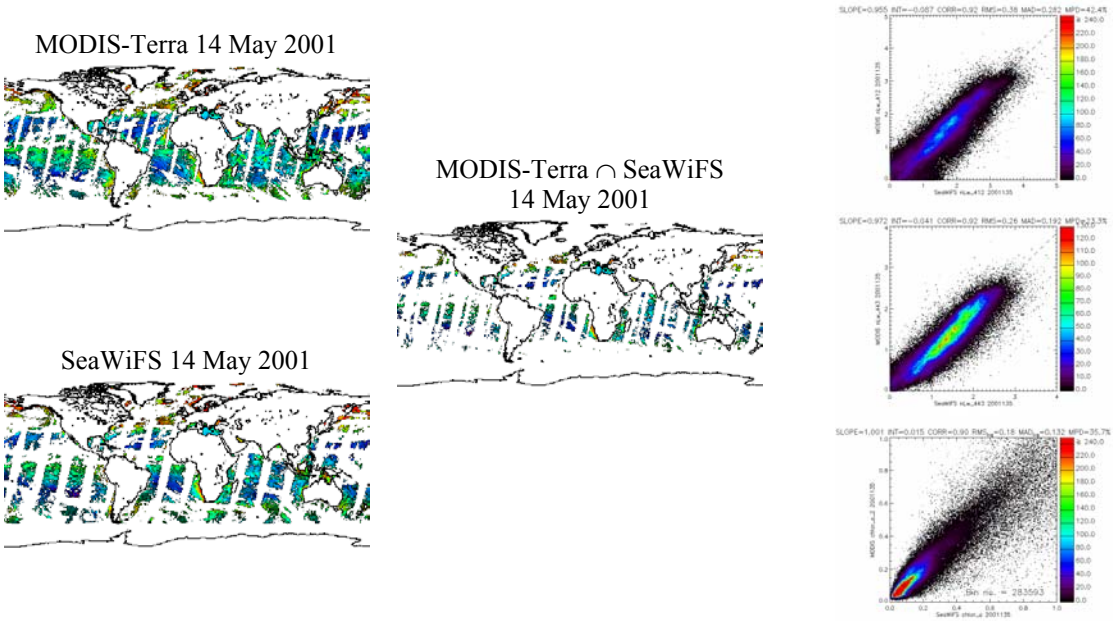


Figure 2.1: MODIS-Terra and SeaWiFS overlapping daily coverage at 9km resolution and a corresponding data scatter-plot for nLw_412, nLw_443, and chlorophyll-*a* concentration.

In regular SeaWiFS L3 binned files, ocean products associated with each bin have the same quality. Consequently, in L3 files, only those bins are present for which all standard SeaWiFS products exist, i.e. have good quality determined by a choice of flags and masks. Unlike SeaWiFS, MODIS L3 bin data may have different quality for different standard ocean products. For example, MODIS chlor_a_2 product may have a good quality and a water-leaving radiance product for the same bin may have a lower quality. In these investigations only good quality MODIS data were used. Additionally, only those MODIS data bins for which all products corresponding to SeaWiFS standard ocean products had a good quality were matched against SeaWiFS. This meant that only those MODIS L3 bins were used for which the following products had a good quality: nLw_412, nLw_443, nLw_488, nLw_531, nLw_551, nLw_667, nLw_678, Tau_865, Eps_78, K_490, and chlor_a_2.

For each data product and date, scatter plots of SeaWiFS vs. MODIS data were created and corresponding matchup statistics were calculated. The statistics were used to investigate time trends in MODIS data as they departed from SeaWiFS measurements. The statistics included slope (SLOPE) and intercept (INT) of the linear fit between SeaWiFS and MODIS data. To calculate the linear fit, an outlier-resistant linear regression was applied based on the robust Tukey's biweight calculated perpendicularly to the bisector of MODIS vs. SeaWiFS and SeaWiFS vs. MODIS data (Press *et al*., 1986). Taking the bisector of the fit took into consideration the presence of uncertainties in both MODIS and SeaWiFS data. The robust bisquare weighting ensured that the slope and intercept parameters were representative of the bulk of the data distribution and were not skewed by a few outlier points. Furthermore, sensor data robust correlation (CORR), root mean squared error (RMS), mean absolute difference (MAD) and mean percentage difference were obtained

$$\mathbf{MAD} = \frac{\sum_{\mathbf{i}=0}^{\mathbf{n}-1} \left| \mathbf{x}_{\mathbf{MODIS}_i} - \mathbf{x}_{\mathbf{SeaWiFS}_i} \right|}{\mathbf{n}},$$

$$\mathbf{MPD} = 100\% \cdot \sum_{\mathbf{i}=0}^{\mathbf{n}-1} \frac{\left| \mathbf{x}_{\mathbf{MODIS}_i} - \mathbf{x}_{\mathbf{SeaWiFS}_i} \right|}{\frac{1}{2}\left(\mathbf{x}_{\mathbf{MODIS}_i} + \mathbf{x}_{\mathbf{SeaWiFS}_i} \right)} \Bigg/ \mathbf{n}.$$

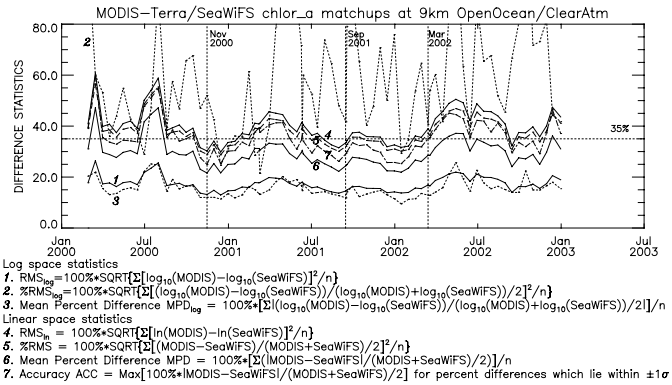
All statistics were calculated in the linear space of nLw and chlorophyll data. Additionally, chlorophyll statistics, except for MPD, were obtained in the logarithmic space because the chlorophyll probability density function has a lognormal distribution (Campbell *et al*., 1995). Linear space MPD was found to be a good estimate of the difference between MODIS chlorophyll and SeaWiFS, however, other chlorophyll error definitions were also determined appropriate and gave errors similar in value and temporal distribution. These definitions included the percent root mean squared error

$$\%\boldsymbol{RMS}_{SeaWiFS-MODIS} = 100\% \cdot \sqrt{\sum_{i=0}^{n-1}\left[\frac{x_{MODIS_i} - x_{SeaWiFS_i}}{\frac{1}{2}(x_{MODIS_i} + x_{SeaWiFS_i})}\right]^2 \Big/ n}$$, and the error $100\% \cdot \ln 10 \cdot \mathbf{RMS}_{\log_{10}} = 100\% \cdot \mathbf{RMS}_{\ln} = 100\% \cdot \sqrt{\sum_{i=0}^{n-1}\left[\ln(\mathbf{x}_{\mathbf{MODIS}_i}) - \ln(\mathbf{x}_{\mathbf{SeaWiFS}_i})\right]^2 \Big/ \mathbf{n}}$, adopted from the estimate of the percent coefficient of variation, %CV, given logarithmic data values, $\%\mathbf{CV} = 100\% \cdot \frac{\boldsymbol{\sigma}(\mathbf{chl})}{\mathbf{chl}} \Leftrightarrow 100\% \cdot \ln 10 \cdot \boldsymbol{\sigma}\left(\log_{10}(\mathbf{chl})\right)$, where $\boldsymbol{\sigma}\left(\log_{10}(\mathbf{chl})\right) = \left(\log'_{10}(\mathbf{chl})\right) \cdot \boldsymbol{\sigma}(\mathbf{chl}) = \log_{10}\mathbf{e} \cdot \frac{1}{\mathbf{chl}} \cdot \boldsymbol{\sigma}(\mathbf{chl})$.

When estimating accuracy relative to *in situ* data or other sensor data assumed to be a benchmark, the MPD and %RMS equations would have the benchmark's data in the denominator, e.g. $\mathbf{MPD} = 100\% \cdot \sum_{i=0}^{n-1} \frac{\left|\mathbf{x}_{\mathbf{MODIS}_i} - \mathbf{x}_{\mathbf{SeaWiFS}_i}\right|}{\mathbf{x}_{\mathbf{SeaWiFS}_i}} \Big/ \mathbf{n}$ if SeaWiFS was assumed to be a benchmark data set. One of the goals of these analyses was the assessment of MODIS data as it differed from SeaWiFS measurements. In this situation, SeaWiFS data were assumed to be a benchmark for the percent difference estimations. Overall, both approaches to error calculations were used in this study, the error calculated as a percent difference from the MODIS and SeaWiFS average and as a percent difference from the SeaWiFS benchmark.

Various authors proposed that the chlorophyll accuracy be calculated either in linear or in logarithmic spaces (Frouin and Gregg, personal communication). The logarithmic space is sensitive to absolute errors in the proximity of 1mg/m^3 of chlorophyll and becomes increasingly more tolerant than the linear space to errors outside of the 1mg/m^3 vicinity, i.e. below ~0.4mg/m^3 of chlorophyll, which accounts for ~90% of the global ocean, and above ~3mg/m^3 of chlorophyll. In the linear space, the accuracy estimates become susceptible for chlorophyll values approaching 0mg/m^3. Bottom limits of chlorophyll for collection 4 MODIS-Terra good-quality data go down to 0.01mg/m^3 and reprocessing 4 SeaWiFS data to 0.001mg/m^3. Figure 2 presents comparisons of different approaches to estimating MODIS-Terra and SeaWiFS relative chlorophyll differences in the linear and logarithmic spaces. To produce the statistics in both plots, identical exclusion criteria were applied to chlorophyll data to eliminate divisions by zero, i.e. only those MODIS and SeaWiFS chlorophyll bins were used for which $chl_{SeaWiFS} > 0.01 mg/m^3$ (for all good quality MODIS data: $chl_{MODIS} > 0.01 mg/m^3$), $|\log_{10}chl_{SeaWiFS}| > 0.01\ \log_{10}(mg/m^3)$, and $|\log_{10}chl_{SeaWiFS} + \log_{10}chl_{MODIS}| > 0.01\ \log_{10}(mg/m^3)$.

Error from sensor average



Error from SeaWiFS benchmark

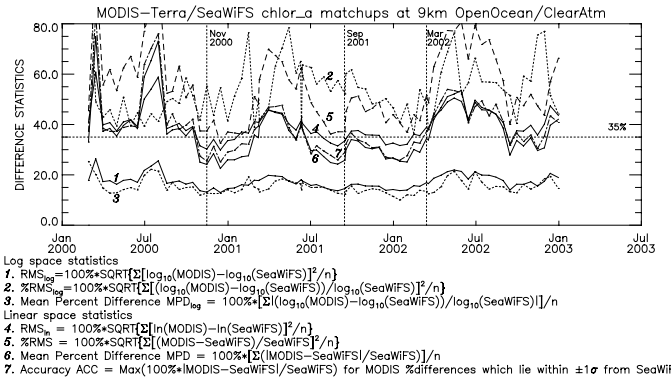


Figure 2.2: Time trends in logarithmic and linear space percentage difference statistics between MODIS-Terra and SeaWiFS chlorophyll data over the chlorophyll range above 0.01mg/m^3 for the three-year period of sensor ocean joint coverage. A standard lower chlorophyll limit for good quality MODIS data is 0.01mg/m^3 and for SeaWiFS data is 0.001mg/m^3.

Relative difference statistics from Figure 2.2 referenced to sensor average and SeaWiFS benchmark and calculated over the MODIS-like chlorophyll range (excluding the data eliminated by the exclusion criteria) are influenced by a bias between MODIS-Terra and SeaWiFS data in low chlorophyll waters. This is because MODIS-Terra chlorophyll in ocean gyres is

generally higher than SeaWiFS, especially in the southern hemisphere. Figure 2 shows that corresponding statistics, such as %RMS and ACC (ACC is defined in Figure 2.2), are not equivalent because relative differences between the sensors are not normally distributed. %RMS estimates are sensitive to values around zero in the denominator for both linear and log statistics. %RMS is less susceptible when the "linear" mean of two sensor data is applied and is very sensitive to chlorophyll around 1mg/m^3 in the log-space statistics. RMS_{ln}, MPD, and ACC estimates provide similar values and temporal distributions of errors between sensor data. Because $RMS_{ln} = \ln 10 * RMS_{log} \approx 2.3 * RMS_{log}$, the RMS_{log} and MPD_{log} log-space statistics give errors which are about 2 times lower than RMS_{ln}, MPD and ACC.

For these investigations, the chlorophyll accuracy was computed in the linear chlorophyll space. The linear space allowed scrutinizing MODIS and SeaWiFS discrepancies effectively within the most prominent ocean provinces, including the gyres. The linear space provided vital information on sensor chlorophyll differences for the merger effort, which could otherwise be unobserved if the logarithmic space had been used. Original SeaWiFS Project requirements limited the chlorophyll range over which the accuracy should be calculated to 0.05mg/m^3 – 50.0mg/m^3 (NASA TM 104566, 1992). Here, the accuracy was calculated within this range as well as within the entire range of MODIS and SeaWiFS chlorophyll.

For detailed studies, the choice of daily overlapping global data was further refined to limit ambiguities in MODIS and SeaWiFS data comparisons and to enable specific analyses focused on extracting sensor and calibration artifacts. The subsetting limited global data to open ocean and clear atmosphere coverage in order to eliminate more ambiguous coastal waters and turbid atmosphere. Clear atmosphere was defined as representing aerosol optical thickness (AOT) conditions below or equal to 0.2.

Data from the western and eastern portions of the MODIS scan were isolated into separate sets to investigate MODIS scan angle dependencies. To separate MODIS scan portions, only those L3 bins were used that were geographically situated between $\pm 40^0$ latitude. This was done in order to eliminate northern and southern high latitude bins that combined a number of data points from consecutive satellite orbits. In the analyses, western and eastern MODIS scan edge data were defined to be those data for which satellite zenith angles were above 25^0. This created MODIS swath edge subsets with widths of about 40^0. This was because collection 4 MODIS-Terra ocean products incorporated data with a maximum sensor zenith angle of 64^0. MODIS quality assurance Satellite Zenith L3 daily binned products were used. In collection 4 MODIS-Terra data, the Satellite Zenith had negative values for data bins on the western portion of the sensor scan and positive values on the eastern portion. In collection 3 MODIS-Aqua data, the Satellite Zenith was negative for data bins on the eastern portion of the sensor scan and positive on the western portion. The extent of MODIS western and eastern scan edge data used in the matchups are displayed in Figure 2.3.

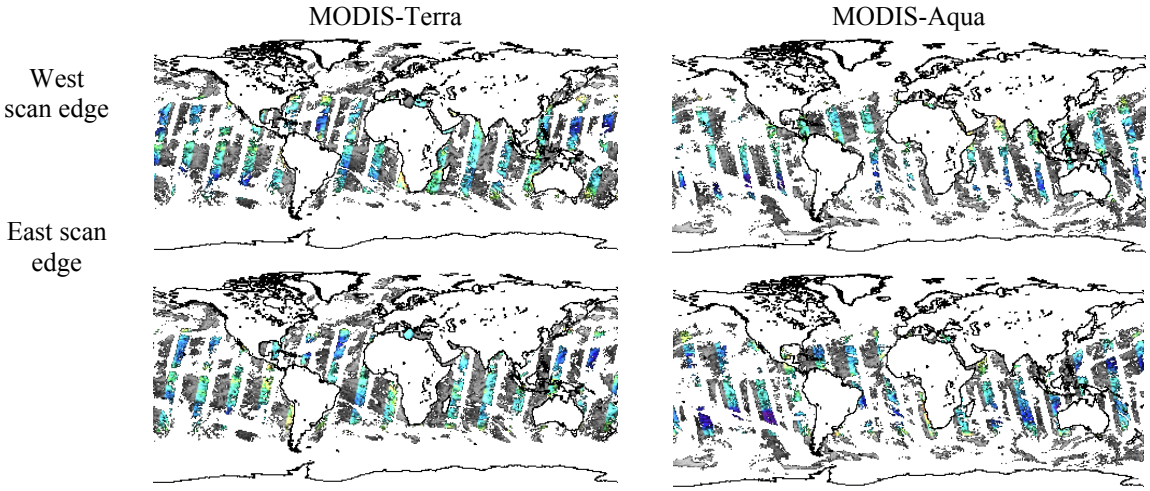


Figure 2.3: Western and eastern scan edge data subsets for MODIS Terra and Aqua.

Data that originated in the northern and southern hemispheres were divided into separate subsets in order to investigate latitudinal dependencies in MODIS ocean products. The northern hemisphere subset encompassed data bins located above 10^0 latitude and the southern hemisphere subset – below -10^0 latitude.

Sensor data were also subsetted depending on ocean chlorophyll contents. Low chlorophyll bins were isolated within MODIS and SeaWiFS daily global coverages, where chlorophyll a product data for SeaWiFS and chlor_a_2 for MODIS were below 0.1mg/m^3. Sensor bin coverage of chlorophyll within the range <0.05mg/m^3, 50.0mg/m^3> was also extracted to allow for chlorophyll accuracy calculations, as defined by the original SeaWiFS Project requirements (NASA TM 104566, 1992).

Individual overlapping chlorophyll and nLw scan lines were compared between MODIS-Terra and SeaWiFS across sensor zenith angles for the dates with the fullest swath overlap between the sensors. Chlorophyll-*a* concentration comparisons used MODIS chlor_a_2 and SeaWiFS chlophyll a products. nLw comparisons were mostly limited to data trend analyses because both sensors operate using different ocean color bands except for the two visible bands in the blue range centered at 412 and 443nm. Following is the summary of the matchup results based on February 2000 to December 2002 collection 4 MODIS-Terra and corresponding reprocessing 4 SeaWiFS data and November 2002 to March 2003 collection 3 MODIS-Aqua data and respective reprocessing 4 SeaWiFS data. The results of the matchups can also be found on the web page http://simbios.gsfc.nasa.gov/~ewa/SeaMODISTerra/seamodis-terra.html.

2.3 DAILY GLOBAL RELATIVE DIFFERENCES BETWEEN MODIS AND SEAWIFS DATA

Preliminary objectives assumed for this investigation for the creation of consistent merged ocean color products followed original accuracy objectives defined for the SeaWiFS instrument (NASA TM 104566, 1992). Multi-sensor data, which were to be incorporated into merged data sets, were required to meet these accuracy and inter-sensor discrepancy limits. The multi-sensor data were expected not to diverge more than the individual sensor accuracy. The requirements were the following:

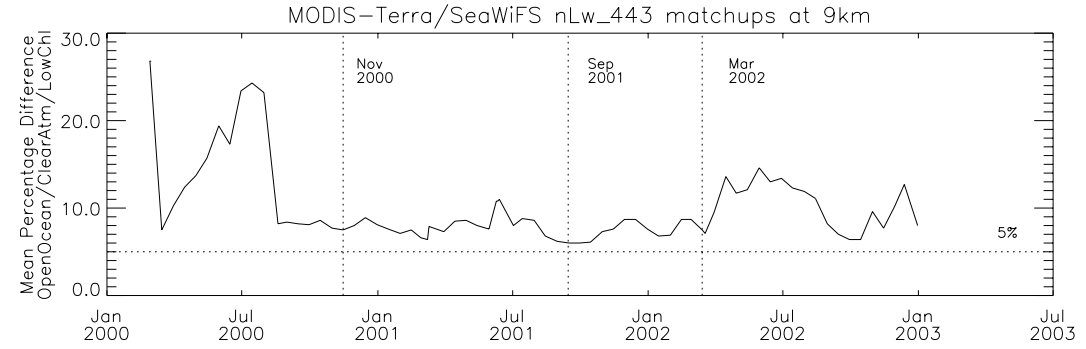
- Water-leaving radiance accuracy to within 5% for each sensor and no more than 5% difference between the sensors.
- Chlorophyll-*a* concentration accuracy to within 35% for each sensor over the range 0.05mg/m^3 to 50.0mg/m^3 and to within 35% difference between the sensors over the same chlorophyll range.
- Temporal stability of sensor data and suppression of data discrepancies between the sensors caused by instrument and calibration artifacts.

Global daily mean percentage differences between MODIS-Terra/Aqua and SeaWiFS conjoint-coverages were calculated for water-leaving radiances at corresponding bands (nLw_412 and nLw_443) and chlorophyll-*a* concentration. The statistics were also obtained within various subsets of daily global MODIS and SeaWiFS data, such as those limited to open ocean and clear atmosphere conditions. Water-leaving radiances were compared in low chlorophyll waters, chlorophyll below 0.1mg/m^3. Chlorophyll matchups were limited to data within the range from 0.05mg/m^3 to 50.0mg/m^3. Both MPD statistics were computed, assuming the error was a difference from the MODIS and SeaWiFS average and the error being a difference from the SeaWiFS benchmark. Figure 2.4 presents the MPD statistics between MODIS-Terra and SeaWiFS.

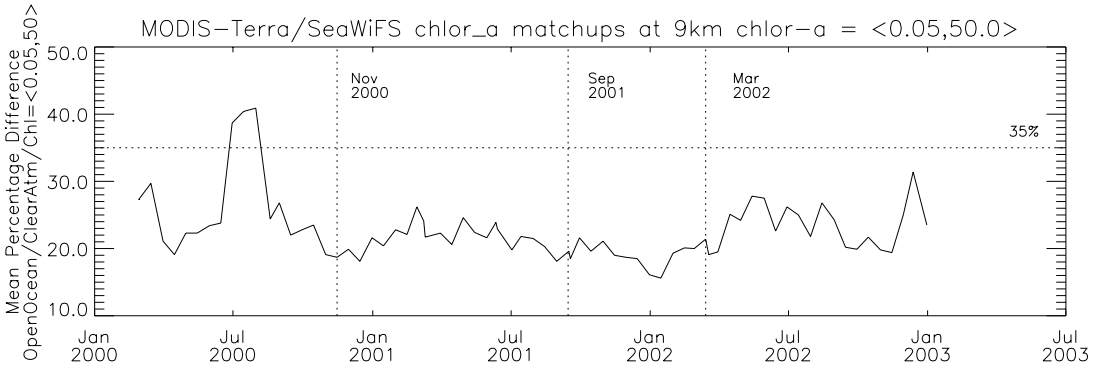
Figure 2.4 shows that the differences in water-leaving radiances between MODIS-Terra and SeaWiFS are persistently higher than the postulated objective of 5%. The discrepancies in chlorophyll-*a* concentration between the sensors meet the objective of 35% for chlorophyll limited to <0.05mg/m^3, 50.0mg/m^3> within the MODIS best calibrated period from November 2000 to September 2001 and all through to the end of 2002. The matchups indicate that MODIS-Terra data are not temporally stable compared to SeaWiFS. Chlorophyll divergence between MODIS and SeaWiFS varies seasonally with the highest differences being in the summer periods. The same seasonal patterns are prominent in the multiple statistics in the linear chlorophyll space displayed in Figure 2.2 and partially discernible from water-leaving radiance comparisons in Figure 2.4.

Both types of statistics in Figure 2.4, the error from the sensor average and the error from the SeaWiFS benchmark, show similar patterns in temporal discrepancies between MODIS-Terra and SeaWiFS data. The MPD values differ between the two calculations because of the presence or absence of MODIS data in the denominator. This is especially apparent for the chlorophyll MPD because globally, open ocean MODIS chlorophyll values are often higher than SeaWiFS, particularly in the southern hemisphere. If there is no requirement to compute errors relative to either sensor, $100\%RMS_{ln}$ can be a stable way to estimate percent discrepancies, although not very intuitive. When the chlorophyll range is not limited to the <0.05mg/m^3, 50.0mg/m^3> domain, such as in the plots in Figure 2.2, chlorophyll statistics are affected by differences in chlorophyll lower bounds in MODIS and SeaWiFS data (MODIS minimum chlorophyll is 0.01mg/m^3 and minimum SeaWiFS chlorophyll is 0.001mg/m^3) and by a sensor bias in low chlorophyll waters.

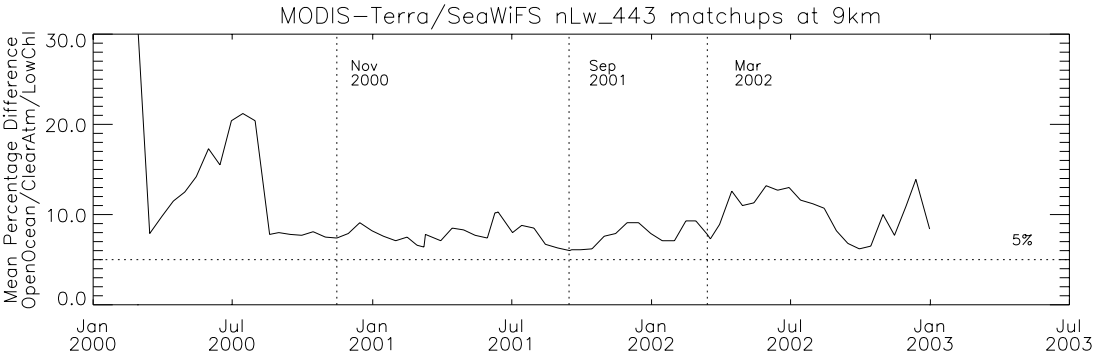
When these investigations were under way, there were only three months of available provisional-quality MODIS-Aqua collection 3 data. Those measurements were compared against SeaWiFS along with MODIS-Terra data for the same time period. Figure 2.5 contains the comparisons of MODIS-Aqua and MODIS-Terra nLw_443 and chlorophyll with corresponding SeaWiFS data. The mean daily global nLw_443 percent differences were calculated in low chlorophyll waters and the chlorophyll differences were limited to chlorophyll within the range <0.05mg/m^3, 50.0mg/m^3>. The statistics represent the error from the SeaWiFS benchmark.



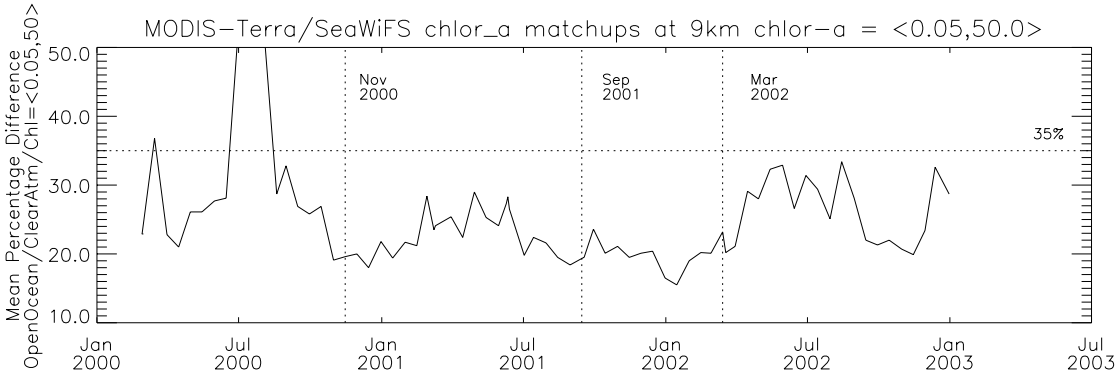
a



b

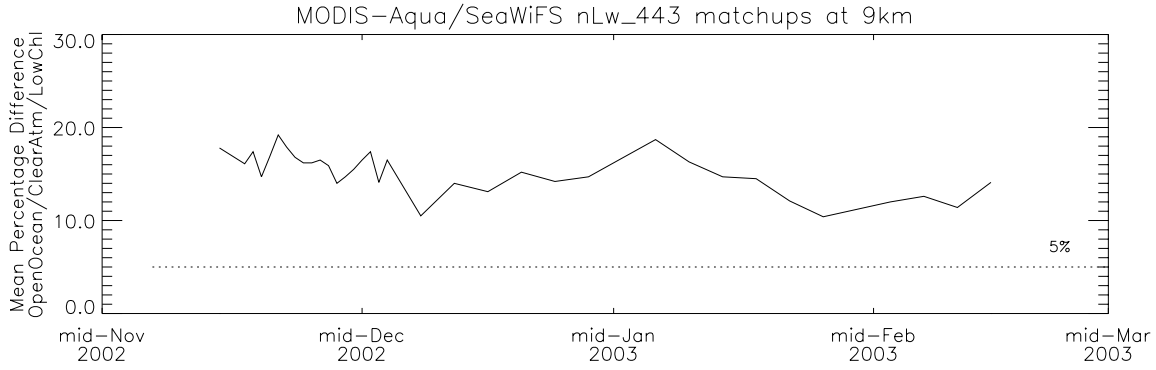


c

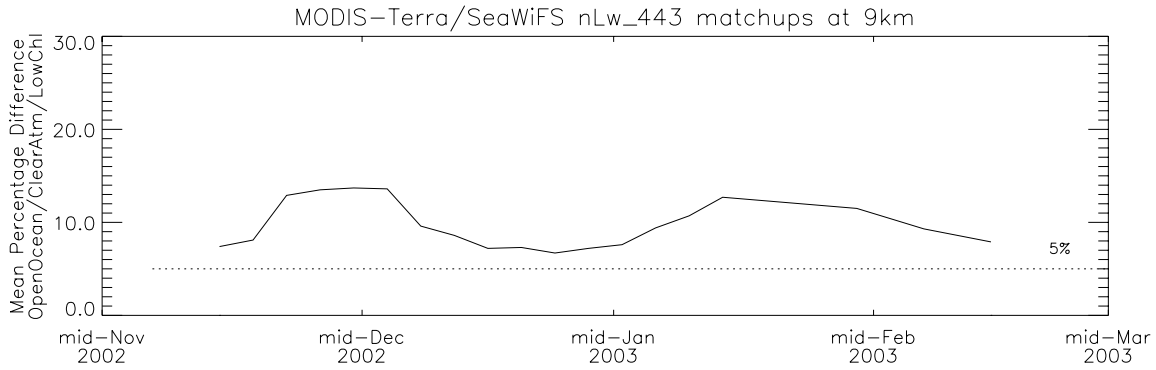


d

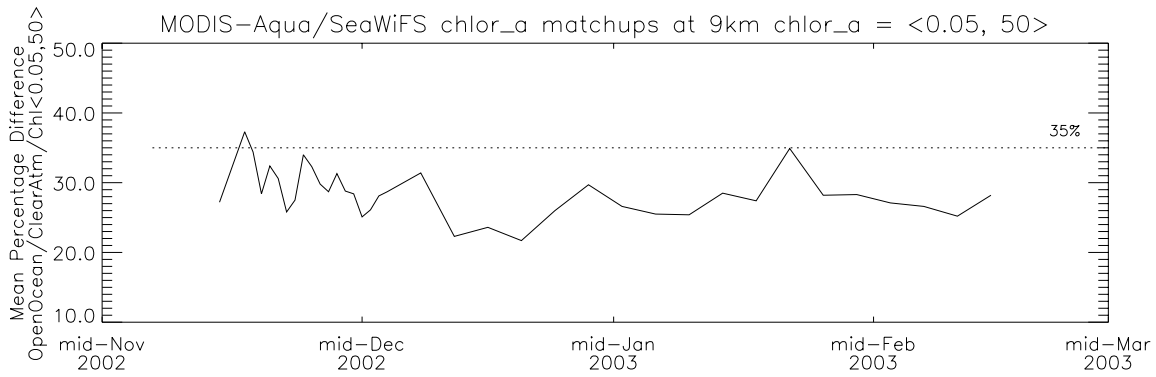
Figure 2.4: Time trends in mean percentage differences between MODIS-Terra and SeaWiFS nLw_443 and chlorophyll calculated over open ocean and clear atmosphere conditions and within corresponding chlorophyll ranges. Panel a & b shows the error from sensor average and panel c & d shows the error from SeaWiFS benchmark.



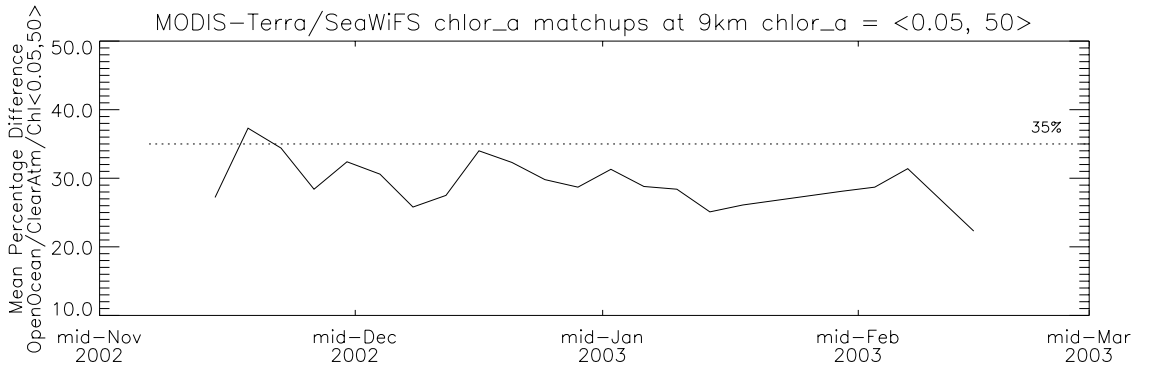
a



b



c



d

Figure 2.5: Time trends in mean percentage differences between MODIS-Aqua/Terra and SeaWiFS nLw_443 and chlorophyll calculated over open ocean and clear atmosphere conditions and within the corresponding chlorophyll ranges.

Figure 2.5 reveals that MODIS Aqua and Terra exhibit similar ocean data discrepancies from SeaWiFS measurements both in value and in temporal variability. However, it was difficult to draw comprehensive conclusions from the comparisons because the time series was relatively short and data from MODIS-Aqua and Terra were of provisional quality.

MODIS Terra and Aqua water-leaving radiances when compared with SeaWiFS did not therefore meet preliminary requirements for the consistent multi-sensor ocean color merged products. The less stringent requirements on chlorophyll-*a* concentration discrepancies between sensors were met within MODIS-Terra's best calibrated period of data. MODIS-Terra also revealed seasonal trends which prohibited the creation of temporally-stable merged data sets.

2.5 MODIS SCAN ANGLE DEPENDENCE

MODIS-Terra, being on a descending orbit, scans the cross-track Earth radiances from West to East. MODIS-Aqua on an ascending orbit scans from East to West. MODIS-Terra response versus scan angle (RVS) was not determined before launch and was found to cause asymmetry in the cross-track radiance profiles in all bands. RVS turned out to vary with the scan angle position within the MODIS field of view as well as with the chosen side of the scanning mirror. Although MODIS-Aqua RVS was measured before launch, it is subject to similar complex and detector-dependent deviations as MODIS-Terra RVS. An effort was made to eliminate RVS effects in MODIS ocean color data and the following study was performed to estimate residual MODIS scan angle dependencies in comparison with SeaWiFS L3 products. To investigate the dependencies, matchups with SeaWiFS were performed separately for MODIS data from the western and eastern parts of the MODIS scan. In the analyses, western and eastern MODIS scan edges were defined using MODIS zenith QA information explained in Section 2.2. SeaWiFS GAC data comprising L3 products used in the matchups were not free from cross-scan dependencies either. SeaWiFS radiances at short wavelengths fell off by up to 6% towards both edges of the GAC cutoff. However, SeaWiFS data corresponding to MODIS western and eastern scan edge coverage were averaged over numerous SeaWiFS angles within the daily global sets making SeaWiFS-viewing geometry statistically neutral. Additional comparisons were performed to validate the effects of SeaWiFS scan angle dependencies on the matchups against MODIS scan edge data. MODIS western scan edge data were matched against SeaWiFS data from western, central and eastern parts of the SeaWiFS scan and the SeaWiFS central part of the scan was compared against MODIS data from the western and eastern MODIS scan edges. The slope of the linear fit between corresponding data sets, the statistic which describes data discrepancies well, was plotted against time for both types of matchups and is shown in Figure 6. Because both SeaWiFS and MODIS daily data sets were limited to specific sensor zenith angles, there were numerous days in which were none or few overlapping bins between the two sensors for these angles. Figure 2.6a & b shows that within the remaining data the slope lines associated with SeaWiFS western, eastern, and central scan parts are similar and crisscross one another all through the time series. The slope lines associated with MODIS western and eastern scan edge data are visibly different through time. SeaWiFS data dependence on scan position was therefore concluded to be insignificant. In the following analyses, MODIS scan angle dependence was investigated against the entire span of SeaWiFS GAC swaths to obtain larger overlapping coverages between the sensors and produce more statistically valid results.

Matchups were specifically tailored to extracting MODIS scan effects and were evaluated against the entire SeaWiFS GAC scan coverage. Individual matchup statistics were calculated for each day, each product, and part of the MODIS scan. Figure 2.7 displays two examples of MODIS western and eastern scan edge data matchups with SeaWiFS.

Figure 2.7 illustrates that scatter plot distributions are different for MODIS data from western and eastern scan parts. These scatter plot differences are well described by matchup statistics, especially the parameters such as slope and intercept of the linear fit. These statistics were plotted through time and their temporal trends were further studied. Figure 8 presents slope and intercept statistics between MODIS-Terra/Aqua and SeaWiFS data for MODIS data subsetted into western and eastern scan edge coverages. Figure 2.8 shows that matchup slopes and intercepts are persistently different between MODIS western and eastern scan edge data for most of both time series: MODIS-Terra and SeaWiFS, and MODIS-Aqua and SeaWiFS. For MODIS-Terra, the differences with SeaWiFS exhibited seasonal patterns. MODIS-Terra western-to-eastern scan edge discrepancies were significant throughout most of the year but ceased being discernible in the summer. Summer period scatter plots of MODIS-Terra and SeaWiFS data were further investigated, such as in Figure 2.9.MODIS eastern scan edge water-leaving radiances in matchups with SeaWiFS for summer periods formed parallel double-scatter distributions, such as shown in Figure 9. These double distributions were separated and plotted as two individual classes of measurements on global standard mapped images. The scatter corresponding to higher in value MODIS East scan data were located in the northern hemisphere and scatter corresponding to lower-in-value MODIS East scan data for the same SeaWiFS radiances were located in the southern hemisphere. This indicated that MODIS data discrepancies from SeaWiFS had latitudinal character in addition to the differences between MODIS western and eastern scan edge data. Concluding, residual scan angle dependence was present in MODIS data and appeared in matchups with SeaWiFS. MODIS-Terra cross-scan variabilities were persistent through time and had seasonal patterns with strong latitudinal trends in summer periods. Limited analyses of MODIS-Aqua data showed scan angle dependencies similar to these in MODIS-Terra data.

2.6 MODIS LATITUDINAL DEPENDENCE

Dependence in MODIS ocean color data on zonal location was investigated through matchups with SeaWiFS for separate global data subsets corresponding to the northern and southern hemispheres. In the analyses, northern and southern hemisphere data were extracted using bin latitude information explained in Section 2.2. Individual matchup statistics were calculated for each day, each product, and for both hemispheres. Figure 2.10 displays two examples of individual northern and southern hemisphere matchups between MODIS-Terra and SeaWiFS data.Figure 2.10 illustrates that scatter plot distributions are different between the northern and southern hemispheres. These scatter plot differences were described well by matchup statistics, especially the parameters such as slope and intercept of the linear fit. These statistics were plotted through time and their temporal trends were further studied. Figure 2.11 presents the slope statistics between MODIS-Terra/Aqua and SeaWiFS data subsetted into northern and southern globe coverages. Figure 2.11 illustrates that matchup slope lines for MODIS-Terra/Aqua and SeaWiFS data are different between the northern and southern hemispheres for most of both time series. This indicated that MODIS data retained latitudinal dependence compared to SeaWiFS measurements. To establish the tendency of the latitudinal bias, mean data differences were extracted between MODIS-Terra and SeaWiFS data for each day, each product, and for both hemispheres. The comparison was repeated by limiting the data to low chlorophyll waters to confirm the results, since it is possible that each hemisphere could have a different proportion of higher chlorophyll ocean. Figure 2.12 contains the two matchups. Figure 2.12 demonstrates that, while northern hemisphere water-leaving radiance differences between MODIS-Terra and SeaWiFS data are generally stable and average close to zero, especially for nLw_443, southern hemisphere discrepancies show pronounced seasonal patterns. In the southern hemisphere, MODIS water-leaving radiances were lower than SeaWiFS in the austral winter and higher than SeaWiFS in the austral summer. The seasonal trend also occurred in chlorophyll differences between the sensors. MODIS-Terra southern hemisphere chlorophyll was typically higher than SeaWiFS and the difference was the largest in the austral winter. In the same period of the austral winter and the northern summer, MODIS northern hemisphere chlorophyll was lower than SeaWiFS.

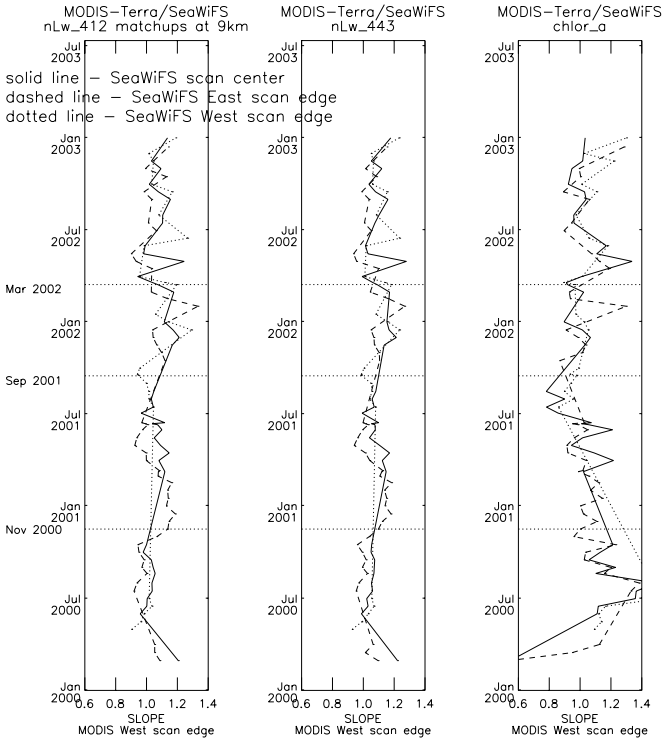


Figure 2.6a: Slope MODIS-Terra and SeaWiFS MODIS western scan edge data matched against SeaWiFS western, central and eastern part of the scan Time trends in the slope of the linear fit between MODIS-Terra and SeaWiFS data for individual parts of MODIS and SeaWiFS scan-angle coverage versus data from the most stable portions of the scan from each sensor. Slope MODIS-Terra and SeaWiFS SeaWiFS central part of the scan data matched against MODIS western and eastern scan edge data.

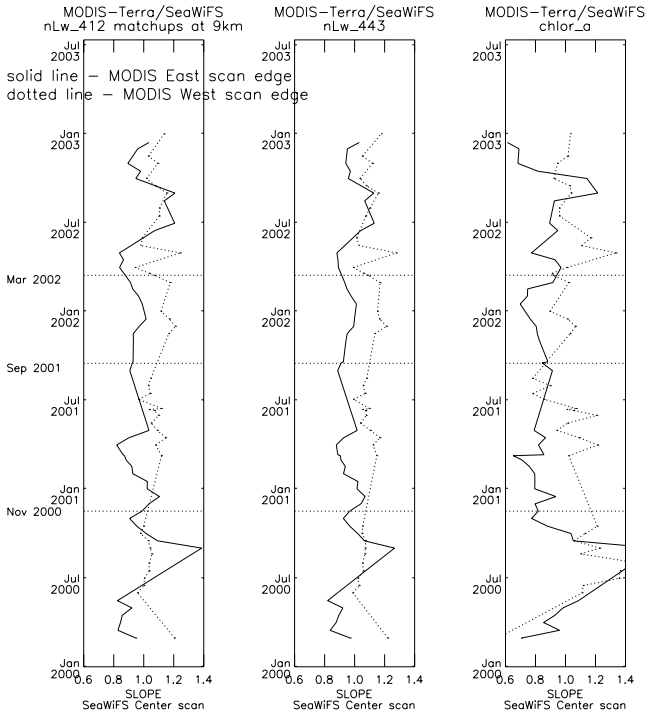


Figure 2.6b: Time trends in the slope of the linear fit between MODIS-Terra and SeaWiFS data for individual parts of MODIS and SeaWiFS scan-angle coverage versus data from the most stable portions of the scan from each sensor.

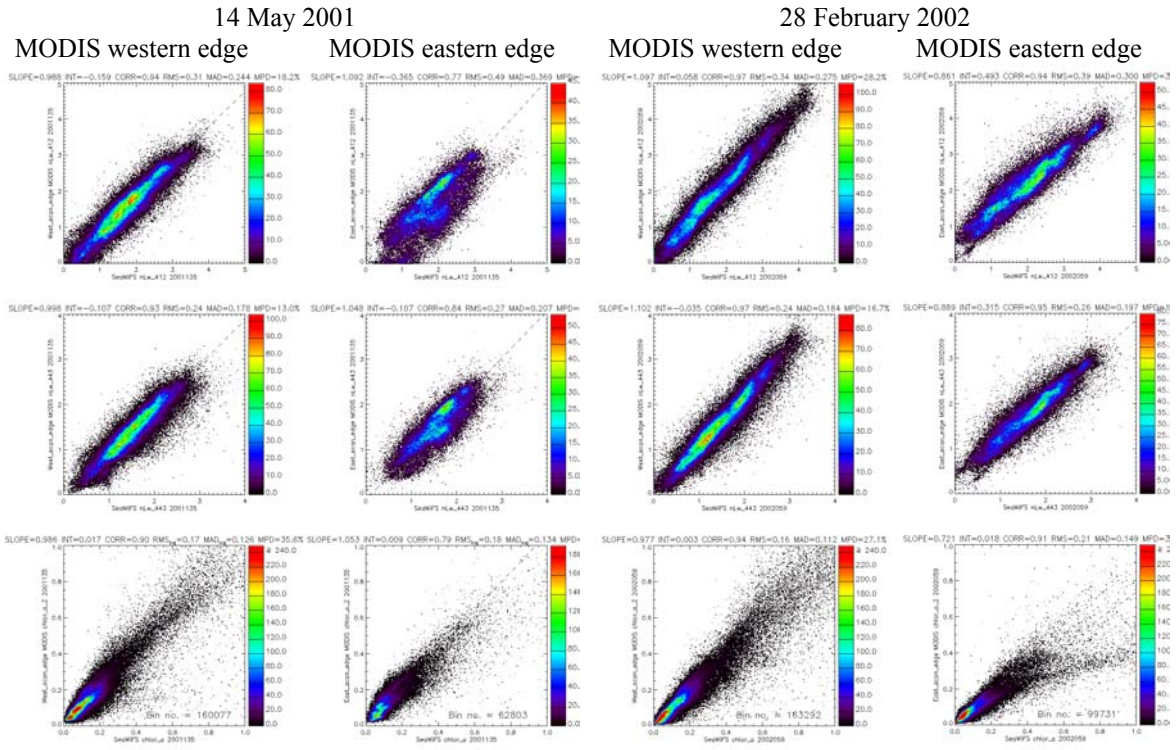


Figure 2.7: Examples of daily matchups with SeaWiFS for MODIS-Terra data from the western and eastern edges of the MODIS scan.

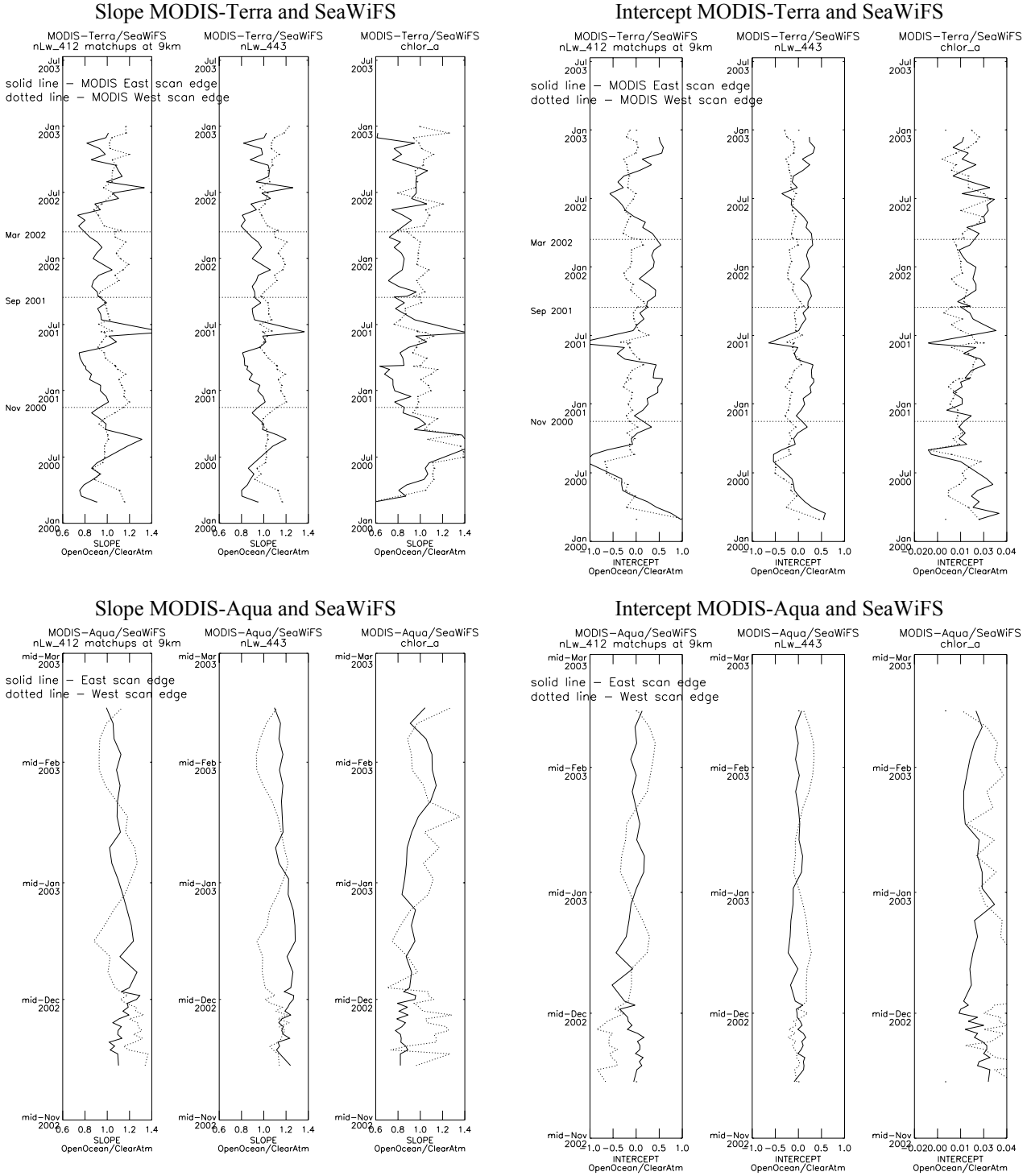


Figure 2.8: Time trends in the slope and intercept of the linear fit between MODIS-Terra/Aqua and SeaWiFS data for separate western and eastern MODIS data scan coverages.

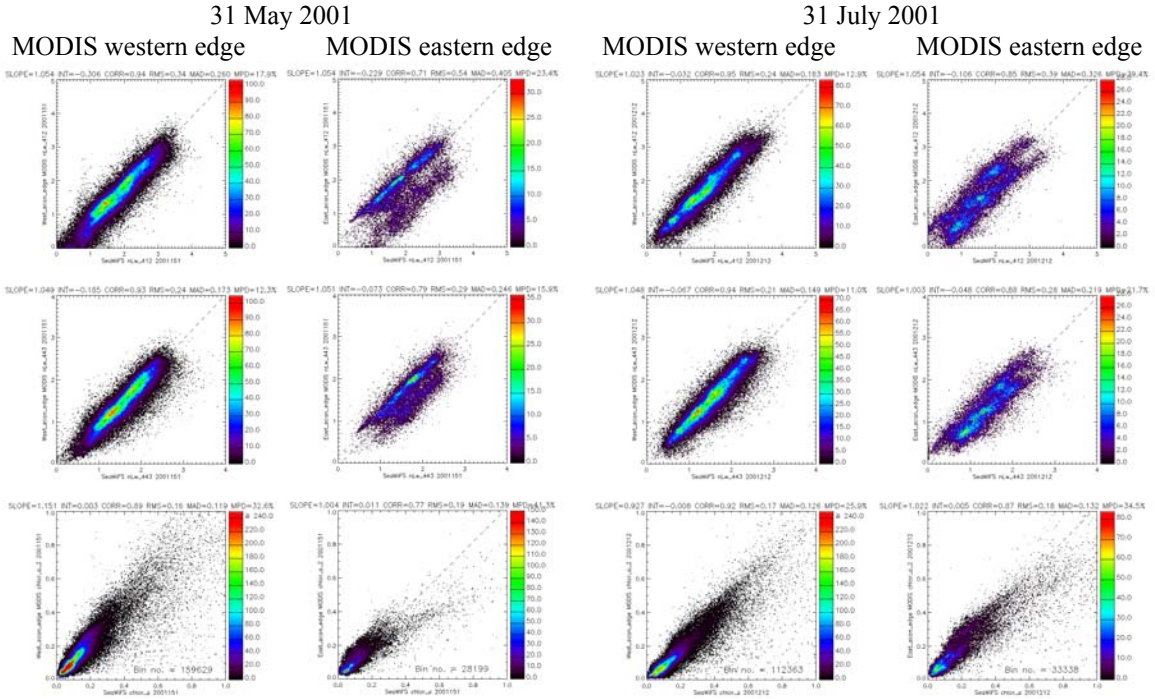


Figure 2.9: Examples of summer period daily matchups with SeaWiFS for MODIS-Terra data from the western and eastern edges of the MODIS scan.

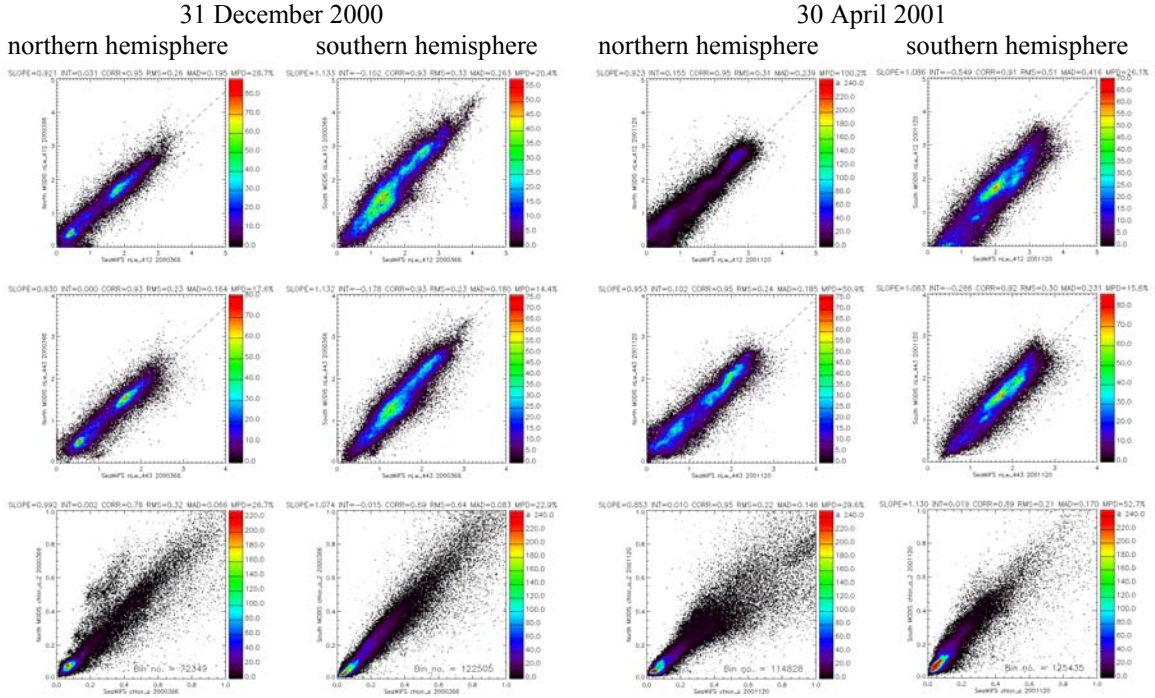


Figure 2.10: Examples of daily matchups of MODIS-Terra and SeaWiFS data from the northern and southern hemispheres.

To complete MODIS scan angle and latitudinal dependence analyses, matchups with SeaWiFS were individually obtained for four global data subsets, western and eastern MODIS scan edge coverage in northern and southern hemispheres. The slope statistics for the four-subset matchups are shown in Figure 2.13.

Figure 2.13 shows that MODIS data - from both Terra and Aqua - compared to SeaWiFS exhibit coupled scan angle and latitudinal dependencies which are all different and should be investigated individually. Trends in MODIS eastern scan edge data varied between the northern and southern hemispheres and so did MODIS western scan edge trends. MODIS-Terra northern hemisphere data in comparison with SeaWiFS demonstrated least degree of seasonal patterns, especially in the eastern part on the scan. The seasonality was, however, prevalent in southern hemisphere MODIS-Terra data and the trends differed between western and eastern scan edge data. From the limited MODIS-Aqua time series, the most stable, both temporally and zonally, in comparison with SeaWiFS were MODIS-Aqua eastern scan edge data for water-leaving radiances and chlorophyll.

Thus, MODIS data compared with SeaWiFS revealed latitudinal dependencies with a strong seasonality present in southern hemisphere water-leaving radiances and chlorophyll. In both MODIS Terra and Aqua data, latitudinal dependencies were coupled with cross-scan variabilites.

2.7 OVERLAPPING SCAN LINES BETWEEN MODIS-TERRA AND SEAWIFS

In order to obtain a more visual and intuitive evaluation of MODIS-Terra and SeaWiFS variability across the scan, individual overlapping scan lines between the two sensors were extracted and compared. This study severely limited the range of aerosol and in-water conditions so as to reliably compare individual scan lines between the sensors. This is because the two instruments are flown 1½ hours apart over changing atmospheric and ocean environments. The scan lines were limited to those taken over open ocean between 15^0 and 30^0 latitude in the northern and southern hemispheres, in areas corresponding to ocean gyres. Open ocean coverage was selected for clear atmosphere, aerosol optical thickness bellow 0.2, and low chlorophyll-*a* concentration, below $0.1 mg/m^3$. Overlapping scan lines between MODIS-Terra and SeaWiFS were extracted for days on which the swath overlay between the sensors was the widest.

MODIS-Terra and SeaWiFS instruments are both flown on descending, sun-synchronous, near-polar, circular orbits at 98.2^0 inclination. MODIS' orbit is kept constant while SeaWiFS' orbit is changing as the satellite's altitude is allowed to degrade slowly with time. Therefore, swath overlap between both sensors varies with time. Figure 2.14 shows MODIS-Terra and SeaWiFS swath phase differences. When the phase difference is near zero, swaths from both sensors overlap the closest. Within the three years of a concurrent MODIS-Terra and SeaWiFS operation, there were a number of days for which swath overlay was the widest between both instruments. Three of these days were investigated: 21 August 2001, 9 January 2002, and 2 June 2002.

To facilitate these analyses, L3 bins, not individual scan lines, were matched between MODIS and SeaWiFS. These but L3 bins were spread across each sensor swath corresponding to cross-track scans from 98.2^0 inclination orbits. The bins were at 4.6km resolution for both MODIS and SeaWiFS, where the original pixel resolution was 1km for MODIS and 4.5km for SeaWiFS GAC. MODIS bin locations within the sensor swath were established from MODIS QA L3 SatelliteZenith data sets available at 4.6km resolution. Those were analyzed concurrently with MODIS L3 ocean color data. To obtain SeaWiFS zenith angle information, sensor viewing angles were binned to L3 4.6km global daily products, senz and sena. The algorithm searched for bins along the scan lines with large amounts of good quality overlapping data present from MODIS and SeaWiFS within the swath. Once extracted, northern and southern hemisphere scan lines were separated and an average linear fit to each sensor and each hemisphere's scan-line data was computed across sensor zenith angles and over all scans. Figure 2.15 exhibits individual northern and southern hemisphere scan lines of water-leaving radiances and chlorophyll between MODIS and SeaWiFS, together with the linear fits in data across sensor zenith angles for 21 August 2001.

The 4.6km data displayed in Figure 2.15 are a product of binning the original 1km-resolution MODIS pixels and 4.5km-resolution SeaWiFS GAC pixels. MODIS binned data therefore are smoother across the scan than SeaWiFS because they are a result of multi-pixel averaging. SeaWiFS bins mostly accumulate just a single GAC pixel and, without the averaging, SeaWiFS cross scan data distribution is more jagged, especially for close-to-zero water-leaving radiances at longer wavelengths and chlorophyll. The atmospheric and ocean conditions for this analysis were chosen to minimize any variation across the swath for MODIS-Terra and SeaWiFS data. Despite that, Figure 2.15 demonstrates that water-leaving radiances and chlorophyll have disparate trends along the scan between the two sensors. The trends are most visible in southern hemisphere water-leaving radiances and to a lesser degree in northern hemisphere data. Linear fit in SeaWiFS water-leaving radiance distribution is relatively flat across the scan. Compared to SeaWiFS, MODIS radiances undergo a biased transition from the western part of the MODIS scan, negative zenith angles, to the eastern part of the scan, positive zenith angles. Analyses of individual overlapping scan lines between MODIS-Terra and SeaWiFS thus confirmed the presence of scan angle dependence in MODIS data. Plots of the scan line distributions for the following two days of MODIS-Terra and SeaWiFS close swath overlay can be found on the web site: http://simbios.gsfc.nasa.gov/~ewa/SeaMODISTerra/overlapp-scanlines.html.

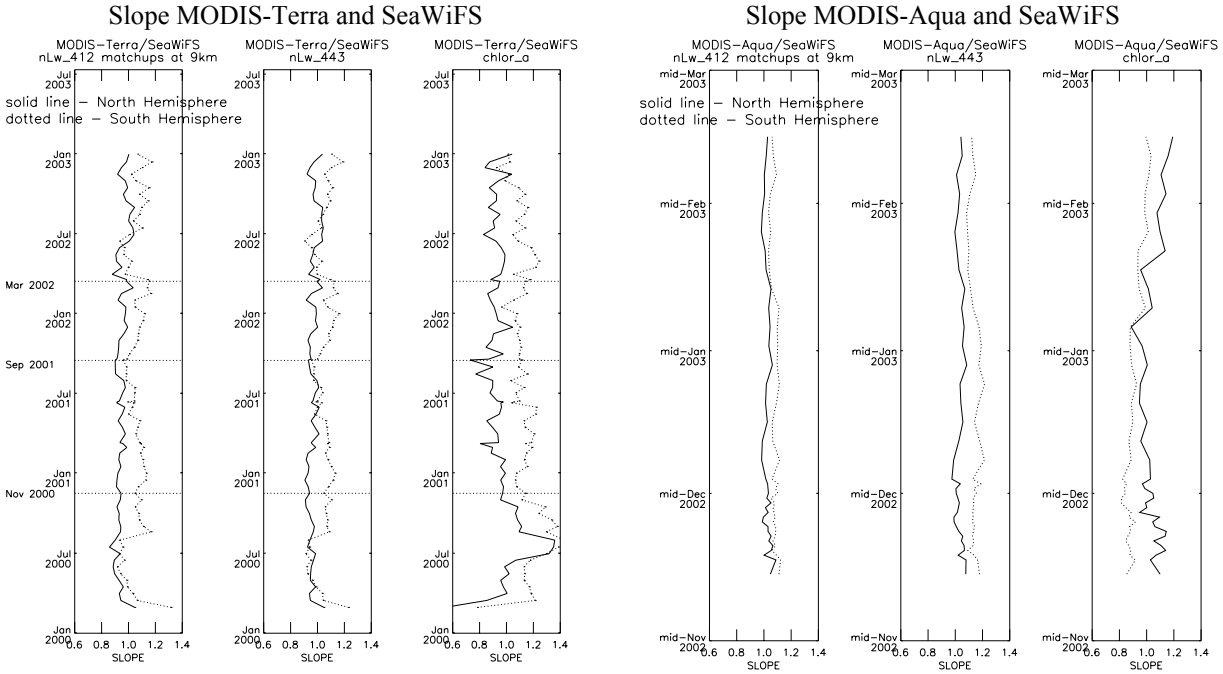


Figure 2.11: Time trends in the slope of the linear fit between MODIS-Terra/Aqua and SeaWiFS data for separate coverages from northern and southern hemispheres.

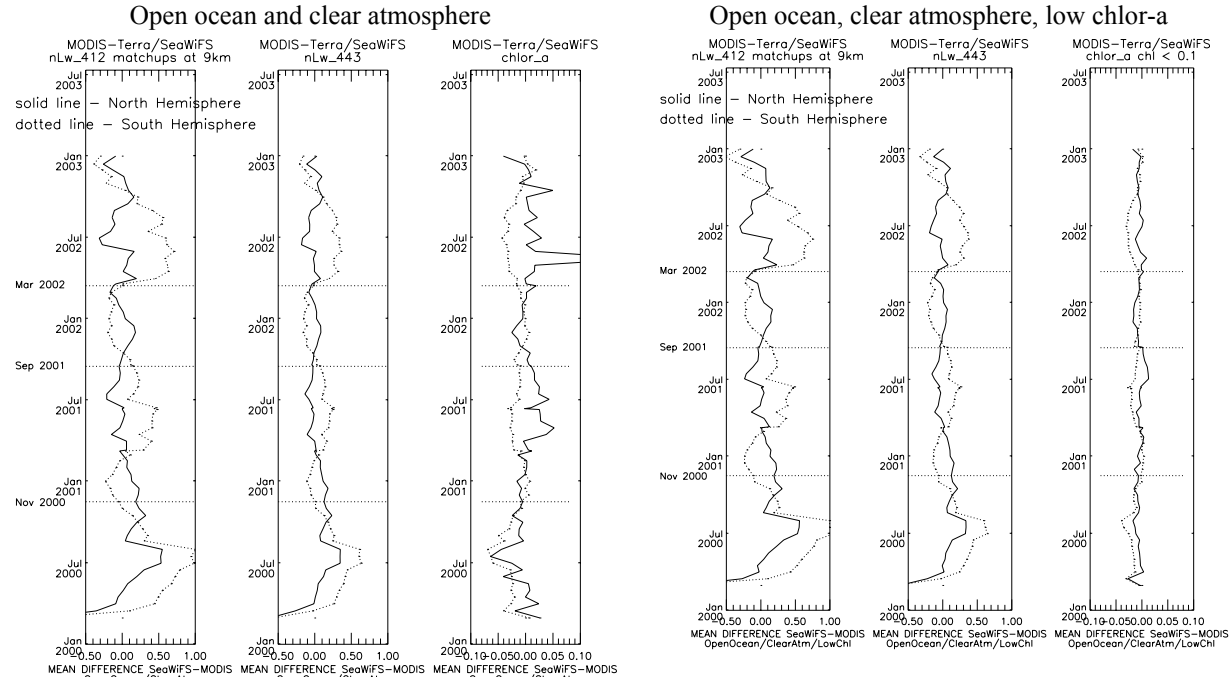


Figure 2.12: Temporal trends in mean differences between MODIS-Terra and SeaWiFS data obtained individually for northern and southern hemisphere coverage. Negative values indicate that MODIS data are higher than SeaWiFS and positive values denote that SeaWiFS data are higher than MODIS.

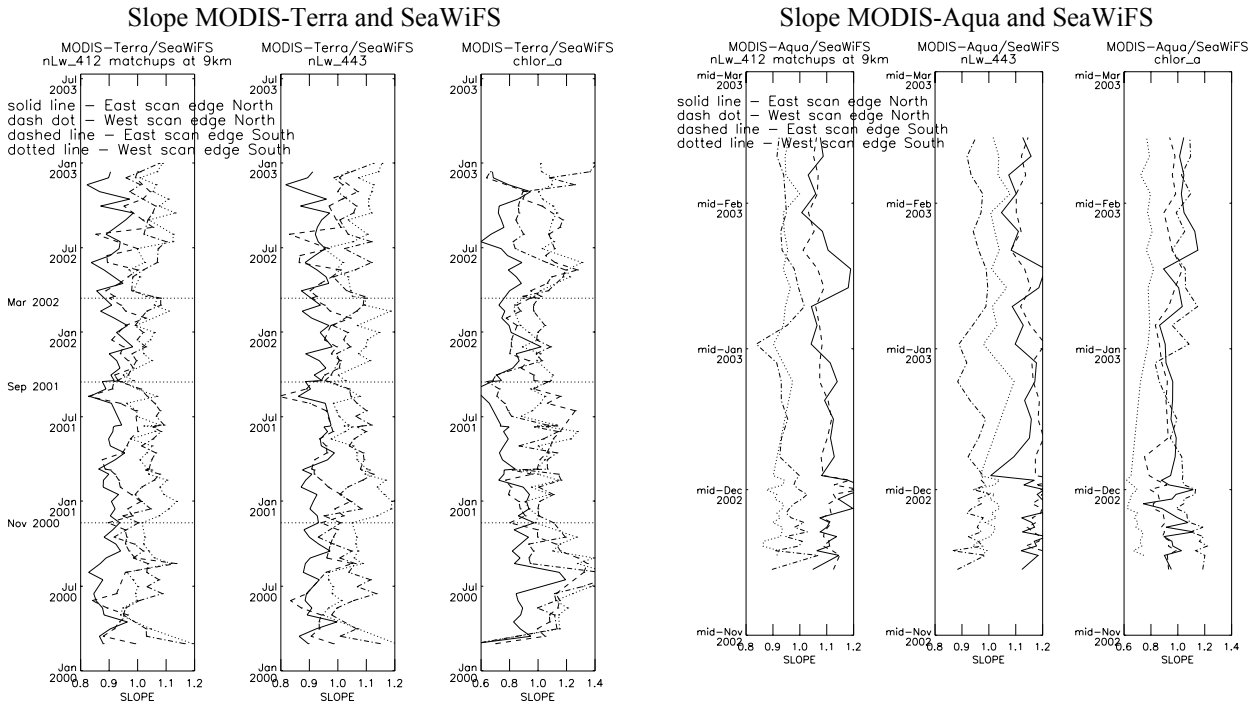


Figure 2.13: Time trends in the slope of the linear fit between MODIS-Terra/Aqua and SeaWiFS data for separate coverages corresponding to MODIS western and eastern scan edge data in the northern and southern hemispheres.

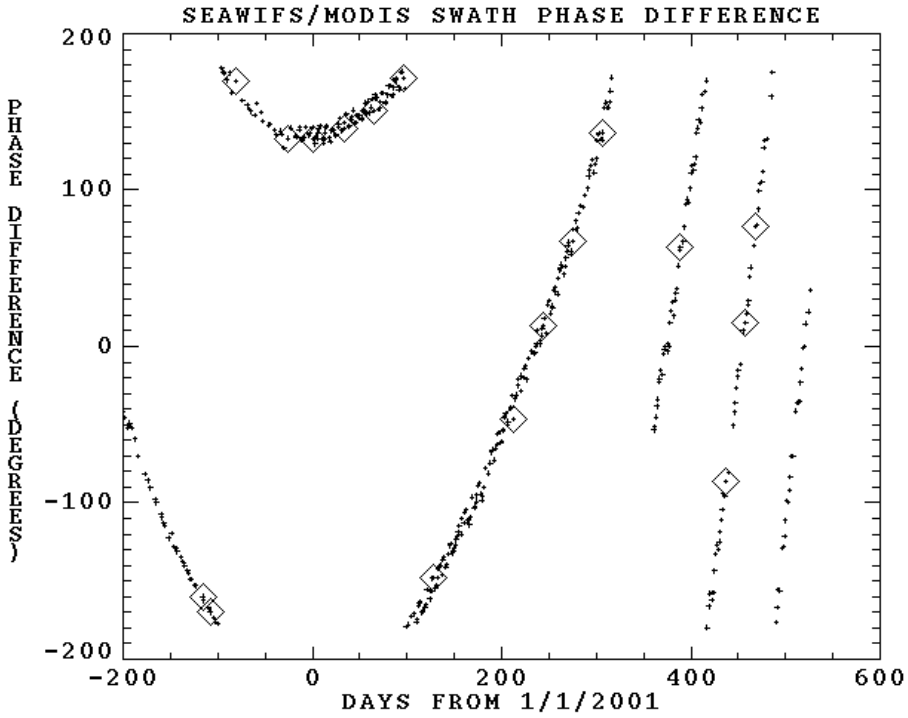


Figure 2.14: Swath phase differences between MODIS-Terra and SeaWiFS across time.

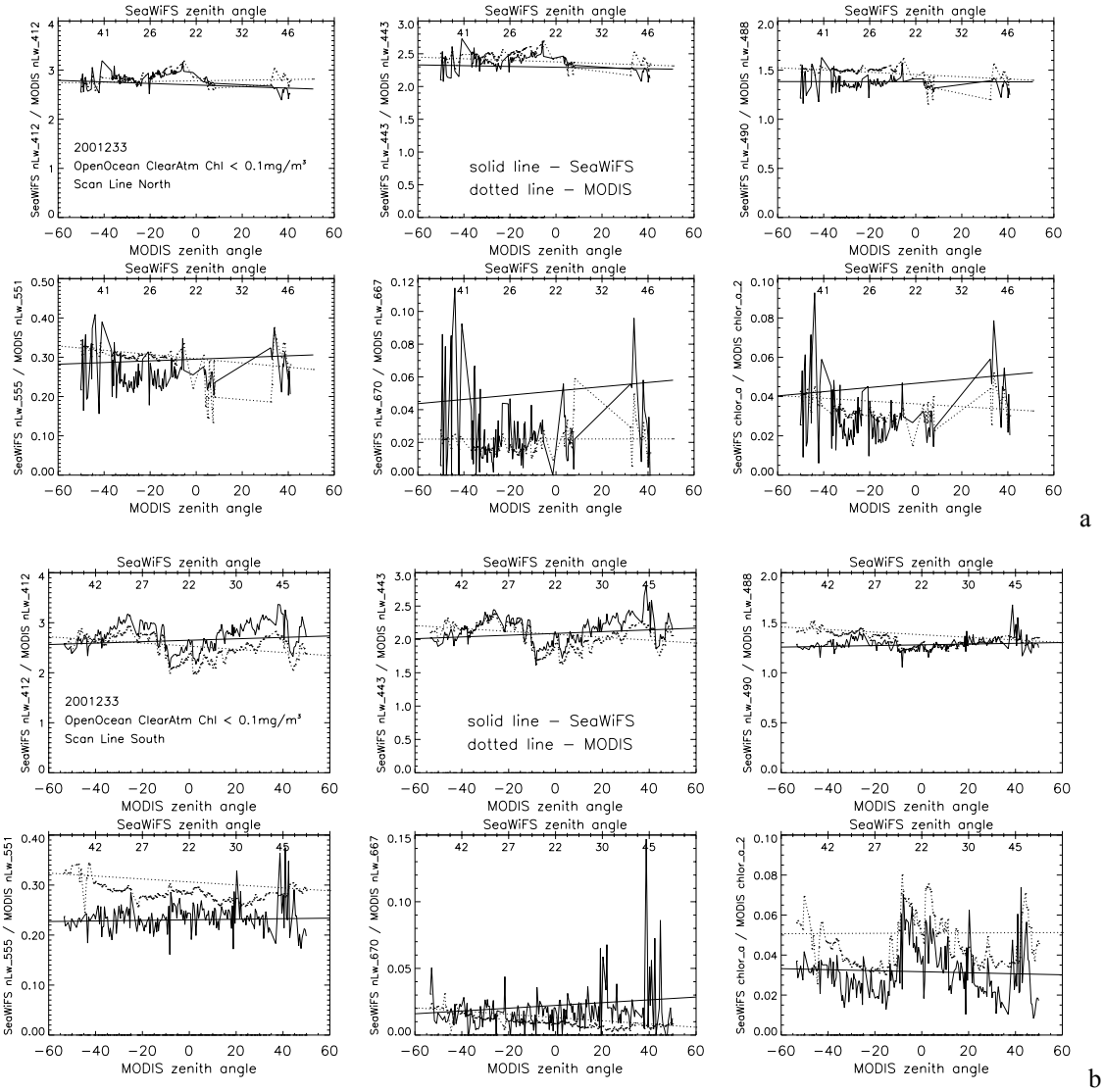


Figure 2.15: Individual MODIS-Terra and SeaWiFS overlapping scan lines at 4.6km bin resolution and linear fit trends across all extracted scan line data for each hemisphere versus sensor zenith angles (a -northern hemisphere; b-southern hemisphere). MODIS data contributing to plotted 4.6km bins were of the original 1km pixel resolution and SeaWiFS GAC data were of 4.5km pixel resolution.

2.8 DAILY GLOBAL COVERAGE IMPROVEMENTS, MODIS-TERRA, MODIS-AQUA, AND SEAWIFS

Improvements in daily coverage of global oceans were investigated by combining MODIS-Terra, MODIS-Aqua, and SeaWiFS data. The improvements were calculated over 9km resolution binned data, where 9km is the SeaWiFS standard bin size. Only good quality, quality 0, MODIS Terra and Aqua data were used. SeaWiFS data were of standard L3-mask quality. Statistics of MODIS chlorophyll coverage used the chlor_a_2 product while SeaWiFS statistics used the corresponding chlor_a product. Two types of statistics were computed: improvement in daily global chlorophyll coverage and improvement in daily global ocean color coverage. The difference between the statistics is caused by the fact that some MODIS bins may contain good-quality chlorophyll data along with water-leaving radiances at certain bands or ocean atmospheric properties of a lesser quality. Consequently, in calculating MODIS daily global ocean color coverage it was assumed that all the following products

had good quality within the bins: nLw_412, nLw_443, nLw_488, nLw_531, nLw_551, nLw_667, nLw_678, Tau_865, Eps_78, K_490, and chlor_a_2.

Adding SeaWiFS data to either MODIS Terra or Aqua chlorophyll coverage improved MODIS daily global chlorophyll coverage by around 40%. Adding SeaWiFS data to MODIS complete ocean color coverage increased MODIS-Terra daily global coverage by around 55% and MODIS-Aqua global coverage by around 45%. The combination of MODIS Terra and Aqua daily global coverage improved single MODIS instrument chlorophyll by 55% and complete ocean color coverage by 65%. Addition of SeaWiFS to combined MODIS Terra and Aqua data increased MODIS Terra ∪ Aqua daily global chlorophyll coverage by around 15% and complete ocean color coverage by 20%. Plots corresponding to each of these statistics can be found on the web site: http://simbios.gsfc.nasa.gov/~ewa/SeaMODISTerra/coverage_improvment.html.

Figure 2.16 shows daily global chlorophyll coverage from MODIS-Terra, MODIS-Aqua, and SeaWiFS instruments. The coverage was obtained relative to the extent of global oceans and inland waters defined by SeaWiFS L3 masks as water bins. At 9km resolution, for which the analysis was performed, there were 3812408 water bins.

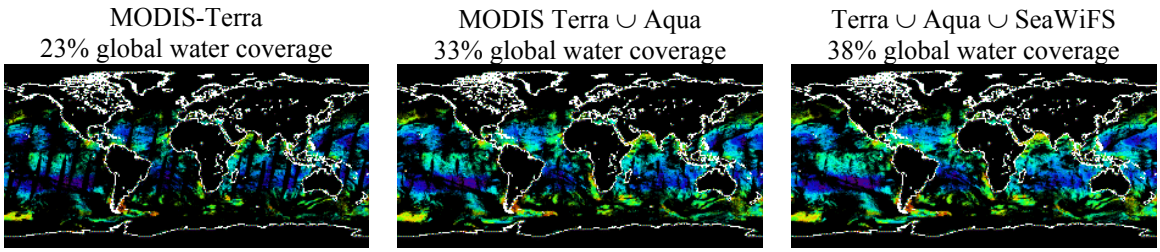


Figure 2.16: Average percentage of daily global ocean and inland water coverage from MODIS-Terra, MODIS-Aqua, and SeaWiFS instruments as defined by SeaWiFS water masks at 9km resolution.

REFERENCES

Kilpatrick, K., E. Kearns, E. J. Kwiatkowska-Ainsworth, and R. L. Evans, 2002:Time Series of Calibrated Ocean Products from NASA's Moderate Resolution Scanning Spectrometer (MODIS), *Proceedings of the Ocean Sciences Meeting*, Honolulu, Hawaii, USA.

Press, W.H.,, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: Numerical Recipes in C.The Art of Scientific Computing, *Cambridge University Press*.

Campbell, J.W., 1995: The Lognormal Distribution as a Model for Bio-optical Variability in the Sea, *J. Geophys. Research*, Vol. **100**, no. C7, 13,237-13,254.

Hooker, S.B., W.E. Esaias, G.C. Feldman, W.W. Gregg, and C.R. McClain, 1992: An Overview of SeaWiFS and Ocean Color. *NASA Tech. Memo. 104566*, Vol. 1, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 24 pp., plus color plates.

Chapter 3

A Long-term Intercomparison of Oceanic Optical Property Retrievals from MODIS-Terra and SeaWiFS

Bryan A. Franz

Science Applications International Corporation, Beltsville, MD

3.1 INTRODUCTION

The work presented here is a comparative analysis of mean global and regional oceanic optical property retrievals from two independent, spaceborne ocean color sensors: the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), and the MODerate resolution Imaging Spectroradiometer (MODIS). The SeaWiFS instrument has been in continuous operation since September of 1997, while the MODIS instrument, flying on the Terra spacecraft, has been collecting data since March of 2000. With the recent reprocessings of both instrument data sets, there now exists over three years of consistently processed, contemporaneous MODIS and SeaWiFS data available through the Goddard Distributed Active Archive Center (GDAAC), providing an unprecedented opportunity for intercomparison of global ocean color retrievals from two independent sources. This study looks at the temporal trends in several ocean color products derived from SeaWiFS and MODIS to evaluate the long-term relative stability between the two sensors and develop an understanding of their similarities and differences. The time-series analysis looks at variations in the mean value of normalized water-leaving radiance and chlorophyll products over the period from 12 March 2000 through 31 December 2002, for both global and regional geographic areas. Results are presented in the form of temporal overlays for common products, as well as product ratios as a function of time.

3.2 DATA SOURCES

The SeaWiFS data used in this analysis were standard, 9-km-resolution, Level-3 time-binned products from the 4th reprocessing, composited over 8-day periods. The MODIS data were standard, 4.6-km-resolution, Level-3 products from MODIS/Terra Oceans Collection 4.0, binned over the same 8-day periods. These Level-3, weekly data products for both SeaWiFS and MODIS are currently available from the GDAAC. It should be noted that some of the MODIS data used in this study are considered provisional. Due to the extensive, on-orbit characterization required to calibrate MODIS for ocean data processing, all data collected after the MODIS Oceans Collection 4.0 reprocessing (after March 19, 2002) are not fully corrected. Data collected prior to November 2000 are also considered provisional, due to the instability of the spacecraft and instrument during the first year of the Terra mission.

Several changes to the MODIS data were required to enable a bin-for-bin match-up with SeaWiFS. The first step was to convert the MODIS files to SeaWiFS-like Level-3 bin format. This was simply a reorganization of the HDF fields, as the SeaWiFS and MODIS formats use the same, sinusoidal binning approach. At this step, specific MODIS products were associated with standard SeaWiFS products, and any necessary unit conversions were performed. Only MODIS quality zero (QL=0) data were retained. The MODIS products chlor_a_2, nLw_412, nLw_443, nLw_488, and nLw_551, were associated with SeaWiFS products chlor_a, nLw_412, nLw_443, nLw_490, and nLw_555, respectively. The band associations are summarized in Table 3.1. Note that the algorithm for the chlor_a_2 product of MODIS (OC3M algorithm, O'Reilly et al., 2000) is very similar to that of the chlor_a product from SeaWiFS (OC4v4 algorithm, O'Reilly et al., 2000). The second step was to reduce the MODIS 4.6-km bin file to 9-km resolution, equivalent to standard SeaWiFS Level-3 bin resolution. This is effectively a 4-to-1 spatial averaging, weighted by the number of observations within each 4.6-km bin. The final step was to reduce the MODIS and SeaWiFS 9-km bin files to common bins. For a given 8-day period, only those bins that were filled in both the MODIS and the SeaWiFS files were retained in the final analysis. Filled bins are those for which one or more QL=0 retrievals were acquired.

3.3 SUBSET DEFINITIONS

With 8-day composited SeaWiFS and MODIS data products in an equivalent form, the data sets were further divided into several geographic subsets. Three global subsets were defined, corresponding to clear water, deep water, and coastal water. The

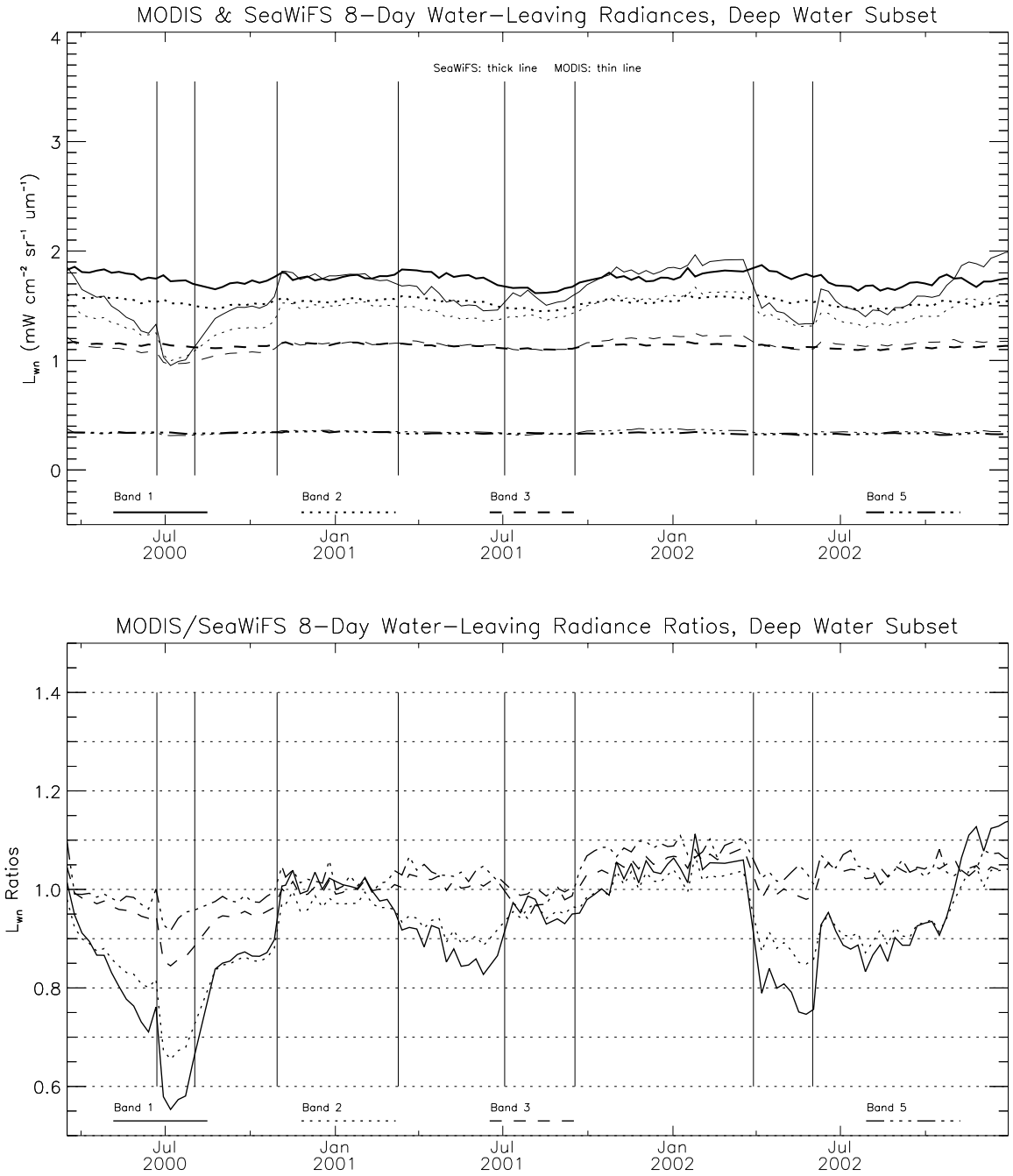


Figure 3.1: MODIS and SeaWiFS normalized water-leaving radiance trends, March 2000 through December 2002. Different wavelength-bands are indicated by different line types. The upper panel shows MODIS and SeaWiFS trends as an overlay, with SeaWiFS indicated as the thick line and MODIS as the thin line. The lower panel shows the ratio of common bands between the two sensors. The solid vertical lines indicate epoch dates in the MODIS Oceans calibration and characterization coefficients.

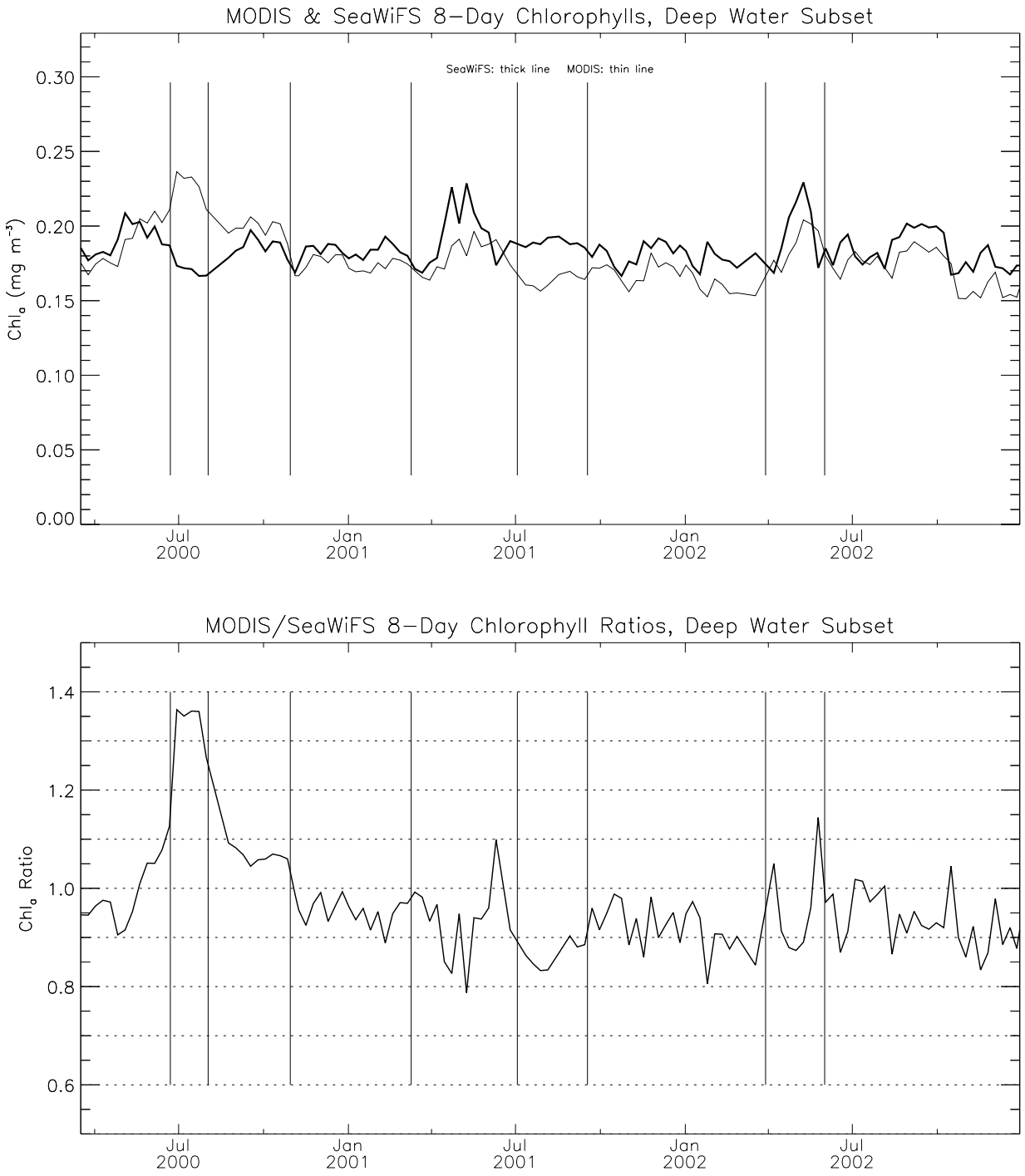


Figure 3.2: MODIS and SeaWiFS chlorophyll trends, March 2000 through December 2002. The upper panel shows the MODIS and SeaWiFS trends as an overlay, with SeaWiFS indicated as the thick line and MODIS as the thin line. The lower panel shows chlorophyll ratio, with MODIS normalized by SeaWiFS. The solid vertical lines indicate epoch dates in the MODIS Oceans calibration and characterization coefficients.

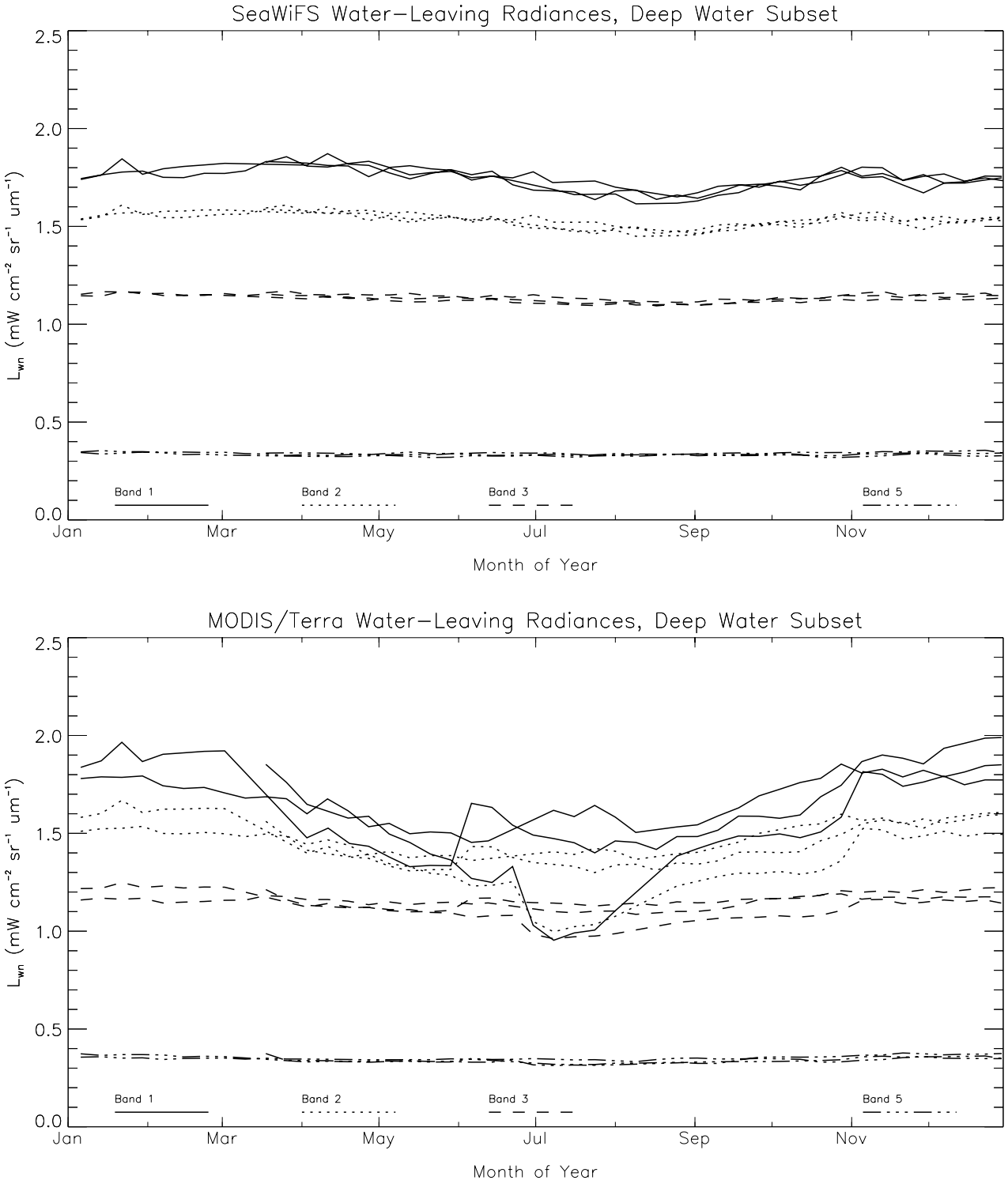


Figure 3.3: MODIS and SeaWiFS normalized water-leaving radiance trends plotted to show the repeatability in the annual cycle. Different wavelength-bands are indicated by different line types. The upper panel shows the SeaWiFS and the lower panel shows MODIS.

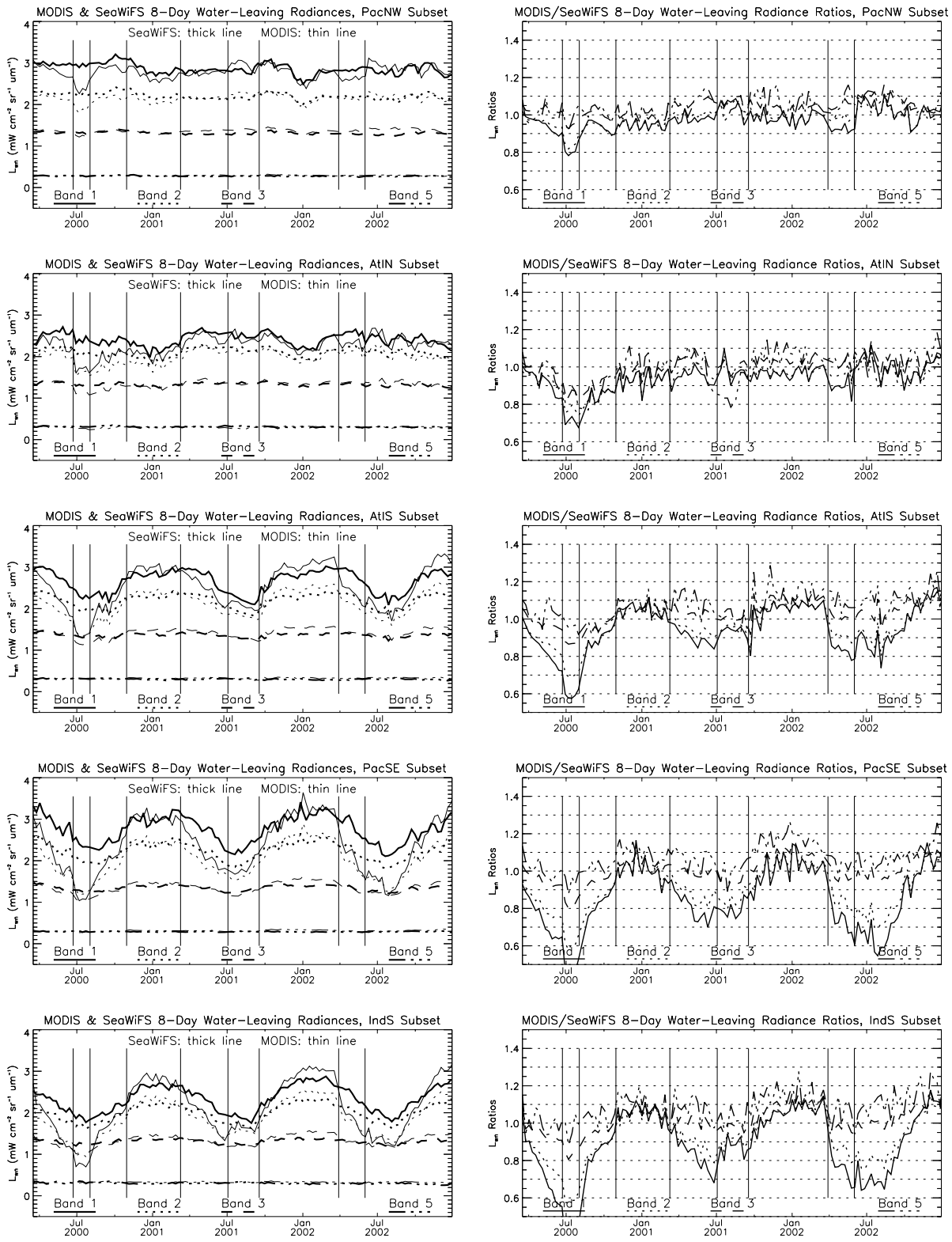


Figure 3.4: MODIS and SeaWiFS normalized water-leaving radiance trends for the regional subsets. The left column shows the overlay of MODIS and SeaWiFS trends, with SeaWiFS indicated by the thicker line. The right column shows the radiance ratios between the two sensors. Wavelength bands are indicated by different line types.

deep-water subset consists of all bins where water depth is greater than 1000 meters. Clear water was defined as deep water where the retrieved chlorophyll is less than 0.15 mg m^{-3}. For the clear-water test, both SeaWiFS and MODIS retrievals were required to be below the chlorophyll threshold. Coastal water was defined as all bins where water depth is between 50 and 1000 meters, as defined by a shallow water mask and the deep water mask. Some caution should be exercised when comparing the clear-water subsetted data, as anomalously high chlorophyll retrievals from either sensor can significantly alter the geographic distribution of selected bins. In contrast, the deep-water and coastal subsets are purely geographic in selection criteria. The coastal subset, however, is more likely to contain regions of significant variability in water structure and atmospheric conditions, as well as Case-2 water types (Morel and Prieur, 1977) for which the bio-optical algorithms are invalid. These effects can be expected to increase retrieval uncertainty and thus result in larger differences between the two sensors. The deep-water subset is, therefore, the most stable subset for cross-sensor comparison of retrieved oceanic optical properties. The geographic extent of all three global subsets will vary, however, with the seasonal change in earth illumination and thus sensor imaging duty cycle.

In addition to the global subsets, six basin-scale subsets were analyzed. These included regions in the northern Pacific (PacN), northwestern Pacific (PacNW), southeastern Pacific (PacSE), northern Atlantic (AtlN), southern Atlantic (AtlS), and the southern Indian Ocean (IndS). In addition, a smaller region near Hawaii was defined. All of these subset regions were adopted from Fougnie et al. 2002, and their locations are listed in Table 3.2. Based on the results of the regional analysis, yet another group of subsets was defined to provide a systematic means for investigating latitudinally-dependent differences between the two sensors. A longitudinal segment of the Pacific from 170W to 150W was divided into 10-deg latitude zones. These zonal subsets are summarized in Table 3.3.

3.4 TRENDING ANALYSIS

For each sensor, for each 8-day product, the filled bins associated with a particular subset were identified and used to compute the mean, standard deviation, and average observation time. Figure 3.1 shows an example of a typical trend plot derived from this analysis. For the upper panel of Figure 3.1, 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. The plot in the lower panel shows the same data as a ratio, with MODIS means normalized by SeaWiFS means. Similarly, Figure 3.2 shows the chlorophyll trends for the same deep-water subset. The solid vertical lines in the temporal trend plots are provided as a reference to indicate the transitions between MODIS Oceans calibration epochs. These epochs are the independent periods over which MODIS/Terra calibration corrections were derived and implemented by the MODIS Oceans group at the University of Miami (RSMAS). In most cases, these periods correspond with the calibration epochs used by the MODIS Calibration Support Team (MCST) for the adjustment of the Level-1B radiances, and they usually correspond with spacecraft safe-hold events or significant instrument state changes.

3.5 DISCUSSION OF RESULTS

On average, the agreement between MODIS and SeaWiFS over the trended time-period is good. It is evident from the deep-water trend plots of Figure 1, however, that MODIS and SeaWiFS radiances deviate considerably in certain time-periods. Table 3.4 shows the mean and standard deviation of the global trends (i.e., the mean and standard deviation of the 8-day subset means). The table serves to illustrate both the good overall agreement and the higher time variability observed with MODIS. Note that at a wavelength of 551 nm, in clear water, MODIS shows a 4% temporal variability over the trend period (standard deviation relative to mean), while the equivalent SeaWiFS variability is just 2%. At shorter wavelengths, the difference is larger. There is significant evidence to suggest that the elevated temporal variability observed in the MODIS products, relative to SeaWiFS, is an artifact of the MODIS characterization and processing. First, the long-term temporal stability of SeaWiFS is well established. The SeaWiFS calibration team makes use of monthly lunar observations to track and correct for time-dependent drifts in detector response. Based on this lunar calibration, the temporal degradation of SeaWiFS has been found to be well characterized as an exponential decay, and the change in responsivity over time has been shown to be highly predictable (Eplee et al., 2003a). Furthermore, the stability of the water-leaving radiance products from either sensor can be independently tested by evaluating the repeatability of the seasonal cycles observed in the temporal trends. While it is possible that the differences between SeaWiFS and MODIS are geophysical, due to the 90-minute difference in node crossing time, it can be expected that such effects (e.g., bi-directional reflectance) would be repeatable from year to year in accordance with the seasonally changing distribution of solar and viewing angles. The instruments may differ from one another, but they should be self-consistent in the absence of any major geophysical event. The deep-water annual repeatability plots presented in Figure 3.3, however, show that, while SeaWiFS is consistent from year to year, MODIS is highly variable. It should also be noted that the deviations between the MODIS and SeaWiFS trends often change character at intervals associated with MODIS Oceans

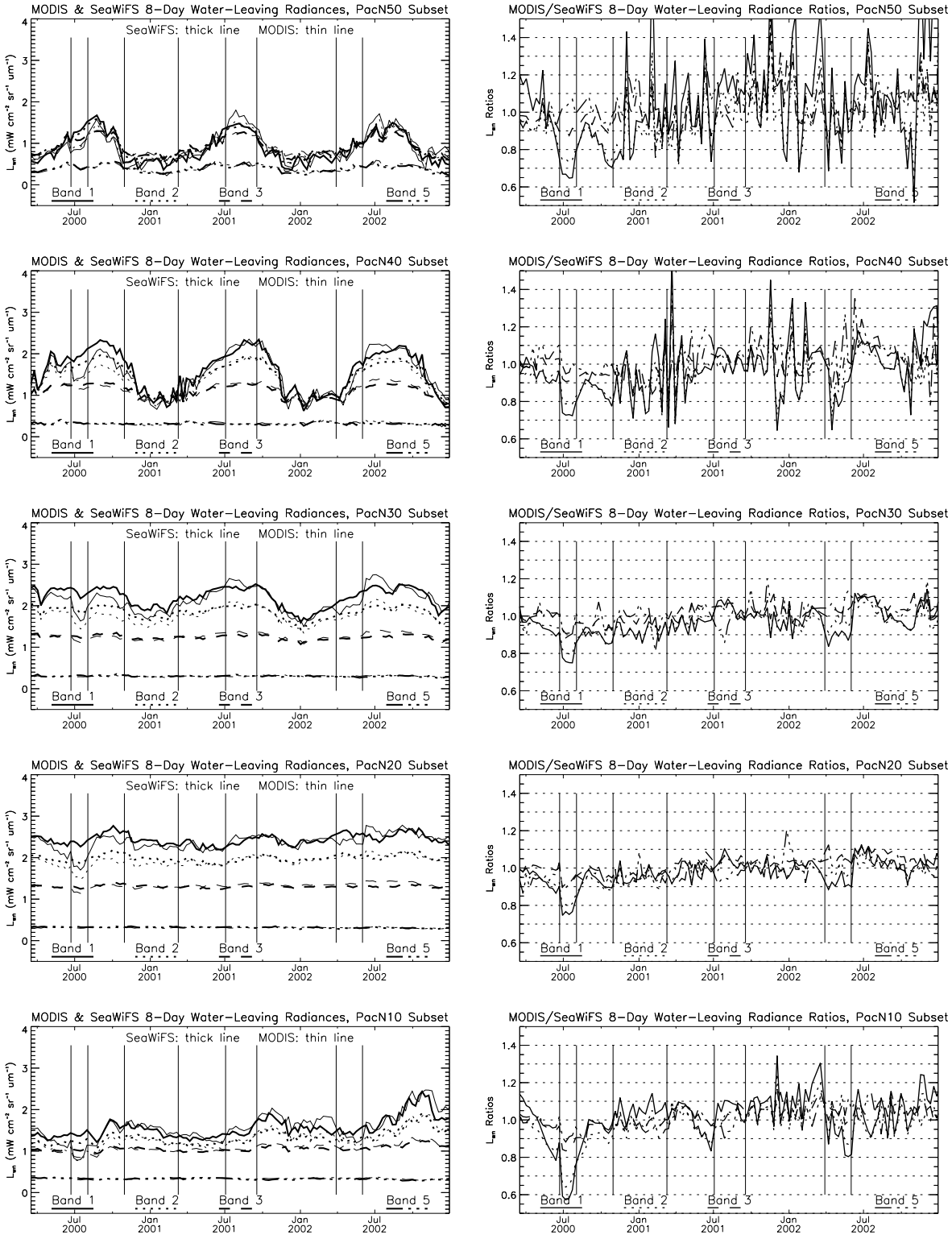


Figure 3.5a: MODIS and SeaWiFS normalized water-leaving radiance trends for the northern latitudinal zones of the Pacific. The left column shows the overlay of MODIS and SeaWiFS trends, with SeaWiFS indicated by the thicker line. The right column shows the radiance ratios between the two sensors. Wavelength bands are indicated by different line types.

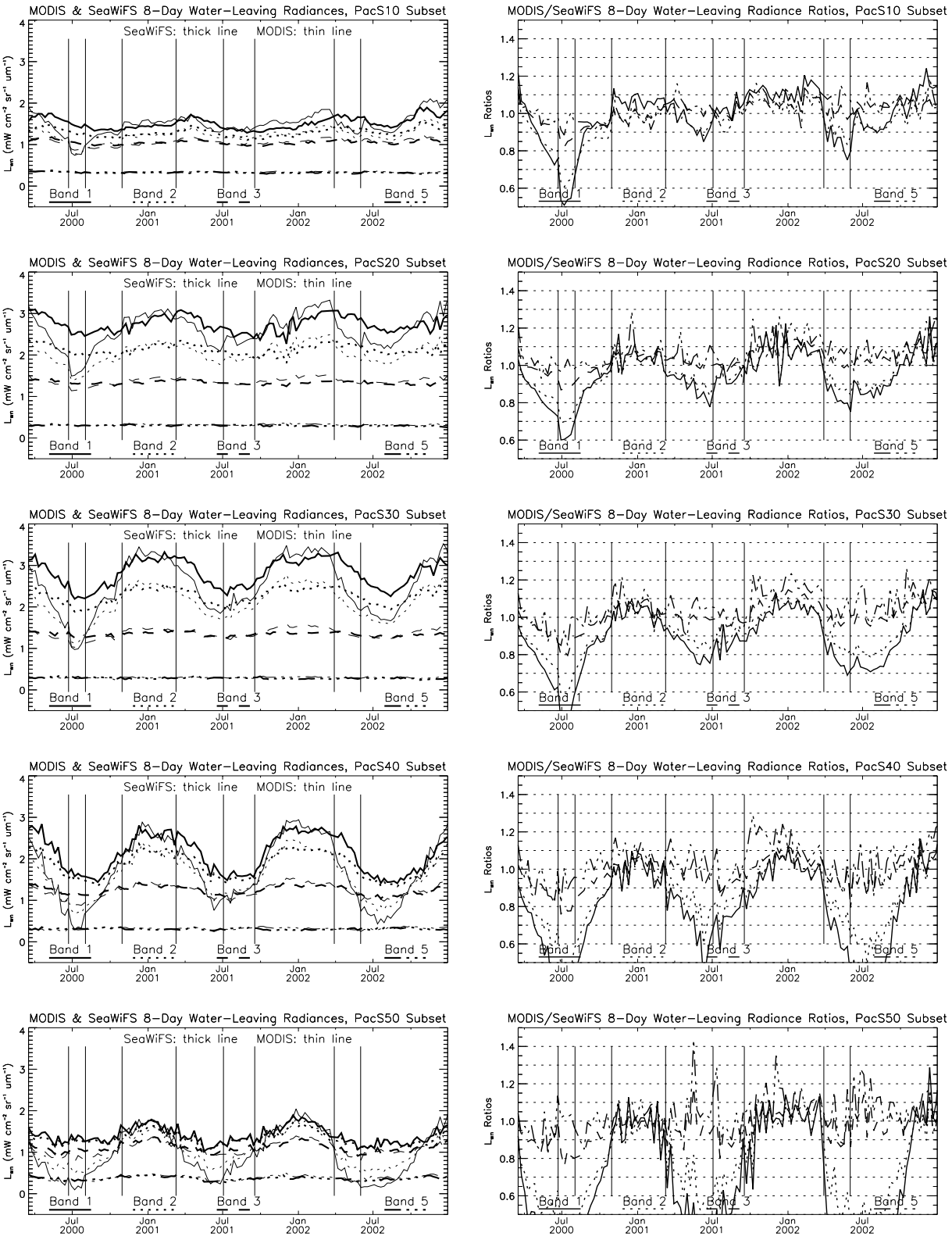


Figure 3.5b: MODIS and SeaWiFS normalized water-leaving radiance trends for the southern latitudinal zones of the Pacific. The left column shows the overlay of MODIS and SeaWiFS trends, with SeaWiFS indicated by the thicker line. The right column shows the radiance ratios between the two sensors. Wavelength bands are indicated by different line types.

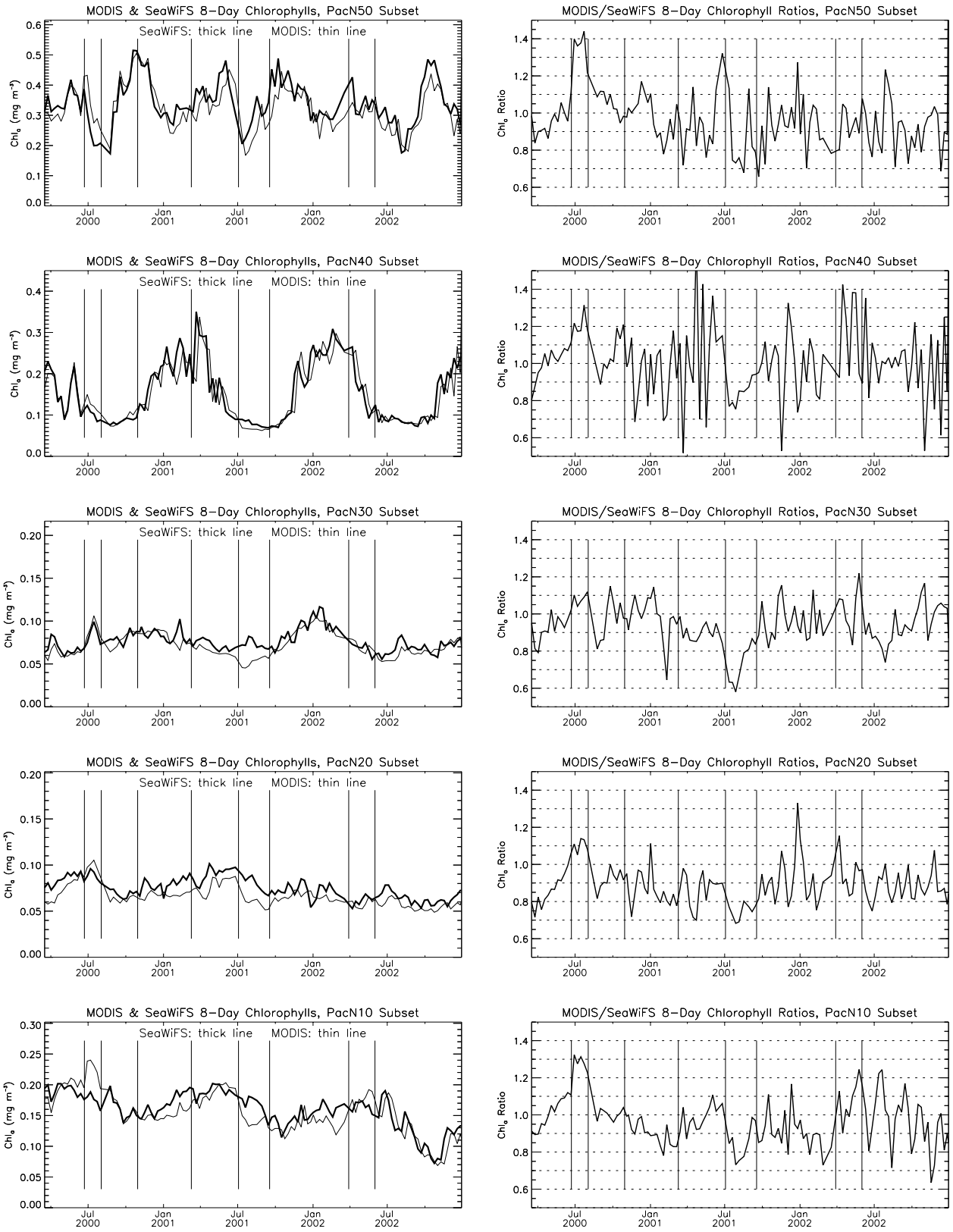


Figure 3.6a: MODIS and SeaWIFS chlorophyll trends for the latitudinal zones of the northern Pacific. The left column shows the overlay of MODIS and SeaWiFS trends, while the right column shows chlorophyll ratios between the two sensors.

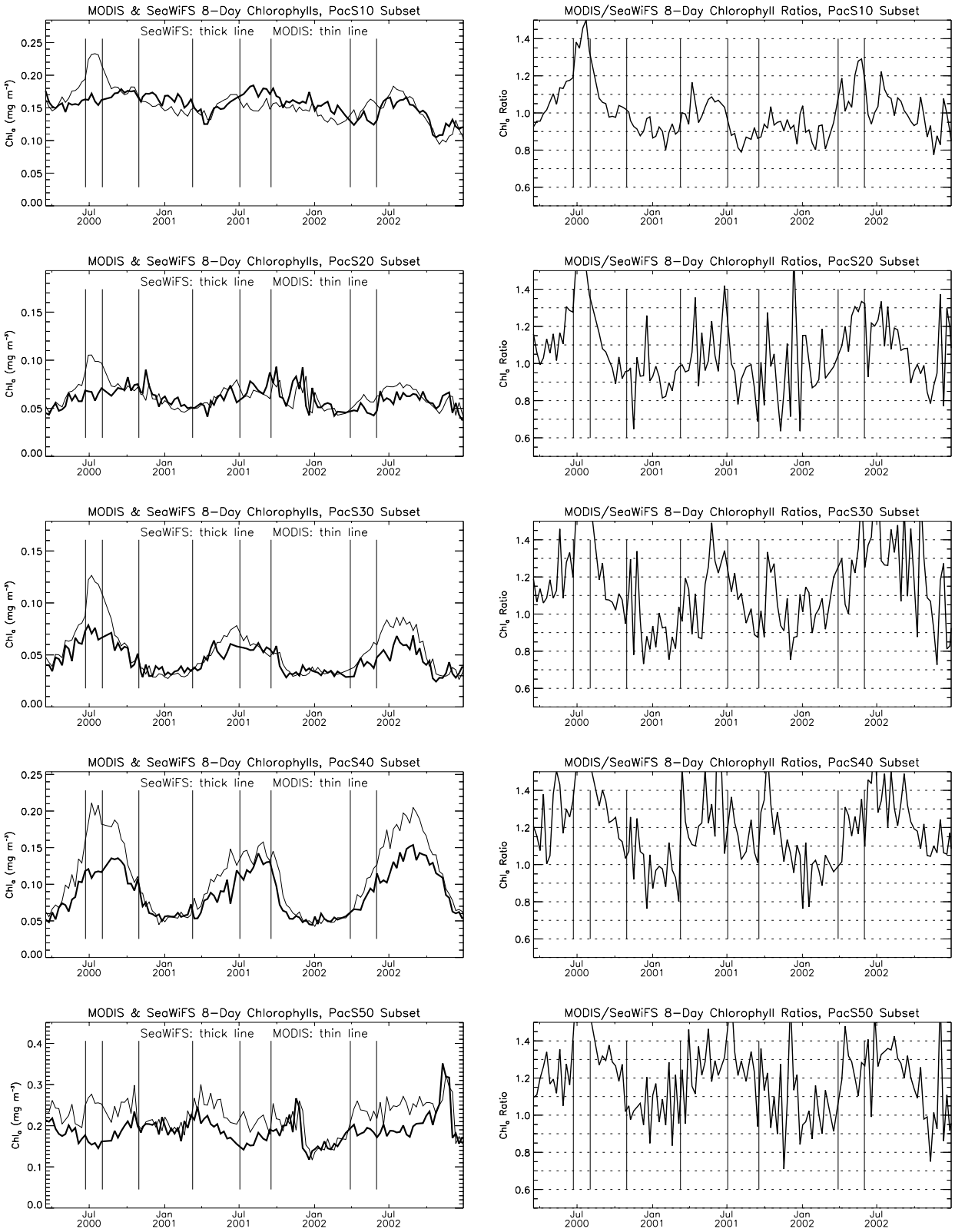


Figure 3.6b: MODIS and SeaWIFS chlorophyll trends for the latitudinal zones of the southern Pacific. The left column shows the overlay of MODIS and SeaWiFS trends, while the right column shows chlorophyll ratios between the two sensors.

calibration epochs. A good example of this coincidence is the discontinuity seen in the ratio trends of Figure 3.1 at the start of November 2000, a date that is directly related to the MODIS transition from A-side to B-side electronics. The large deviation that starts in late March of 2002 and ends at the end of May 2002 is known to be associated with a discrepancy between the measured solar diffuser calibration and the predicted Level-1B calibration factors applied at the time of processing (K. Kilpatrick, RSMAS, personal communication). This inconsistency between measured and predicted Level-1B calibration invalidated the in-flight Level-2 corrections and calibration factors derived previously by RSMAS. Based on the coincidence between MODIS calibration and processing changes and MODIS product deviations relative to SeaWiFS, and the observation that the seasonal cycles in the MODIS trends are inconsistent from year to year, it is likely that a significant portion of the temporal variability observed by MODIS is actually a calibration or instrument characterization artifact.

The regional subset trends provided in Figure 3.4 show that the relative agreement in retrieved water-leaving radiances between the two sensors varies geographically. The best agreement is found within the small Hawaii region and the two larger regions of the northern Pacific (of the three, only the PacNW region is shown in Figure 3.4, for brevity). This may be due to the fact that both SeaWiFS and MODIS are vicariously calibrated to the Marine Optical Buoy (MOBY), which is located near Lanai, Hawaii (Clark et al., 2001). The SeaWiFS calibration makes use of the MOBY measurements to derive a single gain adjustment for each band (Eplee et al., 2003b). The MODIS calibration includes a similar, overall gain correction within each calibration epoch, but it also makes use of the MOBY measurements and the relatively homogeneous waters of the northern Pacific to derive various instrument corrections. These corrections include mirror-side and scan-angle dependencies, and residual detector striping, all of which vary with time in accordance with the calibration epochs (Kearns et al., 2002). The regional trends suggest that the differences between the two sensors increase with distance from the common calibration region, which may be an indication that the MODIS calibration and post-launch instrument characterization is over-tuned to the northern Pacific region. The regions of the northern and southern Atlantic show many similarities with the global deep-water trends, but as the analysis progresses further south to the Indian Ocean region and the south eastern Pacific, the deviations between MODIS and SeaWiFS retrieved radiances increase, and a strong seasonality is evident in the ratio trends.

This apparent latitudinal dependence in the relative agreement between MODIS and SeaWiFS was the motivation for the zonal subset analysis, which is provided in Figure 3.5. The zonal trends clearly indicate that, as the evaluation progresses from the northern latitudes to the southern latitudes, the relative differences between MODIS and SeaWiFS increase and become strongly seasonal, with the largest differences occurring near the austral winter. Furthermore, the effect has a significant spectral dependence, with the blue bands of MODIS being significantly depressed relative to SeaWiFS. Unfortunately, there is little *in situ* data available to determine which instrument is more correct, but the zonal trends do show that, at southern latitudes below 30-deg south, the MODIS normalized water-leaving radiance measurements at 412 nm approach or even fall below the 551-nm radiances every July. Such a spectral dependence is normally associated with very turbid water, which is not common to the open oceans of the southern Pacific. The elevated seasonality in the water-leaving radiances retrieved by MODIS, relative to SeaWiFS, is most likely an artifact of the processing algorithms or instrument characterization. Based on discussion with RSMAS (R. Evans, personal communication), it is believed that the observed seasonal biases of the southern hemisphere may be due to limitations in the pre-launch characterization of polarization sensitivity for the MODIS/Terra mirror, which has significantly degraded since launch. RSMAS is currently exploring ideas for a post-launch re-characterization of the polarization sensitivity for MODIS/Terra.

The spectral dependence observed in the ratio trends indicates that MODIS/Terra Collection 4.0 chlorophylls will be biased high relative to SeaWiFS, for the Southern Hemisphere in the austral winter. This is illustrated in Figure 3.6, which shows the chlorophyll trends for the same zonal subsets. The relative bias in chlorophyll between SeaWiFS and MODIS for July in the southern Pacific is on the order of 50%, or 0.05 mg m^{-3} in absolute terms.

3.6 SUMMARY

A long-term, contemporaneous time-series of global and regional mean normalized water-leaving radiance and chlorophyll retrievals from SeaWiFS and MODIS/Terra was developed and analyzed. The results show that, while SeaWiFS and MODIS products are similar on average, significant differences can be found which correlate with time and location. Some of the largest deviations between the two data sets are directly associated with periods over which the MODIS calibration or processing was changed. This observation, coupled with the fact that the seasonal cycle in global water-leaving radiances measured by SeaWiFS is highly repeatable from year to year while the MODIS seasonal cycle is not, suggests that a significant portion of the temporal variability in the water-leaving radiances measured by MODIS is not geophysical. The regional and zonal trends show that the deviations between MODIS and SeaWiFS increase with increasing southern latitude, and the product ratios show a strong seasonality. The spectral dependence of the radiance ratios results in MODIS chlorophyll retrievals that are as much as 50% higher than SeaWiFS in the Southern Hemisphere, with the greatest differences occurring near the austral winter.

ACKNOWLEDGEMENTS

The author thanks the MODIS and SeaWiFS Projects and the GDAAC for the calibration, processing, and distribution of the MODIS and SeaWiFS data. Thanks also to Robert Evans, Edward Kearns, Katherine Kilpatrick, and Sue Walsh of RSMAS for many useful discussions on the MODIS instrument characterization and MODIS processing algorithms, and Sean Bailey and Jeremy Werdell of the SIMBIOS and SeaWiFS Projects for review of the analysis and many helpful insights. The software that converts MODIS Level-3 formats was developed by Joel Gales of the SIMBIOS Project. Much of the methodologies employed to derive the temporal trends from the Level-3 products were originally developed by Robert E. Eplee, Jr. for the SeaWiFS calibration and validation program.

REFERENCES

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.

Eplee, R.E., Jr., R.A. Barnes, and F.S. Patt, 2003a: Changes to the on-orbit 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, (in press).

Eplee, R.E., Jr., R.A. Barnes, S.W. Bailey, and P.J. Werdell, 2003b: 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, (in press).

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.

Kearns, E.J., R. Evans, K. Kilpatrick, D. Clark, 2002: Early Results from Aqua-MODIS. *Ocean Optics XVI*, Santa Fe, NM, November 18-22, 2002.

O'Reilly, J.E. et al., 2000: SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. *NASA Tech. Memo. 2000-206892,* Vol**. 11**, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, 49 pp.

Morel, A. and L. Prieur, 1977: Analysis of Variation in Ocean Colour, *Limnol. Oceanol*. **22**, 709-721.

Table 3.1: Band Correspondence (nm)

Band	SeaWiFS	MODIS
1	412	412
2	443	443
3	490	488
5	555	551

Table 3.2: Regional Subset Definitions

Region ID	Minimum Latitude	Maximum Latitude	Minimum Longitude	Maximum Longitude
Hawaii	18.0	19.9	-158.5	-156.5
PacN	15.0	23.0	-180.0	-159.4
PacNW	10.0	22.7	139.5	165.6
PacSE	-44.9	-20.7	-130.2	-89.0
AtlN	17.0	27.0	-62.5	-44.2
AtlS	-19.9	-9.9	-32.3	-11.0
IndS	-29.9	-21.2	89.5	100.1

Table 3.3: Zonal Subset Definitions

Region ID	Minimum Latitude	Maximum Latitude	Minimum Longitude	Maximum Longitude
PacN50	40.0	50.0	-170.0	-150.0
PacN40	30.0	40.0	-170.0	-150.0
PacN30	20.0	30.0	-170.0	-150.0
PacN20	10.0	20.0	-170.0	-150.0
PacN10	0.0	10.0	-170.0	-150.0
PacS10	-10.0	0.0	-170.0	-150.0
PacS20	-20.0	-10.0	-170.0	-150.0
PacS30	-30.0	-20.0	-170.0	-150.0
PacS40	-40.0	-30.0	-170.0	-150.0
PacS50	-50.0	-40.0	-170.0	-150.0

Table 3.4: Global Trend Statistics

		Chlorophyll-a		Band 1		Band 2		Band 3		Band 5	
Sensor	Subset	mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
SeaWiS	Clear	0.076	0.0029	2.248	0.0672	1.886	0.0434	1.255	0.0189	0.299	0.0067
MODIS		0.079	0.0054	2.122	0.1971	1.798	0.1126	1.284	0.0443	0.317	0.0115
SeaWiS	Deep	0.185	0.0127	1.746	0.0547	1.533	0.0358	1.133	0.0188	0.336	0.0081
MODIS		0.178	0.0177	1.614	0.2167	1.426	0.1333	1.141	0.0547	0.345	0.0146
SeaWiS	Coastal	0.916	0.1962	0.832	0.0578	0.893	0.0429	0.875	0.0340	0.426	0.0224
MODIS		0.736	0.1216	0.825	0.1387	0.831	0.0926	0.878	0.0513	0.429	0.0261

Chapter 4

Diagnostic data set

Sean Bailey
FutureTech Corporation, Greenbelt, Maryland

At the first organizational meeting for the SIMBIOS program in 1995, a diagnostic data set for ocean color missions was conceived as a way to compare ocean color data across missions. The data set was to be created by each mission as part of routine processing and was to consist of spatial subsets with all relevant information necessary to produce derived products. These subsets were to be produced for a few selected sites. The diagnostic data set concept was revisited at several subsequent SIMBIOS science team meetings. At the third SIMBIOS science team meeting in September 1999, held in Annapolis, Maryland, the diagnostic data set concept took the first steps toward implementation with the selection of a number of proposed sites for the spatial subsets. The IOCCG working group on data merger met in January of 2000 and recommended a more complete list of sites for the data set. The list of sites was finalized at the fourth SIMBIOS Science Team meeting in January 2001.

Two conditions for the selection of a diagnostic data set site were formulated. First, a reliable source of *in situ* data (bio-optical and/or atmospheric) for the site must exist, and second, the principal investigator must be willing to share the *in situ* data with the SIMBIOS project. Sites used as vicarious calibration sources were ranked with the highest priority. Time series sites were ranked as priority 2. All other sites were ranked as priority 3. Several sites were recommended, but did not meet one or both of the defined criteria. Several of the sites were modified, either at the request of an investigator, in order to reduce redundancy or improve coverage (by reducing the amount of land included in the extracted data). The list of sites as currently implemented is found in Table 4.1 and Figure 4.1 shows a map of the locations.

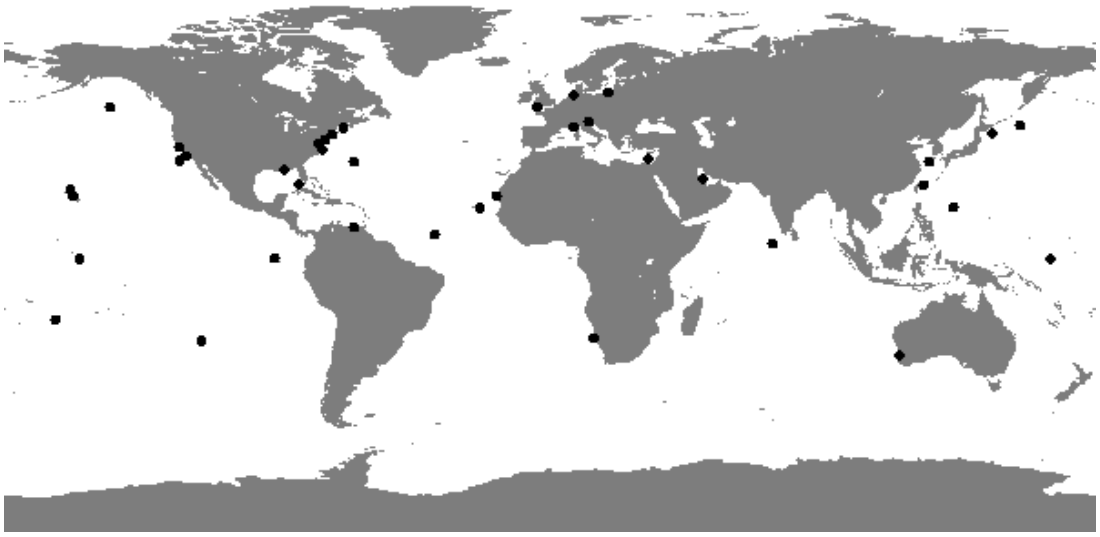


Figure 4.1: Diagnostic data set location.

By midyear 2001, the SIMBIOS project, in conjunction with the SeaWiFS project, had begun production of the L1A and L2 subsets of SeaWiFS LAC resolution data for the list of diagnostic data set sites. Prior to the start of the Collection 4 reprocessing of MODIS (Terra) data in March of 2002, the SIMBIOS project approached the MODIS Oceans Team with the request that MODIS produce a comparable set of extracted L1 and L2 data for inclusion in the diagnostic data set. The MODIS team agreed, however, as a consequence of the flow of data through MODAPS (the MODIS Data Production System), MODIS Oceans provides L1B and L2 extracts, rather than L1A as was the recommendation of the SIMBIOS Science Team and IOCCG working Group. The diagnostic data set files produced by MODAPS are sent to the SIMBIOS project for post

processing. In order to ensure that only useful data are included in the diagnostic data set, a threshold on the number of valid pixels within the region of interest was set. If this threshold (currently 25%) is not met, the files are excluded from further processing. If the threshold is met, a L2 (chlorophyll) browse image and two TAR files are created. One TAR file for the L2 granules and one for the L1B granules. Since the MODIS data are produced in 5 minute granules, the region of interest for a given site may cross the boundary between two granules. When this occurs, all L2 products from both granules are placed in the same TAR file, likewise for the L1B granules. The TAR files are then compressed using gzip compression.

The SIMBIOS and MODIS teams worked with the Goddard Distributed Active Archive Center (GDAAC) to make the diagnostic data set files for SeaWiFS and MODIS (Terra and Aqua) available through the GDAAC. All the necessary mechanisms for the transfer of the dataset to the GDAAC have been put in place, the most critical of which was the creation of 6 new Earth Science Data Type (ESDT) definitions, one for each L1 and L2 data type and the SeaWiFS, MODIS-Terra and MODIS-Aqua data sources. Once the data are archived at the GDAAC, they will be visible to both the GDAAC WHOM search engine and the EOSDIS EDG search engine. The diagnostic data set allows for the rapid processing and testing of atmospheric correction and geophysical product algorithms. The current list of sites cover a wide range of water types and aerosol conditions (see Table 4.1), which will aid algorithm assessments.

Table 4.1: Diagnostic Data Set

Site ID	Location	North latitude	South latitude	West longitude	East longitude	Contact PI
OCAlbron	Alberon Gyre Eastern Med.	33.5N	32.5N	32.0 E	33.0E	
OCA Line	Japan East coast	42.0 N	41.0N	145.283E	146.283E	Tsuda
OCAfrica	Mauritanian Upwelling	21.5N	20N	18W	17W	Carder
OCBahrain	Bahrain, Persian Gulf	26.816 N	25.816N	50.0 E	51.0E	
OCBATS	BATS Bermuda	33.0N	31.0N	65.5W	63.5W	Nelson
OCCALCOF	CALCOFI, California Coast	34.5N	30.5N	124.0W	122.0W	Mitchell
OCCpVerd	Capo Verde, NW African Coast	17.217N	16.217N	23.433W	22.433W	Carder
OCCariac	Cariaco Basin, Venezuela	11.0N	10.0N	65.66W	64.16W	Mueller-Karger
OCChsBay	Chesapeake Bay	39.5N	36.8N	76.8W	75.6W	Harding
OCCook	Cook Island, Western South Pacific	19.5S	20.5S	163.5W	162.5W	
OCDryTrt	Dry Tortugas, Florida Keys	21.1N	24.1N	83.283W	82.283W	Voss
OCEaster	Easter Island, South Pacific Gyre	26S	28S	116W	114W	SIMBIOS Project
OCEqPAC	Eastern Equatorial Pacific	0.5N	0.5S	155.5W	154.5W	Chavez
OCFRONT	Long Island, New York	41.45N	40.45N	72.5W	71W	Morrison
OCGlapgo	Galapagos Islands	2.13N	1.87S	98.81W	88.81W	Feldman
OCHattrs	Cape Hatteras	37.5N	34.5N	76.5W	73.5W	Stumpf & Cota
OCHlglnd	Helgoland, North Sea	54.6N	53.6N	7.3E	8.3E	
OCHOT	HOT Station, Hawaii	23.25N	22.25N	158.5W	157.5W	Letelier
OCKshdoo	Kaashidoo, Maldives Islands	5.45N	4.45N	72.95E	73.95E	Holben & Frouin
OCKNOT	KNOT Station, NW Pacific	44.5N	43.5 N	154.5E	155.5E	Saitoch
OCKorean	Korean seawater	32.5N	31.5N	124.5E	125.5E	Kim

	Monitoring site, East China Sea					
OCLeo 15	LEO 15 Station, New Jersey	40.1	38.5N	74.75W	73.5W	Arnone
OCLigurn	Ligurian Sea, Mediterranean	43.87 N	42.87 N	7.4 E	8.4 E	Antoine
OCLderz	Luderitz Upwelling, Namibian Coast	25.5S	26.5S	14E	15E	
OCMOBY	MOBY Buoy, Hawaii	21.3N	20.3N	157.75 W	156.7W	Clark & Trees
OCMontry	Monterey Bay, California	37N	36.5N	122.75W	121.75W	Chavez
OCNOAAGM	Northern Gulf of Mexico	30N	29N	88W	87W	Arnone
OCNordic	Baltic Sea	55.5N	54.5N	18.8E	19.8E	
OCPAPA	Station PAPA, North Pacific	52 N	48 N	147 W	143 W	
OCPhlipp	Philippines	17.5 N	16.5 N	132.5 E	133.5 E	
OCPlumes	Plumes and Blooms region, Santa Barbara, CA	35.5 N	32.5 N	122 W	118 W	Siegel
OCPlymbdy	PlyMBoDY Mooring, English Channel	50.4 N	49.6 N	4.8 W	3.4 W	Aiken
OCRttnst	Rottnest Island, Western Australia	31.3 S	32.3 S	114.8 E	115.8 E	Lynch
OCSctian	Scotia Prince Ferry Route, Gulf of Maine	43.8N	42.8N	70.25W	65.75W	Balch
OCTahoe	Lake Tahoe	39.671N	38.671N	120.604W	119.604W	
OCVenice	Venice Tower (AAOT) Northern Adriatic	45.6N	44.8N	12.2E	13.4E	Zibordi
OCWrmPol	Warm Pool, Western Equatorial Pacific	0.5N	0.5S	164.5E	165.5E	
OCYBOM	YBOM replacement mooring, East China Sea	24.89N	23.89N	122.77E	123.77E	Ishizaka

Chapter 5

Overview of SeaBASS and MODIS Validation Activity

Jeremy Werdell
Science Systems and Applications Inc., Greenbelt, Maryland

Sean Bailey
FutureTech Corporation, Greenbelt, Maryland

5.1 INTRODUCTION

High quality *in situ* measurements are a prerequisite for satellite data product validation, algorithm development, and many climate-related inquiries. As such, the SIMBIOS and SeaWiFS Projects maintain a local repository of *in situ* bio-optical data, known as the SeaWiFS Bio-optical Archive and Storage System (SeaBASS), to support and sustain regular scientific analyses (Hooker et al. 1994, Werdell and Bailey 2002). This system was originally populated with radiometric and phytoplankton pigment data used in the SeaWiFS Project's satellite validation and algorithm development activities. To facilitate the assembly of a global data set, however, under NASA Research Announcements NRA-96-MTPE-04 and NRA-99-OES-99, SeaBASS was broadened to include oceanographic and atmospheric data sets collected by the SIMBIOS Project. This aided considerably in minimizing spatial and temporal biases in the data while maximizing acquisition rates (Fargion and McClain 2003). To develop consistency across multiple data contributors and institutions, the SIMBIOS Project also defined and documented a series of *in situ* sampling strategies and data requirements that ensure that any particular set of measurements are appropriate for algorithm development and ocean color sensor validation (Mueller et al., 2003).

The SeaBASS bio-optical data set includes measurements of apparent and inherent optical properties, phytoplankton pigment concentrations, and other related oceanographic and atmospheric data, such as water temperature, salinity, and aerosol optical thickness. Data are collected using a number of instrument packages from a variety of manufacturers, such as profilers and handheld instruments, on a variety of platforms, including ships and moorings. As of May 2003, SeaBASS included data collected by research groups at 44 institutions in 14 countries, encompassing over 1,150 individual field campaigns (Figure 5.1). These data include over 300,000 phytoplankton pigment concentrations, 13,500 continuous depth profiles, 15,000 spectrophotometric scans, and 15,000 discrete measurements of AOT. The SIMBIOS Project Office makes use of a rigorous series of submission protocols and quality control metrics that range from file format verification to inspection of the geophysical data values (Fargion et al. 2001, Werdell and Bailey 2002). This ensures that observations fall within expected ranges and do not clearly exhibit characteristics of measurement problems.

5.2 DATA ACCESSIBILITY

The data included in SeaBASS are readily available to members of the MODIS Science Team for use in their validation and algorithm development activities. The SeaBASS World Wide Web site, located at: <http://seabass.gsfc.nasa.gov>, provides a complete description of the system architecture, comprehensive documentation on policies and protocols, and direct access to the bio-optical data set and validation results. Through the use of online search engines, the full bio-optical data set is searchable and available to authorized users via the Web. Note that all online resources described below are linked to the main URL provided above. To protect the publication rights of contributors, access to data collected more recently than 1 January 2000 is limited to SIMBIOS Science Team members, NASA-funded researchers (such as MODIS Science Team members), and regular voluntary contributors, as defined by the SeaBASS access policy (Firestone and Hooker 2001). The remainder of the data is fully available to the general public and, additionally, has been released to the National Oceanic and Atmospheric Administration's (NOAA) National Oceanographic Data Center (NODC) for inclusion in their archive.

Several search engines are available to locate and extract data files and geophysical data values from the bio-optical data set, including the Bio-optical Search Engine, Pigment Locator, and Aerosols Locator. Other search engines are available for compiling metadata relating to the bio-optical data set, such as the General Search Engine and Cruise Search Engine. The latter provide generic information about the data, such as cruise and experiment names, date and location ranges, data parameters collected, and contributor names. For all of the above, visitors may limit queries to particular experiments,

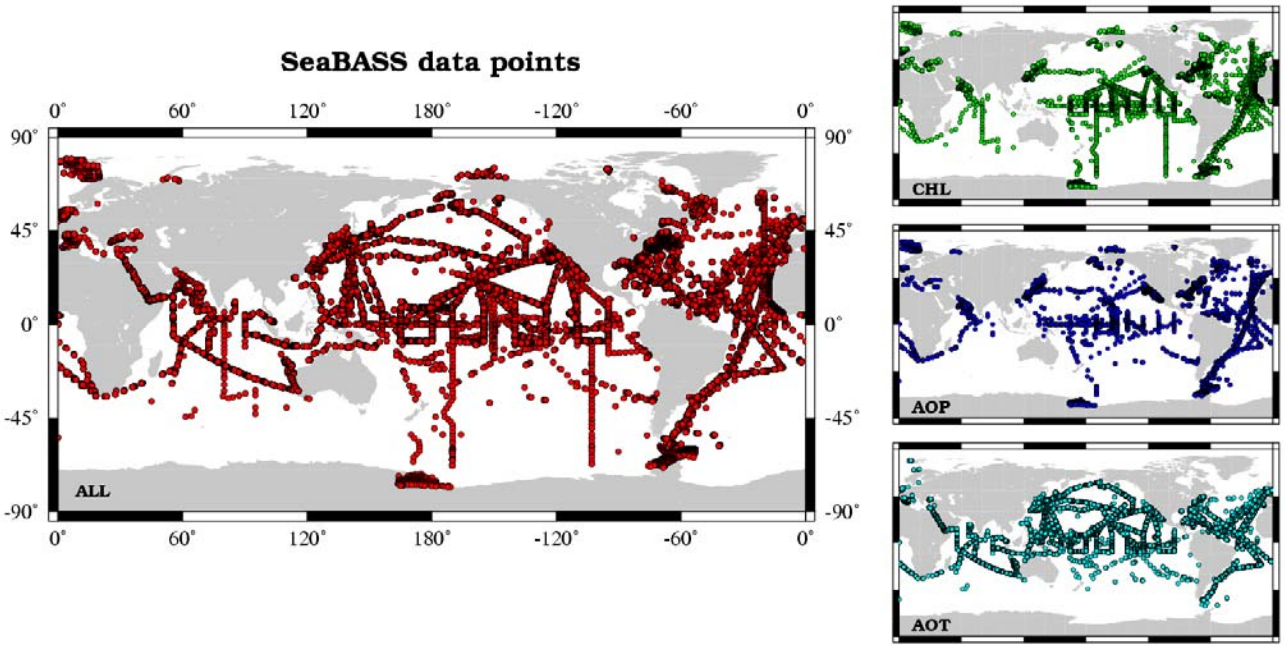


Figure 5.1: The global distribution of data included in the full SeaBASS bio-optical data set, as of May 2003. Clockwise from left: all archived data, chlorophyll a concentrations only (CHL), apparent optical properties only (AOP), and aerosol optical thickness only (AOT).

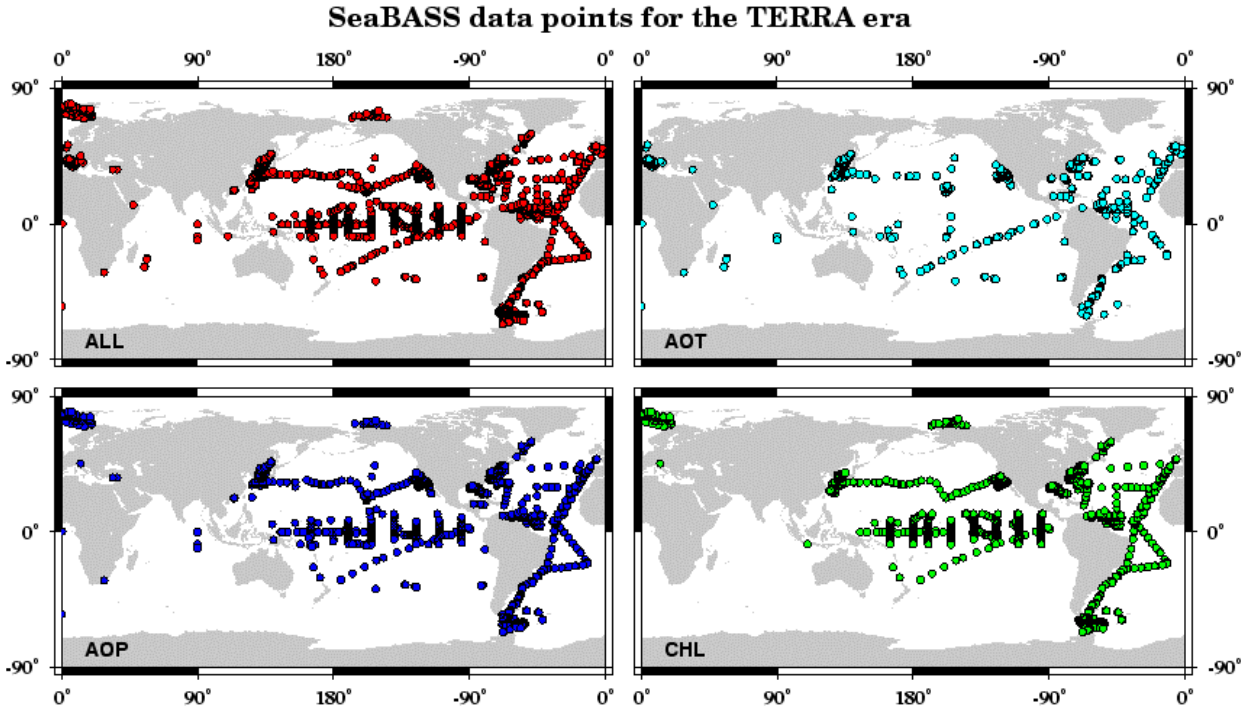


Figure 5.2: The global distribution of MODIS Terra era data included in the SeaBASS bio-optical data set, as of May 2003. Clockwise from upper left: all archived data, chlorophyll a concentrations only (CHL), apparent optical properties only (AOP), and aerosol optical thickness only (AOT). The maps do not include the complete ship track for Fast Rotating Shadowband Radiometer data included in SeaBASS.

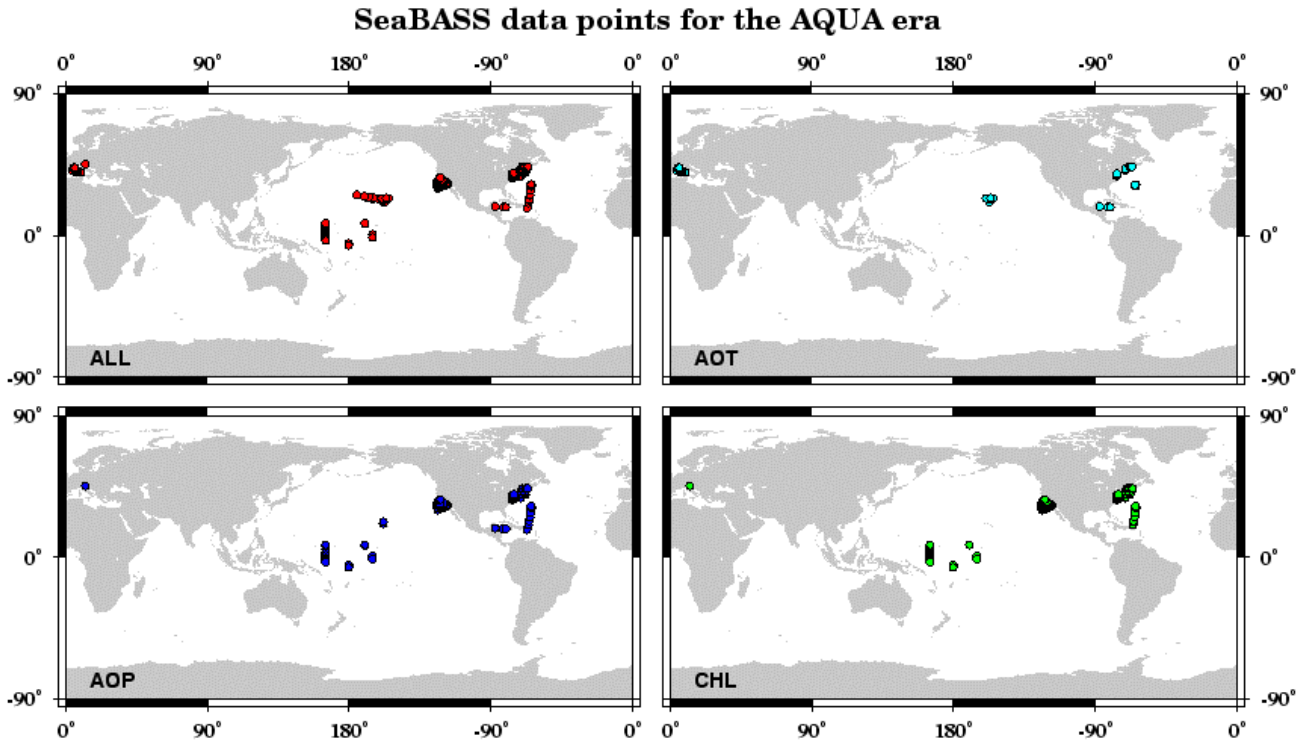


Figure 5.3: The global distribution of MODIS Aqua era data included in the SeaBASS bio-optical data set, as of May 2003. Clockwise from upper left: all archived data, chlorophyll a concentrations only (CHL), apparent optical properties only (AOP), and aerosol optical thickness only (AOT). The maps do not include the complete ship track for Fast Rotating Shadowband Radiometer data included in SeaBASS.

contributors, date and location ranges, and data types (e.g., chlorophyll *a* or water-leaving radiance). A series of supplementary Web pages are linked to each search engine with tables listing additional relevant information, for example, the names of archived experiments or data types, to assist users narrow or tailor their queries. On occasion, JavaScript pop up windows are used to provide definitions or explanations of an online feature.

A third metadata search engine, the Validation Cruise Search Engine, was designed specifically to assist researchers requiring potential validation cruises for their satellite calibration and validation activities. As such, queries may be limited only by satellite mission. Current options include MODIS Terra and Aqua, as well as the Ocean Color and Temperature Scanner (OCTS), SeaWiFS and the Medium Resolution Imaging Spectrometer (MERIS). Queries return a table of potential validation cruises. As for the other metadata search engines, each returned record includes the data contributor(s), the data types collected, the start and end dates, and the center latitude and longitude coordinates for the given cruise.

5.3 DISTRIBUTION STATISTICS AND ADDITIONAL RESOURCES

As of May 2003, 41 research groups outside of the SIMBIOS and SeaWiFS Project Offices have been granted unrestricted access to SeaBASS, including six MODIS Oceans and Atmospheres Science Team groups. In 2002, the 41 groups queried SeaBASS over 950 times and downloaded more than 60,000 data files from the bio-optical data set. The MODIS groups accounted for approximately 30% of the queries and 40% of the file downloads. During the same period, 146 research groups searched the public set 600 times and downloaded over 37,000 files.

Visitors wishing to generate maps of SeaBASS data may do so interactively using the SeaBASS Mapping Utility. The default map is global, however, users are provided the option of customizing latitude and longitude boundaries. Mapped data points may be further limited by user-defined date ranges and specific data types. Several other global maps are linked to this Web site: (1) a map of all data included in the bio-optical data set, and (2) mission-specific maps of pigment, radiometer, and

sun photometer data points. Currently, the latter includes maps for MODIS Terra and Aqua (Figures 5.2 and 5.3), as well as OCTS, SeaWiFS, and MERIS. All of the above are updated daily and are available for download.

The SeaBASS Email Notification Service was designed to announce the arrival of new data to interested parties in a timely fashion. An electronic message listing recently submitted data is sent to members of this list once a week. Only cruises archived in the past week are included in the electronic message. Currently, all 14 members of the MODIS Ocean Science Team are subscribed to this mailing list. This information is readily available to all users via the SeaBASS New and Updates Web site.

5.4 MODIS VALIDATION ACTIVITY

The validation methodology developed for SeaWiFS (Bailey et al., 2000) was used to validate MODIS (Terra). In this context, validation consists of comparing coincidently measured satellite and *in situ* data. The validation, or 'match-up' procedure described in Bailey et al., (2000) require some minor modifications to work with the MODIS dataset. These modifications were necessary to deal with the MODIS file format and data flagging issues. MODIS processing provides a quality flag with values ranging from zero to three. Quality zero indicates the best data quality. The quality flag is a combination of the MODIS common flags and L2 product specific flags. The flagging criteria for data of quality zero most closely approximates the flagging criteria used in McClain et al., (2000) for SeaWiFS validation. Exclusion criteria are applied to the datasets to ensure that a consistent, quality controlled dataset is used in the validation analysis. For the MODIS validation activity, these exclusion criteria include: the use of only quality level zero data, a +/- 3 hour time window centered on the overflight time for *in situ* data to be accepted, and a minimum of 13 pixels in a 5x5 pixel box centered on the *in situ* location is required.

The MODIS data used in this analysis were obtained from the MODIS Oceans Team at Goddard. During the Collection 4 reprocessing, extracted L2 data files were produced based on a list of *in situ* locations and dates that were provided to the MODIS team by the SIMBIOS prior to the reprocessing.

5.5 RESULTS AND DISCUSSION

The results of this validation activity for *in situ* data spanning the time period designated as valid for Collection 4 MODIS Ocean data set forth by the MODIS Science Team (November 2000 through March 2002) are shown in Table 5.1 and Figure 5.4. For the water-leaving radiances, MODIS tends to underestimate coincident *in situ* values by up-wards of 20% for the valid period. Overall, the various chlorophyll product comparisons reasonable. There is much less of a bias with respect to *in situ* data as is seen in the radiance comparisons, and the comparisons are on average within about 40% of *in situ* values. Unfortunately, for the limited available validation data set, only the Chlor_a_2 product is within the goal of +/- 35%. The 15 to 20% difference for the radiance comparisons is well above the expectation of +/- 5% required accuracy for ocean color missions.

Table 5.1: MODIS - *in situ* comparison results: SAT/ENV Med_Ratio is the median value of the ratio of MODIS to *in situ* data, Median % Diff is the median value of the percent absolute difference between MODIS and *in situ* for each individual comparison, CV is the coefficient of variation for the SAT/ENV median ratio, N is the number of observations.

Parameter	SAT/ENV Med_Ratio	Median % Diff	CV	N
Lwn(412)	0.8498	19.93	0.419	53
Lwn(443)	0.8781	17.31	0.375	55
Lwn(488)	0.8649	15.13	0.324	55
Lwn(531)	0.8283	17.17	0.223	34
Lwn(551)	0.8913	16.74	0.380	57
CHLOR_A_2	0.8133	31.82	0.420	31
CHLOR_A_3	0.9740	40.12	0.582	22

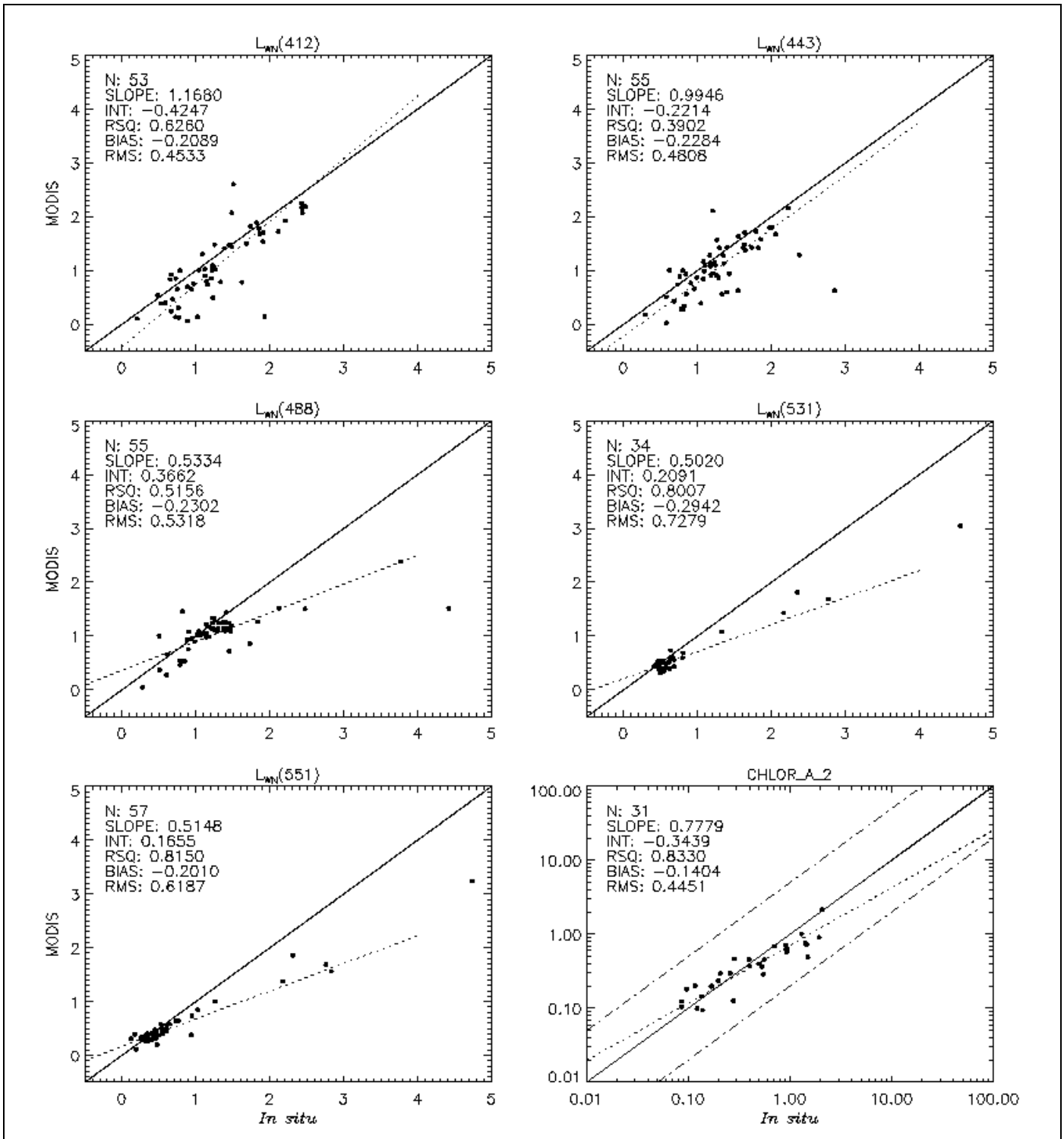


Figure 5.4: Scatter plots of coincident *in situ* MODIS Terra observations for chlorophyll *a* and water-leaving radiance (L_{wn}) at 412,443,488,531,551 nm. The chlorophyll *a* data were transformed to account for their log-normal distribution. A one-to-one line has been included for clarity Collection 4 MODIS Ocean data set forth by the MODIS Science Team (November 2000 through March 2002).

REFERENCES

Fargion, G. S., R. Barnes and C.R. McClain 2001: In Situ Aerosol Optical Thickness Collected by the SIMBIOS Program (1997-2000): Protocols, and Data QC and Analysis. *NASA Tech. Memo. 2001-209982*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 103 pp.

Fargion, G.S. and C.R. McClain, 2003: SIMBIOS Project 2002 Annual Report. *NASA Tech. Memo. 2003-21622*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 157 pp.

Firestone, E.R., and S.B. Hooker, 2001: SeaWiFS postlaunch technical report series cumulative index: Volumes 1-11. *NASA Tech. Memo. 2001-206892 Vol.* ***12*****,** S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 24 pp.

Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E-N. Yeh, and Y. Ge, 1994: The SeaWiFS bio-optical archive and storage system (SeaBASS), Part 1. *NASA Tech. Memo. 104566*, Vol. **20**, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 37 pp.

McClain, C.R., E.J. Ainsworth, R.A. Barnes, R.E. Eplee, Jr., F.S. Patt, W.D. Robinson, M. Wang, and S.W. Bailey, 2000: SeaWiFS Postlaunch Calibration and Validation Analyses, Part 1. *NASA Tech. Memo. 2000-206892,* Vol. **9***,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, 82 pp.

Mueller, J.L., G.S. Fargion and R.C. McClain, 2003: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4, Volumes I-VI, *NASATM-2003-21621/Rev4-Vol. I-VI*, NASA Goddard Space Flight Center, Greenbelt, Maryland.

Werdell, P.J. and S.W. Bailey, 2002: The SeaWiFS Bio-optical Archive and Storage System (SeaBASS): Current architecture and implementation. *NASA Tech. Memo. 2002-211617*, G.S. Fargion and C.R. McClain, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 45 pp.

Chapter 6

Investigation of Ocean Color Atmospheric Correction Algorithms Using *In Situ* Measurements of Aerosol Optical Thickness : Application to MODIS

Christophe Pietras and Giulietta S. Fargion
Science Applications International Corporation, Beltsville, MD

Kirk Knobelspiesse
Science Systems and Applications Inc., Greenbelt, Maryland

6.1 INTRODUCTION

The SIMBIOS pool of sun photometers is composed of three types of instruments. The first is a sun/sky photometer that measures the solar irradiance and the sky radiance. The second is a shadow-band radiometer that measures diffuse and total sky radiances. The third is a Lidar, which measures vertical and horizontal distribution of aerosol backscatter, extinction, and optical depth. Pictures of the pool of instruments are presented in Figure 6.1.

The instruments are deployed by SIMBIOS or NASA Principal Investigators on cruises, and data are archived in SeaBASS (Werdell, 2003). The SIMBIOS Project deploys several sun photometers and radiometers composed of 14 hand-held MicroTops II, and three hand-held SIMBAD and SIMBADA radiometers. The Project has also contributed to the AERONET network by adding 14 stations in coastal regions or islands equipped with CIMEL sun photometers (Holben et al, Fargion et al, 2001). In addition, the Project reviewed and documented the description, characteristics and advantages of each instrument (Fargion et al., 2001). The protocols used for the calibration, operation and data analysis have been continuously reviewed, revisited and updated (Mueller et al., 2003, Porter et al., 2001) and (Knobelspiesse et al., 2003b). A user's guide and instructions are provided to help the Principal Investigators collect measurements according to protocol (Knobelspiesse et al., 2003a).

The Aerosol optical thickness (AOT) values are measured by each sun photometer at several wavelengths in the visible and near infrared. Data collected with hand-held sun photometers deployed at sea have been gathered and are available in the SeaBASS database (http: //seabass.gsfc.nasa.gov). CIMEL sun photometers supported by the SIMBIOS Project were added to the AERONET network and consequently the data are distributed through the AERONET database (http: //aeronet.gsfc.nasa.gov).

Uncertainty analyses were conducted to determine the accuracy of the aerosol retrieval from the hand-held sun photometers (Fargion and McClain, 2003). The analysis was based on the work of (Russell et al., 1993) and on the analysis conducted by the AERONET group on the CIMEL sun photometers (Holben et al., 1998,Eck et al., 1999). Efforts have been made to apply consistent protocols to transfer the calibration from a CIMEL sun photometer to any SIMBIOS sun photometer and to use a consistent algorithm to retrieve the aerosol optical thickness (Fargion et al., 2001). These efforts were conducted to ensure the same quality standards for the atmospheric data held in SeaBASS as the data provided by the AERONET network. Continuous efforts are now carried out to provide uncertainty analysis for the Angström Exponent as well.

Cross calibration analysis is performed for each sun photometer according to the protocols described in prior NASA Technical Memorandums (Fargion et al., 2001, Mueller et al., 2002). A calibration transfer is performed using the master CIMEL that is calibrated at Mauna Loa every 3 months. The interpolated calibration coefficients of the master CIMEL are provided by the AERONET Project. When only the first calibration of the master CIMEL is available, the calibration transfer is performed but the results are considered preliminary, and the campaigns of measurements are processed and submitted to SeaBASS along with a status flag "preliminary". When the final calibration of the master CIMEL is available, the final calibration transfer is performed. The campaigns of measurements are processed again and submitted to SeaBASS along with the status flag "final".

Quality control procedures are conducted to confirm that only the best data are submitted to SeaBASS. Figure 6.2 presents a flowchart of the sun photometer data and quality control analysis. Several plots are created from each AOT file to be used

for qualitative Quality Control (QC). The plots include a map of data locations (to ensure coordinates are located where they should be), a history of calibration coefficients (to ensure they are not changing rapidly over the period of the data), a plot of AOT spectra (to ensure data roughly follow the Junge Law), and various histograms and other statistics (to ensure the data fall within reasonable bounds). QC is used as a tool to determine if there have been problems in the capture, calibration or processing of a set of data, but not as a rigid rule for acceptance or rejection of that data. Plots and statistical analyses created during QC for a file are submitted along with it to SeaBASS.

6.2 METHODOLOGY

Match-up analysis was originally conducted in 2001 with SeaWiFS aerosol products (Fargion et al., 2001). In this technical report we present results of the match-up analysis conducted using MODIS Oceans and MODIS Atmosphere products. Measurements collected by coastal and island CIMEL stations and by SIMBIOS Investigators with the hand-held sun photometers near the CIMEL stations were used for the MODIS match-up analysis.

Figure 6.3 shows the atmospheric data set collected and currently held in SeaBASS. More than 4000 MicroTops II records are shown in blue, more than 5000 SIMBAD records are shown in green and more than 100,000 shadow-band records are shown in red. Figure 4 shows twenty CIMEL stations selected to contribute to our match-up analyses. They were chosen because of a specific interest, such as the presence of the MOBY buoy or regions where satellite retrieval is known to be problematic because of dust or smoke.

MODIS Oceans and Atmosphere Projects distribute level 2 aerosol products. The MODIS Oceans Project distributes level 2 aerosol optical thickness (Gordon and Voss, 1999) at 1 kilometer resolution. These are used to match with *in situ* measurements. Cloud-free, non-land pixels within a 21x21 pixel box are selected to generate a subset of the level 2 product. If more than 50% of the pixels are valid and if the coefficient of variation is less than 10 percent, the subset contributes to the final matchups.

The MODIS Atmosphere Project distributes level 2 aerosol optical thickness (Kaufman and Tanrà 1998) at 10 kilometers resolution over land and over oceans. The aerosol optical thickness derived over oceans was only considered in this study. Cloud-free pixels within a 3x3 pixel box for which the algorithm retrieval is stated “good” or “very good” (Chu et al., 2000) contribute to the final matchups. The scene dimensions are similar for both products (about 20 km) and were chosen according to former match-up analysis performed with SeaWiFS aerosol products (Fargion et al., 2001). Validated MODIS data collected between November 2000 and September 2002 have been considered and analyzed.

6.3 RESULTS

Figure 6.5 shows the match-up results obtained at 865nm for MODIS Oceans-derived AOT’s versus (a) CIMEL AOT’s and (b) hand-held AOT’s. Both match-up results show a slight overestimation by MODIS Oceans-derived AOT’s compared to i*n situ* values. Similar results were obtained and reported in prior analyses with SeaWiFS aerosol products (Fargion et al., 2001, Fargion and McClain, 2003).

A probable cause of the observed overestimation of the AOT derived from space is calibration changes in the near infrared bands, such as was revealed for SeaWiFS by several previous studies (Barnes et al., 2001, Fargion et al., 2001). Other causes were also investigated, such as polarization sensitivity of the sensors (Gordon, 2001, Gordon, 2002) or contribution of the polarization of light to the TOA radiance (Dr Wang personal communication). Nevertheless, further analysis is postponed until after MODIS reprocessing (http: //modis-ocean.gsfc.nasa.gov/).

A similar atmospheric correction algorithm (Gordon and Wang, 1994) is used by both MODIS Oceans and SeaWiFS Projects to derived the aerosol properties at 865nm from radiances measured in the near infrared channels. A different atmospheric correction algorithm is used by the MODIS Atmosphere group to derive aerosol optical thickness over oceans (Kaufman and Tanrà 1998). A similar match-up analysis was conducted using MODIS Atmosphere-derived aerosol products over the same regions where CIMEL stations measured aerosol optical thickness.

Figure 6.6 presents the match-up results obtained at 865nm with MODIS Atmosphere-derived AOT (a) and the corresponding matchups obtained with MODIS Oceans-derived AOT (b). Figure 6 (c) shows MODIS Oceans-derived AOT-865nm versus MODIS Atmosphere-derived AOT-865nm extracted over the same scene. Where AOT’s are lower than 0.3, the MODIS Atmosphere-derived AOT’s compare well with *in situ* AOT’s, while the MODIS Oceans-derived AOT’s are slightly overestimated. Where AOT’s are higher than 0.3, both MODIS Oceans and Atmosphere-derived AOT’s are underestimated. Further analysis is underway.

Match-up analysis was also conducted for MODIS Atmosphere-derived AOT in the visible and for MODIS Atmosphere-derived Angström Exponent and are presented in Figure 7. *In situ* AOT’s measured at 500nm were scaled to 550nm to match the MODIS-derived AOT’s. Figure 6.7a shows that MODIS Atmosphere-derived aerosol optical thickness compares well in

the visible region of the spectrum. Figure 6.7b shows that the range of Angström Exponent measured *in situ* is wider than the range of Angtröm exponents derived by MODIS. Similar results were also obtained in a prior study conducted with SeaWiFS aerosol products (Fargion et al., 2001,Fargion and McClain, 2003).

Figure 6.1: SIMBIOS instrument pool: MicroTops II, SIMBADA & SIMBAD hand-held sun photometers.

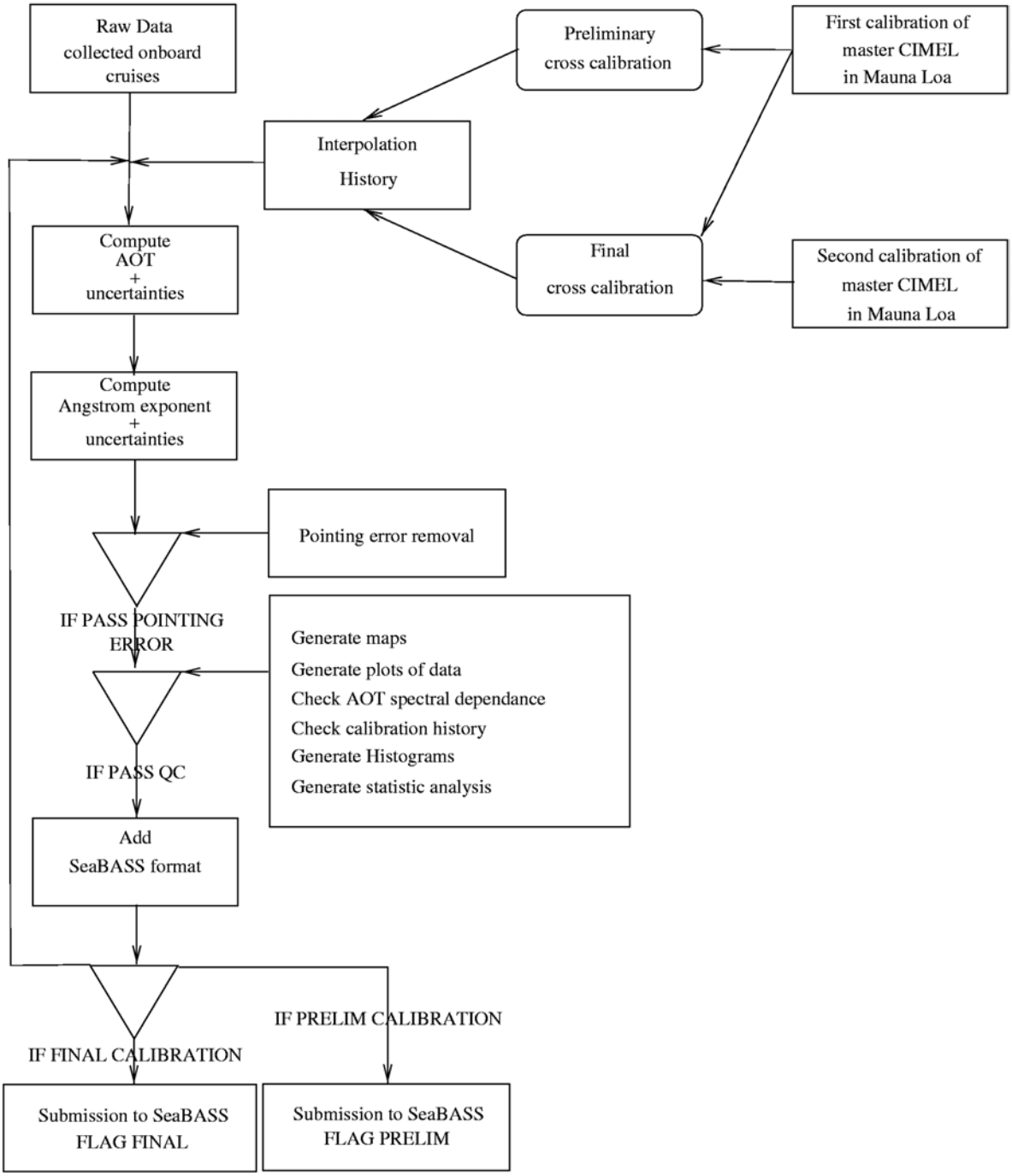


Figure 6.2: Flowchart of the sun photometers data and quality control analysis

Table 6.1: AERONET sites used for aerosol match-up analysis (SIMBIOS Project Office instruments in bold). The sites marked with stars correspond to pre-defined geographic sites for which a diagnostic dataset of ocean color data products is routinely extracted and used for MODIS match-up analysis

Sites#	Site names	Longitude	Latitude	PI
1	**Lanai island***	-156.99	20.83	C. McClain
2	**Coconut Island**	-156.99	20.83	C. McClain
3	**Tahiti**	-149.61	-17.58	C. McClain
4	Nauru Island	166.92	-0.52	M. Miller
5	Bermuda*	-64.70	32.37	B. Holben
6	**Ascension Island**	-14.41	-7.98	C. McClain
7	**Azores**	-28.63	38.53	C. McClain
8	Kaashidhoo Island*	73.47	4.97	B. Holben
9	**Anmyon Island**	126.32	36.52	C. McClain
10	**Puerto Madryn**	-58.50	-34.57	C. McClain
11	Dry Tortugas Island*	-82.80	24.60	K. Voss & H. Gordon
12	Arica	-70.31	-18.47	B. Holben
13	**Bahrain***	50.5	26.32	C. McClain
14	**Wallops***	-75.47	37.94	C. McClain & B. Holben
15	**Erdemli**	34.25	36.56	C. McClain
16	San Nicolas Island*	-119.49	33.26	R. Frouin
17	**Rottnest Island**	115.30	-32.00	Chuck McClain
18	**Dahkla**	-15.95	23.72	C. McClain
19	Venise*	12.50	45.31	G. Zibordi
20	Capo Verde*	-22.93	16.73	D. Tanré
21	**Chinae**	128.65	35.16	C. McClain

REFERENCES

Barnes, R. A., Eplee, R. E. J., Schmidt, G., Patt, F. S., and McClain, C. R., 2001: Calibration Of SeaWiFS. I. Direct Techniques. *Appl. Opt.*, **40**, 6,682 – 6,700.

Chu, K., Strabala, S., Platnick, S., Moody, E., King, M., Mattoo, S., Hucek, R., and Ridgway, B., 2000: Modis Atmosphere Qa Plan. Version 2.2. *NASA Technical Report*, NASA Goddard Space Flight Center, Greenbelt, MD, 44pp.

Eck, T., Holben, B., Reid, J., Dubovik, O., Smirnov, A., O´Neill, N., and Slutsker, I., 1999: Wavelength Dependence Of The Optical Depth Of Biomass Burning, Urban And Desert Dust Aerosols. *J. Geophys. Res.*, **104**, 31,333 – 31,350.

Fargion, G. and McClain, C. R., 2003: SIMBIOS Project 2002 Annual Report. *NASA Tech. Memo*. *2003-211622*, NASA Goddard Space Flight Center, Greenbelt, MD, 157pp.

Fargion, G. S., Barnes, R., and McClain, C., 2001: In Situ Aerosol Optical Thickness Collected by the SIMBIOS Program (1997-2000): Protocols, and Data QC and Analysis. *NASA Tech. Memo*. *2001-209982*, NASA Goddard Space Flight Center, Greenbelt, MD, 107pp.

Gordon, H. and Voss, K., 1999: Normalized Water Leaving Radiance. Version 4.0. ATBD ATBD-MOD-17, NASA Goddard Space Flight Center, Greenbelt, MD, 96pp.

Gordon, H. R., 2001: Modis Terra Normalized Water-Leaving Radiance Data Quality Summary, Data Product: Normalized Water-Leaving Radiance (Mod18), Data Set Version: 3; see http: //modis-ocean.gsfc.nasa.gov/qual.html/ terra/dataqualsum/nlw_qualsum.pdf. Plan, NASA Goddard Space Flight Center, Greenbelt, MD, 4pp.

Gordon, H. R., 2002: Modis Terra Normalized Water-Leaving Radiance Data Quality Summary, Data Product: Normalized Water-Leaving Radiance (Mod18), Data Set Version: Collection 4 Version 4.2 reprocessed; see http: //modis-ocean.gsfc.nasa.gov /qual.html/terra/dataqualsum/nlw_qualsum.v4.pdf. Technical report, NASA Goddard Space Flight Center, Greenbelt, MD., 7pp.

Gordon, H. R. and Wang, M., 1994: Retrieval Of Water-Leaving Radiance And Aerosol Optical Thickness Over The Oceans With Seawifs: A Preliminary Algorithm. *Appl. Opt.*, **33**, 443 – 452.

Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A., 1998: A Federated Instrument Network And Data Archive For Aerosol Characterization. *Rem. Sens. Environ.*, **66**, 1 – 16.

Kaufman, Y. and Tanré, D., 1998: Algorithm For Remote Sensing Of Tropospheric Aerosol From Modis. Products: Mod04, Mod08. ATBD ATBD-MOD-02, NASA Goddard Space Flight Center, Greenbelt, MD, 85pp.

Knobelspiesse, K., Pietras, C., and Frouin, R., 2003a: User's Guide And Instructions To Use The Simbios Sun Photometers And Measurements Protocol, see http: //simbios.gsfc.nasa.gov/Sunphotometers/pdf/ SIMBIOS_sunphoto_protocol.pdf. NASA Goddard Space Flight Center, Greenbelt, MD, 25pp.

Knobelspiesse, K., Pietras, C., and G.S., F., 2003b: Correction For Sea Deployment Of The Microtops II Hand Held Sun Photometer. *J. Atmos. Ocean. Technol.*, **20**(5), 767 – 771.

Mueller, J. et al., 2002. Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 3, volumes 1 and 2. *NASA Tech. Memo*. 2002-210004/Rev3-Vol1 and -2, NASA Goddard Space Flight Center, Greenbelt, MD, 308pp.

Mueller, J., Fargion, G. S., and McClain, C., 2003. Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4, Volume Iii: Radiometric Measurements And Data Analysis Protocols. *NASA Tech. Memo*. *2003-210004*, NASA Goddard Space Flight Center, Greenbelt, MD, 78pp.

Porter, J., Miller, M., Pietras, C., and Motell, C., 2001. Ship-Based Sun Photometer Measurements Using Microtops Sun Photometers. *J. Atmos. Ocean. Technol.*, **18**, 765 – 774.

Russell, P., Livingston, J., Dutton, E., Pueschel, R., Reagan, J., DeFoor, T., Box, M., Allen, D., Pilewskie, P., Herman, B., Kinne, S., and Hoffmann, D., 1993. Pinatubo And Pre-Pinatubo Optical-Depth Spectra: Mauna Loa Measurements, Comparisons, Inferred Particle Size Distributions, Radiative Effects, And Relationships To Lidar Data. *J. Geophys. Res.*, **98**(D12), 22,969 – 22,985.

Werdell, J., et al., 2003: Unique Data Repository Facilitates Ocean Color Satellite Validation. *EOS, Trans. AGU*. (accepted).

Chapter 7

Operational Merging of MODIS and seaWiFS Ocean Color Products at Level-3

Bryan A. Franz and John G. Wilding
Science Applications International Corporation, Beltsville, Maryland

Joel M. Gales
Futuretech Corp., Lanham, Maryland

7.1 INTRODUCTION

One of the primary goals of the SIMBIOS Project was to develop and evaluate methods for the merging of data products from multiple ocean color missions. This merging can be done at the level of observed radiances, water-leaving radiances, or derived products such as chlorophyll. Various techniques have been developed and evaluated, and several are discussed in other chapters of this document. This chapter discusses what is perhaps the simplest form of merger to implement, which is the averaging of Level-3 products from multiple sensors into geo-located, equal area bins. As a demonstration of this technique, software and procedures were developed within the SIMBIOS Project to generate merged Level-3 products from SeaWiFS and MODIS. In coordination with MODIS/Terra oceans collection #4 reprocessing, the SIMBIOS Project began to receive daily Level-3 binned chlorophyll products, and to merge the MODIS products with SeaWiFS Level-3 chlorophyll products within the framework of the SeaWiFS Data Processing System (SDPS). When the first daily binned chlorophyll products from MODIS/Aqua became available, these were immediately incorporated into the merging process as well.

7.2 IMPLEMENTATION

The SeaWiFS products used in the merging are standard 9-km resolution bin files composited over one-day periods. The MODIS products are standard Level-3 daily binned files at 4.6-km resolution. The specific ocean color parameters used are the chlor_a product of SeaWiFS, which is the chlorophyll concentration derived using the OC4V4 algorithm (O'Reilly, 2000), and the chlor_a_2 product of MODIS (Terra & Aqua), which is the chlorophyll concentration derived with the OC3M algorithm (O'Reilly, 2000). The MODIS product suite includes multiple chlorophyll products, but the chlor_a_2 product is considered to be the SeaWiFS-analog.

Both the MODIS and SeaWiFS bin file formats use a sinusoidal distribution of equal area bin elements; however, the MODIS products are generated at a higher resolution than SeaWiFS. The first step in the merging process is to convert the MODIS products to the 9-km resolution of SeaWiFS, to achieve a 1-to-1 mapping of the MODIS and SeaWiFS bins. The SIMBIOS Project developed two pieces of software to accomplish this task. The first, modbin2seabin, converts the MODIS format to a SeaWiFS-like format at the original MODIS bin resolution. This is just a slight reorganization of the Hierarchial Data Format (HDF) fields. The second program, reduce_bin_resolution, is essentially a modified version of the SeaWiFS time binning code which performs spatial compositing of the input bins. For the MODIS files, reduce_bin_resolution effectively averages four 4.6-km bins into a single 9-km bin. The averaging is weighted by the square root of the number of observations within each 4.6-km bin, which is the same approach used for standard temporal compositing of MODIS and SeaWiFS (Campbell et al., 1995).

Once the MODIS products have been converted to SeaWiFS-like format and resolution, the standard SeaWiFS temporal binning code, timebin, is employed to composite the files from both missions into daily, weekly, and monthly Level-3 bin products at 9-km resolution. Again, the time binner performs a weighted average, with weights computed as the square root of the number of observations within each input bin. These binned products are then mapped using the standard SeaWiFS mapping software, smigen, and the mapped files and browse images are distributed through the SeaWiFS Standard Mapped Image browser.

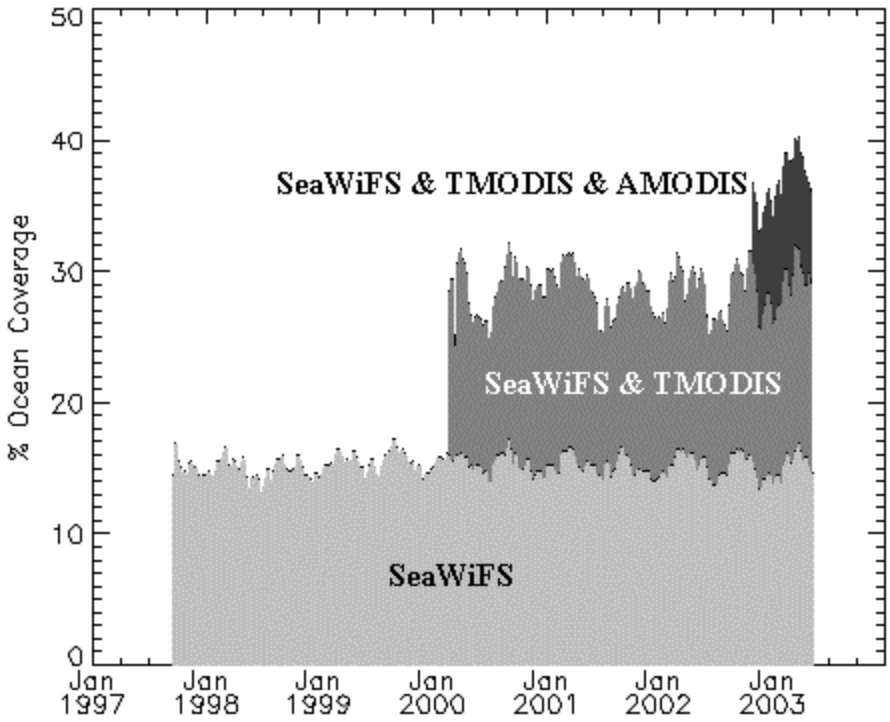


Figure 7.1: Multi-mission time-series showing the total coverage attained each day through the operational merging of SeaWiFS, MODIS/Terra (TMODIS), and MODIS/Aqua (AMODIS). Coverage is computed as a percentage of total area over the world's oceans. The plot also indicates the time at which each mission began operational data production.

7.3 SPATIAL COVERAGE

One of the primary reasons for merging data from multiple missions is to produce a consistent ocean color data set which maximizes the amount of ocean observed within one day (Gregg, 1998a). Many factors contribute to limit the amount of ocean surface visible from any single spaceborne sensor. Beyond the basic engineering limitations associated with the spacecraft orbit, sensor scan width, and imaging duty cycle, the biggest obstacle to viewing the ocean is obstruction and contamination of the line-of-sight by clouds. Another major factor is specular reflection of the Sun by the sea surface, or sun glint, which severely contaminates the ocean color signal.

Prior to the MODIS/Terra and SeaWiFS launches, a study was performed by Gregg (Gregg 1998b), which determined through simulation that the amount of ocean coverage, expressed as a percentage of ocean area, should be between 15.2% and 16.5% for SeaWiFS and between 17.8% and 18.4% for MODIS/Terra. The expected coverage varies seasonally, with the minimum occurring in the Northern Hemisphere winter and the maximum occurring in the summer. Figure 1 shows the actual ocean coverage attained by SeaWiFS over the life of the mission, as well as the coverage attained by merging with MODIS/Terra and the total coverage achieved through the merger of SeaWiFS, MODIS/Terra, and MODIS/Aqua. The SeaWiFS data covers the period from 15 September 1997 through 3 May 2003. The MODIS data covers the entire Terra mission-period from 4 March 2000 through 3 May 2003, with MODIS/Aqua data beginning on 31 October 2002. The average coverage attained for each individual sensor, and for various combinations of the three sensors, is summarized in Table 7.1. Results for SeaWiFS alone show 15% daily coverage on average, with a standard deviation of 1.2%, which is in good agreement with the predictions by Gregg. In general, MODIS coverage is expected to be higher than SeaWiFS because of the wider swath width of MODIS. Average daily coverage from either MODIS sensor is about 21%. This is 3% higher than Gregg predicted, which may be due to the fact that the MODIS processing algorithm is obtaining ocean color retrievals deeper into the glint-contaminated regions than originally anticipated. Similarly, cloud contamination thresholds in the MODIS processing may be higher than expected. It has also been suggested that cloud cover is generally lower in the morning than at noon, so the

10:30 a.m. orbit of MODIS/Terra might be more favorable than the noon orbit of SeaWiFS, and this diurnal variation in cloud cover was not included in Gregg's model. However, the 1:30 p.m. orbit of MODIS/Aqua yields very similar coverage to that of MODIS/Terra, suggesting that, if diurnal variation in cloud cover is a significant factor, it would have to be symmetric about noon.

The merging of either MODIS sensor with SeaWiFS nearly doubles the daily coverage, from 15% for SeaWiFS alone to 28% for SeaWiFS and MODIS. This is 5% better than expectation. Adding MODIS/Aqua to the list has a less dramatic but still significant impact, with average daily coverage reaching 36%. To date, the maximum daily ocean coverage of the merged products occurred in the spring of 2003, with totals exceeding 40%.

Figure 7.1 also shows a seasonal cycle in the coverage, with maxima in the Northern Hemisphere spring and fall. This differs from Gregg's analysis, which anticipated maximum coverage in the summer. The seasonal cycle is largely driven by the amount of ocean area that is visible within the sunlit portion of the Earth's surface, as well as seasonally persistent cloud cover. The peak-to-peak variability is on the order of 2-3% for a given sensor. The same cycle is evident in both SeaWiFS and MODIS, so the effect is amplified in the merged product.

7.4 DATA ACCESS & DISTRIBUTION

The SDPS continues to generate merged chlorophyll products from MODIS and SeaWiFS, as new Level-3 products are received. This includes the full merged product of MODIS/Terra, MODIS/Aqua, and SeaWiFS, as well as merged pairs such as MODIS/Terra with SeaWiFS, MODIS/Aqua with SeaWiFS, and MODIS/Terra with MODIS/Aqua. In addition to the daily products, the system also generates weekly (8-day) and monthly products. These merged Level-3 bin products are further processed to SeaWiFS Standard Mapped Image (SMI) format, which is an HDF file containing a global image in a 0.1-degree plate carre projection. The SMI files are distributed through the Mapped Image browser on the SeaWiFS webpage. Since these files are in a standard SeaWiFS format, they can be viewed and manipulated using pre-existing SeaWiFS tools and utilities such as SeaDAS. In addition to the merged products, the SDPS also makes available the individual MODIS/Terra and MODIS/Aqua Level-3 daily, weekly, and monthly products in SMI format, thus allowing the ocean color community to work with MODIS data using familiar tools and procedures developed for SeaWiFS data. The SMI products for MODIS are available from the MODIS Oceans web site, as well as from the SeaWiFS Mapped Image browser.

Table 7.1: Mean daily coverage attained by SeaWiFS, MODIS/Terra (TMODIS), and MODIS/Aqua (AMODIS), both individually and in combination. Coverage is computed as a percentage of total area over the world's oceans.

Sensor Mix	Coverage (%)	Std. Dev. (%)
SeaWiFS	15.0	1.20
TMODIS	20.9	1.88
AMODIS	20.7	1.94
TMODIS/SeaWiFS	28.1	2.44
AMODIS/SeaWiFS	27.9	2.11
TMODIS/AMODIS	32.3	2.25
TMODIS/AMODIS/SeaWiFS	36.3	2.30

REFERENCES

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

Gregg, W.W., W.E. Esaias, G.C. Feldman, R. Frouin, S.B. Hooker, C.R. McClain, R.H. Woodward, 1998a: Coverage Opportunities for Global Ocean Color in a Multimission Era, *IEEE Trans. Geosci. and Remote Sens*., Vol. **36**, no. 5, pp. 1620-1627.

Gregg, W.W. and R.H. Woodward, 1998b: Improvements in coverage Frequency of Ocean Color: combining Data from SeaWiFS and MODIS*, IEEE Trans. Geosci. and Remote Sens*., Vol. **36**, no. 4, pp. 1350-1353.

O'Reilly, J.E., et al, 2000: SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. *NASA Tech. Memo. 2000-206892,* Vol. **11***,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, 49 pp.

.

Chapter 8

Ocean Color Data Merger

Ewa Kwiatkowska
Science Applications International Corporation, Beltsville, MD

8.1 INTRODUCTION

The objective of ocean color data merger is to create a consistent series of systematic ocean color measurements from multi-instrument, multi-platform and multi-year observations based on accurate and uniform calibration and validation over the lifetime of the measurement. The most obvious benefit of data merger is improvement in spatial and temporal ocean color coverage. Single sensor daily coverage is severely limited by gaps between consecutive swaths and gaps caused by clouds, sun glint and other phenomena which hinder the extraction of ocean color (Gregg *et al.*, 1998; Gregg and Woodward, 1998). For example, merged data from three global satellite sensors, MODIS on the Terra and Aqua platforms and SeaWiFS, provide only about 40% of global ocean and inland-water coverage at 9km resolution within a single day (Section 2.6). The other critical benefit is an increase in statistical confidence in extracted bio-optical parameters. Merger algorithms can utilize sensor-varying attributes, such as spectral, spatial, temporal, and ground coverage characteristics. Merger is the ultimate tool for the creation of ocean-color climate data records.

There are many difficulties associated with ocean color data merger. Sensors have varying designs and characteristics. There are disparate instrument calibrations, data processing algorithms, and validation accuracies. The same ocean color quantities can be derived using different spectral bands and different algorithms which may cause dissimilarities in mission standard products. Discrepancies in sensor characteristics, calibrations, and data processing create relationships between data products from different instruments which may show temporal trends and dependencies on sensor observation conditions. These relationships may also be noisy, indefinite and sometimes contradictory. Data especially susceptible to noisiness are those contaminated by clouds, dust, other types of turbid atmosphere, coastal waters, and mixed pixel representations. Another type of ambiguity arises from the fact that sensors are flown over the same regions at different times of a day. Natural changes in bio-optical conditions of the global ocean occurring over these time spans are hard to establish because they are difficult to discriminate from instrument and calibration artifacts.

Detailed objectives for the creation of a consistent series of multi-instrument and multi-year ocean color observations and related ocean Climate Data Records have not yet been defined. An objective way to assess accuracy of ocean color data is through comparisons, called matchups, with *in situ* measurements (Bailey *et al.*, 2001). However, ocean color matchups against *in situ* ship-born measurements are relatively sparse. This is because of the difficulties in acquisition of *in situ* observations and uncertainties involved in comparing *in situ* measurements against satellite-derived data. Over the sensors' lifetime, there have been 250 chlorophyll-a concentration matchup points, strictly screened for quality, for SeaWiFS and 34 for MODIS-Terra (http://seabass.gsfc.nasa.gov/matchup_results.html). Therefore, matchups with *in situ* observations are mostly used for intermittent validation of sensor data in concert with spatial and temporal data consistency analyses. The other approach to validation is matchup of ocean color data between sensors (Chapter 2; Kilpatrick *et al.*, 2001). This method assesses discrepancies between sensors over global to local zones and daily to seasonal time scales. Such assessments are vital for the data merger because they enable extraction of disparate trends and trend dependencies in data from different instruments.

There have been a number of methods developed to merge ocean color data. These methods include averaging and weighted averaging of data within the sensor overlapping coverage (Section 4.1). Blending algorithms have been applied which fit a function over shape-of-the-field defining data from one sensor given an internal boundary condition delimited by data from the other sensor or *in situ* observations (Gregg and Conkright, 2001). A semi-analytical optical algorithm has been developed which uses combined nLw retrievals within overlapping coverage at different sensor-specific wavelengths to calculate chlorophyll concentrations, combined detrital particulate and dissolved absorption coefficients, and particulate backscattering coefficients (Maritorena *et al.*, 2002; Maritorena *et al.*, 2000).

The major data merger effort undertaken by the SIMBIOS Project Office focused on integrating ocean color data from global sensors at a daily temporal resolution. MODIS-Terra and SeaWiFS data were used to study methodologies to create a consistent series of long-term observations from sensors of different design, characterization, processing algorithms, and calibrations. The information derived from MODIS-Terra and SeaWiFS comparisons, described in Chapter 2, was used to derive an ocean color sensor cross-calibration strategy to eliminate pronounced data temporal discrepancies between the

sensors and MODIS data artifacts. Statistical objective analysis was investigated to spatially and temporally interpolate MODIS-Terra (cross-calibrated with SeaWiFS) and SeaWiFS data onto daily global ocean color maps using individual sensor accuracies and producing error bars for each data point on the map. Additional research was performed to support local-area data merger applications, for instance in coastal zones. These applications utilized ocean color data of different spatial resolutions and *in situ* measurements. The multiresolution merger focused on enhancement of oceanic features in lower resolution imagery using higher resolution data.

8.2 CROSS-CALIBRATION OF MULTI-SENSOR AND MULTI-YEAR DATASETS TO CREATE A CONSISTENT AND CALIBRATED GLOBAL OCEAN COLOR BASELINE

Described in Chapter 2 results from collection-4 MODIS-Terra and reprocessing-4 SeaWiFS data comparisons showed that over three years of concurrent operation, there were significant discrepancies in ocean color measurements between the two sensors. Differences in water-leaving radiances and chlorophyll-*a* concentrations between the two instruments exceeded the maximum uncertainties preliminarily established for the creation of consistent merged ocean color products. These discrepancies furthermore varied temporally and spatially and were dependent on MODIS-Terra characterization problems for features such as the side of the optical mirror, response versus scan angle, and polarization sensitivity (Esaias *et al.*, 1998).

To create a consistent series of ocean color measurements from multi-instrument, multi-platform and multi-year observations, spatial, temporal and instrument artifact-driven discrepancies between sensor data had to be eliminated. To accomplish it a sensor cross-calibration approach was proposed. The goal of the ocean color sensor cross-calibration was to bring multi-instrument and multi-year data to a single well-calibrated and consistent baseline representation (Kwiatkowska, 2003). When data were cross-calibrated, daily global sensor observations could be combined to provide a joint ocean color coverage which was consistent through time and space and supported a number of applications, including the creation of ocean Climate Data Records.

Cross-calibration of ocean color sensors has been a typical approach to adjusting retrievals from instruments which do not have sufficient on-board calibration capabilities, such as Ocean Scanning Multispectral Imager (OSMI) (Franz and Kim, 2001). Sensors with operational calibrations have been also validated and cross-calibrated. Modular Optoelectronic Scanner (MOS) was vicariously calibrated using three overlapping scenes with SeaWiFS (Wang and Franz, 2000). SeaWiFS-obtained aerosol models and normalized water-leaving reflectances and MOS-derived aerosol concentration and molecular scattering estimates were used to calculate MOS top-of-the-atmosphere (TOA) radiances. For spatial fields of relatively uniform SeaWiFS TOA radiances and normalized water-leaving reflectances, MOS band calibration gains were obtained as a ratio of SeaWiFS to MOS TOA radiances. Ocean Color and Temperature Scanner (OCTS) and Polarization and Directionality of the Earth's Reflectances I (POLDER) sensors were vicariously recalibrated with common *in situ* nLw data and a consistent atmospheric correction to produce relatively comparable ocean color products with no obvious bias differences (Wang *et al.*, 2002). With more complex ocean color sensors currently on orbit, such as MODIS, the cross-calibration task is more difficult and a higher temporal and spatial accuracy of ocean retrievals is expected.

In this implementation, SeaWiFS was consequently chosen as the ocean color baseline data set because it had a long history of calibration and validation efforts (Barnes *et al.*, 2001; Eplee *et al.*, 2001), its data proved stable and self-consistent through the years, and it was invariably used as a standard against which other sensors were cross-calibrated. The cross-calibration aimed to bring MODIS-Terra product data to SeaWiFS-like values or, in other words, to emulate SeaWiFS response given MODIS data. This was accomplished by deriving a comprehensive machine learning approach. Machine learning methodology was attractive because the amount of required a priori knowledge about detailed sensor characterization, response changes, and the uncertainties of the radiative transfer modeling, calibration, and algorithms, was minimal. The machines learned from examples using abundant data from global overlapping coverage between MODIS and SeaWiFS. They worked as regularity detectors to discover statistically salient properties of investigated data.

Machine learning techniques have been used extensively in ocean color data processing. Various sensor calibration and atmospheric correction parameters were derived by means of regression (Barnes *et al.*, 2001; Eplee *et al.*, 2001; Gordon and Wang, 1994). The empirical algorithm associating chlorophyll concentration with water-leaving radiances was a result of the polynomial regression using *in situ* measurements (O'Reilly *et al.*, 1998). Multivariate optimization techniques were applied to ocean color data to simultaneously retrieve in-water biophysical conditions and aerosol optical properties in atmospheres containing weakly and strongly absorbing aerosols (Gordon *et al.*, 1997; Chomko and Gordon, 1998).

The cross-calibration strategy employed here was based on two foundations. Firstly, it defined a variety of MODIS data products and parameters necessary to alleviate the effects of MODIS temporal and spatial trends and artifact-driven dependencies in data. Secondly, large amounts of joint MODIS and SeaWiFS data were applied to serve as examples of these dependencies in order to determine the most appropriate regression functions from MODIS to SeaWiFS data. Three types of machine learning techniques were used in the current implementation: evolutionary computation, clustering, and regression.

8.3 MACHINE LEARNING TECHNIQUES FOR CROSS-CALIBRATION

Machine learning techniques were employed to define MODIS inputs to the cross-calibration which were essential in resolving the temporal and spatial trend and artifact-driven dependencies in MODIS data. The space of possible MODIS data products and parameters was searched for the most effective set of features enabling data dependence decorrelation and cross-calibration with SeaWiFS. These features needed to provide information on MODIS sensor and measurements that most unambiguously defined MODIS data and made possible one-to-one mapping to SeaWiFS. A set of possible features encompassed a variety of information describing MODIS data including water-leaving radiances at all visible ocean bands, chlorophyll-*a* concentration, K_490, atmospheric parameters such as AOT and ε, MODIS viewing and solar geometry, geographical situation, day in the time sequence, and ancillary meteorological setting of each MODIS data point. Conventional statistical and non-parametric rank-statistical correlation algorithms (Press *et al.*, 1992) were found insufficient to extract dependencies in MODIS data because of the nonlinearity and fuzziness of the cross-calibration problem. Consequently, an evolutionary-computation search mechanism was applied through a genetic algorithm (Goldberg, 1989). The genetic algorithm evaluated and propagated the fitness of various combinations of MODIS feature inputs through generations of non-parametric regression neural networks which mapped these inputs to SeaWiFS chlorophyll. The neural networks were trained on a multi-day MODIS and SeaWiFS data set which was scaled down for fast processing.

Clustering was employed in the cross-calibration process to partition the feature space of multi-day, multi-year, global overlapping MODIS-Terra and SeaWiFS data into clusters of similar feature values. Separate cross-calibration processes were then performed on each cluster data. This limited the complexity and improved the accuracy of the overall sensor cross-calibration. In this implementation, a Linde-Buzo-Gray clustering algorithm was applied (Linde *et al.*, 1980).

Regression was the actual tool that determined the relationship between MODIS and SeaWiFS data needed by the cross-calibration. The pre-merger validations presented in Chapter 2 showed that the cross-calibration problem was highly nonlinear and highly multidimensional with many dependencies present among data variables. Therefore, the cross-calibration depended closely on a suitable choice of MODIS input features, adequate clustering which spanned the entire space of MODIS and SeaWiFS feature data, and on the regression algorithm, which needed to be effective with highly nonlinear and ambiguous problems. Although the mapping between MODIS and SeaWiFS data could be performed using conventional linear or non-linear regression, the use of artificial neural networks or support vector machines was preferred (Pao, 1989). Neural networks and support vector machines could deal with flawed, biased, and cross-dependent sensor data because they are distribution free and can support highly nonlinear decision boundaries in the feature spaces.

Artificial neural networks have been widely employed in remote sensing, mainly however to solve classification problems, such as land cover or cloud type categorization (Atkinson and Tatnall, 1997; Ainsworth and Jones, 1999; Gross *et al.*, 2000; Tanaka *et al.*, 2000). In the current implementation, neural networks were initially employed to perform the regression between MODIS and SeaWiFS data (Kwiatkowska and Fargion, 2002a). However, with the training sets becoming massive and encompassing a large number of MODIS input features and multi-day, multi-year, global joint sensor coverages, the neural networks turned out too slow in training. Support vector machines replaced the neural network regression. Support vector machines are learning kernel-based systems that use a hypothesis space of linear functions in high dimensional feature spaces (Cristianini and Shawe-Taylor, 2000). Unlike neural networks, which try to define complex functions in the input feature space, the kernel methods perform a non-linear mapping of the complex data to high dimensional feature spaces and then use linear functions to create decision hyperplanes or regression functions (Schölkopf, 2000). The problem of choosing an architecture for a neural network is replaced by the problem of choosing a suitable kernel. Kernel functions project the data into high dimensional feature spaces to increase the computational power of the linear learning machines. The advantages of support vector machines over neural networks are that they train significantly faster, are better suited for work with high dimensional data, they often have a single minimum to search, and they allow for scaling the importance of outliers. Support vector machines have been applied in remote sensing to solve classification problems (Azimi-Sadjadi and Zekavat, 2000; Gualtieri *et al.*, 1999; Fukuda and Hirosawa, 2001). Applications of support vector machines in remote sensing to solve regression problems are sparse (Brown *et al.*, 2000) and there are no reported implementations involving ocean color data.

8.3.1 Machine Learning Implementation

The ocean color merger approach defined here was aimed at minimizing spatial and temporal discrepancies between MODIS and SeaWiFS data, and scan angle and latitudinal dependencies in MODIS data to produce consistent daily global coverages from both sensors. The goal of the sensor cross-calibration was to reproduce uniform SeaWiFS baseline response from MODIS data. To obtain the cross-calibration, support vector machines were used to approximate the regression functions from examples of MODIS and SeaWiFS concurrent coverages. Support vector machines learned complex relationships between MODIS and SeaWiFS ocean data and were required to extend this knowledge to unknown cases.

In this implementation, the regression was obtained between MODIS data and SeaWiFS chlorophyll to merge both sensor data into a time series of daily global chlorophyll baseline data sets. However, the regression could also be performed to generate baseline data sets for water-leaving radiances and other products. The regression functions were extracted using data from daily overlapping global bin coverages between MODIS-Terra and SeaWiFS over a significant time series. L3 binned data were used at 36km resolution. The 36km bin size was sufficient to extract generalized temporal and spatial trends and instrument and calibration artifacts from MODIS data. The total data set encompassed 44 days of joint MODIS and SeaWiFS coverage spanned through time from October 2000 to July 2002. This set was comprised of 1,672,188 examples and was subsetted in various ways into training and testing sets for the support vector machine learning. The results presented in this chapter were obtained by training on 42 days of overlapping MODIS and SeaWiFS coverage, where, for each day, data from alternating bins were placed into the training and testing sets. The remaining two days where used as an additional testing set to evaluate the machine's performance at extrapolating its knowledge through time to unknown days of data.

Before applying the regression, input data had to be prepared to facilitate the support vector machine training. Elimination of linear trends in data, seasonal components, and slow variation could be important data preprocessing tasks because the algorithm might ignore decisive subtle information present in the data in favor of large variations exhibited by a trend (Masters, 1994). The seasonal trends, North-to-South variations, and MODIS scan-angle dependencies shown by MODIS and SeaWiFS comparisons in Chapter 2 were however ambiguous and intertwined. For the current implementation, none of these deterministic components were removed from data. Instead, the regression was made to exploit indirect information about these conditions included in input feature data. The other data preprocessing consideration was scaling. Global chlorophyll concentration inherently forms lognormal probability density functions (Campbell, 1995). MODIS and SeaWiFS chlorophyll feature data were therefore passed through the logarithm function to provide the most effective spread of their distributions. All input feature data, including the chlorophyll logarithms, were then scaled and translated so that they were within the limits from 0 to 1 in value. The scaling simplified the regression because data values were then distributed within a small known range and encouraged equitable spread of importance among input feature elements.

The genetic algorithm found 16 MODIS input features to be the most effective in decorrelating MODIS data dependencies and these features were consequently used to perform the cross-calibration from MODIS to SeaWiFS data:

- nLw_412, nLw_443, nLw_488, nLw_531, nLw_551, nLw_678,
- chlor_a_2,
- Tau_865,
- Eps_78,
- satellite zenith angle,
- solar zenith angle,
- longitude and latitude,
- ozone amount,
- humidity, and
- date.

After initial regression training exercises, it was determined that a single support vector machine would be slow in learning decision functions on large data sets. Extensive data sets were however needed to present the regression algorithm with a large variety of possible feature distributions and relationships between MODIS and SeaWiFS data, especially considering the noisiness and fuzziness of these relationships and complex dependencies in data. To increase the efficiency and the accuracy of the regression, input training data were clustered and then separate support vector machines were trained on their corresponding data clusters. From plots and statistics it was established that the clusters adequately spread data distributions for all input features. The distributions were dependent on feature data probability densities within the training set. Parallel training on smaller data sets decreased support vector machine learning times by a few orders of magnitude, depending on the numbers of clusters and sizes of the training sets. Applying multiple support vector machines trained on their individual clusters of data increased the accuracy of the regression, which was critical to make the cross-calibration useful. One thousand twenty four (1024) clusters were used to obtain the results presented in this chapter, although different numbers of clusters were also investigated.

The regression emulated the response from the SeaWiFS baseline given data from MODIS. To obtain the regression-based cross-calibration, support vector machines were trained on clusters of MODIS feature data to reproduce corresponding SeaWiFS chlor_a values. There was one support vector machine for each cluster. The machines discovered the relationships between MODIS data and SeaWiFS chlorophyll and stored them in their support vectors and coefficients. The training was automatic and once accomplished, the stored vectors and system parameters could produce regression output for new data with no significant processing. The convergence of support vector machines depended on the choice of a kernel function. This study

found the radial basis function to perform better than linear, polynomial, and ANOVA kernels in support of regression between MODIS and SeaWiFS data (Schölkopf *et al.*, 1996). The results presented later in this chapter were obtained using radial basis kernels:

$$\text{radial basis function} = e^{-\gamma\|x-y\|^2},$$

where the γ parameter was equal to 1.0. A capacity parameter used in support vector machines measured the richness or flexibility of the regression functions and provided the protection against overfitting. An initial search for the capacity parameter, which improved the generalization accuracy of the learning system, resulted in selecting the capacity value of 8.4 for the subsequent investigations. Another parameter used in the training of support vector machines was ε. ε controlled the insensitivity of the regression by setting the predictions which lied within ε distance of their true values to be sufficiently accurate (Hearst, 1998). ε was chosen to be equal to 0.001.

8.3.2 Machine Learning Cross-Calibration Results

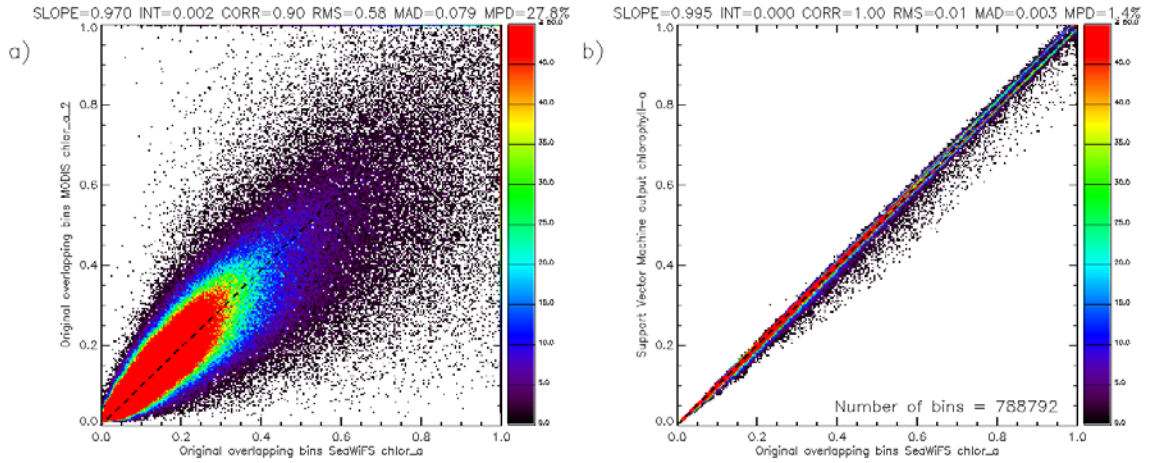
The MODIS and SeaWiFS cross-calibration approach evolved over dozens of regression training trials initially using neural networks and then support vector machines, applying different parameters and data sets for training and testing. Some of these trials were described in Kwiatkowska (2003). The results presented here were obtained by a recent version of the cross-calibration strategy.

A large set comprising 42 days of global overlapping MODIS and SeaWiFS data was used to create a representative training set for the support vector machine sensor cross-calibration. The dates spanned in time from October 2000 to July 2002 in about half-a-month intervals. For each day, data from every second overlapping bin between the sensors were placed into the training set and the remaining data were included in the testing set. There were 788,792 examples in the training set and 788,772 examples in the testing set. The additional testing set comprised complete 2 days of global overlapping data, which were separate from the 42 days used in training, and contained 94,624 examples. The training data was clustered into 1024 clusters and individual support vector machines were trained on each cluster.

To evaluate the regression, scatter plots, called matchups, were created for SeaWiFS chlorophyll versus the result of support vector machine regression from MODIS data. These plots were presented alongside the scatter plots of SeaWiFS versus original MODIS chlorophyll. Corresponding statistics were calculated for both data matchups and displayed with the plots. The statistics were used to quantify the improvements introduced by the machine learning cross-calibration to the time series of MODIS chlorophyll distribution in comparison with the SeaWiFS baseline. The statistics included slope (SLOPE) and intercept (INT) of the linear fit between SeaWiFS and MODIS data. To calculate the linear fit, an outlier-resistant linear regression was applied based on the robust Tukey's biweight calculated perpendicularly to the bisector of MODIS vs. SeaWiFS and SeaWiFS vs. MODIS data (Press *et al.*, 1992). The robust bisquare weighting ensured that the slope and intercept parameters were representative of the bulk of the data distribution and were not skewed by a few outlier points. Furthermore, matchup data robust correlation (CORR), root mean squared error (RMS), mean absolute difference ($\mathbf{MAD} = \frac{\sum_{i=0}^{n-1} \left| \mathbf{x}_{\mathbf{MODIS}_i} - \mathbf{x}_{\mathbf{SeaWiFS}_i} \right|}{\mathbf{n}}$), and mean percentage difference ($\mathbf{MPD} = 100\% \cdot \frac{\sum_{i=0}^{n-1} \frac{\left| \mathbf{x}_{\mathbf{MODIS}_i} - \mathbf{x}_{\mathbf{SeaWiFS}_i} \right|}{\mathbf{x}_{\mathbf{SeaWiFS}_i}}}{\mathbf{n}}$) were obtained. Statistics were calculated in the linear space of chlorophyll data. The linear space allowed an effective scrutiny of the MODIS and SeaWiFS cross-calibration statistics within most prominent ocean provinces, including low chlorophyll waters and the gyres, and high chlorophyll coastal zones. Because the chlorophyll probability density had a lognormal distribution (Campbell, 1995), matchups and corresponding statistics were also displayed in the logarithmic chlorophyll space, except for the MPD which was calculated in the linear space.

Support vector machines were accurately trained on the 42-day training set. Figure 8.1 shows a scatter plot of original SeaWiFS and MODIS chlorophyll values from overlapping daily coverage contained in the training set, a), and a scatter plot of original training SeaWiFS chlorophyll against the chlorophyll obtained by regression from MODIS data, b). In an ideal case, data within plot b) should lie on the y=x line and the training data are very close to this distribution. The support vector machines therefore learned the training set almost perfectly.

Linear scale



Logarithmic scale

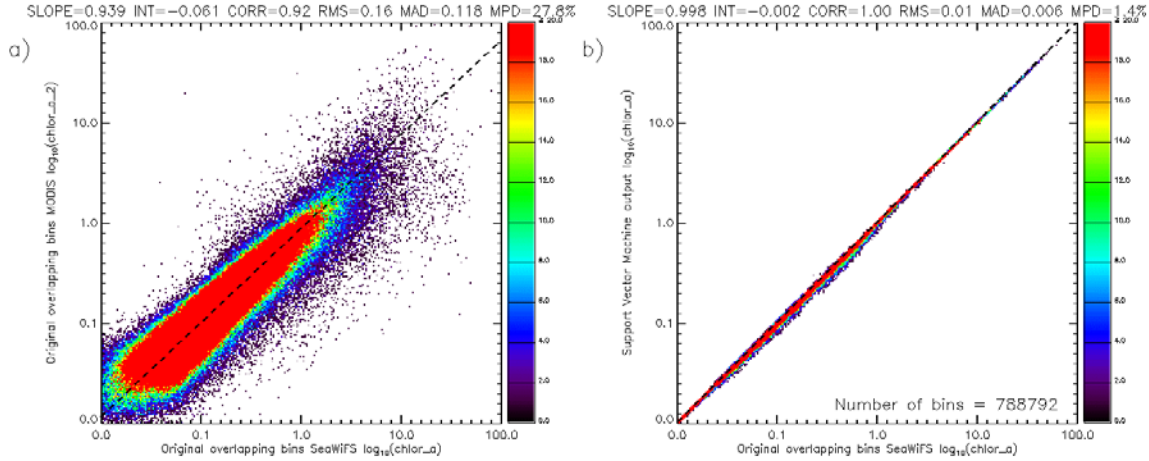


Figure 8.1: Scatter plots for the 42-day training data set a) SeaWiFS against MODIS training set chlorophyll and b) SeaWiFS training set chlorophyll against chlorophyll regressed from MODIS data. The support vector machines were trained highly accurately to reproduce SeaWiFS chlorophyll data.

The support vector machines were then evaluated against the 42-day testing set data. The result of the evaluation is displayed in Figure 8. 2. Figure 8.2b shows that the cross-calibration introduced substantial improvements to the general trends in MODIS chlorophyll distribution when compared to the original MODIS and SeaWiFS discrepancies from Figure 8.2a. The improvements are present in all statistical parameters obtained from the matchups, including enhanced slope and intercept of the linear fit in data, increased correlation, and decreased mean absolute differences. The mean percent difference between MODIS and SeaWiFS chlorophyll was decreased by over 10% across the two-year daily global time series. For the purpose of a consistent series of multi-instrument and multi-year ocean color observations, this decrease in overall sensor discrepancies was significant. Although the cross-calibration demonstrated the ability to correct the distribution of the majority of MODIS chlorophyll, the support vector machines also misclassified some examples which appear in plot b) as noise away from the y=x line. The low density of the misclassified points indicated that they were infrequent. The misclassifications were caused by the overall noisiness of the data set, the complexity and inconsistency of sensor data relationships, and the difficulty to establish a fine boundary between regression function overfitting and overgeneralization.

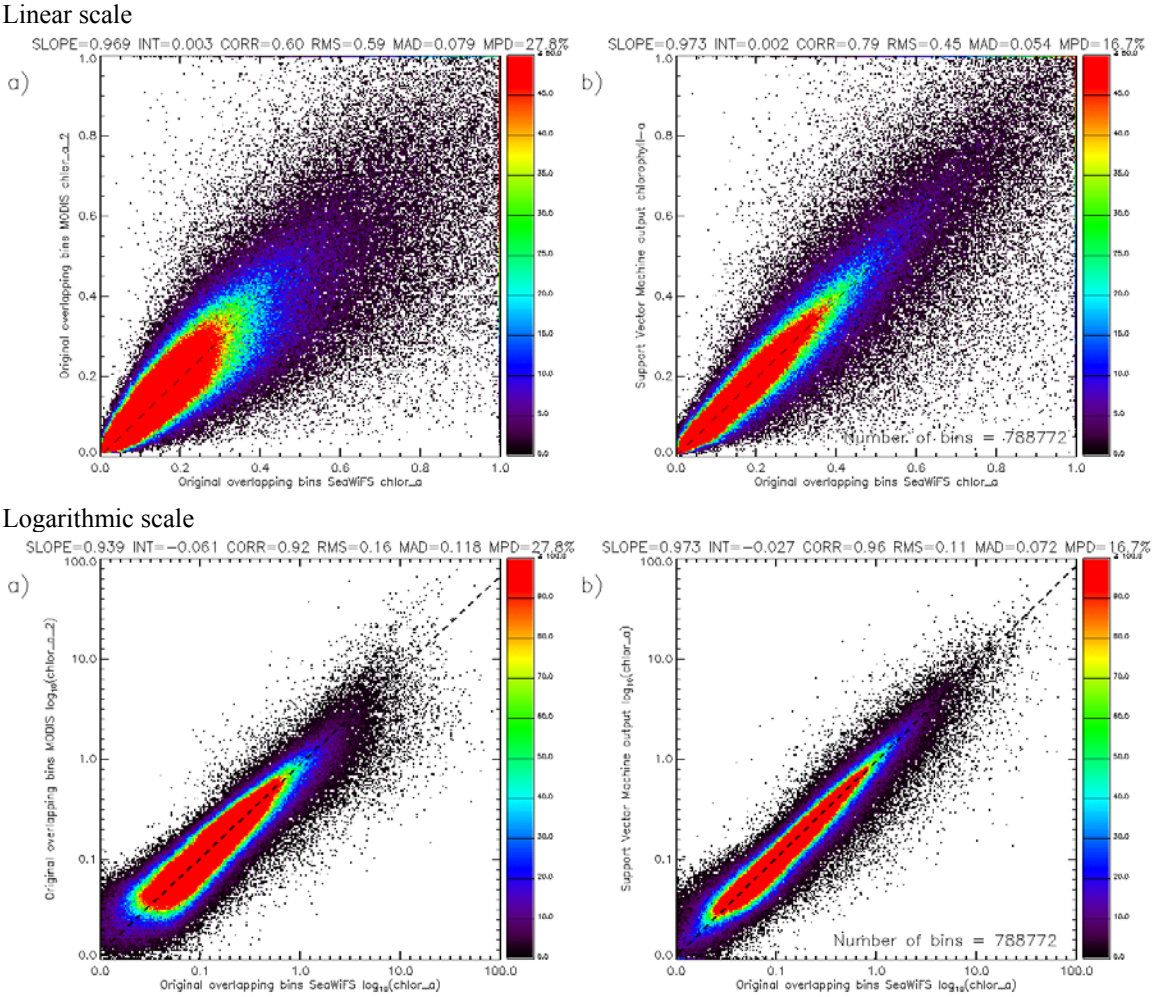


Figure 8.2: Scatter plots for the 42-day testing data set: a) SeaWiFS against MODIS testing set chlorophyll, and b) SeaWiFS testing set chlorophyll against chlorophyll regressed from MODIS data. The support vector machine regression substantially improved the distribution of the bulk of MODIS chlorophyll data compared to SeaWiFS chlorophyll.

The significance of the improvements obtained through MODIS and SeaWiFS cross-calibration was demonstrated by the reduction of artifacts present in MODIS data. To evaluate the changes in MODIS data distribution in more detail, the regression's impact on MODIS data temporal trends and scan-angle and latitudinal dependencies was investigated. The testing results of support vector machine cross-calibration were separated into the individual days and analyzed independently. For each day, matchups were created between original MODIS and SeaWiFS testing chlorophyll and between SeaWiFS chlorophyll and the result of the support vector machine regression. The MPD and the slope of the linear fit gave a good estimation of the discrepancies between the corresponding data sets (Chapter 2) and were used to evaluate the performance of the mapping.

Figure 8.3 illustrates MPDs, calculated against the SeaWiFS benchmark, plotted for each day and connected by lines through 42 days of the time series. The MPDs were obtained between SeaWiFS and MODIS chlorophyll and between SeaWiFS chlorophyll and the result of the regression mapping from MODIS data. Throughout the time series, the MPDs are consistently and substantially smaller for the support vector machine chlorophyll than for MODIS chlorophyll. Figure 8.3 also shows that the MPDs between SeaWiFS and original MODIS chlorophyll vary seasonally. These seasonal sensor chlorophyll disparities disappear in the MPD plot for SeaWiFS and the regression result. The support vector machine cross-calibration therefore eliminated the seasonal trends in MODIS and SeaWiFS data discrepancies and decreased the differences in chlorophyll data between the sensors.

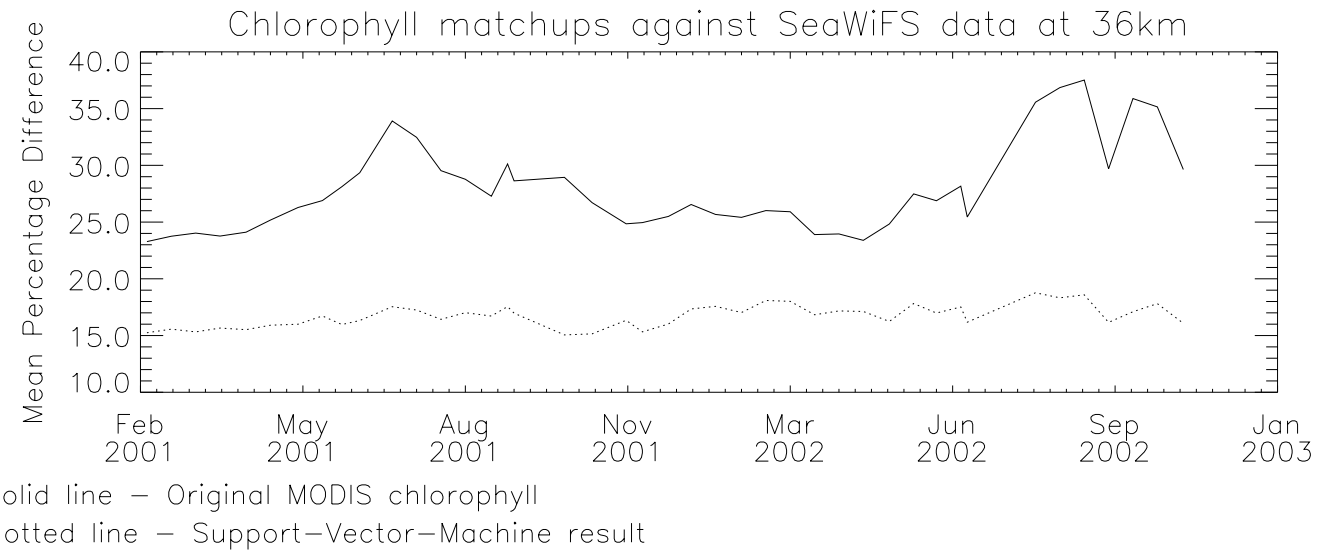


Figure 8.3: Time trends in daily mean percent differences between SeaWiFS and original MODIS chlorophyll and between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data.

From MODIS and SeaWiFS comparisons of daily global data sets described in Chapter 2 it was established that individual matchups with SeaWiFS for MODIS western and eastern scan-part data revealed distinct patterns in MODIS data behavior at different scan angles. To investigate the change in MODIS scan angle dependencies obtained by MODIS and SeaWiFS cross-calibration, MODIS daily global data were divided into subsets corresponding to their scan angle coverages and the subsets were individually matched against SeaWiFS chlorophyll. The two subsets of main concern corresponded to data located on the western and eastern scan edges of the MODIS swath. The scan edge widths constituted almost 40^0 of MODIS zenith angle coverage for each part of the scan. For each day within the 42-day time series, MODIS chlorophyll data from the testing set, located within these two scan-edge coverages, were separately matched against SeaWiFS test set chlorophyll. Similarly, the output of the support vector machine tests corresponding to MODIS western and eastern scan edges was compared against SeaWiFS test set chlorophyll. It was demonstrated in Chapter 2 that slope and intercept of the linear fit between the matched data gave a good estimate of scan angle dependencies in MODIS products. Figure 8.4 contains plots of slope values obtained in the two comparisons between testing matchup data coinciding with MODIS western and eastern scan-edge coverages. To produce the figure, the slope values were connected across the 42-day time series into slope lines. Scan angle dependencies in original MODIS chlorophyll appear in Figure 8.4a as systematically different slope values for western and eastern MODIS-scan test data in matchups with SeaWiFS. The slope lines obtained from matchups against support vector machine testing results in Fig. 4b run almost conjointly. This provided evidence that MODIS scan angle dependency was adequately eliminated from the regression result by the sensor data cross-calibration.

Comparisons of daily global MODIS and SeaWiFS data described in Chapter 2 demonstrated that sensor data discrepancies exhibited latitudinal dependencies. To investigate change in the latitudinal pattern produced by MODIS and SeaWiFS cross-calibration, daily global testing set data were divided between the northern and southern hemispheres and the two coverages were investigated independently. Fig. 8.5 illustrates slope in the linear fit between SeaWiFS and original MODIS chlorophyll and between SeaWiFS and the result of support vector machine regression obtained separately for the northern and southern hemispheres over the 42-day time series of testing set data. Fig. 8.5 demonstrates that, while the northern-to-southern hemisphere discrepancies were very prominent in original chlorophyll data in plot a), the discrepancies largely disappeared from the cross-calibrated data in plot b). Support vector machine regression therefore eliminated latitudinal dependencies in MODIS data in comparison with SeaWiFS chlorophyll.

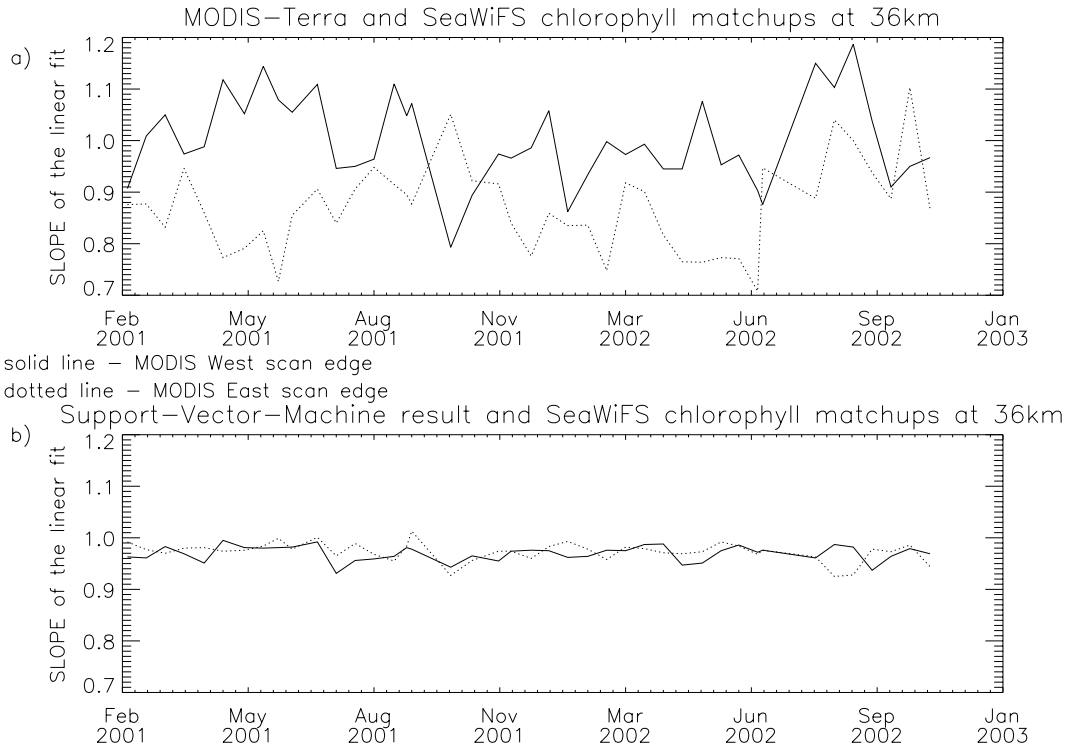


Figure 8.4: Time trends in the slope of the linear fit between a) SeaWiFS and original MODIS chlorophyll for separate western and eastern MODIS scan-angle coverages and b) between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data for the same MODIS scan coverage.

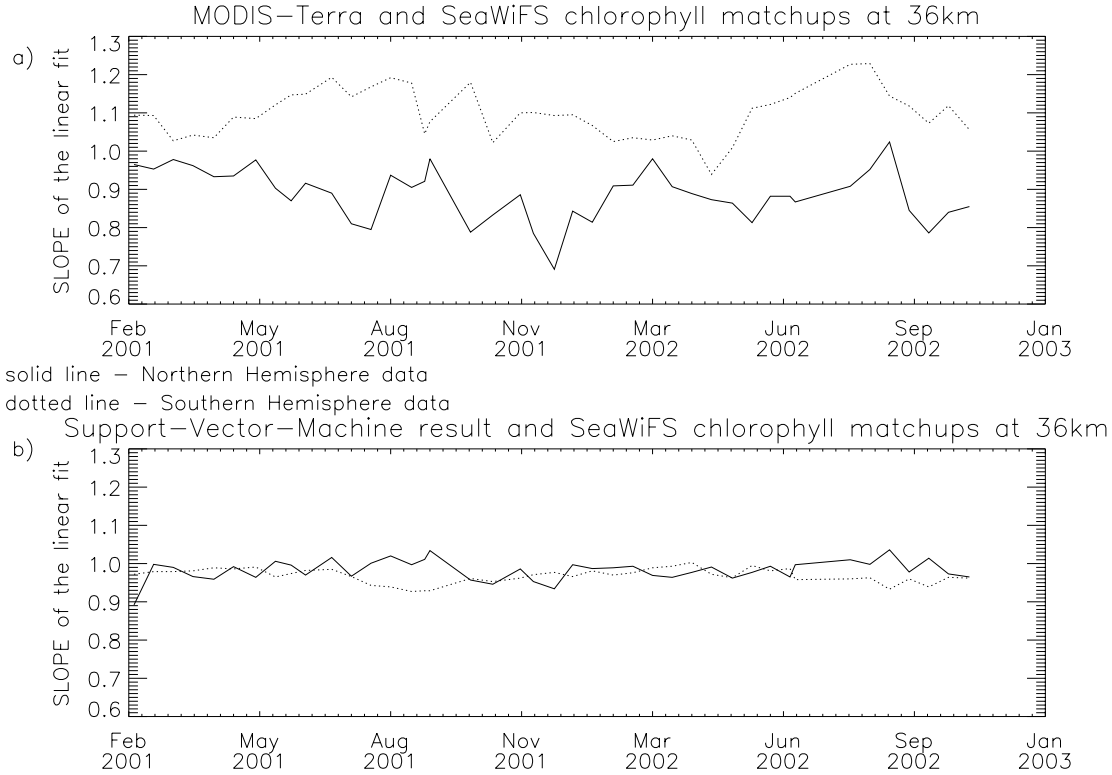


Figure 8.5: Time trends in the slope of the linear fit between a) SeaWiFS and original MODIS chlorophyll for separate northern and southern hemisphere coverages and b) between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data for the corresponding hemispheres.

MODIS and SeaWiFS ocean color comparisons from Chapter 2 demonstrated that discrepancies between the two sensor data manifested significant temporal and spatial variabilities and involved many dependencies among data variables. It was therefore challenging to extrapolate the knowledge of MODIS and SeaWiFS data relationships gained over a sparse time series onto the entire two-year period of the concurrent sensor coverage. There were two additional days in the data set which were not applied in the support vector machine training. Their purpose was to serve as a supplementary testing set to verify the support vector machine capabilities to extrapolate their knowledge through time to unknown dates. The days were inside the October 2000 to July 2002 period used in the training. Fig. 6 shows the scatter plots between SeaWiFS and original MODIS chlorophyll for these two dates, a), and between SeaWiFS chlorophyll and the cross-calibration result, b), where the cross-calibration support vectors were obtained from the 42-day training set. Fig. 6b illustrates that the cross-calibration knowledge gathered within the 42 days of combined MODIS and SeaWiFS coverage transferred relatively well to the new data dates. The bulk of the support vector machine testing result closely approximated SeaWiFS chlorophyll. This was evident from improved scatter plot distributions. The MPD in plot b) was not as low as the value obtained on the 42-day testing set, which showed an average of 16.7% in Figure 8.2, but it decreased from the original MODIS and SeaWiFS MPD in plot a). All other statistics were comparable or better than those achieved on the 42-day testing set, including a substantial improvement in the slope of the linear fit between SeaWiFS and the cross-calibrated chlorophyll.

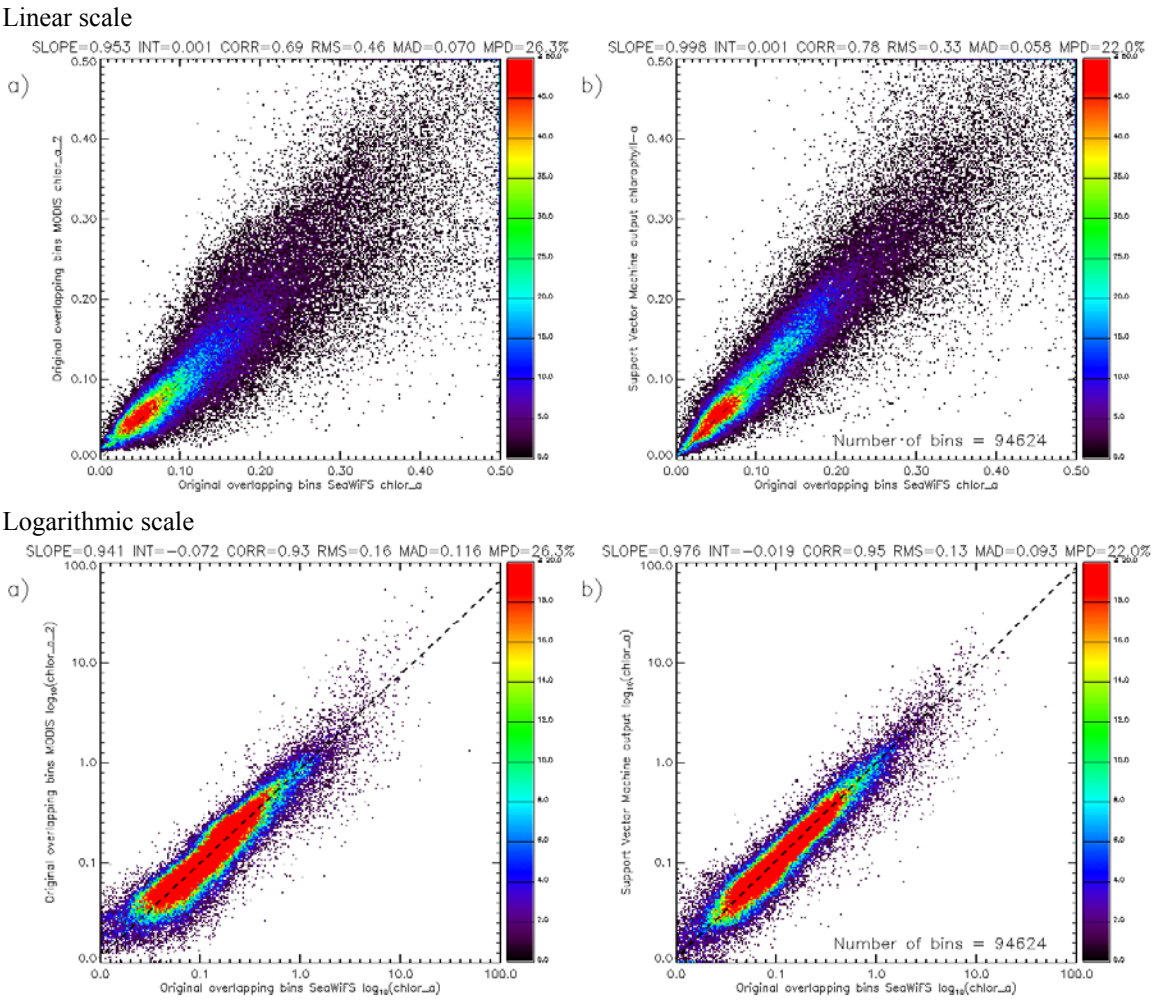


Figure 8.6: Scatter plots for the 2-day testing data set separate from the 42-day time series used in the support vector machine training: a) SeaWiFS versus MODIS testing set chlorophyll, and b) SeaWiFS testing set chlorophyll versus chlorophyll regressed from MODIS data. The regression was able to extrapolate its cross-calibration knowledge through time to new data days and substantially improve the distribution of the bulk of MODIS chlorophyll data compared to SeaWiFS chlorophyll.

The cross-calibration was also investigated for its ability to extrapolate its knowledge of scan angle and latitudinal dependencies in MODIS data onto unknown data dates. The 2-day testing set was sub-sampled into western and eastern MODIS scan edge coverages and northern and southern hemispheres. The results of the corresponding matchups are displayed in Figure 8.7. The figure demonstrates that the support vector machines were still effective at eliminating scan angle dependencies and northern-to-southern hemisphere discrepancies in MODIS data. Consequently, their cross-calibration experience gained with a limited time series could be transferred to the complete time span of data.

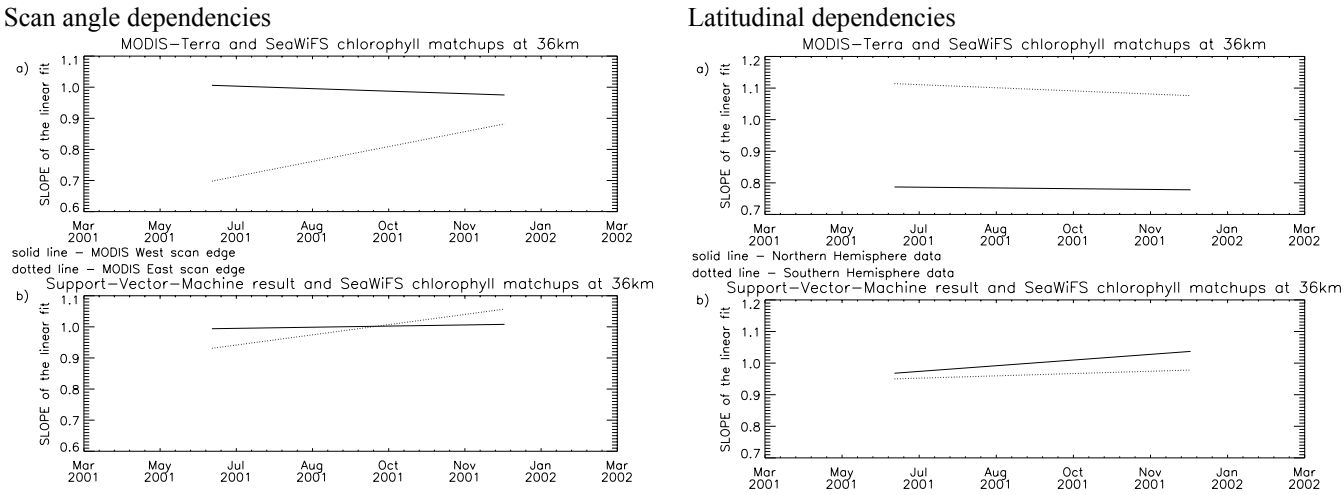


Figure 8.7: Time trends in the slope of the linear fit for data corresponding to separate western and eastern parts of the MODIS scan and northern and southern hemisphere coverages between a) SeaWiFS and original MODIS chlorophyll and b) between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data.

The goal of sensor cross-calibration to reproduce uniform SeaWiFS baseline response from MODIS data was consequently accomplished. Spatial and temporal discrepancies between MODIS and SeaWiFS data and scan angle and latitudinal dependencies in MODIS data were significantly reduced. Support vector machines were also able to extend their knowledge of complex relationships between MODIS and SeaWiFS ocean data to new unknown cases. A consistent series of daily global chlorophyll measurements could then be produced from MODIS and SeaWiFS by using MODIS data cross-calibrated with SeaWiFS. Fig. 8 contains merged MODIS and SeaWiFS global chlorophyll for 14 May 2001, where MODIS unique coverage for this day was regressed to the SeaWiFS baseline.

3.3.3 Machine Learning Conclusions

The objective of creating global ocean color data sets from multiple satellite sensors is important in the era of many concurrent ocean-observing missions. This goal is, however, hampered by incompatibilities in product data between the missions. MODIS-Terra and SeaWiFS ocean color data sets revealed significant discrepancies, described in Chapter 2, which were dependent on sensor calibrations and operational characteristics. These discrepancies inhibited the creation of consistent daily global merged data sets from both sensors. To bring MODIS ocean data to the SeaWiFS baseline, the application of machine learning cross-calibration was investigated. Support vector machines were trained to emulate SeaWiFS baseline chlorophyll from MODIS data. The ultimate objective was to produce joint MODIS and SeaWiFS daily global coverages which had the accuracy and the spatial and temporal consistency of SeaWiFS data sets.

Machine learning regression turned out to be a promising tool for the data merger. Support vector machines were able to accurately learn complex relationships between MODIS and SeaWiFS data and to effectively reduce sensor data discrepancies and eliminate MODIS artifacts, such as seasonal trends, scan angle dependencies, and spatial variation. Overall, the machines performed well within the time series on which they were trained and also proved the capability to extrapolate their knowledge to the entire time span of concurrent operations of the instruments. The performance of the machines can be improved by forming training sets that are more representative of the total MODIS and SeaWiFS time series and by reducing the noise in the data. Also, the support vector machine regression can be further investigated for parameters and implementation additions to make it more robust and accurate.

Although the machine learning approach presented in this paper regards cross-calibration of MODIS and SeaWiFS global chlorophyll-*a* products, the mapping into other sensor product data can also be performed. For example, MODIS data can be

used to predict SeaWiFS nLw measurements at various wavelengths. Chlorophyll concentration can then be calculated from these radiances using the standard SeaWiFS OC4v4 algorithm (O'Reilly et al., 1998). Top of the atmosphere reflectances can be mapped between the instruments given different sensor and solar geometries and atmospheric paths. Also, various other sensor parameters, such as adjustments to sensor calibration gains, can be mapped using the machine learning methodology.

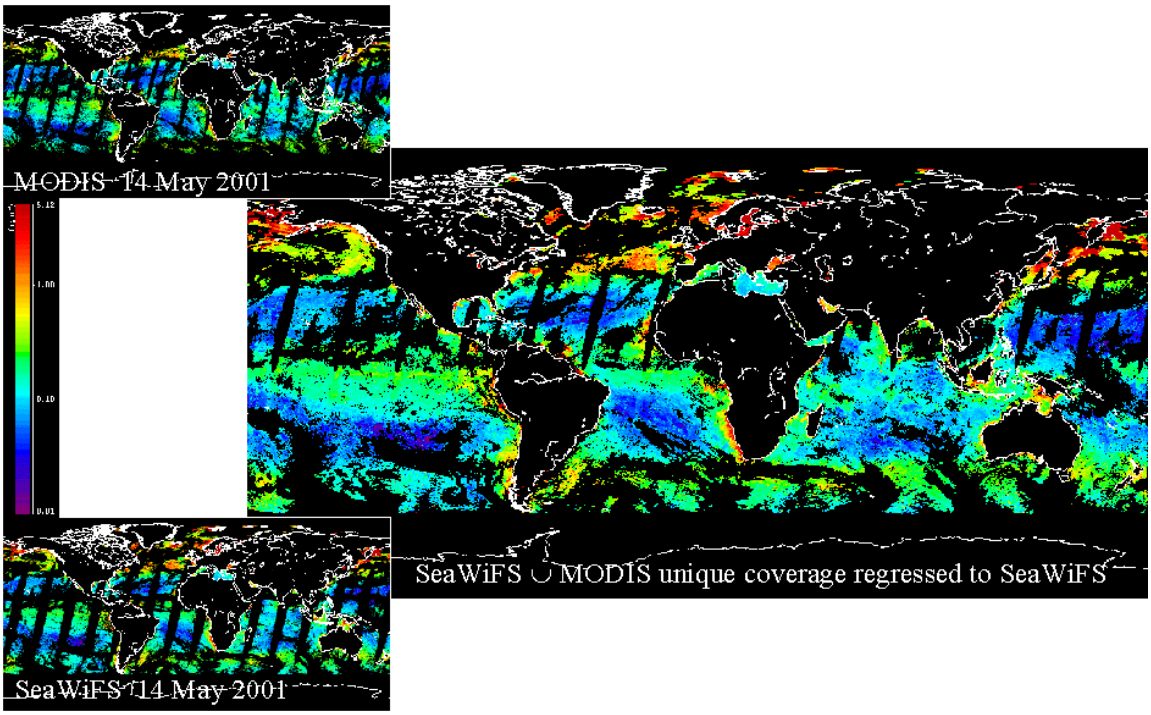


Figure 8.8: Merged MODIS and SeaWiFS daily global chlorophyll baseline.

8.4 STATISTICAL OBJECTIVE ANALYSIS: SPATIAL AND TEMPORAL INTERPOLATION OF MULTI-SENSOR OCEAN COLOR DATA ONTO DAILY GLOBAL BINNED COVERAGE

Spatial and temporal interpolation can be applied to merge multi-sensor daily ocean color data onto global coverage grids (Kwiatkowska and Fargion, 2002b). The interpolation can be used in two ways. Firstly, single sensor data can be adeptly interpolated onto a global grid and to fill sensor's gaps in ocean coverage. Multi-sensor data can then be integrated by joint binning, a technique that is comparable to the "big-bin" approach (Section 4.1) or by optical algorithms using combined nLw retrievals at sensor-specific wavelengths (Maritorena *et al.*, 2002; Maritorena *et al.*, 2000). This would avoid the circumstances which limit the spatial consistency of the merged coverage, when some bins contain data from only one of the sensors and some bins contain a mix of data from multiple sensors. Secondly, multi-sensor data can be concurrently integrated onto a global grid using corresponding sensor data accuracies. Both methods of interpolation should preferably work with temporally and spatially stable ocean color measurements because applying time and space dependent instrument error estimates may be ambiguous and impractical.

For the purpose of this ocean color merger application, spatial and temporal interpolation was envisioned as a binning approach for multi-sensor data. The binning could fill many gaps present in daily global ocean color coverage depending on space and time distances between gaps and existing data. The binning was designed to operate on ocean color output products, such as chlorophyll-*a* concentration. Interpolation considered a spatial and temporal correlation structure of the chlorophyll field, which was dependent on the local area natural variabilities. Prior sensor cross-calibration was performed, such as in Section X.1, to bring the multi-sensor data to a consistent baseline and eliminate sensor temporal trends and data artifacts. Sensor data accuracy used as a weighting factor in the interpolation was calculated from matchups with *in situ* measurements.

A well-known method to perform interpolation of environmental data is statistical objective analysis (Thiebaux and Pedder, 1987). Objective analyses use numerical methods to estimate geophysical field variables on surfaces and on three- and four-dimensional grids from data that are available at discrete locations and times. The method was first applied in meteorology to ground and satellite measurements. It calculates an interpolated grid-point value as a weighted linear combination of observations (Thiebaux, 1973). In an empirical linear interpolation, weights are either a function of separation

between analysis and observing locations or a function of accuracy of one observation relative to another. A distance weighting with weight normalization is the most common. In statistical objective analysis, the approximation is obtained by making additional use of the ensemble spatial correlation structure of the whole field, i.e. the spatial distribution of observations relative to one another (Julian and Thiebaux, 1975). The analysis considers instrument errors and other variations in data so that an interpolated value of a field variable does not have to be identical with an observed value at corresponding space/time coordinates, but it is intended to coincide with the signal component of the observed variable.

The practical requirement for the use of this algorithm is that there has to exist a preliminary "prediction" of the signal, or a first-guess field, and the objective analysis corrects this prediction by interpolating the signal with a single or multiple passes of the algorithm. At successive corrections, non-zero weights are given to observed increments only if the observations lie within a prescribed distance, known as the influence radius, of the grid point being considered. This influence radius may be decreased with successive passes of the algorithm. The analyzed grid point value is written as $\boldsymbol{Y_o} = \sum_{j=1}^{m} \boldsymbol{h}_j (\boldsymbol{X}_j^o - \boldsymbol{X}_j^f) + \boldsymbol{X}_o^f$, where m is a number of observing locations, Y_o is the interpolated value of a grid point, X^o_j is an observation value at point j, X^f_j is a first guess for the jth observation point, X^f_o is the first guess for the analyzed grid point, and $\boldsymbol{h}$ is the weight vector. The weights are obtained by minimizing the ensemble average of the squared difference between the analysis value and the true value of the field signal (Thiebaux and Pedder, 1987). The solution to this minimization problem is a covariance array for the joint distribution: $\begin{pmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_m \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{01} \\ \vdots \\ \sigma_{om} \end{pmatrix}$, where σ_{ij} is a covariance between an ith and jth observation.

Statistical objective analysis has been applied to create NOAA's real-time global sea surface temperature (SST) maps (Reynolds, 1988; Reynolds and Smith, 1994). The maps are produced weekly on a one-degree grid. The analysis uses buoy, ship, and satellite SST data, and SST's simulated by sea-ice coverage. The approach applies individual sensor errors and a globally averaged space-lag correlation structure of the SST field.

8.4.1. Statistical Objective Analysis Implementation

For this study, the covariances between chlorophyll data points were expressed in terms of space and time-lag correlation functions, ρ(s), which were calculated from ocean color data. This assumed that the variance of the chlorophyll concentration truth-value was a constant σ^2 at all locations, the noise variance of chlorophyll was a constant η^2 and independent of location, the space-lag covariance of the chlorophyll-truth was isotropic, and the covariances of the noise at different locations were zero (Thiebaux and Pedder, 1987). The weighting scheme for the chlorophyll-interpolated truth-value was then written as: $\begin{pmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \\ \vdots \\ \boldsymbol{h}_m \end{pmatrix} = \begin{pmatrix} 1+\gamma & \rho(\boldsymbol{s}_{12}) & \cdots & \rho(\boldsymbol{s}_{1m}) \\ \rho(\boldsymbol{s}_{12}) & 1+\gamma & \cdots & \rho(\boldsymbol{s}_{2m}) \\ \vdots & & \ddots & \vdots \\ \rho(\boldsymbol{s}_{1m}) & \rho(\boldsymbol{s}_{2m}) & \cdots & 1+\gamma \end{pmatrix}^{-1} \begin{pmatrix} \rho(\boldsymbol{s}_{01}) \\ \rho(\boldsymbol{s}_{02}) \\ \vdots \\ \rho(\boldsymbol{s}_{0m}) \end{pmatrix}$, where s_{ij} is a distance between the ith and jth points and γ^{-1} is the signal-to-noise ratio, σ^2 / η^2, of the chlorophyll concentration product.

The assumptions for this equation were not met for chlorophyll data and modifications were introduced into the algorithm. Chlorophyll truth and noise variances vary depending on the amount of chlorophyll concentration which fluctuates through five scales of magnitude from 0.001mg/m^3 to 100mg/m^3. Because chlorophyll has a lognormal distribution, analyzing chlorophyll in a logarithmic space shrinks the range of data values to a single scale of magnitude (Campbell, 1995). To use this property, chlorophyll and chlorophyll-error variances were derived from *in situ* matchups in the logarithmic chlorophyll space and the logarithmic signal-to-noise ratio was assumed constant at all locations. For each interpolation grid point, chlorophyll values were then converted to the logarithmic space to tie in with the logarithmic statistics.

Modeling space-lag correlation functions is a subject of extensive research (Julian and Thiebaux, 1975). The correlation is expressed as a function of the spatial separation of locations of points in geographic coordinates. For this study, a 3-dimensional statistical objective analysis was investigated using time as the third dimension and with data points separated by a day or a number of days to interpolate a given grid location. For the 3-dimensional analysis, space-lag correlations were derived for data at different distances and 0 to 7 days apart. The space and time-lag correlation functions were initially

calculated globally over daily chlorophyll concentration fields. Afterward, non-isotropic space and time-lag covariance of the chlorophyll-truth was investigated. Space and time-lag correlation of the chlorophyll field was assumed to be dependent on local area spatial and temporal variabilities. Chlorophyll variabilities were modeled using a standard deviation function. The variabilites were derived using 9-day and biweekly MODIS-Terra and SeaWiFS global L3 chlorophyll maps at an initial spatial resolution of 36km, chosen to limit processing time. Standard deviations were dependent on the radius of the local area under investigation and on the average chlorophyll magnitude within the area. Standard deviations were approximated for different chlorophyll magnitudes and 13 classes of ocean variabilities were defined based on the standard deviation functions. Each ocean data point, a bin, on the global map was then assigned to a vector of chlorophyll variability classes. The vector was composed of chlorophyll variability classes at consecutive distance ranges from 0 to 1000km at 10km intervals from the point under consideration. Space and time-lag correlations could then be calculated in a manner dependent on the classes of chlorophyll variability in the ocean. Correlations were obtained for chlorophyll value increments from the first-guess background field at different distance ranges. Separate increment correlations were calculated spatially within a single given day and between the given-day chlorophyll and chlorophyll a number of days away. The background field was assumed to be a global chlorophyll 9-day mean. To compute the correlations, data were applied from days which followed the week used as the first-guess estimate. The correlations across consecutive distance ranges were calculated separately for each chlorophyll variability class. The components for the increment correlation functions came from the global chlorophyll maps and, for each point, used its variability classes at corresponding distance intervals. Fig. 9 shows preliminary results of space-lag correlations for increments from the 9-day global chlorophyll first-guess field at the same day and 2-day time intervals. The results of the calculations were approximated by exponential functions. The 13 correlations are shown in different colors corresponding to their classes of chlorophyll spatial variabilities, from low variance in dark blue to high variance in red. The figure also displays global ocean variability maps for the 13 spatial variability classes with the same color-coding. The first map shows classes of variability within a 100km radius and the second – classes of variability within a 600km radius.

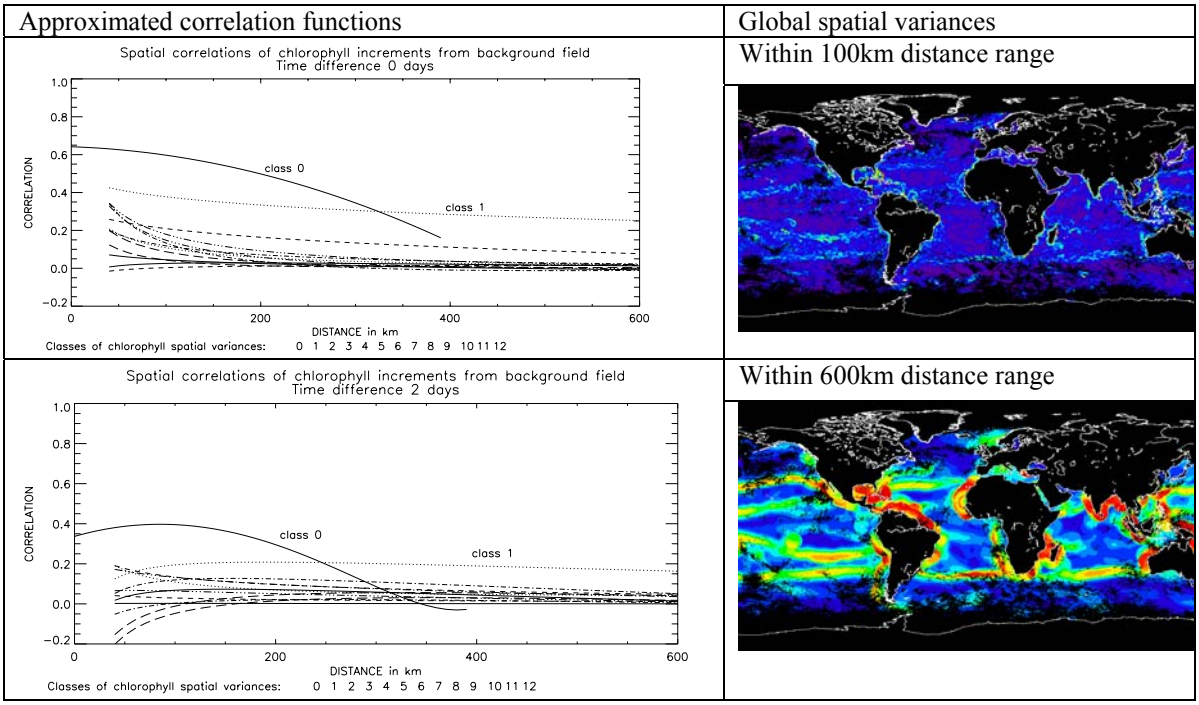


Figure 8.9: Space-lag correlation functions for chlorophyll increments from the first-guess 9-day mean field averaged over global data sets and within 10km distance intervals between the points. The 13 functions shown in different colors correspond to different spatial variability classes, from low variance in dark blue to high variance in red. Global spatial variability distributions are shown on the right where the variances where calculated within 100km and 600km distance ranges, correspondingly. The correlations were approximated from MODIS-Terra L3b 36km global time series.

Figure 8.9 illustrates that natural spatial variability of phytoplankton limits the extent of spatial and temporal interpolation of ocean color data. The spatial-lag correlations for chlorophyll increments were relatively low. The class number 1 at time difference of 0 days had correlation values around 0.4 for short distances which did not decrease to 0 for longer distances. High

variability classes formed increment correlation functions not higher than 0.4 and approximating 0 at short distances. This apparent spatial diversity of chlorophyll concentration increments, averaged over global scales, showed that chlorophyll data differed from meteorological data, which were better correlated and for which the statistical objective analysis was originally created (Thiebaux and Pedder, 1987). The correlation functions were dependent on the definition of the first-guess background field. A relatively up-to-date and complete chlorophyll map had to be chosen to initialize the analysis. Eventually, the first-guess field would be the previous day global chlorophyll coverage obtained by the preceding step of the interpolation. Therefore, the correlation functions could change somewhat when a more appropriate first-guess field is applied. The influence radius for the analysis was defined to be equal to 600km, following information about the shape of the space-lag correlation functions for chlorophyll increments. With low correlation values and the short influence radii, the statistical objective analysis cannot interpolate chlorophyll grid points which lie relatively far from valid data.

8.4.2 Statistical Objective Analysis Results

Because the current investigations of the statistical objective analysis were preliminary, only single-sensor-based interpolation was performed to fill sensor's gaps in ocean coverage (Kwiatkowska and Fargion, 2002b). A simple logarithmic signal-to-noise ratio of chlorophyll data was calculated by dividing the mean value of chlorophyll by the chlorophyll variance both derived from the matchups with *in situ* measurements. Two-dimensional and 3-dimensional forms of the analysis were tested only for globally averaged chlorophyll increment correlations without considering local area variability ranges. A globally averaged space-lag correlation function for chlorophyll-*a* concentration in single-day increments is illustrated in Fig. 10. The function was created from chlorophyll-increment correlations from a weekly mean and was approximated using an exponential function: $\mathbf{y} = 0.582708 \cdot 0.984943^{x} + 0.0189976$, where X was a distance between points expressed in kilometers. The function was obtained using SeaWiFS L3b daily global chlorophyll time series at 36km resolution. A value of 1 was assumed when points within the same-day overlapped spatially and a value of 0 was assumed when the points were outside the radius of influence, which was set at 600km. Space-lag correlation functions were similarly approximated for points separated by up to 3 days from the interpolated day.

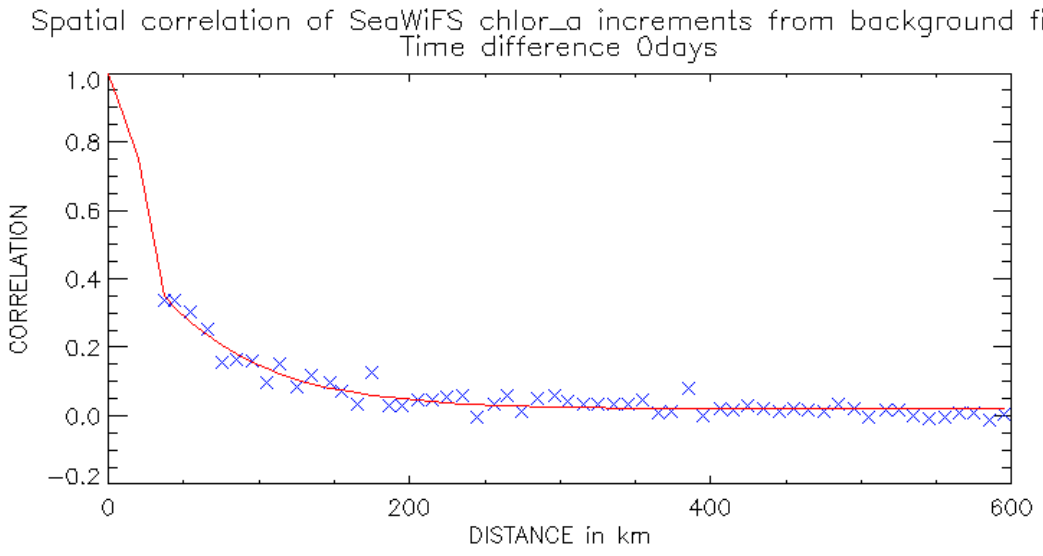


Figure 8.10: Chlorophyll increment space-lag correlation function averaged over a single-day global data set and within 10km distance intervals between the points. The first-guess background field was the weekly global chlorophyll mean.

Figure 8.11 shows the result of the spatial statistical objective analysis on SeaWiFS L3 binned daily global chlorophyll coverage at 36km resolution for 8 April 2001. Only those SeaWiFS grid points were interpolated which, for this day, coincided in coverage with MODIS bins containing valid data. If a similar interpolation was done using this day's MODIS data, MODIS and SeaWiFS chlorophyll concentration products could be merged by means of averaging or weighted averaging, where all valid bins would contain data from both sensors. For the analysis to be applied with the semi-analytical optical algorithm (Maritorena *et al.*, 2002), the interpolation had to be done on MODIS and SeaWiFS nLw products on the bands used for

chlorophyll extraction. Ultimately, the interpolation would be performed jointly on MODIS and SeaWiFS chlorophyll using corresponding statistics for both sensors and, preferably, as a binning scheme beginning with the L2 products. MODIS data would then be first cross-calibrated with SeaWiFS because, otherwise, MODIS trends and artifacts would be inseparably intertwined with SeaWiFS data in the merged data set.

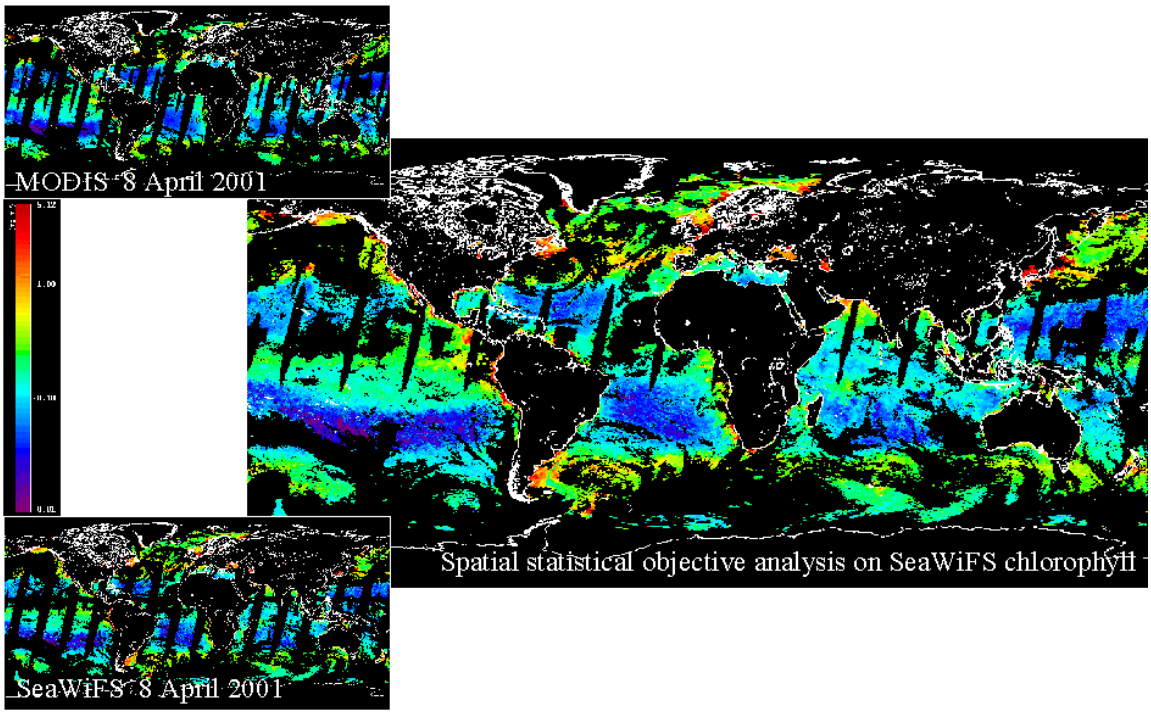


Figure 8.11: Original MODIS and SeaWiFS 36km binned chlorophyll concentration data sets for 8 April 2001 and the result of the 2-dimensional statistical objective analysis on the SeaWiFS chlorophyll bins coinciding with the MODIS coverage.

For the interpolation, problematic areas in ocean color daily imagery were those where gaps in global coverage were large in spatial and temporal terms, such as below persistent clouds, sun-glint, or SeaWiFS tilt change. It was observed that within these gaps the result of the analysis looked realistic but some interpolated coverage had values very close to the first-guess field. Eventually, each output point would be assigned a confidence level which would be dependent on accuracies of the original data sets and on the distances from the existing points used for the interpolation.

The statistical objective analysis was computationally involved. It considered ensemble spatial distributions of observations relative to one another which were contained within a radius of influence of an investigated grid point. This resulted in large covariance matrices whose size depended on the number of valid data points within the radius of influence from the interpolated point. At 9km, which is the ultimate resolution for the data merger, the quantity of L3 bins will be an order of magnitude higher than at the resolution of 36km. When ultimately operating on L2 data, the amounts of points analyzed inside the radius of influence could be massive. The analysis involved inverting square matrices of covariances for each investigated grid point. To ease the computational effort, an effective strategy was designed in which radii smaller than the influence radius were first searched to determine whether they contained sufficiently high proportions of valid data bins to perform the interpolation. If there were enough valid data inside the smaller radius, these data were then used to interpolate the grid point. Because the covariance matrices were symmetric and positive definite, an efficient Cholesky decomposition was used to solve the matrix equation (Press *et al.*, 1992).

8.4.3 Conclusions

Statistical objective analysis was introduced as a spatial and temporal interpolation approach to combine multi-sensor ocean color data sets onto daily global grids using corresponding sensor accuracies and an ensemble correlation structure of the global chlorophyll field. The interpolation was envisioned as a binning approach for multi-sensor data beginning with the L2 products. The binning would effectively combine sensor data using pixels surrounding the bin grid points in space and across

time and would consequently fill many gaps in daily global ocean color coverages. The ensemble correlation structure of the chlorophyll field was established individually for all global coverage grid points and made dependent on the local area natural variabilities. Preceding the interpolation, sensor cross-calibrations would be performed to bring the multi-sensor data to a consistent baseline and eliminate sensor temporal trends and data artifacts. Sensor data accuracy was used as a weighting factor in the interpolation and came from matchups with *in situ* measurements.

The initial results were obtained from the 2- and 3-dimensional statistical objective analysis of daily global SeaWiFS L3 binned chlorophyll data. The analysis interpolated selected missing SeaWiFS bin coverage for this day. The analysis demonstrated to be a useful tool for ocean color data merger. However, more research would be needed to make the statistical objective analysis more effective in terms of the choice of the first-guess background fields and associated space-lag correlation functions and influence radii. The statistical objective analysis was also computationally involved in operational processing of multi-sensor data. Therefore, means for improvement of its efficiency could be investigated. Finally, the capabilities of the analysis to provide error bars for all interpolated data points could be further studied.

8.5 LOCAL AREA APPLICATION OF DATA MERGER: ENHANCEMENT OF OCEANIC FEATURES IN LOWER RESOLUTION IMAGERY USING HIGHER RESOLUTION DATA

This study examined ocean color merger opportunities at local spatial scales to provide useful tools for scientists interested in smaller-size geophysical phenomena and in complex environments such as coastal zones. The feasibility of merging ocean color data from sensors of different spatial resolutions was studied for cases where there was overlapping ground coverage for individual scenes (Kwiatkowska-Ainsworth, 2001; Kwiatkowska and Fargion, 2002a). The prospect of enhancing oceanic features in lower resolution imagery through the use of higher resolution data was also investigated. The algorithm operated on L2 ocean color data products and was based on a signal processing approach — wavelet multiresolution analysis (Rioul and Vetterli, 1991). The wavelet transform enabled an image to be examined at different frequency and scale intervals (Mallat, 1989). This corresponded to image analyses at variable frequency and spatial resolutions.

The resolution of an image, corresponding to a measure of detail information in the scene, was defined and changed by a combination of high pass and low pass filtering operations. The scale of an image was altered by downsampling and upsampling operations. The wavelet transforms used in this analysis therefore functioned as power-of-2 operators for subsampling and resolution change. Fig. 8.12 illustrates the process of decomposition of a one-dimensional signal by a discrete wavelet transform (DWT). The figure also shows changes in scale and frequency contents of the filtered output.

The wavelet merger algorithm thus operated on scenes from sensors of different spatial resolutions (Núñez *et al.*, 1999; Blanc *et al.*, 1998). The high-frequency, low-scale spatial detail in the higher resolution scene was extracted using the high pass filters of the wavelet transform. The result of the low pass filtering of the higher resolution image was completely replaced by the lower resolution scene. This modified wavelet transform of the higher-resolution image was then reversed. For the lower resolution scene this process resulted in the increased spatial resolution and added high frequency variation. The enhanced spatial resolution was gained without altering the mean magnitudes of lower-resolution ocean color values. This made the wavelet method particularly useful when the quality of data was different between the sensors and the measurement accuracy of the lower resolution sensor had to be preserved. To perform merger of data from both scenes, the result of the low pass filtering of the higher resolution image, instead of being completely substituted by the lower resolution scene, would be replaced with its weighted average with the lower resolution scene. The reversal of the transform would then produce a merged image where the merger was performed on the level corresponding to the lower-resolution coverage from both sensors and the lower scale detail was added from the higher resolution scene.

8.5.1 Wavelet Transform Implementation and Results

The wavelet algorithm was tested using chlorophyll-*a* concentration imagery from SeaWiFS and MOS. The SIMBIOS Project cross-calibrated SeaWiFS and MOS missions, processed their data uniformly, and analyzed them for overlapping concurrent ground-coverage (Wang and Franz, 2000). SeaWiFS L2 HRPT and LAC scenes used in the analysis had a native resolution of 1.1km and MOS imagery had the resolution of 0.5km. A significant obstacle was to co-register scenes from both sensors so that the accuracy of the overlay was within 0.5km and to define resolutions and scene sizes to be in power of 2 for the multi-resolution analysis. A basic strategy was designed where SeaWiFS data were binned at 1km and MOS data were binned at 0.5km. Bins were then projected onto a rectangular longitude/latitude grid map to facilitate image processing. Because the dimensional sizes of the rectangular grids were limited to powers of 2, the mapped scenes had to be padded with zeros to fill the grid. The size of the SeaWiFS grid was half the size of the MOS grid. To preserve the spatial resolution of the bins, the projection spread the bins longitudinally according to the longest row for each scene. Bins from all other rows were then fitted into the grid given their longitude distance from the longest-row longitudes. This technique was only applicable to

local coverage scenes used in this analysis, which were of the LAC and HRPT size, and it would not be appropriate for global imagery. Any missing grid points caused by the mapping of spherical coordinates onto a rectangular grid were approximated. The approximation used a wavelet-based iterative algorithm that minimized high frequency anomalies associated with the missing data points. The preprocessing therefore resulted in reformatted mapped scenes from both sensors with the desired size and resolution for the wavelet analysis.

Because the resolution ratio between SeaWiFS and MOS scenes was equal to 2, only one level of wavelet decomposition was required for the processing. This single pass of the wavelet filtering was applied to the MOS image. The transform extracted pixel-to-pixel spatial detail from MOS data in its high-frequency components and higher-scale background in its low-frequency components. The transform also subsampled the MOS scene by 2 so that its high and low frequency components corresponded to 1km spatial resolution, the same as the SeaWiFS scene resolution. The SeaWiFS image was concurrently preprocessed to bring the magnitude of its chlorophyll values to the level corresponding to a single application of the low pass filter. Then in the two scenarios, the MOS low-pass filter result was replaced either by the entire preprocessed SeaWiFS scene or by its weighted ratio with the preprocessed SeaWiFS scene. The ratio depended on the established relative accuracies of the chlorophyll products from each instrument. An inverse wavelet transform was subsequently applied which produced an increased 0.5km resolution SeaWiFS image or a 0.5km SeaWiFS image merged with MOS data. The enhanced SeaWiFS scene inherited its low-scale spatial detail from the high-frequency contents of the MOS scene. To generate the final product, flags and masks from SeaWiFS and MOS chlorophyll data were also merged and applied to the subsequent image. Fig. 13 shows an example of the wavelet multiresolution merger of SeaWiFS and MOS scenes of Mallorca and Menorca in the Mediterranean Sea.

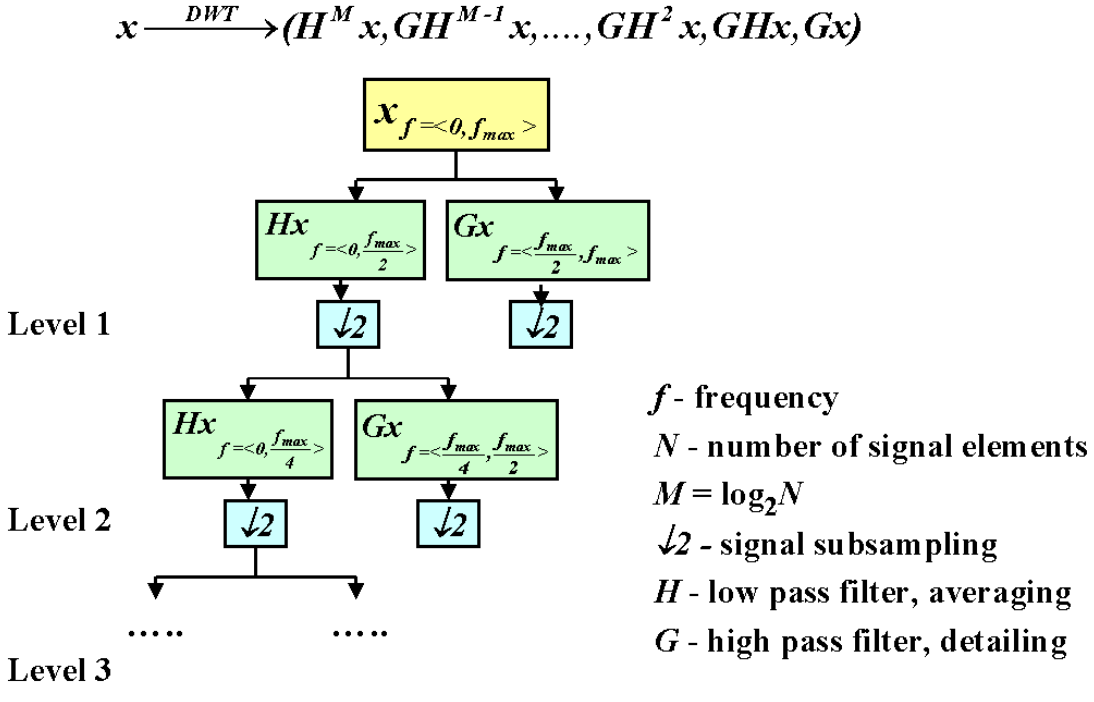


Figure 8.12: Consecutive levels of wavelet transform decomposition of a one-dimensional signal. At the first decomposition level, the signal is passed through the high pass and low pass filters, followed by subsampling by 2. The output of the low pass filter is then passed at level 2 through the same low pass and high pass filters for further decomposition. The process is repeated at the subsequent levels. Changes in output frequency and scale are indicated.

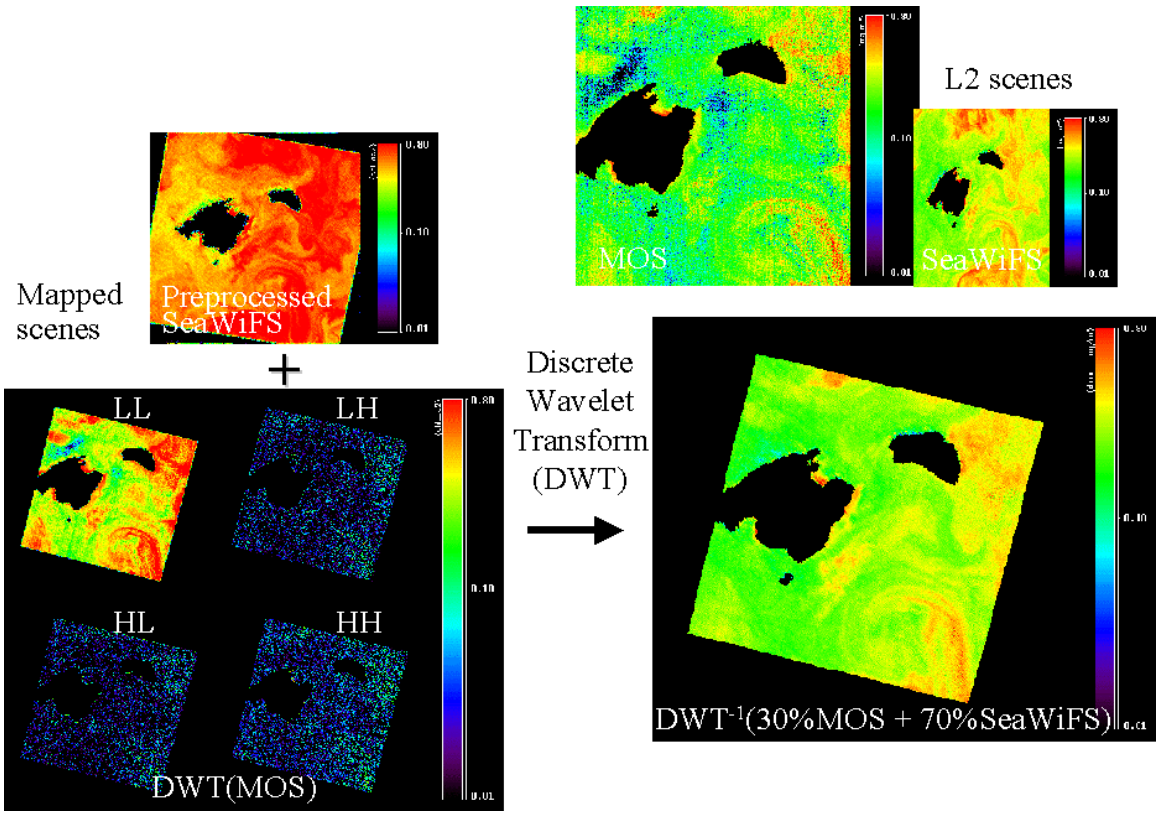


Figure 8.13: Original L2 MOS and SeaWiFS chlorophyll-*a* concentration scenes and the process of wavelet-based merger of both data sets mapped to a rectilinear grid. The SeaWiFS scene was preprocessed and the MOS scene underwent a single level of wavelet decomposition. The 2-dimensional DWT separately passed high (H) and low (L) pass filters through the scene at first across the rows with column subsampling, and then across the columns with row subsampling. The MOS row and column low-pass coefficients (LL) were replaced by their weighted ratio with the preprocessed SeaWiFS scene. The wavelet transform was reversed, thus producing the merged output image at 0.5-km resolution.

To validate the wavelet algorithm, the original MOS scenes were compared against the SeaWiFS scenes enhanced to 0.5km resolution using the wavelet method and using bilinear interpolation. The bilinear interpolation on its own did not provide the benefits of the higher-frequency feature extraction which enabled SeaWiFS imagery to acquire spatial detail inherent in MOS data. Quantitatively, the correlation of bilinearly interpolated SeaWiFS imagery with original MOS imagery was considerably smaller (~10%) than the correlation for the wavelet-enhanced SeaWiFS scenes. Qualitatively, the gain in spatial detail obtained by the wavelet approach was consequential and unique. Merger of original MOS scenes with bilinearly interpolated 0.5km resolution SeaWiFS data was also compared against the result of the wavelet multiresolution analysis. This merger did not, however, allow the preservation of the magnitudes of SeaWiFS lower-resolution ocean color values and complete high-frequency spatial variation from MOS data. Overall, the wavelet algorithm performed superior to other approaches.

The application of the wavelet approach brought also some difficulties. MOS data were inherently noisy. Although the wavelet-merged scenes appeared sharper, there was a degree of high-frequency noise introduced from MOS scenes. As it happened, wavelets also provided a means for denoising speckled imagery (Donoho, 1995). Therefore, denoising was implemented as an option in the algorithm. The implementation was based on soft-thresholding of wavelet coefficients which was equivalent to removing Gaussian noise from an image. Additionally, manipulation of wavelet coefficients caused undesirable ringing effects in images because of the presence of high frequency features. To limit the ringing, a selected number of transformed solutions based on different wavelet functions was averaged. Daubechies_20, Coiflet, Haar, and spline functions were examples of the wavelet functions used.

8.5.2 Conclusions

This study examined possible applications of ocean color data merger at local spatial scales. It investigated integration of data from sensors of different spatial resolutions. It also determined the ability to generate merged products of the resolution equal to that of the higher resolution sensor. Inherent in the technique was an option to preserve in the merged output, mean high-scale ocean color values from the lower resolution coverage. This corresponded to an ability to enhance the lower-resolution ocean color baseline with low-scale spatial detail from the higher resolution data. Simultaneously, the baseline was able to retain its calibration quality. This would provide the useful tools for the investigation of smaller-size geophysical phenomena and complex environments, such as coastal zones. Wavelet-based multiresolution analysis was used to extract high-frequency low-scale features from high resolution imagery and transfer them to lower resolution scenes.

It would be of interest to apply the wavelet algorithm to the merger of overlapping scenes between MODIS and SeaWiFS GAC so that SeaWiFS imagery could be enhanced by the spatial detail contained in MODIS data. A useful application would also be to combine MODIS or SeaWiFS ocean color products at 1km or 4km resolution with high frequency spatial information contained in MODIS high-resolution bands, such as 500m and 250m.

8.6 LOCAL AREA APPLICATION OF DATA MERGER: MERGER OF SATELLITE AND *IN SITU* MEASUREMENTS

An approach was developed to merge L2 ocean color data with *in situ* measurements. The major purpose was to provide a utility to demonstrate changes in remotely sensed chlorophyll or nLw range and distribution when collected *in situ* measurements were overlaid upon local area scenes. The algorithm was intended for use in local area applications to verify remotely sensed ocean color data and provide a change visualization tool. The merger of satellite and *in situ* data was dependent on the spatial and temporal correlation structure of the ocean color field, which was by itself, contingent upon local area spatial and temporal variabilities, as shown in Section 8.2.

Merger was based on the application of the wavelet transform which spatially extended *in situ* data point values onto corresponding areas in satellite scenes (Kwiatkowska and Fargion, 2002a; Mallat, 1989). These areas were defined by a radius of influence and depended on the geographical location of *in situ* measurements (Barnes, 1964). The radius of the area of influence was defined using local texture estimates, such as the spatial variability classes defined in Section 8.2. The more irregular the texture was around the *in situ* measurement point, the smaller the radius; the smoother the texture, the bigger the radius. The Hann window function was applied to scale the effects of the *in situ* data points away from the area centers (Press *et al.*, 1992). Ultimately, a space-lag correlation function for a given area spatial-variability class would be used. The degree of change introduced by *in situ* measurements onto ocean color satellite scenes also depended on the established relative accuracies assigned to *in situ* and satellite data.

The methodology behind the wavelet merger was the following: Because *in situ* measurements were screened for quality, they were assumed representative of generalized ocean color conditions within their area. *In situ* data points were typically intended not to affect the local area low-scale spatial variabilities and not to change shapes of ocean patterns within the scenes. Each *in situ* observation was therefore associated with a low-frequency background ocean-color value corresponding to its coverage point. The low frequency background was extracted by the low-pass wavelet filter. The original wavelet coefficients of each scene were then replaced with the coefficients updated with the *in situ* data point and the point's values scaled smoothly towards the edges of the area of influence. The magnitude of the correction also depended on the estimates of the relative accuracies of satellite and *in situ* measurements. The wavelet thus forced the resulting satellite pixels to be interpolations of *in situ* data only within the low-resolution representation of the scenes. The high frequency coefficients of the updated imagery were left unchanged to preserve the original high-resolution spatial variabilities within the areas of influence and to protect spatial structures in the scenes.

8.6.1 Satellite and In Situ Merger Implementation and Results

Merger of satellite and *in situ* chlorophyll-*a* concentration observations was analyzed using SeaWiFS and California Cooperative Oceanic Fisheries Investigation (CalCOFI) data for the years 1997 and 1998. From the experience with ocean color validation, it was known that there was a significant scarcity of contemporaneous satellite and *in situ* observations, mainly because of the presence of clouds, sun glint, coverage gaps between satellite orbits, and other satellite viewing and meteorological conditions. Merger was dependent on the time difference between the satellite overflight and *in situ* data collection. The maximum time span between SeaWiFS and *in situ* observations was set for 12 hours, although the ultimate time span would be dependent on the local area spatial variability and the corresponding time-lag correlation of the chlorophyll field. Within the 12-hour time difference, there were just 13 SeaWiFS L2 LAC and HRPT files with concurrent satellite and

CalCOFI measurements for which the merger could be performed. One file out of the 13 contained three points within the scene. The low number of matchups was principally caused by the presence of cloud cover.

To limit the cases where small clouds (a few pixels long) and other conditions caused ocean color pixels to be masked out from the imagery, a gap-filling algorithm was implemented. Its goal was to preserve spatial patterns of chlorophyll distributions in ocean color scenes without smoothing. The algorithm was based on an iterative reduction of the total of high frequencies associated with missing pixels in the analyzed scene. The high frequency content of a pixel was established by inverting a wavelet transform output of the scene where the inversion was limited to the result of the forward transform high-pass filter. The iterations were initialized by filling the gaps with values corresponding to the lower frequency representation of the scene. To eliminate local minima, a random perturbation was introduced to the best values for the gap pixels which were found by the recursive search. The gap-filling approach was implemented in combination with the satellite and *in situ* measurement merger to eliminate small clouds within the areas of influence of *in situ* data points. This produced an increase in matchups of about 10%.

A sequential processing algorithm was implemented for all *in situ* data points extracted from selected SeaBASS records (Werdell and Bailey, 2002) and a corresponding list of SeaWiFS L2 files. The algorithm processed image subscenes encompassing areas of influence of consecutive *in situ* points and fused the points into the images. Examples of the *in situ* and satellite data merger are displayed in Figure 8.14.

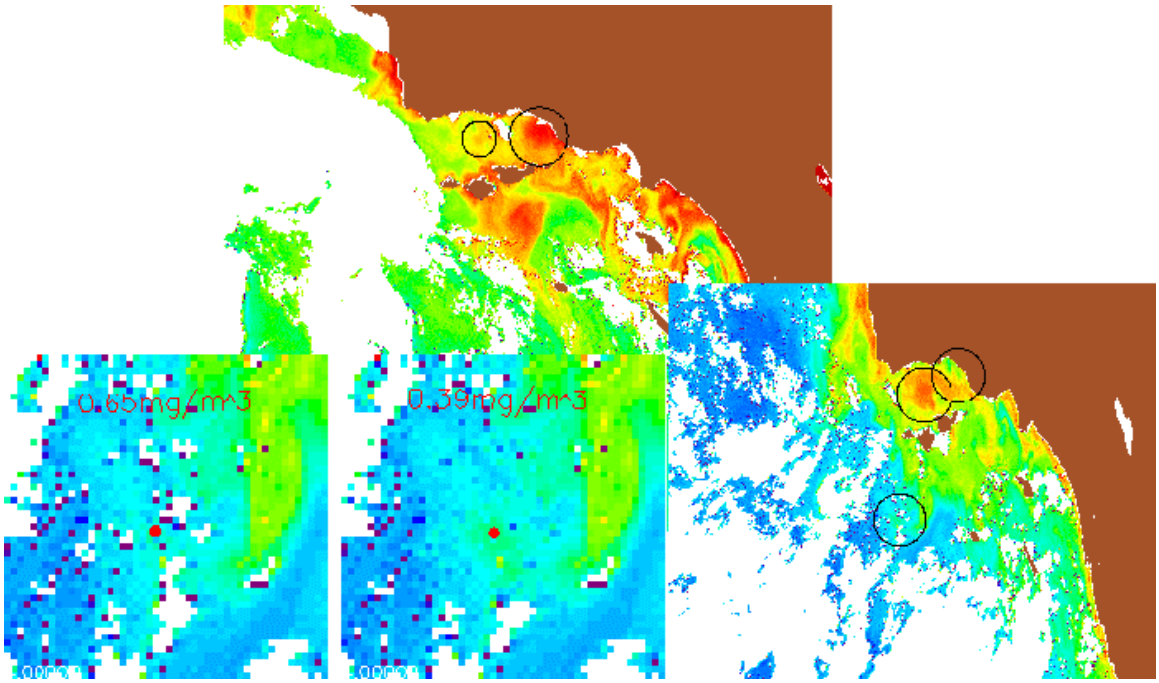


Figure 8.14: Merger of CalCOFI *in situ* chlorophyll measurements with SeaWiFS L2 data using the wavelet multiresolution approach. The option of filling small cloud gaps in ocean color satellite data was applied in the bottom left-hand side scene.

8.6.2 Conclusions

A tool was implemented to investigate differences and local-field distribution changes in ocean color imagery when overlaying *in situ* data onto satellite scenes. The implementation was based on wavelet multiresolution analysis. The wavelets enabled the spatial spread of *in situ* values onto the imagery without smoothing the ocean color fields. Spatial variability and chlorophyll structures were also preserved. The merger application was designed to be ultimately dependent on local area spatial and temporal variabilities and data correlations. During the study, it was determined that concurrent *in situ* and satellite observations were scarce, even when data from recurrent oceanic surveys were applied. Therefore, it was intended that *in situ* data were presently used for validation of ocean color imagery and not for broad application in the merger efforts to complement satellite data.

REFERENCES

Ainsworth, E. J. and I. S. F. Jones, 1999: Radiance Spectra Classification from the Ocean Color and Temperature Scanner on ADEOS. *IEEE Trans. Geoscience and Remote Sens.*, **37**, issue 3, 1645-1656.

Atkinson P. M. and A. R. L. Tatnall, 1997: Neural Networks in Remote Sensing. *Int. J. Remote Sens*., **18**, no. 4, 699-709.

Azimi-Sadjadi M. R. and S. A. Zekavat, 2000: Cloud Classification Using Support Vector Machines. *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, **2**, 669-671.

Bailey, S. W., J. P. Werdell, and C. R. McClain, 2001: Validation of Satellite-Derived Ocean Color: Theory and Practice. *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, **82**, no. 47, F676 (OS52A-0520).

Barnes, R. A., R. E. Eplee Jr., G. M. Schmidt, F. S. Patt, and C. R. McClain, 2001: Calibration of SeaWiFS. I. Direct Techniques. *Applied Optics*, **40**, no. 36, 6682-6700.

Barnes, S. L., 1964: A Technique for Minimizing Details in Numerical Weather Map Analysis. *J. Applied Meteorology*, **3**, 396-410.

Blanc, P., T. Blu, T. Ranchin, L. Wald, and R. Aloisi, 1998: Using Iterated Rational Filter Banks Within the ARSIS Concept for Producing 10m Landsat Multispectral Images. *Int. J. Remote Sens*., **19**, no. 12, 2331-2343.

Brown, M., H. G. Lewis, and S. R. Gunn, 2000: Linear Spectral Mixture Models and Support Vector Machines for Remote Sensing. *IEEE Trans. Geoscience and Remote Sens.*, **38**, no. 5, 2346-2360.

Campbell, J. W., 1995: The Lognormal Distribution as a Model for Bio-optical Variability in the Sea. *J. Geophys. Research*, **100**, no. C7, 13,237-13,254.

Chomko R. M. and H. R. Gordon, 1998: Atmospheric Correction of Ocean Color Imagery: Use of the Junge Power-Law Aerosol Size Distribution with Variable Refractive Index to Handle Aerosol Absorption. *Applied Optics*, **37**, no. 24, 5560-5572.

Cristianini N. and J. Shawe-Taylor, 2000: *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press.

Donoho, D. L., 1995: De-Noising by Soft-Thresholding. *IEEE Trans. on Information Theory*, **41**, no. 3, 613-627.

Eplee Jr., R. E., W. D. Robinson, S. W. Bailey, D. K. Clark, P. J. Werdell, M. Wang, R. A Barnes, and C. R. McClain, 2001: Calibration of SeaWiFS. II. Vicarious Techniques. *Applied Optics*, **40**, no. 36, 6701-6718.

Esaias, W. E., M. R. Abbot, I. Barton, O. B. Brown, J. W. Campbell, K. L. Carder, D. K. Clark, R. L. Evans, F. E. Hoge, H. R. Gordon, W. P. Balch, R. Letelier, and P. J. Minnett, 1998: An Overview of MODIS Capabilities for Ocean Science Observations. *IEEE Trans. Geosci. and Remote Sens*., **36**, no. 4, 1250-1265.

Franz B. A. and Y. Kim, 2001: A Comparative Study and Intercalibration Between OSMI and SeaWiFS. *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, **82**, no. 47, F661 (OS42D-09).

Fukuda S. and H. Hirosawa, 2001: Polarimetric SAR Image Classification Using Support Vector Machines. *IEICE Trans. Electron*., **E84-C**, no. 12.

Goldberg, D. E., 1989: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

Gordon H. R. and M. Wang, 1994: Retrieval of Water-Leaving Radiance and Aerosol Optical Thickness over the Oceans with SeaWiFS: a Preliminary Algorithm. *Applied Optics*, **33**, 443-452.

Gordon, H. R., T. Du, and T. Zhang, 1997: Remote Sensing of Ocean and Aerosol Properties: Resolving the Issue of Aerosol Absorption. *Applied Optics*, **36**, no. 33, 8670-8684.

Gregg, W. W., W. E. Esaias, G. C. Feldman, R. Frouin, S. B. Hooker, C. R. McClain, R. H. Woodward, 1998: Coverage Opportunities for Global Ocean Color in a Multimission Era. *IEEE Trans. Geosci. and Remote Sens*., **36**, no. 5, 1620-1627.

Gregg W. W. and R. H. Woodward, 1998: Improvements in coverage Frequency of Ocean Color: combining Data from SeaWiFS and MODIS. *IEEE Trans. Geosci. and Remote Sens*., **36**, no. 4, 1350-1353.

Gregg W. W. and M. E. Conkright, 2001: Global Seasonal Climatologies of Ocean Chlorophyll: Blending *In Situ* and Satellite Data for the CZCS Era. *J. Geoph. Research – Oceans*, **106(C2)**, 2499-2515.

Gross, L., S. Thiria, R. Frouin, and B. G. Mitchell, 2000: Artificial Neural Networks for Modeling the Transfer Function Between Marine Reflectance and Phytoplankton Concentration. *J. Geophysical Research*, **105**, no. C2, 3483-3495.

Gualtieri, J. A., S. R. Chettri, R. F. Cromp, and L. F. Johnson, 1999: Support Vector Machine Classifiers as Applied to AVIRIS Data. *Summaries of the Eight JPL Airborne Earth Science Workshop*, USA.

Hearst, M. A., 1998: Support Vector Machines. *IEEE Intelligent Systems*, Jul./Aug., 18-28.

Julian P. R. and H. J. Thiebaux, 1975: On some Properties of Correlation Functions Used in Optimum Interpolation Schemes. *Monthly Weather Review*, **103**, 605-616.

Kilpatrick, K., E. Kearns, E. J. Kwiatkowska-Ainsworth, and R. L. Evans, 2002: Time Series of Calibrated Ocean Products from NASA's Moderate Resolution Scanning Spectrometer (MODIS). *Proceedings of the Ocean Sciences Meeting*, Feb., Honolulu, Hawaii, USA.

Kwiatkowska-Ainsworth, E. J., 2001: Merger of Ocean Color Information of Different Spatial Resolution: SeaWiFS and MOS. *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, **82**, no. 47, F675 (OS52A-0514).

Kwiatkowska E. J. and G. S. Fargion, 2002a: Merger of Ocean Color Information from Multiple Satellite Missions under the NASA SIMBIOS Project Office. *Proceedings of the Fifth International Conference on Information Fusion*, Annapolis, MD, USA, **1**, 291-298.

Kwiatkowska E. J. and G. S. Fargion, 2002b: Merger of Ocean Color Data from Multiple Satellite Missions within the SIMBIOS Project. *Proceedings of SPIE Symposium – Remote Sensing of the Atmosphere, Ocean, Environment, and Space*, Oct., Hangzhou, China, Ocean Remote Sensing and Applications, **4892**, 168-182.

Kwiatkowska, E. J., 2003: Application of Machine Learning Techniques Towards the Creation of a Consistent and Calibrated Global Chlorophyll Concentration Baseline Dataset Using Remotely Sensed Ocean Color Data. *IEEE Trans. Geosci. and Remote Sens*., (accepted).

Linde, Y., A. Buzo, and R. M. Gray, 1980: An Algorithm for Vector Quantizer Design. in IEEE Trans. Commun., COM-28, **1**, 84-95.

Mallat, S. G., 1989: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **11**, no. 7, 674-693.

Maritorena, S., D. A. Siegel, A.R. Peterson and M. Lorenzi-Kayser, 2000: Tuning of a Pseudo-analytical Ocean Color Algorithm for Studies at Global Scales. *Eos Trans. American Geophysical Union*, Jan., San Antonio, USA, OS12M-05.

Maritorena, S., D. A. Siegel, and A. Peterson, 2002: Optimization of a Semi-Analytical Ocean Color Model for Global Scale Applications. *Applied Optics*, **41**, no. 15, 2705-2714.

Masters, T., 1994: *Signal and Image Processing with Neural Networks. A C++ Sourcebook*. John Wiley & Sons, inc.

Núñez, J., X. Otazu, O. Fors, A. Prades, V. Palà, R. Arbiol, 1999: Multiresolution-Based Image Fusion with Additive Wavelet Decomposition. *IEEE Trans. Geosci. and Remote Sens.*, **37**, no.3, 1204-1211.

O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. R. McClain, 1998: Ocean Color Chlorophyll Algorithms for SeaWiFS. *J. Geoph. Research*, **103**, no. C11, 24,937-24,953.

Pao, Y.-H., 1989: *Adaptive Pattern Recognition and Neural Networks*. Reading, MA, Addison-Wesley.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press.

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

Reynolds R. W. and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929-948.

Rioul O. and M. Vetterli, 1991: Wavelets and Signal Processing. *IEEE SP Magazine*, 14-38.

Schölkopf, B., K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, 1996: Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers. *Massachusetts Institute of Technology Artificial Intelligence Laboratory Memorandum*, **1599**, C.B.C.L. paper no. 142.

Schölkopf, B., 2000: Statistical Learning and Kernel Methods. *Microsoft Research Technical Report*, no. MSR-TR-2000-23, 1-27.

Starck J.-L. and F. Murtagh, 1994: Image Restoration with Noise Supression Using The Wavelet Transform. *Astron. Astrophys.*, **288**, 342-348.

Tanaka, A., H. Kobayashi, M. Kishino, T. Oishi, R. Doerffer, H. Schiller, and T. Kubota, 2000: Neural Network Algorithm for Simultaneious Retrieval of Aerosol and Water Constituents Based on Atmospshere-Ocean Coupled Radiative Transfer Model, *Proceedings of the Ocean Optics Conference*, no. XV.

Thiebaux, H. J., 1973: Maximally Stable Estimation of Meteorological Parameters at Grid Points. *J. Atmospheric Science*, **30**, 1710-1714.

Thiebaux H. J. and M.A. Pedder, 1987: *Spatial Objective Analysis with Applications in atmospheric Sciences*. Academic Press.

Wang M. and B. A. Franz, 2000: Comparing the Ocean Color Measurements between MOS and SeaWiFS: A Vicarious Intercalibration Approach for MOS. *IEEE Trans. Geosci. and Remote Sens.*, **38**, no. 1, 184-197.

Wang, M., A. Isaacman, B. A. Franz, and C. R. McClain, 2002: Ocean-Color Optical Property Data Derivation from the Japanese Ocean Color and Temperature Scanner and the French Polarization and Directionality of the Earth's Reflectances: a Comparison Study. *Applied Optics*, **41**, no. 6, 974-990.

Werdell J. P. and S. W. Bailey, 2002: The SeaWiFS Bio-optical Archive and Storage System (SeaBASS): Current Architecture and Implementation. *NASA Technical Memorandum 2002-211617*, G. S. Fargion and C. R. McClain, Eds., 50pp.

Chapter 9

Current SeaDAS Support for MODIS Products

Mark Ruebens and Wang Xiao-Long
Science Applications International Corporation, Beltsville, MD

Providing a variety of services to the user community has been a primary objective of the SeaWiFS Project since its inception at the NASA Goddard Space Flight Center. These free services include rapid and easy access to all SeaWiFS data products, comprehensive documentation of Project activities (i.e., the SeaWiFS Technical Memorandum Series), maintenance of an extensive website (http://seawifs.gsfc.nasa.gov/seawifs.html), and user-friendly data processing and display software i.e., the SeaWiFS Data Analysis System (SeaDAS; Baith et al., 2001). This philosophy and approach was a result of the research community's experience with the proof of concept Nimbus7/Coastal Zone Color Scanner (CZCS) mission. The CZCS data set was not exploited by the research community until several years after the launch in 1978 because the data was not available from an on-line data archive system until around 1990 and processing and display software was not generally available. Processing software for CZCS data was developed by individual researchers. The first PC version of one such package, SEAPAK (McClain et al., 1989), was distributed to the research community in 1989 and UNIX versions were subsequently released. SEAPAK provided the foundations for the SeaDAS development effort.

SeaDAS is a comprehensive image analysis software package for the processing, display, analysis, and quality control of ocean color data from multiple satellite sensors (SeaWiFS, CZCS, OCTS, MOS and OSMI) and is designed to serve a wide range of users, including scientists, SeaWiFS ground stations, and operational or commercial users. SeaDAS is designed to accurately replicate the operational data products, e.g., geophysical fields and data formats, generated by the SeaWiFS Project by using the default input values, but to also allow processing flexibility in the algorithms applied, the map projections used, and other aspects of processing and analyses that allow users to customize their data products.

Flexibility is enhanced by providing executable programs for those who only need the basic capabilities as well as source code for those who wish to modify the code to insert alternative algorithms. The SeaDAS development group is co-located with the SeaWiFS Project to help ensure close coordination with the SeaWiFS Project's development activities. The SeaDAS software is freely available for download from the SeaDAS website(http://seadas.gsfc.nasa.gov). Since SeaDAS development began in 1993, versions of SeaDAS have been released periodically even before the launch of SeaWiFS in order to prepare the community for SeaWiFS data. Version 4.4, the most current version, was released in March 2003. During its development SeaDAS has been expanded and generalized to provide processing for four additional satellite sensors: the Coastal Zone Color Scanner (CZCS), the Ocean Color Temperature Sensor (OCTS), the Ocean Scanning Multispectral Imager (OSMI), and the Modular Optoelectronic Scanner(MOS), as well as display and analysis support for the Moderate Resolution Imaging Spectroradiometer(MODIS) ocean data products and Advanced Very High Resolution Radar sea surface temperature (AVHRR SST) data. The support of international ocean color data sets is possible because of the Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Project (McClain and Fargion, 1999), also co-located with the SeaWiFS Project, which provides the processing code.

The Interactive Data Language (IDL) software product from Research Systems, Inc. is an integral part of SeaDAS. It is a high level interpretive programming language that is portable and can be used to develop GUI's, scientific graphics or any standard analysis application quickly and with a small amount of maintainable code and development time, relative to that required by low level programming languages such as C. Users do not need to know the IDL programming language in order to run SeaDAS. However, users who do know IDL can use their own IDL programs within the SeaDAS batch scripting environment.

The SeaDAS software package containsboth processing programs as well as a full suite of interactive display and basic analysis tools. SeaDAS is not designed to provide an extensive data analysis capability because these applications can be easily developed using IDL. Instead, the SeaDAS effort focuses primarily on satellite data processing and display. The majority of the underlying processing programs are C and/or FORTRAN programs developed by the SeaWiFS and SIMBIOS Projects and are the same programs as used in the operational processing of the SeaWiFS data. As a convenience to the user, SeaDAS provides GUI's from which to run these processing programs interactively as well as command mode capability and detailed documentation. The SeaDAS tool kit includes many navigation, display, analysis, and output functions. Navigation functions include data registration, map projections, overlaying of coastlines, plotting of *in situ* data, and latitude/longitude point

location. General display functions include data scaling, color bar definition, annotation, zooming, roaming, and color palette manipulation. General analysis functions include bathymetry generation, simple arithmetic functions, contour plots, profile plots, scatter plots, and histograms. Output functions allow outputting either data or latitude/longitude values (ASCII, HDF, and binary flat file formats) or displayed images (PNG and Postscript formats).

Rather than supporting old versions of operating systems, SeaDAS historically has maintained official support for the most current operating systems for each supported platform, as well as the most current versions of IDL. Currently, SeaDAS provides support for SUN Solaris 2.6, 2.7, and 2.8, SGI Irix 6.5, PC RedHat Linux 7.1, 7.2, and 7.3 platforms, as well as IDL 5.4 and 5.5. Efforts are currently underway to port the SeaDAS software to RedHat Linux 9.0, Mac OS X, and IDL 5.6. If compiling from scratch, then the vendor C and FORTRAN compilers are also required for SGI and SUN, and the commercial Absoft FORTRAN compiler(www.absoft.com) is required for the PC environment. Recommendations on minimum hardware requirements are SGI O2, Sun UltraSPARC, or a PC with 350 Mhz Pentium II processor. All platforms require a minimum of 192 megabytes of memory(384 megabytes if processing HRPT data) and suggested 9 gigabytes of disk space. The SeaDAS installation only requires about 330 megabytes without demo files, and 1.2 gigabytes with the optional demo files. However, a 9 gigabyte drive is suggested in order to accomodate the user's data processing needs. A 19-inch monitor with a resolution of at least 1024x1280 is suggested and X terminals need at least 20 megabytes of local memory.

The SeaDAS development team has always worked closely with the user community to provide timely detailed user support. Suggestions from the oceanographic user community have always driven the SeaDAS development efforts. The development of PC support as well as the embedded IDL license option were both development efforts driven by user demand. Historically, SeaDAS has released a major version of the software approximately once every six to eight months to coincide with new operating system releases, feature enhancements, and reprocessings of the SeaWiFS data. As an additional convenience to the user, SeaDAS provides optional interim updates between releases. These updates allow users to apply enhancements and bug fixes without waiting for a major release. These updates are announced via e-mail to the user community and are posted on the SeaDAS website. The SeaDAS user base is primarily comprised of oceanographic institutes and universities in more than 45 different countries. The total user base has historically been split equally between domestic US users and foreign users, however with the release of SeaDAS 4.3, foreign users now exceed domestic users by about 14%. This appears to be attributed to the release of the embedded IDL run-time license option.

Current support for MODIS ocean data products include the display of Terra and Aqua Level 1A, Level 1B, Level 2, Level 3 Binned, Level 3 Mapped, and Level 4 Productivity data products. All of the standard data analysis tools mentioned above can be applied to the MODIS ocean data products. In addition, specific tools have been developed to allow the display of MODIS Level 2 flags, common flags, and quality levels. The following table 9.1 shows the MODIS Terra/Aqua data products that SeaDAS can display by level and product name.

REFERENCES

Baith, K., R. Lindsay, G. Fu, and C. R. McClain, SeaDAS: A data analysis system for ocean color satellite sensors, *Eos, Trans. Am. Geophys. U.*, 82, 202, 2001.

McClain, C., G. Fu, M. Darzi and J. Firestone, PC-SEAPAK User's Guide, Version 1, NASA Goddard Space Flight Center, Greenbelt, Maryland, 159 pp., 1989.

McClain, C. R., and G. Fargion, SIMBIOS Project 1999 Annual Report, NASA Tech. Memo. 1999-209486, NASA Goddard Space Flight Center, Greenbelt, Maryland, 128 pp., 1999.

Table 9.1. Current support for MODIS ocean data products include the display of Terra and Aqua Level 1A, Level 1B, Level 2, Level 3 Binned, Level 3 Mapped, and Level 4 Productivity data products.

MOD1	MOD021KM	MODOCL2	MODOCL2A	L3 Binned/Mapped D, W, M SST	L4
ev_412	EV_1KM_RefSB_08	nLw_412	CZCS_pigment	nLw_412	Productivity
ev_443	EV_1KM_RefSB_09	nLw_443	chlor_MODIS	nLw_443	P1_Quality
ev_488	EV_1KM_RefSB_10	nLw_488	pigment_c1_total	nLw_488	P2_Quality
ev_531	EV_1KM_RefSB_11	nLw_531	chlor_fluor_ht	nLw_531	P1
ev_551	EV_1KM_RefSB_12	nLw_551	chlor_fluor_base	nLw_551	P2
ev_667	EV_1KM_RefSB_13lo	nLw_667	chlor_fluor_effic	nLw_667	MLD
ev_678	EV_1KM_RefSB_13hi	nLw_678	susp_solids_conc	nLw_678	PAR
ev_748	EV_1KM_RefSB_14lo	Tau_865	cocco_pigmnt_conc	Tau_865	P1_Number_o
ev_870	EV_1KM_RefSB_14hi	Eps_78	cocco_conc_detach	Eps_78	f_Obs
	EV_1KM_RefSB_15	aer_model1	calcite_conc	aer_model1	P2_Number_o
	EV_1KM_RefSB_16	aer_model2	K_490	aer_model2	f_Obs
	EV_1KM_RefSB_17	eps_clr_water	phycoeryth_conc	eps_clr_water	
	EV_1KM_RefSB_18	common_flags	phycou_conc	CZCS_pigment	
	EV_1KM_RefSB_19	L2_flags	common_flags	chlor_MODIS	
	EV_1KM_RefSB_26	quality	L2_flags	pigment_c1_total	
	EV_1KM_Emissive_20		quality	chlor_fluor_ht	
	EV_1KM_Emissive_21			chlor_fluor_base	
	EV_1KM_Emissive_22		MODOCL2B	chlor_fluor_effic	
	EV_1KM_Emissive_23			susp_solids_conc	
	EV_1KM_Emissive_24		chlor_a_2	cocco_pigmnt_conc	
	EV_1KM_Emissive_25		chlor_a_3	cocco_conc_detach	
	EV_1KM_Emissive_27		ipar	calcite_conc	
	EV_1KM_Emissive_28		arp	K_490	
	EV_1KM_Emissive_29		absorp_coef_gelb	phycoeryth_conc	
	EV_1KM_Emissive_30		chlor_absorb	phycou_conc	
	EV_1KM_Emissive_31		tot_absorb_412	chlor_a_2	
	EV_1KM_Emissive_32		tot_absorb_443	chlor_a_3	
	EV_1KM_Emissive_33		tot_absorb_488	ipar	
	EV_1KM_Emissive_34		tot_absorb_531	arp	
	EV_1KM_Emissive_35		tot_absorb_551	absorp_coef_gelb	
	EV_1KM_Emissive_36		common_flags	chlor_absorb	
			L2_flags	tot_absorb_412	
			quality	tot_absorb_443	
				tot_absorb_488	
				tot_absorb_531	
				tot_absorb_551	

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2003	Technical Memorandum

4. TITLE AND SUBTITLE
MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003

5. FUNDING NUMBERS
970

6. AUTHOR(S)
G. Fargion, C.R. McClain

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS (ES)
Goddard Space Flight Center
Greenbelt, Maryland 20771

8. PEFORMING ORGANIZATION REPORT NUMBER
2003-02854-0

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS (ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
TM—2003–212249

11. SUPPLEMENTARY NOTES
Giulietta S. Fargion, SIMBIOS Project, Science Applications International Corp., Greenbelt, MD
Charles R. McClain, Goddard Space Flight Center, Greenbelt, MD

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unclassified–Unlimited
Subject Category: 48
Report available from the NASA Center for AeroSpace Information,
7121 Standard Drive, Hanover, MD 21076-1320. (301) 621-0390.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this technical report is to provide current documentation of the Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) Project activities, satellite data processing, and data product validation. This documentation is necessary to ensure that critical information is related to the scientific community and NASA management. This critical information includes the technical difficulties and challenges of validating and combining ocean color data from an array of independent satellite systems to form consistent and accurate global bio-optical time series products. This technical report focuses on the SIMBIOS Project's efforts in support of the Moderate-Resolution Imaging Spectroradiometer (MODIS) on the Earth Observing System (EOS) Terra platform (similar evaluations of MODIS/Aqua are underway). This technical report is not meant as a substitute for scientific literature. Instead, it will provide a ready and responsive vehicle for the multitude of technical reports issued by an operational project.

14. SUBJECT TERMS
SIMBIOS, Moderate-Resolution Imaging Spectroradiometer (MODIS,Earth Observing System (EOS)

15. NUMBER OF PAGES
80

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL