# Ocean-Colour Data Merging

Reports of the
International Ocean-Colour
Coordinating Group

REPORT NUMBER 6

IOCCG

An Affiliated Program of SCOR
An Associate Member of CEOS

# Reports of the International Ocean-Colour Coordinating Group

An Affiliated Program of the Scientific Committee on Oceanic Research (SCOR)
An Associate Member of the Committee on Earth Observation Satellites (CEOS)

## IOCCG Report Number 6, 2007

# Ocean-Colour Data Merging

Editor:
Watson W. Gregg (NASA/Goddard Space Flight Center, Greenbelt, MD, USA)

Report of an IOCCG working group on ocean-colour data merging, chaired by Watson W. Gregg and Paula Bontempi and based on contributions from (in alphabetical order):


| | |
|---|---|
| James Aiken | Plymouth Marine Laboratory, UK |
| Watson Gregg | NASA Goddard Space Flight Center, USA |
| Ewa Kwiatkowska | NASA Ocean Biology Processing Group, USA |
| Stéphane Maritorena | Institute for Computational Earth System Science (ICESS), University of California, Santa Barbara, USA |
| Frédéric Mélin | Joint Research Centre, European Commission |
| Hiroshi Murakami | JAXA Earth Observation Research Center, Japan |
| Simon Pinnock | European Space Agency, ESRIN, Italy |
| Claire Pottier | Laboratoire d'Etudes en Géophysique et Océanographie Spatiales, Toulouse, France - Collect Localization Satellites, Ramonville Saint-Agne, France |


Series Editor: Venetia Stuart

Correct citation for this publication:

*IOCCG (2007). Ocean-Colour Data Merging. Gregg, W. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 6, IOCCG, Dartmouth, Canada.*

The International Ocean-Colour Coordinating Group (IOCCG) is an international group of experts in the field of satellite ocean colour, acting as a liaison and communication channel between users, managers and agencies in the ocean-colour arena.

The IOCCG is sponsored by NASA (National Aeronautics and Space Administration), NOAA (National Oceanic and Atmospheric Administration), ESA (European Space Agency), JAXA (Japan Aerospace Exploration Agency), CNES (Centre National d'Etudes Spatiales), CSA (Canadian Space Agency), JRC (Joint Research Centre, EC), GKSS Research Centre (Geesthacht, Germany), BIO (Bedford Institute of Oceanography, Canada) and SCOR (Scientific Committee on Oceanic Research).

http://www.ioccg.org

Published by the International Ocean-Colour Coordinating Group,
P.O. Box 1006, Dartmouth, Nova Scotia, B2Y 4A2, Canada.

ISSN: 1098-6030 ISBN: 978-1-896246-57-4



Printed by Transcontinental, Dartmouth, Canada.

# Contents

# Preface

Visible Spectral Radiometry (VSR, often referred to as ocean colour) is a highly quantitative methodology that yields accurate and precise fields of an important marine geophysical variable (chlorophyll concentration) of high significance for understanding the planetary carbon cycle. It is a quantity in ocean biogeochemistry and has many other applications including management of marine resources. It provides our only window into the marine ecosystem on synoptic scales. Of particular importance is the application to climate research: chlorophyll resides in phytoplankton, which use carbon dioxide in photosynthesis leading to a potential reduction in the atmospheric concentration of this greenhouse gas. In the climate context, it is vital that a seamless time series be constructed using data from all available missions. The merged data set protects the temporal continuity of the data stream and optimises the spatial coverage.

Despite the undoubted importance of VSR to earth observation, the data stream was allowed to lapse at the close of the pioneer mission (Coastal Zone Color Scanner, CZCS) in 1986. The hiatus in the record was to continue for another ten years. During this period, when many important publications from the CZCS data revolutionised ocean biogeochemistry, it became clear that VSR was a key variable for earth observation of great significance to climate research, and agencies in several countries began to plan second-generation VSR missions of varying scope from regional to global. Ironically, even during a decade with no VSR data whatsoever, space agency managers expressed concern about the perceived future redundancy of ocean-colour missions.

Against this background, the IOCCG was formed in 1996. Optimising the return through synergy of the plurality of missions was an early item of business for the committee. A related task was to dispel the perception of excess represented by the suite of planned missions in VSR. The second IOCCG report, "Status and Plans for Satellite Ocean-Colour Missions: Considerations for Complementary Missions" addressed the issues squarely. The panel found that by combining data from the missions, significant improvements in spatial coverage and temporal resolution could be achieved. Based on knowledge of scientific requirements and estimates of observational capabilities from six of the nearest future missions, using modelling studies, the panel concluded that three concurrent ocean-colour sensors were needed to offset loss of coverage due to clouds, inter-orbit gaps, and sun glint encountered in the polar orbits planned by all the prospective missions. The finding that three satellites could provide 60% ocean coverage in four days was considered to be a requirement for future missions. Although the panel recognised the complexities arising from radiometric, spectral and other factors of combining data from the different missions, it was assumed that such data merging efforts could and would be undertaken as a matter of priority. However, the ways and means to achieve this goal were not yet available.

At the time of writing this report, there are eight functioning VSR missions, from six different international sponsors, of which five are global in scope. But data merging from this suite of missions is not yet routine. We can count only three publications on the subject, and all are very recent. It appeared that progress was blocked by lack of a consensus on the issues involved. In this context, the IOCCG formed a working group to examine the issues surrounding merging of VSR data. A panel of leading experts in the field was convened in a workshop in May 2005 to discuss the problem. This report details the findings of the working group, focusing on issues associated with data merging, why it is necessary, and what is needed to construct high quality data sets of merged ocean colour.

Trevor Platt Chairman, IOCCG 1996-2005

# Chapter 1

# Introduction

Satellite ocean-colour observations are now widely recognized as an important component of international remote sensing programs. Chlorophyll, the primary product of ocean-colour sensors, is a measure of marine phytoplankton biomass. Phytoplankton are responsible for approximately half the global photosynthetic uptake of carbon (Field *et al*., 1998). In response to the potential importance of phytoplankton in the global carbon cycle and the lack of comprehensive data, the international community has established high priority satellite missions designed to acquire and produce high quality global ocean-colour data (Figure 1.1).

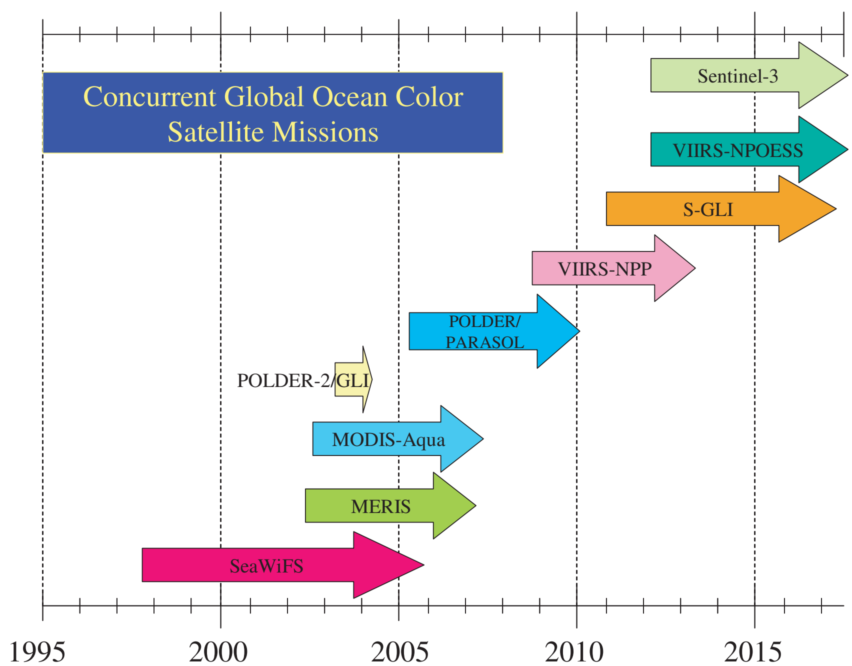


**Figure 1.1** Global ocean-colour missions that overlap at least one other mission. MODIS-Terra is not included because ocean-colour data processing has stopped for this sensor. All currently flying missions are assumed to have a 5-year lifetime.

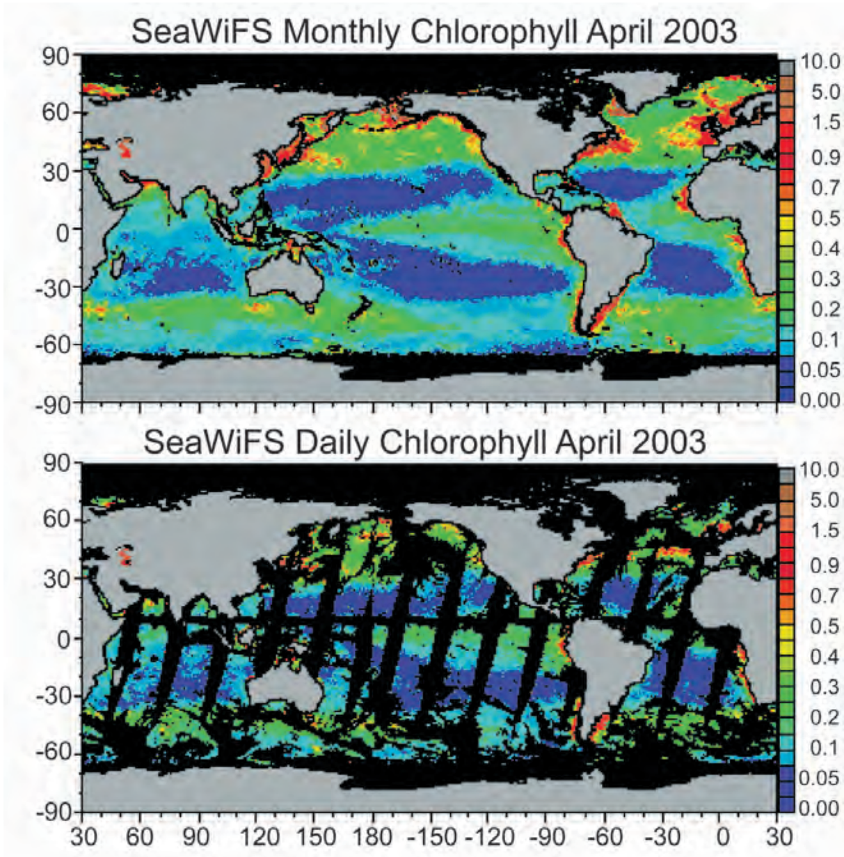


**Figure 1.2** Top: Monthly mean image of SeaWiFS chlorophyll for April 2003, showing nearly complete global coverage. Bottom: Daily image showing data gaps arising from inter-orbit gaps, sensor tilt changes, and clouds.

This proliferation of global missions, all of them in polar-orbiting configuration, presents an opportunity for ocean-colour science and applications. This is because a single polar-orbiting ocean-colour satellite does a poor job of sampling the ocean on short time scales. While we are familiar with the nearly-complete coverage of chlorophyll distributions in a typical monthly image, a less frequently shown daily image illustrates the issue (Figure 1.2), and these are at 1-degree spatial resolution. At 1-km resolution, the native resolution of MODIS-Aqua, less than 5% of the global ocean surface is observed (see Chapter 2). Clouds, inter-orbit gaps, sun glint, and thick aerosols prevent good sampling on a daily basis. Since phytoplankton populations can increase their biomass by more than double in a single day under favourable circumstances (Eppley, 1972; Doney *et al.*, 1995), this lack of sampling can have importance to our understanding of their dynamics, and their relationship with natural variability.

Previous efforts have shown that combining, or merging, data from coincident multiple satellites can greatly improve the daily coverage of the global ocean (*e.g.*, Gregg *et al.*, 1998, Gregg and Woodward, 1998), and has led to the acknowledgement that multiple missions are needed to complement and enhance ocean-colour science (IOCCG, 1999).

The case for merging ocean-colour data from multiple missions is established. Unfortunately data merging is an enormous challenge, both technically and politically. Ocean-colour sensors are very complex, with a very small signal relative to noise sources, and require a massive amount of effort to get high-quality observations. Couple these difficulties with requirements for international cooperation and collaboration, and the magnitude of the problem begins to emerge. Similar efforts with other remote-sensing observations, such as sea surface temperature (SST; *e.g.*, Reynolds and Smith, 1994), altimetry (*e.g.*, Le Traon and Ogor, 1998) or clouds (Rossow and Schiffer, 1991) have encountered similar obstacles, but have overcome them with reasonable success. The problem is relatively new for ocean colour.

This report describes the opportunities and potential benefits for ocean-colour data merging, the specific complexities involved, methods already undertaken with assessment of strengths and weaknesses, knowledge requirements, success criteria, and finally a discussion of what needs to be done in the future. We hope to provide a basis from which ocean-colour data merging activities can proceed, leading eventually to high quality archives of merged products.

In this report data merging refers to the process of combining coincident data from more than one satellite sensor. The main goals are improved temporal resolution and coverage and, to a lesser extent, improved accuracy. These goals do not necessarily support construction of ocean-colour time series, which is the essential component of Climate Data Records (CDR). The primary requirement for CDR’s is consistency (precision) and may require inclusion of observations that are not coincident. Since the objectives are different, so may be the requirements. Antoine *et al.* (2005) and Gregg *et al.* (2002) have argued that time series construction requires that all sensors involved have similar atmospheric and bio-optical algorithms. This is to avoid confusing trends, the main objective of a time series, with methodological differences. Data merging does not require similar methods, but rather seeks to improve an overall data set where coincident observations occur.

# Chapter 2

## Benefits of Merging

**Ewa Kwiatkowska, Simon Pinnock and Stéphane Maritorena**

The goal is to merge multi-instrument and multi-year observations of the oceans into consistent daily, global, high-resolution data records based on accurate and uniform calibration and validation over the lifetime of the missions. From the procedural point of view, the merger is a natural progression from multi-sensor data inter-comparisons, cross-validations and calibrations. The inter-comparisons support better scientific and technical understanding of ocean remote sensing processes and optical instrument operation. They are aimed at identifying and eliminating sensor or algorithm-originated biases and trends in data, which are impossible to detect solely by comparison with sparse *in situ* measurements. The inter-comparisons consequently bring multi-instrument time series to a common consistent ocean-colour baseline. Immediate benefits of the subsequent merger are:

- increase in daily ocean-colour global coverage which facilitates enhanced spatial and temporal resolution of ocean processes, and
- improvement in the statistical confidence in extracted bio-optical parameters due to expanded sampling rate for each location.

The added benefit of data merger is that scientists and other data users will have a single access point to multi-sensor ocean-colour data holdings, which are currently spread worldwide among different space agencies. The users will also take advantage of the consistent quality of merged datasets. This is because the data merger task will further stimulate and expand characterization efforts for individual instruments, their calibration, sensor inter-comparisons, cross-calibrations, and validations. With multi-sensor ocean-colour datasets, the users will be able to advance science and make educated decisions based on the most comprehensive information available. Subsequent scientific findings and operational applications derived from data from multiple sources will carry with them more gravity and credibility compared to those employing only a single mission's worth of data.

The advantage of the data merger will then be a shift from mission-centered to measurement-centered ocean-colour data utilization. Datasets, processing, analyses, and the resulting science, all of which are currently largely specific to individual missions and sensors, will be seamlessly coalesced and extended to form a unified ocean-colour discipline. The required data merger effort, which includes knowledge of each sensor operation, characterization, calibration as well as involved sensor inter-comparisons, calibrations, validations, and the merger itself, significantly exceeds the level of involvement desired by most individual data users. When the merger becomes operational, the measurement-centered data utilization will enable the users to exploit coalesced ocean-colour observations without requiring the detailed knowledge of individual instruments that compose the data. The users will be confident that they are using the most comprehensive and consistent data available.

A generic baseline for ocean-colour merged products has the following characteristics:

- Daily global coverage
- High accuracy - consistent and seamless in space and time
- High spatial resolution.

No single ocean-colour sensor is able to provide all three of these characteristics concurrently. Daily coverage by a single sensor is constrained by practical design

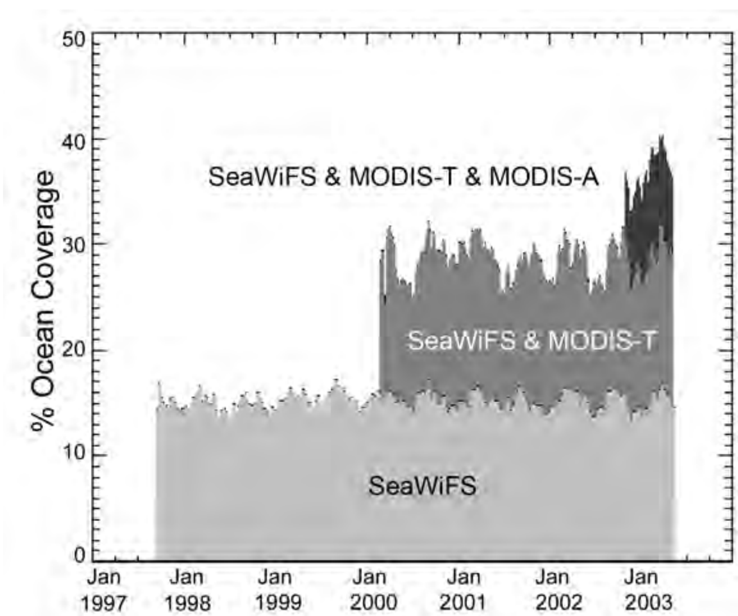


**Figure 2.1** Multi-mission time-series showing the total coverage attained each day at 9-km resolution through the operational merging of SeaWiFS-GAC (4-km sub-sampled data stored onboard), MODIS-Terra (MODIS-T), and MODIS-Aqua (MODIS-A). Coverage is computed as a percentage of total area over the world's oceans. The plot also indicates the time at which each mission began operational data production.

limits and by atmospheric path phenomena that naturally restrict observations of the ocean surface. Daily global coverage can be substantially improved by combining observations from instruments on different satellite platforms. This will fill gaps between swaths, shift sun glint patterns, and increase the chance of observing through a cloud-free atmosphere when clouds patterns move through the day. Figure 2.1 shows the increase in daily global ocean coverage by combining SeaWiFS, MODIS-Terra, and MODIS-Aqua Level-3 binned datasets at 9-km resolution. The result comes from a 2003 study when MODIS Terra data were still processed for ocean colour. Data merger can reduce both stochastic (*i.e.* random) errors, by increasing the number of measurements, and deterministic (*i.e.* systematic) trends, by sensor inter-comparisons and cross-calibrations, to provide improved quantitative accuracy and user confidence in data. Table 2.1 gives approximate daily percentage of global ocean coverage by SeaWiFS and MODIS-Aqua. Joint observations between the two sensors reach 29% of the ocean at 9-km resolution and nearly 20% at 4-km. With more instruments involved in the data merger, higher daily coverage is expected. As seen from the table, data merger is of particular benefit for users wanting to investigate ocean colour at higher spatial resolutions.

**Table 2.1** Percentage of the global ocean observed by MODIS-Aqua (v1.1) and SeaWiFS (v5.1) sensors on day 21 June 2003. GAC is Global Area Coverage, which is 4-km sub-sampled data stored onboard SeaWiFS. MLAC is Local Area Coverage 1-km data received by downlink stations around the world (does not provide global coverage).

| **Binned Data Resolution** | **MODIS Aqua** | **SeaWiFS GAC** | **SeaWiFS GAC MLAC** | **MODIS-Aqua + SeaWiFS GAC** | **MODIS-Aqua + SeaWiFS GAC MLAC** |
|---|---|---|---|---|---|
| 1 km | 4.651 | 0.516 | 6.233 | 5.058 | 9.639 |
| 4 km | 8.461 | 7.977 | 14.885 | 13.862 | 19.206 |
| 9 km | 13.264 | 16.059 | 23.108 | 23.433 | 28.908 |

Ocean-colour data merging can support a wide variety of users and applications and increase scientific and operational output from ocean-colour measurements. The users include scientists as well as education and outreach institutions, environmental agencies (*e.g.* conservation, coastal planning), policy makers, and industries (*e.g.* tourism, fisheries, oil). To understand the benefits of merging ocean-colour data from different missions it is instructive to identify these users' applications and how the merged data can meet their extended requirements better than any single-sensor dataset. Table 2.2 provides a summary of users and applications of ocean-colour information. The users have been divided into three rather arbitrary and somewhat overlapping classes: Marine Operations, including commercial and military users; Environmental Protection, including government agencies and commercial users who are obliged by law to responsibly manage their environmental impact; and finally, Earth System Science. Many of the user requirements arise from obligations under various international and national policies, such as the International Convention for the Prevention of Pollution from Ships, the European Bathing Water Directive (1976), the European Convention on Climate Change.

To effectively meet the goals of ocean-colour remote sensing, data accuracy, product specifications, and merger methodology need to match the application requirements desired by the community. Depending on whether these applications are for Earth science data records or near-real time operational use for open ocean or coastal studies, they may need different or modular approaches. Each application listed in Table 2.2 may require specific data accuracy, spatial and temporal resolution, time series length, grid type, output products, and operational data delivery. The full benefits of data merger will then depend on how these merged datasets meet the user requirements and become an effective science and decision tool. For example, climate modelling will benefit from increased global ocean-colour coverage and enhanced statistical confidence in derived bio-optical variables. The merger will address the data accuracy by combining multi-sensor observations to improve the signal-to-noise ratio and by promoting further sensor calibrations, algorithm development, and inter-comparisons whose improvement will greatly enhance the consistency of the data sets. Additionally, the merged data will support better resolution of high frequency events and improved identification of biological and physical phenomena and their variability. These more extensive and frequent local and regional coverages will increase the operational use of ocean-colour data within monitoring and forecasting systems.

Increased global coverage, improvement in derived parameter accuracy, and data standardization will expand scientific and operational output from ocean-colour observations. These advances will provide broader public recognition for ocean-colour measurements as well as an increased scope for ocean-colour applications to the benefit of the scientists, international space and environmental protection agencies, marine operations, and industries. The enhanced ocean-colour datasets will promote our understanding of the global biogeochemical cycles.

**Table 2.2** Users and applications of ocean-colour information

| Domain | Users | Applications |
|---|---|---|
| Marine Operations | - Industries concerned with marine engineering (such as offshore oil and gas), marine survey, shipping, aquaculture, and fisheries<br>- Port authorities<br>- Naval operations | Ocean colour is particularly useful for monitoring surface currents and eddies where thermal homogeneity precludes the use of SST. Near real time (NRT) information is required for forecasting to aid short term operational decision making, while climatologies are required as input to the design of new structures and for determining efficient shipping routes. As well as current flow information, naval operations require information on water clarity.<br>Information on sediment concentration and transport is required to help combat the silting-up of harbours and shipping terminals.<br>Aquaculture requires NRT monitoring of the occurrence and transport of harmful algal blooms and sediment plumes.<br>Operators of fishing vessels require phytoplankton concentrations to help in fish finding, particularly for tuna and sardine purse seine and pelagic long line fishing |
| Environmental Protection | - Government agencies with responsibility for coastal erosion, marine pollution (including oil spills), tourism and climate change assessment<br>- Industries which are required to establish environmental impact management systems | Monitoring of coastal erosion and industrial pollution requires information on sediment concentrations and transport. Algal blooms need to be monitored for marine ecosystem and fisheries protection.<br>Offshore engineering and shipping operators are required to systematically monitor water quality as part of their environmental protection responsibilities.<br>Oil spills can be detected and monitored through a combination of ocean colour and synthetic aperture radar observations.<br>Tourism can benefit from monitoring of sediment, dissolved organic pollutants and red-tides for the management of bathing water quality.<br>Climate change forecasts depend on global ocean-colour data in order to accurately assess the oceanic uptake of atmospheric carbon. |
| Earth System Science | - Scientists: oceanographers, carbon cycle modellers, marine biologists | Ocean-colour merged information is required in order to understand marine biogeochemical cycles by providing more observations, and therefore better signal to noise - which is important, because over time spans of 10-years the climate change signal can be very weak |

# Chapter 3

# Coincident Global Ocean-Colour Missions

This report emphasizes global missions as candidates for data merging (Table 3.1), because of opportunities for data and knowledge associated with global missions, and the very large reward in terms of benefits. The IOCCG Data Merging Working Group, however, also encourages data merging investigators to seek out and use data from the smaller scale missions. All global ocean-colour missions that have temporal overlap with at least one other mission are listed in Table 3.1. A listing of non-overlapping global missions, and all of the smaller scale missions is provided in the Appendix. Further information can be obtained from the IOCCG web site http://www.ioccg.org/sensors_ioccg.html.

Since all but one of the merging methods surveyed in this report utilized Level-3 data (see Chapter 4), we consider Level-3 products to be highly recommended. Level-3 is defined as products mapped onto an Earth projection. For ease of use by data merging scientists, we recommend that all global missions follow standard guidelines in the development of Level-3 data sets (see IOCCG, 2004).

**Table 3.1** Level-3 data product information for current global missions. Only missions that overlap in time with at least one other mission at some point during their lifetime are shown. MODIS-Terra is not considered because ocean-colour data processing has been discontinued (there is a possibility that it may be resumed in the future, although there are no such plans at the time of writing). N/A indicates not available.

| Mission | Launch date | L3 Products | L3 Grid | Data Availability Period |
|---|---|---|---|---|
| SeaWiFS | Aug. 1997 | Chlorophyll | 4-km, 9-km, daily, 8-day, monthly, yearly | Sep. 1997-present |
| | | LwN (412, 443, 490, 510, 555, 670) | | |
| | | K490 | | |
| | | PAR | | |
| MODIS-Aqua | May 2002 | Chlorophyll | 4-km, 9-km, daily, 8-day, monthly, yearly | Jul. 2002 -present |
| | | LwN (412,443,488,431,551,667) | | |
| | | K490 | | |
| GLI | Dec. 2002 | Chlorophyll | 9-km daily, 8-day, monthly | Mar 19-22, 2003, Apr. 2-Oct. 24, 2003 |
| | | LwN (380, 400, 412, 443, 460, 490, 520, 545, 565, 625, 666, 680, 678, 710) | | |
| | | CDOM at 440nm | | |
| | | K490 | | |
| | | Total suspended matter | | |
| | | PAR (planned) | | |
| POLDER-2 | Dec. 2002 | Chlorophyll (Case-1) | 9-km daily, 10-day, monthly | Apr. 2-Oct. 24, 2003 |
| | | Marine diffuse reflectance (443, 490, 565) | | |
| MERIS | Mar. 2002 | N/A but many Level-2 products: reflectance at 412.5, 442.5, 490, 510, 560, 620, 665, 681.25, 705, 753, 775, 865, 885 nm, 2 pigment indices, total suspended matter, yellow substance absorption, instantaneous PAR (free software tools are available to produce Level-3 products) | 9-km, monthly, yearly | May 2002-present |

**Table 3.2** Level-3 data product information for future global missions.

| Mission | Launch date | L3 Products | L3 Grid | Status (approved or planned/ pending approval) |
|---|---|---|---|---|
| VIIRS-NPP | 2008 | Not yet defined | Not yet defined | Planned |
| SGLI | 2011 | Chlorophyll | 4-km, 9-km, daily, 8-day, monthly | Planned/pending approval |
| | | LwN (380, 412, 443, 490, 530, 565, 670) | | |
| | | CDOM | | |
| | | Suspended sediment | | |
| | | PAR | | |
| VIIRS-NPOESS | 2012 | Not yet defined | Not yet defined | Planned |
| GMES Sentinel-3 | 2012 | Not yet defined | Not yet defined | Planned |

# Chapter 4

# Survey of Ocean-Colour Data Merging Methods

This chapter contains a survey of data merging methodologies used in ocean-colour research. The list is not necessarily exhaustive, but includes all methods known by the IOCCG Data Merging Working Group at the time of publication of this report. Nevertheless, it provides a look at the breadth and depth of data merging activities that have been explored. Although they are all very different, there are some common features among most of them. First, all but one of the methods surveyed use Level-3 data. Level-3 is the preference because of ease of use, availability, and because one of the main advantages of merging is spatial coverage, which is inherently defined on an Earth grid. All have so far involved only chlorophyll.

Additionally, all methods are error-correcting in nature, and many are bias-correcting. Most are statistical methods, but the bio-optical methods are quite different and are placed in a separate class. Data assimilation uses statistical methods with an underlying numerical model of ocean biological and sometimes optical, processes, and is also a separate class as follows:

## Statistical Methods

**Random Error-Correcting**

- Binning
- Averaging
- Error-weighted averaging

**Bias-correcting**

- Subjective Analysis
- Blended Analysis
- Optimal Interpolation
- Objective Analysis
- Wavelet Analysis
- Machine Learning Analysis

**Bio-Optical Methods**

- Spectral Bio-Optical Modeling

**Numerical Model-Based Methods**

- Data Assimilation into numerical models

## 4.1 Binning (Watson Gregg)

The binning method for data merging is the only method in the survey that requires Level-2 data. It treats multiple satellite data the same as individual mission data, starting with Level-2 data and producing Level-3 data by placing within pre-defined bins. Although all data are treated equally, the Level-3 merged result will be biased in favour of data with the highest native Level-2 resolution. For example, if SeaWiFS-GAC and MODIS-Aqua data are binned to a 9-km Level-3 grid, MODIS-Aqua data will predominate because its native Level-2 resolution is 1-km, whereas SeaWiFS is 4-km. A 1-km Level-3 grid resolution data set will have an order of magnitude more (or significantly more) bins occupied by MODIS-Aqua than by SeaWiFS.

## 4.2 Averaging (Watson Gregg)

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

$$C_{ij} = \frac{\sum_s C_{ijs}}{\sum_s n_{ijs}} \tag{4.1}$$

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

Overall the averaging method is best for data with no biases, because it is simple, objective, and computationally fast. If there are biases in either or both data sets that are uncorrected or unrecognized, this method will propagate these errors into the merged field, and produce a poor quality data set. Knowledge of biases in the new versions of each sensor is presently lacking, and requires substantial effort.

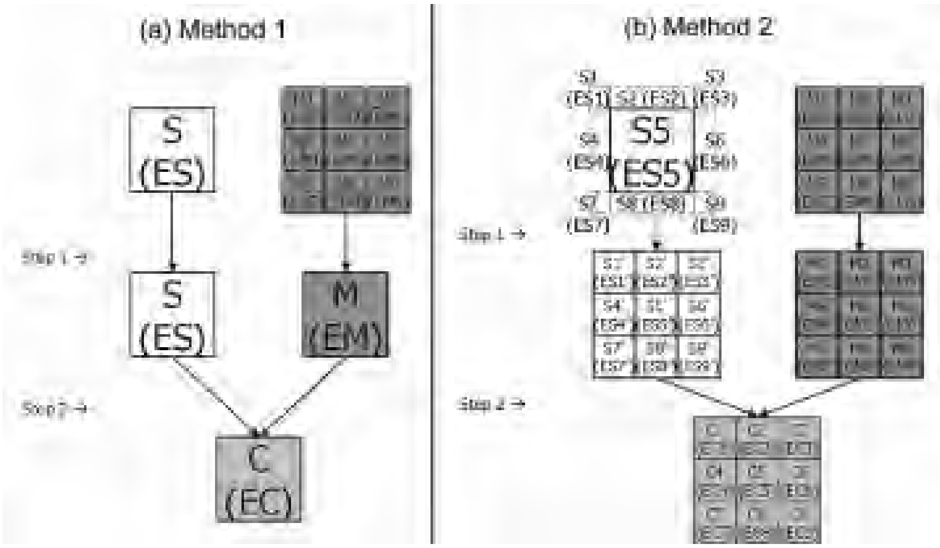


**Figure 4.1** (a) Description of the first method of weighted averaging. $X$ and $EX$ are the pixel's chlorophyll concentration and the associated error respectively (in mg m$^{-3}$ for un-transformed values or unitless for log-transformed values). $X$ is “S" for SeaWiFS, “M" for MODIS-Aqua and “C" for the combined data. (b) Description of the second method of weighted averaging. Same notation as for (a), (Figure taken from Pottier *et al.* (2006).

## 4.3 Error-Weighted Averaging (Claire Pottier)

The **error-weighted averaging** method was developed by Pottier *et al.* (2006). They used a map of chlorophyll concentrations for a given day plus a map of the measurement errors for each sensor (the root mean square computed from the match-ups between *in situ* and satellite data) to obtain a map of the combined chlorophyll for that particular day and a map of the associated errors. The value for a combined pixel, centered at a given longitude and latitude, equals the weighted averaging of the pixels of each sensor at the same location.

As an example, the merging of SeaWiFS and MODIS-Aqua data is examined using SeaWiFS and MODIS-Aqua data obtained from the NASA Goddard Earth Science (GES), Distributed Active Archive Center (DAAC). Daily Level-3 binned chlorophyll concentration data at $1/12°$ equal-angle resolution for SeaWiFS and $1/24°$ for MODIS-Aqua were used in this example. Because of the log-normal distribution of chlorophyll (Campbell, 1995) and the difference in the resolution of both sensors, four cases are conceivable: merging un-transformed or log-transformed values to obtain combined data at $1/12°$ or $1/24°$ resolution (Methods 1 and 2 respectively, Figure 4.1).

In the following section, “candidate" refers to the pixels obtained after the first step of the weighted averaging, *i.e.* the pixels used to compute the combination between the sensors' products. The first step of Method 1 (see Figure 4.1a) consists of undersampling MODIS-Aqua pixels to bring them to the same resolution as that of SeaWiFS. For computation on un-transformed data, MODIS-Aqua's candidate $M$ is obtained with the following formula:

$$\begin{aligned} M &= 1/4M_5 + 1/8(M_2 + M_4 + M_6 + M_8) + 1/16(M_1 + M_3 + M_7 + M_9) \\ &= \sum_{\mathrm{i}=1}^{9} \lambda_\mathrm{i} M_\mathrm{i} \text{ with } \sum_{\mathrm{i}=1}^{9} \lambda_\mathrm{i} = 1 \end{aligned} \tag{4.2}$$

To compute the associated error, $EM$, a distribution with a mean of $M$ and a standard deviation of $EM$ can be considered as a linear combination of the eight distributions of a $M_\mathrm{i}$ mean and an $EM_\mathrm{i}$ standard deviation (*i.e.* the error associated with the $M_\mathrm{i}$ pixel). Since no specific rule for log-normal distributions exists, $EM$ was computed as follows: first, a log-normal distribution $i(i = \{1, \ldots, 9\})$ was constructed with a $M_\mathrm{i}$ mean and an $EM_\mathrm{i}$ standard-deviation; then their linear combination (the weights are the $\lambda_i$) was computed; $EM$ is the standard deviation of the new distribution obtained. The second (and last) step of Method 1 is the combination of SeaWiFS and MODIS-Aqua's candidates. The combined pixel, C, is computed as follows:

$$\mathrm{C} = \left(1 - \frac{\%ES}{\%EM + \%ES}\right) S + \left(1 - \frac{\%EM}{\%EM + \%ES}\right) M \tag{4.3}$$

where $\%EX = EX/X$. Actually, the weights of $S$ and $M$ represent the percentage confidence in the corresponding pixel in comparison with the others. The same method as step 1 was used to compute the associated error.

To obtain MODIS-Aqua's candidate for log-transformed values, the "$X$" values were replaced by "$\log X$" values. Since the geometric mean was used, Equation 4.2 was replaced by Equation 4.4 below:

$$M = 10^m \text{ with } m = \sum_{\mathrm{i}=1}^{9} \lambda_\mathrm{i} \log M_\mathrm{i} \tag{4.4}$$

In this case $EM_\mathrm{i}$ is the error associated with $\log(M_\mathrm{i})$. Since $\log(M_\mathrm{i})$ has a normal distribution, $EM$ can be expressed as follows:

$$EM = \sum_{\mathrm{i}=1}^{9} \lambda_\mathrm{i} EM_\mathrm{i} \tag{4.5}$$

To compute the combined pixel, Equation 4.3 was replaced by the following:

$$C = 10^{m'} \tag{4.6}$$

where $m' = \left(1 - \frac{\%ES}{\%EM+\%ES}\right) \log S + \left(1 - \frac{\%EM}{\%EM+\%ES}\right) \log M$

As in the first step, the associated error $EC$ is the linear combination of the errors of $\log(S)$ and $\log(M)$.

For Method 2 (see Figure 4.1b) the first step is to increase the number of SeaWiFS pixels to the same resolution as MODIS-Aqua. For untransformed values, SeaWiFS' candidates were then obtained as follows:

$$
\begin{aligned}
S1' &= 1/4(S1 + S2 + S4 + S5)\\
S2' &= 1/2(S2 + S5)\\
S3' &= 1/4(S2 + S3 + S5 + S6)\\
S4' &= 1/2(S4 + S5)\\
S5' &= S5\\
S6' &= 1/2(S5 + S6)\\
S7' &= 1/4(S4 + S5 + S7 + S8)\\
S8' &= 1/2(S5 + S8)\\
S9' &= 1/4(S5 + S6 + S8 + S9)
\end{aligned}
\tag{4.7}
$$

The associated errors were obtained in the same way (by replacing "*S*" with "*ES*"). The combination of SeaWiFS and MODIS-Aqua's candidates is the same as for Method 1. Regarding log-transformed values, Equations 4.7 are derived using the same techniques as those used in Method 1: the combinations are then similar.

**Table 4.1** Coefficient of determination, $r^2$, RMS errors, and bias between combined data obtained by weighted averaging and *in situ* data in the North Atlantic basin. N is the number of points used for the computation (from Pottier *et al.*, 2006).

| | | N | Statistics on decimal values (mg m$^{-3}$) | | | Statistics on log values (unitless) | | |
|---|---|---|---|---|---|---|---|---|
| | | | $r^2$ | RMS | Bias | $r^2$ | RMS | Bias |
| Method 1 | Decimal values | 1288 | 0.471 | 5.481 | 0.826 | 0.884 | 0.292 | 0.164 |
| | Log values | 1288 | 0.471 | 5.221 | 0.758 | 0.882 | 0.292 | 0.161 |
| Method 2 | Decimal values | 2268 | 0.472 | 3.755 | 0.530 | 0.862 | 0.304 | 0.178 |
| | Log values | 2268 | 0.468 | 3.956 | 0.554 | 0.863 | 0.301 | 0.174 |

Both methods were applied to untransformed and log-transformed data from the North Atlantic basin (Pottier *et al.*, 2006). For each of the four cases, match-ups between the combined and *in situ* data were examined and RMS and bias were computed on the untransformed and log-transformed values (Table 4.1). The quality of the combined data does not differ much between the applications using untransformed and log-transformed values. Nevertheless, the decimal RMS and bias shows that Method 2 (combined data using MODIS-Aqua resolution) gives better results than Method 1 (combined data using SeaWiFS resolution).

Regarding CPU times, the application using log-transformed values is naturally the fastest one, taking up to 10 minutes (using a 700 MHz processor) for both methods to produce one daily map of the North Atlantic basin. In contrast, the application using untransformed values is much slower: 20 to 30 minutes for Method 1 and 10 to 20 minutes for Method 2. As an example, Figure 4.2 shows the results

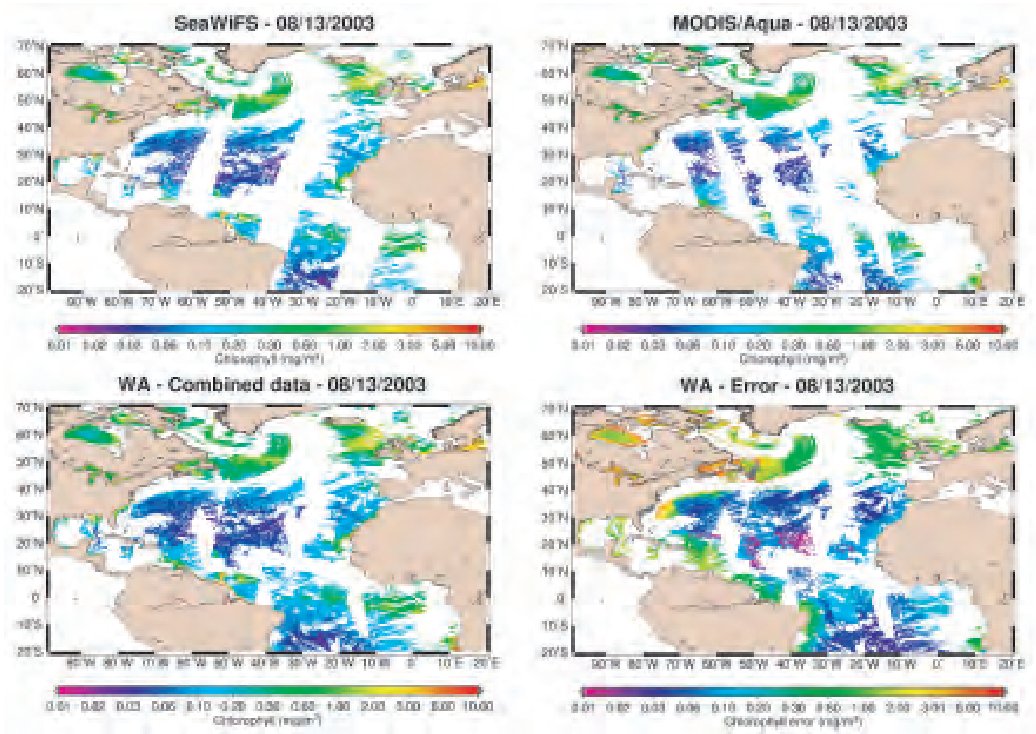


**Figure 4.2** Application of the weighted averaging Method 2 for 13 August 2003. Top: SeaWiFS and MODIS-Aqua chlorophyll. Bottom: map of the combined data (left), and the map of the associated error (right).

of the application of Method 2 on untransformed values for 13 August 2003. The spatial coverage of the combined data is 34%, which is an improvement of 11% over the initial SeaWiFS coverage (23%). The associated error depends on the amount of overlap of the sensors.

## 4.4 Subjective Analysis (Watson Gregg)

In the subjective analysis, quantitative information about the quality of the sensors is used to develop a system of weighting functions $W$, that enable the production of an enhanced merged data set, at least in principle

$$C_{ij} = \frac{\sum_s W_s C_s}{\sum_s W_s} \tag{4.8}$$

where $W_s = W_s(\theta, \theta_o, s, L_{g,\ldots})$ and $\theta$, $\theta_o$, s, $L_g$ represent satellite zenith angle, solar zenith angle, sensor behaviour, and sun glint, respectively, and are intended to be a small subset of the possible variables that may determine superior performance of one sensor over the others. This quality information can vary from sensor to sensor over the Level-3 grid. For example, sun glint will impact different sensors in different locations on the Earth depending upon the orbit and observation characteristics. If no quality information is available, this method can default to equal weighting, and is thus identical to the averaging method.

A comprehensive quantitative weighting system is a difficult task that exceeds the abilities of a single investigator. Thus its successful application requires detailed information about different missions from the mission representatives, the SIMBIOS Project, and the scientific community at large. It is probably unlikely that complete definitions can be made for all missions. However, it is widely used in missions now, at least in some form. For example, a mask is a condition in which W= 0. A flag is a qualitative data quality indicator.

The advantage of this method is that it relies entirely on scientific and engineering information to produce the highest quality merged data product. Conceptually it is superior to all of the other methods because fundamental information about sensor performance is explicitly incorporated into the final product. It suffers from the informational demands required. Not only must one know the reasons for relative performance and the errors resulting from the influences, but also how to quantitatively represent them. It is likely that this method can be used most effectively in combination with other methods.

## 4.5 Blended analysis (Watson Gregg)

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

$$\nabla^2 C^b = \Psi \tag{4.9}$$

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

$$C_{ibc} = I \tag{4.10}$$

where the subscript "ibc" indicates internal boundary condition (IBC) and $I$ is the *in situ* value of chlorophyll. Thus *in situ* data appear un-adjusted in the final blended product. In its application to multiple ocean-colour data sets, *in situ* data would be replaced by a determination of superior performance by one of the sensors data, and utilized as the IBC. This could occur across the domain for an individual sensor, if its calibration was considered superior, for example. Or it could occur by location as the environmental conditions provide for better performance of one sensor over the others (*e.g.*, location of sun glint, individual scan problems, etc.).

Where one sensor data could be established as superior, it would serve as the IBC. If no distinction could be provided, the data could simply be merged using one or more of the other methods. Then the final merged product would be blended, so that the internal boundary conditions are upheld, and the rest of the Level-3 field is adjusted according to the spatial variability of the other sensors. This can provide a bias correction to the non-IBC points, while retaining their spatial structure, and potentially produce an overall enhanced data set.

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

The blended method is effective at eliminating biases if a "truth field" can be identified. In the analyses done so far, we assumed SeaWiFS to be a truth field unilaterally, and MODIS was the data blended to produce the final merged product. The effectiveness of the bias-correction capability of the blended analysis is quite well known in *in situ*-satellite data merging, but not in satellite-satellite merging. Our results indicate that significant differences in satellite data quality coupled with the very large coverage of both sensors, results in over-correction by the blended method.

## 4.6 Optimal Interpolation (Watson Gregg)

In this method the weights, W, are chosen to minimize the expected error variance of the analyzed field (Daley, 1991). It differs from the spatial analysis method by allowing error correlations to determine the effective separation distance, and from the blended analysis by use of a statistical approach for defining the weights. A weight matrix W represents the error correlations, and is referred to as the error co-variance matrix:

$$C_{ij} = C_{sij} + W_{ijkm}(C_{s+1,km} - C_{s,km}). \tag{4.11}$$

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

The optimal interpolation (OI) method has many of the advantages of the blended method in bias-correction. However, the over-correction behaviour of the blended method is reduced to the point that it is not readily apparent in the resulting merged field. The method suffers from computational complexity and is very slow.

## 4.7 Objective Analysis (Claire Pottier, Ewa Kwiatkowska)

Statistical objective analysis is a well-known method used to perform spatial interpolation of environmental data onto global 2-, 3-, and 4-dimensional coverage grids (Thiebaux and Pedder, 1987). The approach was first applied in meteorology to ground and satellite measurements and was introduced to oceanography by Bretherton *et al.* (1976). Statistical objective analysis is currently used operationally to create NOAA's real-time global sea surface temperature (SST) maps (Reynolds, 1988; Reynolds and Smith, 1994) and to integrate altimeter data (Le Traon *et al.*, 1998). Other promising spatial and temporal assimilation algorithms include Kalman filtering (the error statistics evolve with time), Bayesian approaches, and nudging.

Statistical objective analysis has been investigated in the merger of chlorophyll concentrations from MODIS-Terra and SeaWiFS onto global coverage Level 3-bin grid (Kwiatkowska and Fargion, 2002b; Kwiatkowska, 2003c) and from MODIS-Aqua and SeaWiFS onto a grid covering the North Atlantic (Pottier *et al.*, 2006). The weights of the interpolation are dependent on the ensemble spatial (and temporal) correlation structure of the environmental field, *i.e.* the spatial distribution of observations relative to one another (Julian and Thiebaux, 1975). The correlation structure of the field is expressed in terms of space-lag (and time-lag) correlation functions. Modeling space-lag correlation functions has been a subject of extensive research. Since ocean-colour data varies significantly spatially, different categories of correlation functions are extracted for either *a priori* or statistically defined ocean provinces. The weights also incorporate information on instrument errors, biases and the signal-to-noise ratio of the measurements. These are calculated from sensor match-ups with *in situ* observations. The analysis applies a preliminary "prediction" of the field, a first-guess field, for which past research used a climatology and a weekly average. Eventually, the first-guess field will be the previous day global ocean-colour coverage derived from the analysis. To determine the value of a field $\theta$ at a point $x$ in space and time, a single or multiple pass of the least squares descent algorithm is used to minimize the ensemble average of the increments between the multi-sensor measurements, $\Phi_{\mathrm{obs}^i}$ (with i, $\epsilon$ [1,N]), and the first-guess field, $\Phi_{\mathrm{i}}$, applying the correlation structure of the field. The minimum variance solution is given by:

$$\theta_{\mathrm{est}}(x) = \sum_{\mathrm{i}=1}^{N}\sum_{\mathrm{j}=1}^{N} A_{\mathrm{ij}}^{-1} C_{\mathrm{xj}} \Phi_{\mathrm{obs}^{\mathrm{i}}} \tag{4.12}$$

with $\Phi_{\mathrm{obs}^i} = \Phi_{\mathrm{i}} + \epsilon_{\mathrm{i}}$, where $\epsilon_{\mathrm{i}}$ is the measurement error (assumed to be uncorrelated with the signal), $A$ is the covariance matrix for the observations:

$$A_{\mathrm{ij}} = \langle \Phi_{\mathrm{obs}^{\mathrm{i}}} \Phi_{\mathrm{obs}^{\mathrm{j}}} \rangle = \langle \Phi_{\mathrm{i}} \Phi_{\mathrm{j}} \rangle + \langle \epsilon_{\mathrm{i}} \epsilon_{\mathrm{j}} \rangle \tag{4.13}$$

and $C$ is the covariance vector for the observations and the field to be estimated:

$$C_{\mathrm{xj}} = \langle \theta(x) \Phi_{\mathrm{obs}^{\mathrm{j}}} \rangle = \langle (x) \Phi_{\mathrm{j}} \rangle \tag{4.14}$$

The associated error variance is given by:

$$e^2 = C_{xx} - \sum_{i=1}^{N}\sum_{j=1}^{N} C_{xi}C_{xj}A_{ij}^{-1} \tag{4.15}$$

The objective analysis used by Le Traon *et al.* (1998) takes into account the biases between sensors and *in situ* data, by expressing $\langle\epsilon_i, \epsilon_j\rangle$ in the following form:

$$\langle\epsilon_i, \epsilon_j\rangle = \delta_{ij} b^2 \text{ for points (i, j) from different sensors,}$$
$$\langle\epsilon_i, \epsilon_j\rangle = \delta_{ij} b^2 + E \text{ for points (i, j) from the same sensor,}$$

where $b^2$ is the variance of the measurement noise and $E$ is the variance of the bias.

With successive corrections, non-zero weights are given to observed increments only if the measurements lie within a prescribed distance, known as the influence radius, of the grid point $x$ being considered. Consequently, statistical objective analysis extends the multi-sensor merged coverage over areas which are within reach of the influence radius. To merge chlorophyll concentration, data have to be log-transformed to avoid ill-conditioning of the covariance matrix that defines the weights of the interpolation.

Statistical objective analysis is computationally intensive. To produce a daily map of the North Atlantic on a regular grid of $0.1^\circ \times 0.1^\circ$, the required CPU time is about 1 h 30 min on a 700 MHz processor. The processing speed depends on the extent of the multi-sensor coverage which varies seasonally (Pottier *et al.*, 2006).

Figure 4.3 shows the result of statistical objective analysis applied within the North Atlantic basin using the combination of SeaWiFS and MODIS-Aqua chlorophyll on 13 August 2003. Spatial coverage of the merged data is 71% on a $0.1^\circ \times 0.1^\circ$ grid, while the original SeaWiFS coverage was 23% and MODIS-Aqua was 20%.

## 4.8 Wavelet Analysis (Ewa Kwiatkowska)

Data merger opportunities at local spatial scales may provide useful tools for scientists interested in smaller-size geophysical phenomena and in complex environments such as coastal zones. An example of a local merger application is multi-resolution analysis which can be applied in cases where there is coincident ground coverage between two sensors of different spatial resolutions (Núñez *et al.*, 1999). If the more accurate sensor has lower spatial resolution, discrimination of oceanic features in its imagery can be enhanced using higher-resolution data from other instruments without changing the overall magnitude of the bio-optical field. This has been accomplished using a discrete wavelet transform (Kwiatkowska-Ainsworth, 2001; Kwiatkowska and Fargion, 2002a).

The wavelet transform examines an image at different frequency and scale intervals, which correspond to measures of spatial detail and resolution in the image

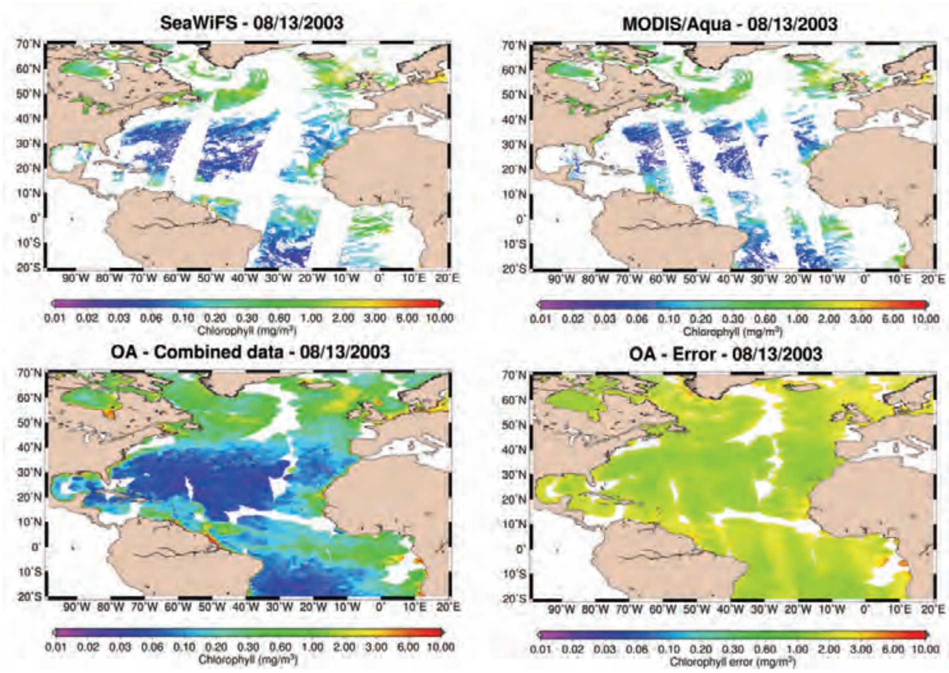


**Figure 4.3** Application of the objective analysis for 13 August 2003. Top: SeaWiFS and MODIS-Aqua chlorophyll. Bottom: the map of the combined data (left), and the map of the associated error (right).

(Mallat, 1989). In the published implementation, geographically overlapping scenes are extracted from higher and lower resolution sensors (Kwiatkowska, 2003b). A combination of high pass and low pass filtering extracts high frequency detail and low frequency background for the scene from the high-resolution sensor. This is done iteratively with the scale decreasing in powers of two until the down-sampled resolution matches the scale of the lower resolution image. Then, the low frequency background is replaced by the overlapping lower resolution image and the wavelet transform is inverted. The inversion produces a new high-resolution image where the magnitude of the general background field corresponds to bio-optical values derived from the lower resolution sensor and the spatial detail is contributed from the higher resolution instrument. Alternatively, the low frequency background extracted from the high-resolution scene can be averaged in a certain proportion with the overlapping lower resolution image. This creates a weighted merger of data from both sensors.

The wavelet algorithm was tested using chlorophyll concentration imagery from SeaWiFS and MOS. SeaWiFS Level 2 LAC scenes have a nadir resolution of 1.1 km and MOS imagery has the resolution of 0.5 km. Temporally concurrent and geographically coincident scenes between the sensors were co-registered by binning at 1 and 0.5km, respectively, and by projecting the bins onto a rectilinear grid to facilitate wavelet processing. Only one level of wavelet decomposition of the MOS scene was required

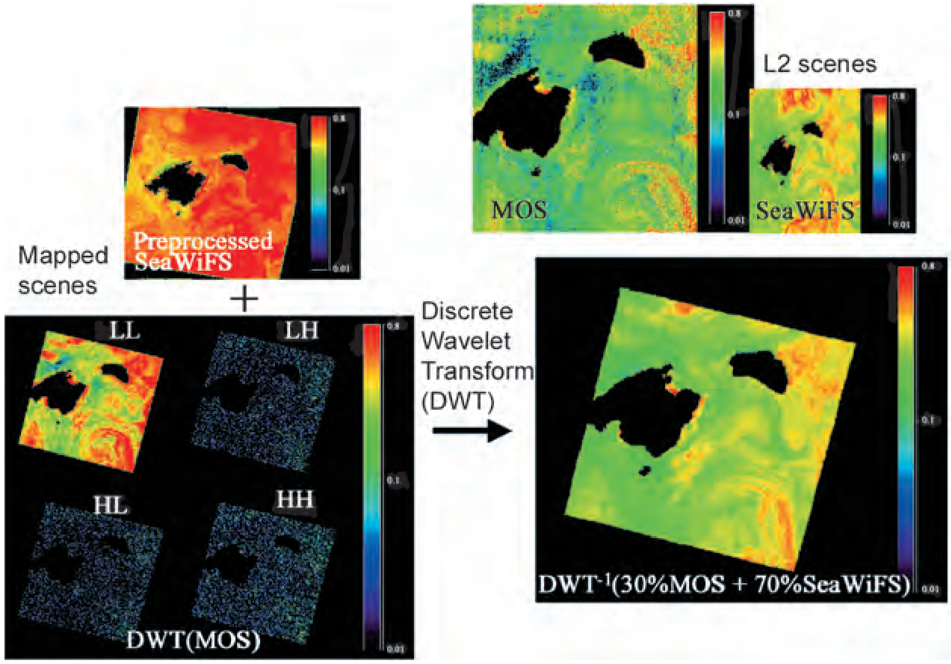


**Figure 4.4** Original Level 2 MOS and SeaWiFS chlorophyll concentration scenes and the wavelet-based merger process for both data sets after they have been mapped to a rectilinear grid. The SeaWiFS scene was pre-processed and the MOS scene underwent a single level of wavelet decomposition. High (H) and low (L) pass filters were applied to the MOS scene across the rows with column sub-sampling and across the columns with row sub-sampling. The MOS row and column low-pass coefficients (LL) were replaced by their weighted ratio with the pre-processed SeaWiFS scene. The wavelet transform was then inverted which produced the merged output image at 0.5km resolution.

due to the factor-of-two difference in spatial resolution between the sensors. An example of the MOS and SeaWiFS multi-resolution merger is displayed in Figure 4.4.

## 4.9 Machine Learning Analysis (Ewa Kwiatkowska)

There are many difficulties associated with ocean-colour data merger. Sensors have varying designs, characterizations, and data processing approaches, including atmospheric correction and bio-optical algorithms. Limitations in instrument characterization may cause calibration biases, temporal trends, and other issues, such as variable response versus scan angle (RVS), detector-to-detector calibration errors, and polarization sensitivity dependencies. It is a challenge to separate these instrument artefacts from the uncertainties in the processing algorithms. Further ambiguities in data from different sensors arise from the fact that missions are flown over the same regions at different times of a day. During the intervening period, natural changes occur in the observed atmospheric and Earth surface conditions.

Cross-calibration of concurrently on-orbit satellite sensors plays a significant part in the effort to create a long-term global consistent time series of ocean-colour climate data records from multi-mission observations. Sensor cross-calibration provides an important risk mitigation capability when instrument calibration and characterization problems cannot be fully addressed through re-examinations of pre-launch and on-board measurements or by vicarious calibrations. Cross-calibrations offer useful insights into sensor calibration and processing algorithms and, therefore, advance scientific and technical understanding of ocean-colour remote sensing and individual instrument operation. Cross-calibrated instrument data can be subsequently merged to provide ocean-colour coverage which is accurate, consistent through time and space, and suitable for oceanic climate data records.

Sensor cross-calibrations are performed using matching coverage from two instruments, which are close in time, overlap geographically, and meet clear case-1 water and clear atmosphere conditions. It is often a challenge to obtain such quality spatial and temporal overlap along with independent evaluation information. Described below are three approaches to ocean-colour cross-calibration which were researched and implemented in the context of sensor re-calibration or data merger.

1. The cross-calibration can be performed directly on the top-of-the-atmosphere (TOA) by comparing individual instrument radiances. To extract scalar gains, only the data with the same viewing and solar geometries between the sensors are applicable. The method is useful for following relative trends over time in instrument radiances. To obtain absolute gains, the differences in band spectral response functions need to be fully accommodated. This cross-calibration approach was applied to MODIS-Aqua and SeaWiFS as polarization sensitive and insensitive instruments. The occurrences of matching geometry between these two sensors are sparse. The analysis of relative TOA radiance trends achieved only limited success; mainly verifying MODIS pre-launch polarization sensitivity measurements in the visible bands.
2. Another cross-calibration approach is based on inverse atmospheric correction. The method uses normalized water-leaving radiance, a quantity that is theoretically independent of geometry and convertible between sensor bands using bio-optical models (Morel and Maritorena, 2001). Case-1 water-leaving radiances and aerosol models from a sensor used as a source of the calibration are brought to the TOA using the geometry of the sensor which is being calibrated. Consequently, original sensor TOA radiances and the radiances simulated from the calibrator sensor are compared, analyzed and the calibration gains are extracted. This technique is currently operational only for the cross-calibration of sensor visible bands. It is equivalent to the vicarious calibration of visible bands using *in situ* measurements such as the ones from MOBY. The method has also been successfully applied to the cross-calibration of OSMI and SeaWiFS (Franz and Kim, 2001), MOS and SeaWiFS (Wang and Franz, 2000) and to the calibration of POLDER (Wang *et al.*, 2002). The approach can

be expanded to trace instrument characterization problems, such as detector striping, RVS, and polarization and temperature sensitivities. It can also be used to cross-calibrate an instrument with *in situ* observations where both the in-water and atmospheric measurements are available.

3. The following approach is a machine learning multivariate regression applied in place of conventional scalar gain derivation in ocean-colour sensor cross-calibration. The regression is defined using large numbers of examples of overlapping data between instruments however, unlike the other methods, it does not require in depth understanding of sensor characterization and algorithm issues. Before the regression, a set of features associated with ocean-colour measurement is defined, which is most effective and efficient in de-correlating characterization and calibration artefacts in the data from the sensor that is being calibrated. This set will compose the inputs to the cross-calibration. In the published implementation (Kwiatkowska and Fargion, 2003), the features were extracted using a genetic algorithm and they included instrument radiances, atmospheric parameters, and viewing and solar geometries for each data point. The subsequent set of inputs consists of corresponding point radiances from the sensor used as the calibration baseline. Machine learning cross-calibration then extracts functional relationships between sets of features from both sensors by deriving a multivariate regression function that maps one sensor data into the other. The algorithm is important when sensor characterization, calibration, and processing artefacts are intertwined and hard to separate since *a priori* knowledge about them is limited. These learning machines apply examples using global overlapping coverage between sensors, and work as regularity detectors to discover statistically salient properties of investigated data. The regression method found most suitable for ocean colour is based on support vector machines. The machines are distribution free and cope with noisy, biased, and cross-dependent sensor data. Support vector machines are learning kernel-based systems (Cristianini and Shawe-Taylor, 2000). They use nonlinear kernel functions to project complex data to high dimensional feature spaces where it is adequate to form simple linear decision hyperplanes to perform the regression. This corresponds to highly nonlinear decision boundaries for the original data. The machine learning cross-calibration was implemented to calibrate MODIS-Terra using SeaWiFS. Radial basis functions, $e^{-\gamma||x-y||^2}$, were applied as kernels, where the $\gamma$ parameter was set equal to 1.0. The goal was to obtain a consistent series of merged MODIS and SeaWiFS chlorophyll measurements (Kwiatkowska, 2003a). The cross-calibration largely decreased the bias in chlorophyll concentration between the sensors, which is shown in Figure 4.5. It also eliminated seasonal trends in MODIS chlorophyll and latitudinal drifts that are both associated with changing solar geometries. It additionally significantly reduced MODIS RVS instabilities. Although machine

learning regression was tested using final ocean-colour products, it can also be applied to cross-calibrate sensor data on the level of water-leaving radiances and TOA radiances.

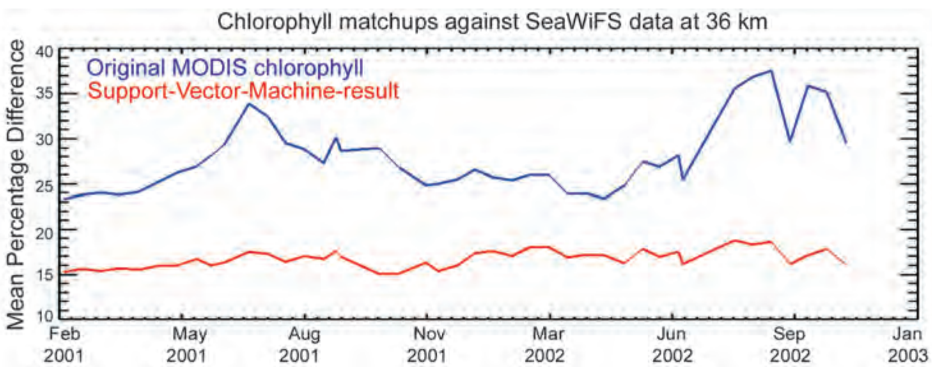


**Figure 4.5** Time trends in daily mean percent differences between SeaWiFS and original MODIS chlorophyll, in blue, and between SeaWiFS chlorophyll and the result of the support vector machine regression from MODIS data, in red.

## 4.10 Spectral Bio-Optical Modeling (Stéphane Maritorena and Frédéric Mélin)

Bio-optical merging models are based on the inversion of a semi-analytical model: inputs are the spectral normalized water-leaving radiances from available sensors. This is a completely different approach to methods that combine end-products such as chlorophyll. Model-based merging methods can generate several biogeochemical products simultaneously and they have several other benefits such as ensuring that retrievals are simultaneous and consistent by using a single bio-optical algorithm. In addition, they work with single or multiple data sources regardless of their specific bands, exploiting spectral band redundancies and band differences. Model-based approaches can also account for uncertainties in the input $L_{wN}(\lambda)$ data and provide confidence intervals for the retrieved products.

### 4.10.1 The GSM merging model

The GSM (Garver-Siegel-Maritorena) model (Garver and Siegel, 1997; Maritorena *et al.*, 2002) is the core of this merging approach and is based on a quadratic relationship between the normalized water-leaving radiance, $L_{wN}(\lambda)$, and the absorption, $a(\lambda)$, and backscattering, $b_b(\lambda)$, coefficients (*e.g.*, Gordon *et al.* 1988) or

$$\hat{L}_{wN}(\lambda) = \frac{tF_0(\lambda)}{n_w^2} \sum_{i=1}^{2} g_i \left( \frac{b_b(\lambda)}{b_b(\lambda) + a(\lambda)} \right)^i \qquad (4.16)$$

where $t$ is the sea-air transmission factor, $F_0(\lambda)$ is the extraterrestrial solar irradiance, $n_w$ is the index of refraction of the water and $g_i$ are geometrical factors. Absorption and backscattering coefficients are partitioned into sub-components and the non-water terms are all parameterized as a known shape function of an unknown magnitude. In its fully developed form the model is expressed as

$$\hat{L}_{wN}(\lambda) = \frac{tF_0(\lambda)}{n^2}\sum_{i=1}^{2} g_i \left( \frac{b_{bw}(\lambda) + b_{bp}(\lambda_0)(\lambda/\lambda_0)^{-\eta}}{b_{bw}(\lambda) + b_{bp}(\lambda_0)(\lambda/\lambda_0)^{-\eta} + a_w(\lambda) + \mathrm{Chl}\, a^*_{ph}(\lambda) + a_{cdm}(\lambda_0)\exp(-S(\lambda-\lambda_0))} \right) \quad (4.17)$$

where $a_w(\lambda)$ is the water absorption coefficient, $a^*_{ph}(\lambda)$ is the chlorophyll-specific absorption coefficient, $a_{cdm}(\lambda)$ is the sum of the dissolved and detrital particulate absorption coefficients, $b_{bw}(\lambda)$ is the water backscattering coefficient, $b_{bp}(\lambda)$ is the suspended particulates backscattering coefficient, $S$ is the spectral decay constant for CDOM absorption (Bricaud *et al.*, 1981), $\eta$ is the power law exponent for particulate backscattering coefficient, and $\lambda_o$ is a scaling wavelength (443 nm). In Equation 4.17, $a_w(\lambda)$, $b_{bw}(\lambda)$, $F_o(\lambda)$, $n_w$, $t$ and $g$ are constants taken from the literature whereas the values of $\eta$, $S$, $a^*_{ph}(\lambda)$ were determined by "tuning" the model against a large *in situ* data set (Maritorena *et al.*, 2002). The three remaining unknowns, namely the chlorophyll-a concentration, Chl, the CDOM absorption coefficient ($a_{cdm}(443)$), and the particulate backscatter coefficient ($b_{bp}(443)$), are then retrieved by applying a non-linear least-squares technique to fit Equation 4.17 to $L_{wN}(\lambda)$ or $R_{rs}(\lambda)$ data collected at four or more wavelengths.

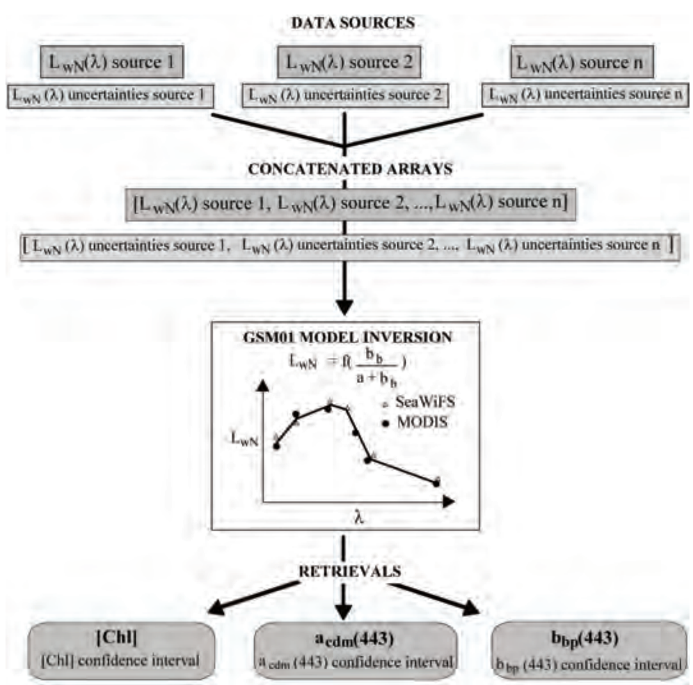


**Figure 4.6** Schematic of the GSM merging inversion scheme.

The application of the GSM01 model to the merging of ocean-colour data from different sensors is exactly the same as it is with a single sensor, making the merging of multiple data sets straightforward. If multiple data sets are available for a particular target (*i.e.* a pixel), the $L_{wN}(\lambda)$ data from all data sources are pooled together. The combined $L_{wN}(\lambda)$ observations are then used in the model inversion procedure to produce merged data products (Figure 4.6). When different sensors have the same spectral $L_{wN}(\lambda)$ observations, these data are used individually, "as is", without any averaging or other transformation. When data sources have different bands, the merging procedure takes advantage of the improved spectral resolution. The present merging procedure can also weight individual $L_{wN}(\lambda)$ observations differentially to ensure that the best observations are given a higher weight in the fitting procedure that generates the retrievals. The GSM model also provides confidence interval estimates for each of the retrievals (for details see Maritorena and Siegel, 2005). The GSM merging model is currently operational and is routinely used to merge SeaWiFS and MODIS-Aqua data. An example of merged products from SeaWiFS and MODIS-Aqua water-leaving radiances using the GSM merging model is shown in Figure 4.7.

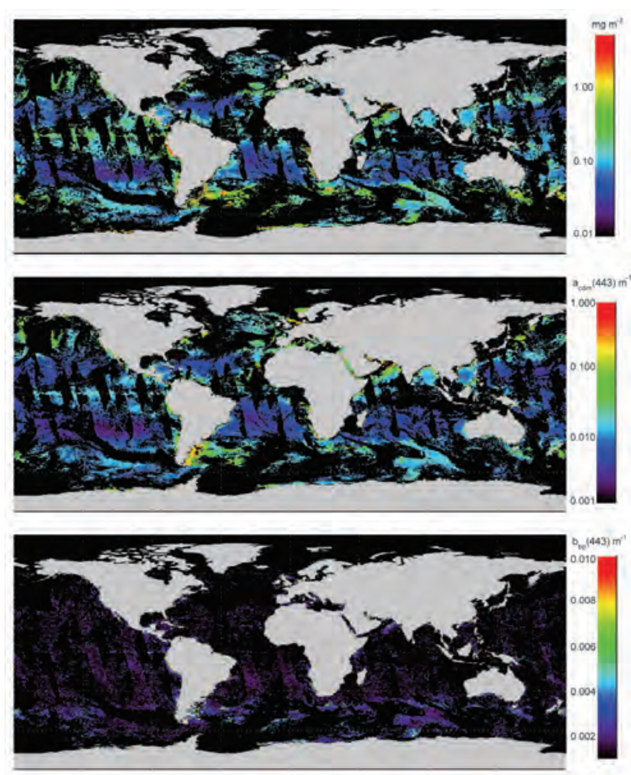


**Figure 4.7** Example of daily merged products from SeaWiFS and MODIS-Aqua Level-3 $L_{wN}(\lambda)$ observations (1 March 2003) using the GSM merging model (top: chlorophyll; middle: $a_{cdm}(443)$; bottom: $b_{bp}(443)$)

### 4.10.2 Optical merging applied to optical properties (Frédéric Mélin)

In the previous section, the merging technique ingests the normalized water leaving radiance spectra from various sensors into a semi-analytical bio-optical model to produce a set of optically significant constituents (Chl-a) and inherent optical properties. Such an approach can be generalized for merging any spectral quantities that can be represented by an optical model, either empirical or analytical, to produce a merged spectrum of an equivalent quantity. An optical model here is meant to be a spectral representation of the phenomenon under study, that is defined by a number of parameters, some of which are usually the target of an inversion, and others are fixed.

Let us consider that each sensor, i, provides a spectral quantity $(X_{i,j})_{j=1,ni}$ made up of a varying number $n_i$ of wavelengths $(\lambda_{i,j})_{j=1,ni}$ (see Figure 4.8 below). The value and number of the wavelengths can differ between sensors. The sets constituted by the $(X_{i,j})_{j=1,ni}$ and the associated wavelengths $(\lambda_{i,j})_{j=1,ni}$ are the inputs to the optically-based merger, in a similar fashion to Maritorena and Siegel (2005). A generalized optically-based merging would be a two-step procedure, that first makes the inversion using all available spectral information $(X_{i,j})_{j=1,ni}$ in order to derive the free parameters of the optical model, and then the output $X_m(\lambda)$ is recomputed with the same model used in forward mode. This approach can be applied to the merging of normalized water leaving radiance $L_{wN}$, but also to spectra of inherent optical properties. For instance, the merging of chromophoric dissolved organic matter absorption spectra derived from various satellite missions could be achieved in a straightforward way, with the selection of an (empirical) exponential behaviour of the spectra and an inversion aiming at the slope of exponential decay and the absorption value at a reference wavelength.


**Figure 4.8** General concept of optically-based merging.

In the case of the merging of normalized water leaving radiance $L_{wN}$, the spectral representation would be made through a bio-optical model that defines $L_{wN}$ as a function of inherent optical properties (IOPs), which in turn are expressed as a function of the major optically significant constituents (*i.e.* Chl-*a* for a Case 1 bio-optical model, or more generally, Chl-*a* and/or inherent optical properties expressed at one reference wavelength). In both cases, some bio-optical parameters have to be assumed (like the specific absorption of phytoplankton, or spectral shapes of IOPs) to completely define the model. The bio-optical model is inverted and then reapplied in forward mode to find $L_{wN}$ at any wavelength, representing the term $X_m(\lambda)$.

There are several advantages to such a method. First, it combines all spectral information available from different sensors in a consistent way assuming a realistic spectral model, as explained in the previous section. Moreover, the set of output wavelengths can be selected either to match the sensor specific channels, the wavelengths of field radiometers (for validation purposes) or wavelengths appropriate for dedicated bio-optical algorithms. This spectral flexibility should however be handled with caution since the spectral resolution of the output is limited by that of the inputs. Natural waters display a large variability in bio-optical parameters that are usually fixed in semi-analytical models (like the chlorophyll specific phytoplankton absorption spectrum, or the slope of exponential decay of the dissolved or detrital coloured material). By producing a merged spectrum of $L_{wN}$, the two-step procedure keeps the door open to the subsequent application of the most appropriate bio-optical model deemed suitable for a given region. On the other hand, it relies on the assumption that the final output $L_{wN}$ does not depend significantly on the fixed parameters of the model (a dependence that can be easily quantified). Another important point is that such an approach should ideally be assessed on the basis of radiometric match-ups common to several sensors, and the corresponding match-up data base is likely to be very sparse.

The output of the merger is a spectrum of a physical quantity directly related to the inputs, but not necessarily identical: an output expressed in reflectance might be preferred (derived in a way fully consistent with the bio-optical model chosen). Finally it is interesting to note that such a merger also acts as a filter that can remove noise in the spectral fields (the two-step procedure implies the reduction of an input, usually with 5 or more channels, to an intermediate value with typically 3 degrees of freedom).

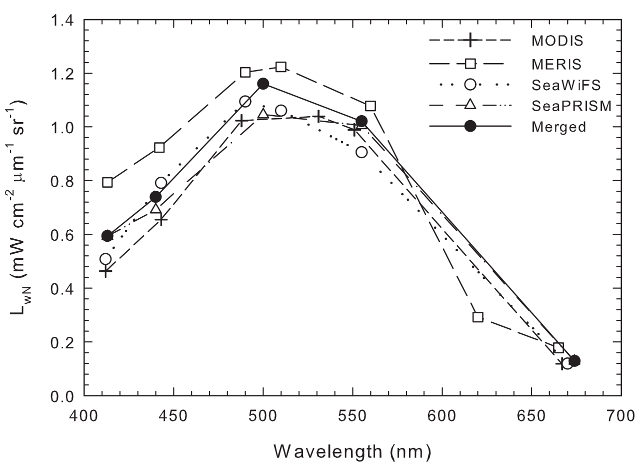


**Figure 4.9** Contemporaneous spectra of $L_{wN}$ for MERIS, MODIS-Aqua and SeaWiFS on 23 March 2003, with field values provided by the above-water radiometer SeaPRISM located at a northern Adriatic site (AAOT), and the results of the optically-based merging approach (the wavelengths of the merged spectrum are those of the field radiometer).

Figure 4.9 illustrates the method by showing an example where MERIS, MODIS and SeaWiFS $L_{wN}$ were available for the same day, together with field measurements for a location in the northern Adriatic Sea (Zibordi *et al.* 2004) and with the merged spectrum output at the field radiometer wavelengths.

## 4.11 Data Assimilation into a Numerical Model (Watson Gregg)

A conceptually different approach to ocean-colour data merging is to use classical data assimilation methods. This typically means combining satellite data and a numerical model that produces estimates of physical and biological processes using statistical methods. Many of the methods commonly used in data assimilation have already been discussed previously in the context of data merging, such as optimal interpolation and objective analysis. Most often data assimilation is used to improve model results, which has the added benefit of enabling short-term prediction by improving the model's initialization state prior to advancing forward. However, in the data merger context, merging by assimilation is an added benefit in that it can fill all gaps in ocean-colour data in a physically/biologically consistent manner without resorting to statistics. An added benefit is that data can be assimilated at the time of observation rather than on a daily basis.

An example of data assimilation into a numerical model used an existing coupled general circulation, biogeochemical, and radiative model of the global oceans (Gregg *et al.*, 2003) as a platform to assimilate SeaWiFS chlorophyll data products. SeaWiFS chlorophyll data were assimilated using the blended analysis discussed before. Data assimilation was performed daily, to remove biases associated with sampling by SeaWiFS (*i.e.*, cloud cover, sun glint, inter-orbit gaps), that are incorporated in 8-day and monthly data products. Assimilation occurred at model midnight.

The blended analysis (Equation 4.9) is computationally very fast, so much so that there is nearly negligible additional processing time in its use. However, it is very strongly weighted toward the data. Thus data errors are a very important problem in its application. For this reason, data errors must be minimized to the extent possible. In the present application, data error minimization efforts involved:

1. All daily SeaWiFS chlorophyll > 2 times the monthly mean were excluded
2. Monthly mean SeaWiFS data weighted 25% to 75% daily data
3. SeaWiFS data occurring within a model grid point containing ice were excluded
4. Regional weighting of model and SeaWiFS chlorophyll, loosely based on the global and regional evaluation of SeaWiFS data by Gregg and Casey (2004).

The assimilation model exhibited major benefits in daily coverage (Figure 4.10), to the extent that it is only visually meaningful to compare daily assimilation with monthly SeaWiFS data. The improvement to the free-run (not assimilated) model by assimilating SeaWiFS is also readily apparent.

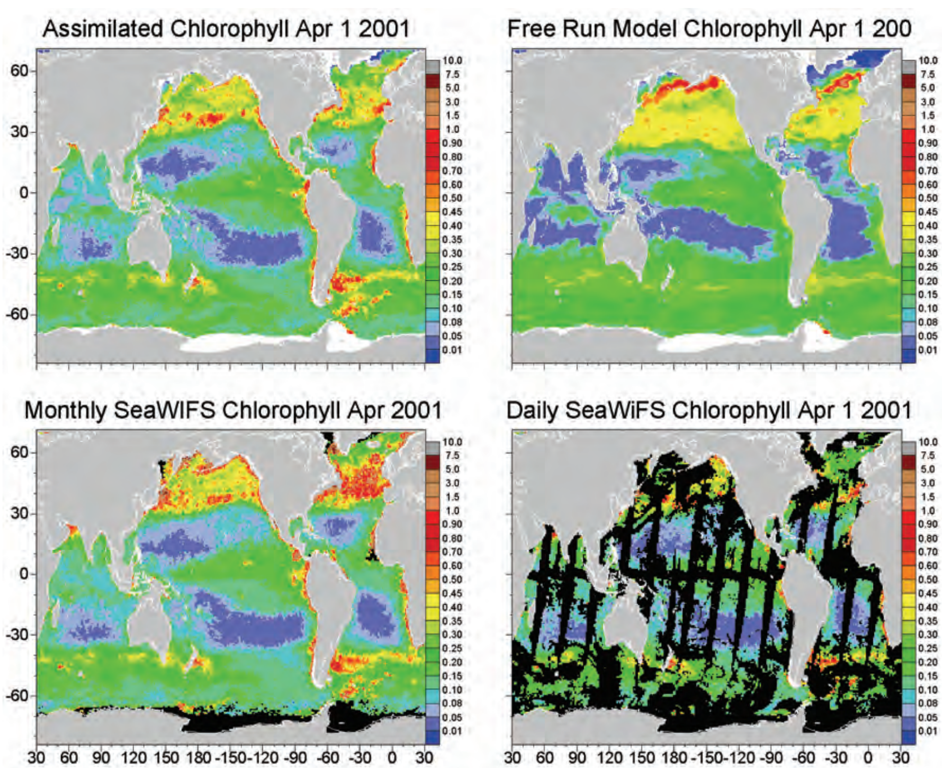


**Figure 4.10** Comparison of chlorophyll (mg m$^{-3}$) from the assimilation model, the free-run model, and SeaWiFS. The assimilation and free-run chlorophyll distributions represent simulations for April 1, 2001. SeaWiFS data for the same day are shown for comparison, along with the monthly mean. Grey indicates land and coast, black indicates missing data, and white indicates sea ice.

# Chapter 5

# Knowledge Requirements for Ocean-Colour Data Merging

Knowledge of ocean-colour remote-sensor characteristics is more important for successful data merging than high quality data. This is because all of the merging methods investigated have an error-reduction capability, and some even have a bias-correction capability (Chapter 4). Some of the data quality aspects can be determined by individual efforts of data merging scientists and the general ocean-colour community at large. However, there are at least several types of information that are at a high level that are required to even begin merging:

1. Availability of Level-3 data. We reiterate the recommendation of IOCCG Report No. 4 (IOCCG, 2004) that Level-3 data must be produced and made available. This is the primary source of data for ocean-colour data merging (see Chapter 4).
2. Native resolution Level-3 data. All of the global modern missions except the three POLDER sensors produce Level-2 data at or near 1-km spatial resolution. These data are typically binned onto 9-km or 4-km grids at Level-3, partially to improve ocean coverage by the individual sensors. This is what data merging does best: improve ocean coverage. In a data merging perspective, ocean coverage is improved by utilizing data from two or more sensors. This alleviates the need to trade spatial resolution for ocean coverage as is done for individual mission data sets. The highest possible spatial resolution (native resolution) should be produced at Level-3 for data merging efforts.
3. Availability of daily data. Daily merged products represent the maximum advantage for data merging, since the greatest increase in coverage results from using data from two or more satellites on a daily basis (Gregg *et al.*, 1998).
4. Level-3 grid structure details. These should follow the recommendation in IOCCG Report No.4 (IOCCG, 2004). A common grid is very important to data merging. Otherwise merging will result in loss of effective resolution since the bin locations will not match exactly. This is only a concern for high resolution merging efforts (at or very near the native spatial resolution of the sensors) since at larger grids loss of resolution is expected.

5. Product definition with processing version number, including as much detail on the algorithms, flags, masks, and procedures as possible.
6. Dates of Level-3 data availability.
7. Estimates of sensor performance derived from comparisons from *in situ* data. These should include, at a minimum, root-mean-square (RMS) error and bias (see Chapter 6, section 6.1). For chlorophyll and absorption coefficients (if provided), transformation to logarithm (base 10) is required, because they are log-normally distributed in the natural oceans (Campbell, 1995). Application of RMS and bias (if represented as the mean error) without first transforming will result in erroneous representation of results, since these and many statistical methods only have meaning for normally-distributed variables.

Additional error and performance representations of sensors are of extraordinary importance to data merger scientists. Such information can help guide the choice of data merger methods and implementations. A comprehensive estimate of errors is strongly urged. We recognize that this is potentially a major effort for mission personnel, and may not always be achievable given funding and delivery constraints. For these reasons, a fully comprehensive error analysis is not required but is rather, strongly urged. A comprehensive error analysis includes:

- Global and regional means and standard deviations (or variances)
- Scan angle dependencies
- Polarization
- Global and regional trends
- On-board calibration
- Cross /vicarious calibration
- Stray light
- Detector sensitivity correction error
- Channel spectral response
- Out-of-band response
- Bright target recovery
- Geometric correction error
- Bidirectional Reflectance Distribution Function (BRDF)

Chapter 6

# Data Merger Success Criteria

**Frédéric Mélin and Stéphane Maritorena**

## 6.1 Validation of Merged Products

The validation of products generated by data merging should follow the same kind of procedures as those used for single missions. Most frequently, product validation relies on match-up analyses where *in situ* measurements collected at known locations and times are compared to satellite products for that same location from an overpass within a few hours of the field measurement. The level of agreement between the *in situ* and satellite data is generally assessed by statistical comparison of the two sets of data. Matchup analyses are as critical for merged products as for data from individual missions.

Once the match-up pairs are selected, the metrics usually employed for quantifying the distance between satellite products, $(y_{\mathrm{i}})_{\mathrm{i}=1,N}$, and field measurements, $(x_{\mathrm{i}})_{\mathrm{i}=1,N}$ (log-transformed or independent of quantity), are, RMS (Equation 6.1), bias (Equation 6.2) or average ratio (Equation 6.3), coefficient of determination ($r^2$, Equation 6.4), and slope and intercept for linear regressions.

$$RMS = \sqrt{\frac{1}{N}\sum_{\mathrm{i}=1}^{N}(y_{\mathrm{i}} - x_{\mathrm{i}})^2} \tag{6.1}$$

$$\text{Bias} = \frac{1}{N}\sum_{\mathrm{i}=1}^{N}(y_{\mathrm{i}} - x_{\mathrm{i}}) \tag{6.2}$$

$$\text{Average Ratio} = \frac{1}{N}\sum_{\mathrm{i}=1}^{N}\frac{y_{\mathrm{i}}}{x_{\mathrm{i}}} \tag{6.3}$$

$$r^2 = \frac{\left(\sum_{\mathrm{i}=1}^{N}(x_{\mathrm{i}} - \overline{x})(y_{\mathrm{i}} - \overline{y})\right)^2}{\sum_{\mathrm{i}=1}^{N}(x_{\mathrm{i}} - \overline{x})^2 \sum_{\mathrm{i}=1}^{N}(y_{\mathrm{i}} - \overline{y})^2} \tag{6.4}$$

These validation statistics should be given for all match-ups and also for classes of match-ups based on the number of input sensors (possibly including a class of match-ups with no input sensors, *i.e.*, resulting from an extension of the original coverage by the merging technique). These statistics could then be compared with the results of the validation of single mission products over the same sets of match-up points. The statistics obtained using match-ups associated with products from several sensors can also be compared.

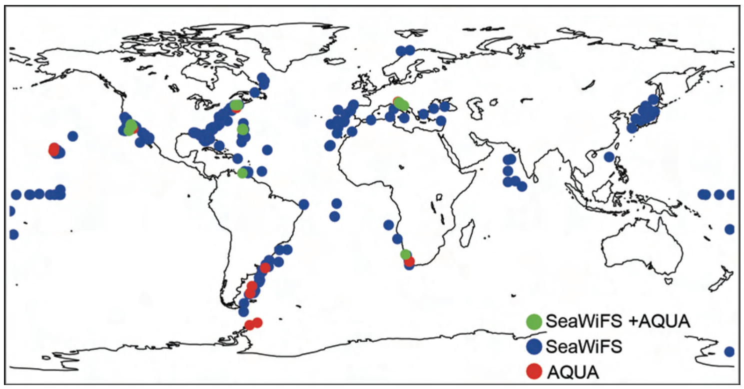


**Figure 6.1** Distribution of *in situ* stations that are coincident with SeaWiFS observations (blue dots), MODIS-Aqua (red dots), and those observed by both SeaWiFS and MODIS-Aqua (green dots).

However, an analysis based on *in situ* match-ups suffers several kinds of inevitable limitations. Frequently, the number of available good quality *in situ* datasets is limited because sampling in the open ocean is expensive and time consuming. The ability to represent oceanic diversity is poor because ships and buoys cannot cover large areas in short periods of time. Large regions of the world’s ocean are poorly sampled, or not sampled at all (Claustre and Maritorena, 2003), particularly with respect to their optical properties, and thus are not represented in the match-up data sets (Figure 6.1). Moreover, *in situ* match-up datasets are generally not well suited to examine temporal trends because “at-sea" time-series are few and do not cover all possible bio-optical provinces.

The problem is even more important for merged data since for a given date, field measurements must come from areas covered by two or more sensors (6% of the ocean surface area in the case of SeaWiFS + MODIS merging at 9-km resolution).

Because validating optical remote sensing products implies the comparison of measurements characterized by very different spatial scales (point field measurements and pixel-scale integrations), possibly collected at different times of the day, these differing scales add an inherent uncertainty to validation exercises. Match-up analyses for merged products have their own specific issues, some of which are addressed below.

Merging products from different satellite passes ultimately relies on the combination of radiance data collected in different conditions, as well as the sensor-specific characteristics (wavelengths, spectral response, *etc.*). Ideally, the result of the atmospheric correction process, represented by the spectrum of water leaving radiance, should be independent of the conditions of geometry (illumination and observation), atmospheric content and water content (within certain bounds). In practice, it is difficult to analyze how differences in environmental conditions may impact the performance of the atmospheric correction scheme, and to what extent they account for the discrepancies observed between satellite products.

Fortunately, it seems that the aerosol diurnal variability is low for oceanic regions, on average ±10% departure from the daily value (Smirnov *et al.*, 2002; Ichoku *et al.*, 2005), but counter-examples can be found, particularly in coastal areas close to urban centers and/or influenced by coastal atmospheric dynamics (Léon *et al.*, 2001; Mélin and Zibordi, 2005) and formation processes in the illuminated tidal zone (a major source for new marine aerosol particles; O'Dowd, 2002). The temporal variability of the atmospheric component thus potentially induces a level of uncertainty on the final merged product. Additionally, these diurnal aspects would need to be taken into account for the merging of aerosol products, and they would be a disturbing element for merging schemes taking the top-of-atmosphere radiance as input (merging at Level-1).

## 6.2 Considerations of the Spatio-Temporal Scales and Variability of the Ocean-Colour Signal

Spatio-temporal variability also affects the marine component in different ways and needs to be examined when merging ocean-colour products resulting from measurements collected at different times with different resolutions. Level-2 images are clearly associated with a time. On the other hand, a Level-3 merged product can no longer be associated with a particular time; rather, it could be considered a representative value of the geophysical variable of interest over a certain period, typically the central part of the day (say from 9h00 to 15h00 local time). Field measurements are also made on different time frequencies (from single stations to automated high-frequency measurements). In the context of validation of merged products, the local scales of variability should be kept in mind to characterize both terms of the comparison.

Temporal variability in optical properties with time scales shorter than a day have been documented in various contexts, for laboratory phytoplankton (Stramski and Reynolds, 1993; Claustre *et al.*, 2002), and for actual communities, for instance in the equatorial Pacific Ocean (Binder and DuRand, 2002, and references therein; Neveux *et al.*, 2003), the Arabian Sea (Gardner *et al.*, 1999), and the Sargasso Sea (Wiggert *et al.*, 1999). Field measurements describing this high-frequency optical variability are often based on continuous attenuation measurements. More specifically, Gardner *et al.* (1999) measured particulate beam attenuation (at 660 nm) variations as high as 70% in the North Atlantic and the equatorial Pacific surface waters, a figure which is in the high range of those reported in literature. In North Pacific subtropical waters, Siegel *et al.* (1989) have shown peak-to-peak differences that reach 0.01 $m^{-1}$ for the beam attenuation coefficient at 600 nm, approximately 15% of the signal (pure seawater excluded). This variability can stem from cellular abundance and growth (and thus size) that are driven by photosynthesis during the day, as well as by dynamic processes, like diurnal mixing and horizontal advection. Such variability in the optical properties is more likely in frontal regions and coastal waters where coastal currents or tidal cycles and stirring play an important role (*e.g.*, Bowers *et al.*, 1998; Chang *et al.*, 2002). Admittedly, this variability is much reduced if the time interval separating near-noon satellite overpasses is considered, as in the case of MODIS and SeaWiFS. For instance, Zibordi *et al.* (2006) calculated variations of a few percentages for hourly triplets of water leaving radiance obtained from automated above-water radiometry. In any case, a part of the difference between products is real.

Simultaneously, the ocean-colour signal is modulated by spatial variability, the scales of which might be relevant for validating merged products. The availability of the ocean-colour global record has spurred recent studies of its mesoscale variability (Doney *et al.*, 2003; Uz and Yoder 2004). Conversely, the grid resolution of SeaWiFS global data has put a lower limit to the scales that can be pictured by statistical techniques, and there is evidence that finer scales could be highlighted (Denman *et al.*, 1977; Abbott *et al.*, 1995; Mahadevan and Campbell 2002) if appropriate gridded products were available. It is worth noting that merging ocean-colour products brings a more frequent sampling of the ocean; also, the possibility of safeguarding the finest resolution of the input products (Kwiatkowska, 2003b) is particularly enticing for the study of spatial variability. Therefore, the quality of the merged product should take into account its ability to capture spatio-temporal scales as far as possible.

Possible temporal and spatial variations associated with the comparison of remote sensing products and field measurements called for the adoption of protocols for selecting match-ups. Besides enforcing some flags documenting the output of the ocean-colour processor, validation studies usually added selection criteria based on time and spatial variability. Typically, the time difference between field measurements and satellite pass is restricted to ±3 or 4 hours (Mélin *et al.*, 2003; Carder *et al.*, 2004; Wang *et al.*, 2005) and might include a limit on the spatial variations as computed for a $3 \times 3$ or $5 \times 5$-pixel square (Bailey *et al.*, 2000; Hooker and McClain, 2000; Eplee *et al.*, 2001; Wang *et al.*, 2002). In the case of automated measurements, shorter time scales might be considered (Pinkerton *et al.*, 2003; Zibordi *et al.*, 2004). Global scale studies using large data bases have taken the day as the comparison window (Gregg and Casey, 2004).

These protocols for selecting the match-up pairs need to be adapted to the particular time scale represented by the merged product mentioned above. For instance, comparisons could be made between the satellite products and field values collected between the first and last time of the satellite passes with an extra time interval on both sides. It is also recommended that an indicator of spatial variability and the difference between sensor-specific products be traceable for each match-up. Field measurements covering the central part of the day at regular intervals are ideal (albeit rare) data sets for validating merged products, since they allow the comparison of mean as well as variability.

## 6.3 Other Validation Tools and Measure of Success

Merged products can also be evaluated by comparison with the products from single missions assuming the original data sources are validated and/or not strongly biased. The frequency distribution of the merged products should not depart significantly from that of the data sources (for the bins they have in common) and match-up statistical results for the merged product should not be worse than the worst of the single unmerged sources. Global and local averages and standard deviations should be very close in the merged and non-merged products and local estimates should clearly represent major ocean biogeochemical features, such as mid-ocean gyres, equatorial upwelling regions, and high latitude seasonal blooms. Merged products should not introduce spatial or temporal trends that cannot be attributed to an increased number of observations.

Transition between areas covered by different sets of data sources should be seamless. In other words, merged products should not show discontinuities caused by changes in the number of available data sources. Second order derivatives, gradient calculations or other techniques should be used to check for discontinuities. Directionality in the discontinuities can also be tracked as they may provide clues to the cause of the discontinuities

## 6.4 Metrics for the Improvements Resulting from Data Merging

Beside the validation of merged products, it is also important to quantify some of the benefits that result from data merging. Improvement in the sampling frequency can easily be demonstrated and should exist regardless of the merging method used. An objective metric should be proposed, such as the percentage of grid points covered by a season and region on a daily basis; this enhanced coverage improves the representation of ocean variability and the significance of a time average, and enables a better characterization of the time scales that can be studied. Increase in spatial coverage is another obvious improvement and can be easily calculated (for those techniques that "create" data the validation should discriminate between original and extended coverage).

Merging techniques may have to address the combination of products mapped with a different grid size. This problem can be ameliorated if all of the missions follow the recommendations of IOCCG Report No. 4 (IOCCG, 2004). The problem can be eliminated by use of a common grid, with the same dimensions in x, y, and time. This is a recommendation of the IOCCG Data Merging Working Group.

Beside the validation and the demonstration of the benefits of data merging, the ultimate measure of success will be determined by whether or not ocean-colour data users will use a unified merged product in preference to data from one or more individual missions.

# Chapter 7

# Merged Ocean-colour Products

## 7.1 Candidate Merged Ocean-colour Products

Candidate merged products are drawn from the list of Level-3 standard products produced by each of the global satellite sensors (Table 3.1) and that are common to two or more sensors. Chlorophyll and Level-3 water-leaving radiances at all common sensor wavelengths are primary candidates. Although differences in algorithms for all the products may exist among sensors, uniformity of methods is not required for data merging. As long as the differences in methodology, performance, and expected application are known, data merging methods can be developed to accommodate them. An example might be a turbid water chlorophyll algorithm proposed for MERIS (Moore *et al.*, 1999), which may be of great value in coastal waters. Such an algorithm may not have an analog for the other sensors, because of lack of required bands. It is possible that a clever merging algorithm would select the turbid water chlorophyll algorithm in areas where it is most needed and combine the chlorophyll products elsewhere. Possible products for merging are listed below.

1. Chlorophyll is the most widely used product derived from remotely-sensed data and despite the sensor differences, is the primary candidate for a merged-product.
2. Level-3 normalized water leaving radiances are the primary physical remotely sensed products for natural waters. Level-3 radiances are primary candidates for merging. Differences in center wavelengths (see Table 3.1) and bandwidths represent formidable challenges for data merging. But their role as the basic inputs to bio-optical algorithms and the potential information inherent in them (although not yet fully exploited) make them important candidates for data merging.
3. Photosynthetically available radiation (PAR) is routinely produced by Sea-WiFS and is planned for production by GLI. PAR derived from polar-orbiting satellites is subject to diurnal variability in clouds, and a single sensor will only observe at a single time at a given location. Multiple observations from satellites in different orbits will lend a measure of diurnal variability, potentially resulting

in an improved product.

Other common products that are not standard products, such as Inherent and Apparent Optical Properties (IOP and AOP, respectively) have great potential for merging. Recent algorithms (*e.g.*, Maritorena and Siegel, 2005) using semi-analytical bio-optical techniques (see Chapter 4) provide a means of deriving IOP's from remotely-sensed radiances determined from spaced-based sensors, including absorption and backscattering coefficients. Using data from multiple missions with non-overlapping spectral bands is of great use in such studies, which require as much spectral information as possible.

## 7.2 Note on Aerosol Products

The ocean-colour processing chains potentially provide the spectrum of aerosol optical thickness $\tau_a(\lambda)$. Operationally, this has been represented by $\tau_a$ at one near-infrared wavelength (around 865 nm) and the Ångström exponent between this wavelength and the mid-visible (around 500 nm). Merging techniques could be applied to these aerosol distributions as for other geophysical products derived from ocean colour. Even though there can be an excellent agreement with simultaneous ground optical measurements, the $\tau_a(\lambda)$ distribution as a by-product of the atmospheric correction of ocean-colour data suffers from some limitations, such as excluding areas of high aerosol load. Moreover, there is a large diversity of processing chains specifically dedicated to deriving the aerosol load and type over land and ocean (see King *et al.*, 1999 for a review). Some processing chains make use of spectral bands out of the range associated with ocean colour (*e.g.*, TOMS with the ultraviolet or MODIS with the infrared), or they make use of characteristics specific to particular sensors (multi-directionality and/or polarization, in the cases of ATSR-2, MISR or POLDER) and, more generally, they include platforms that are not related with ocean colour (*e.g.*, AVHRR, METEOSAT, TOMS). These elements thus indicate that the topic of merging aerosol remote-sensing products can not be adequately covered by this report and should be addressed in a wider forum.

## 7.3 What are the Minimum Requirements for Data Product Quality before Merging can Begin?

It is necessary to set requirements before beginning merging implementation and analysis to restrict efforts. Furthermore, all of the statistical methodologies surveyed are error-correcting in nature, and many are bias-correcting. For example, binning, averaging, objective analysis, and optimal interpolation can reduce random errors but not biases. Many of the other statistical methodologies have a capability for bias correction. The blended analysis is well known for its bias-correction capability, as are machine-learning methods, and the error-weighted averaging methodology could be extended to incorporate biases as well. Knowledge of data and model errors and biases is a critical component of data assimilation into numerical models, and different assimilation methodologies can adapt in different ways to the circumstances. Use of the blended analysis for assimilation, as surveyed here, is an explicit bias-correcting scheme, and the examples provided here also accounted for some data errors and biases. The bio-optical methods inherently contain a measure of error-correction when handling multiple radiances in a best-fit to a semi-analytic algorithm. Additional correction can be obtained by data weighting upon input. In all of these cases, knowledge of the sensor data errors, biases, and distribution of the errors/biases is extremely important.

Thus decisions about when to merge data sets are best left to the data merging scientist who, given knowledge about errors and biases, can choose and adapt methods to best account for and accommodate them. The IOCCG Data Merging Working Group recommends analysis of merging results in a consistent and comprehensive manner, as outlined in Chapter 6 of this report, and the decisions on individual sensor data quality be evaluated by the data merging scientist's choice of method and the final quantitative results. In fact, the panel encourages innovative research into merging methodologies in pursuit of merged data sets with improved accuracy and reduced uncertainty.

## 7.4 Diurnal Variability

Merging of data products as discussed in this report assumes the temporal overlap of observations. Temporal overlap is assumed to be at daily frequency. This enables maximization of benefits (*i.e.*, improved daily coverage and improved number of simultaneous observations) and utilization of daily Level-3 sensor data. The missions considered for merging here (Table 3.1) are in very similar orbits, with equatorial crossing times at or near noon. The maximum difference in local time of observations is two hours or less for all pairs at all locations, except for MODIS-Aqua.

MODIS-Aqua is different in that its 13h30 ascending node orbit provides observations often at substantially different times of day than the other missions, all of which have noon (SeaWiFS) or morning descending orbits. The maximum difference in time of day observations occurs between MODIS-Aqua and MERIS. Here observation time differences range from about 4 hours at 60°N, to 7 hours at 60°S, reaching a maximum of $>$ 15 hours at the northern polar extreme (Gregg *et al.*, 1998). SeaWiFS and MODIS-Aqua have smaller extremes, but represent the next most divergent case.

It is not clear how much diurnal variability in ocean chlorophyll and other optical properties exists in the natural ocean, and where and when it is an issue. There is evidence that it can be substantial, at least under certain conditions (see Chapter 6).

Production of merged data sets using the descending node missions in Table 3.1, that is, all but MODIS-Aqua, is likely to be independent of variability associated with diurnal patterns, because of the small range of local observation times. However, use of MODIS-Aqua data may introduce variability derived from sub-daily processes and needs to be considered when merging data. We recommend research into the potential impact of diurnal variability and different observation times from different satellite missions.

# Chapter 8

# What is Needed to go Forward?

## 8.1 Data Access and Sensor Knowledge

The prerequisite for ocean-colour data merger is access to multi-sensor data themselves. SeaWiFS and MODIS ocean-colour data are easily accessed through the NASA Ocean Color Web interface (http://oceancolor.gsfc.nasa.gov), free of charge and without restriction for scientific purposes. MERIS data is also free of charge and without restriction, and can be obtained, after registration, through the ESA Principal Investigator Portal (http://eopi.esa.int/). Data is available in NRT via an FTP rolling archive, via a web interface for archived data, and via delivery on physical media. POLDER data is also available via the PARASOL-POLDER products distribution center (http://parasol-polder.cnes.fr/). It cannot be overemphasized that access to remote-sensing mission data is critical.

We have shown in earlier chapters of this report that much more than satellite data is required for successful merging. *In situ* data are necessary to validate the end products. As pointed out in Chapter 6, this requirement is even more difficult for merged data than for individual data sets, because *in situ* data at mission data spatial overlaps are required. There is a role for the IOCCG to play here, by coordinating and providing links to *in situ* data sets maintained by international space agencies and other sources. Such a webpage has already been established by the IOCCG (see http://www.ioccg.org/data/insitu.html), where links to 11 international data sets are listed. This is an excellent start, but more needs to be done. Oceanographic centers and researchers are therefore encouraged to help update this webpage by submitting links to valid *in situ* data sets.

Going beyond data, detailed information about the ocean-colour sensor characteristics and performance is necessary to promote successful data merging. A description of knowledge requirements for merging is provided in Chapter 5 of this report, and detailed information of various international ocean-colour sensors is provided on the IOCCG website at http://www.ioccg.org/sensors_ioccg.html.

## 8.2 Data Set Stability

Frequent data reprocessing events are a major impediment to successful data merging. Most merging methods require knowledge of sensor data performance, both absolute and relative. Given this knowledge, methods and adaptations can be developed and evaluated to maximize data set quality in the merged products. If the data set quality is constantly changing, data merging efforts cannot proceed beyond hypothetical analyses. Results are always provisional in such an environment.

Unfortunately, from the data merging perspective, ocean-colour mission reprocessing events are all too frequent. The entire SeaWiFS archive was processed 5 times in its first 7 years of existence. MODIS-Terra was processed 4 times in 4 years, and MODIS-Aqua 3 times in 2.5 years. GLI has been processed twice in 1.5 years and MERIS is scheduled for its third processing in 2006, in only its third active year. Each of these complete archival processing events has produced a major improvement in data quality, but with new and radically different errors and biases and error distributions. Each one was necessary to improve or correct problems in the previously available data set.

Each processing event presents a challenge to a data merging scientist. It is difficult enough to keep track of an individual sensor, but when contemplating several data sources the problems multiply. The probability at any given moment that one of several sensors has recently undergone, or is currently in the process of, or planning, a major processing event, is nearly 100%. The frustration for data merging is immense. Methods that may work under one set of archived versions may fail in a new one. Hopefully the process is one for which less radical error-correction methods are needed for merging because of improvements and greater consistency in the individual missions, but there is no experience to demonstrate this.

Unfortunately, there is no simple solution to this problem. An archive freeze will solve the problem but it introduces new problems, such as allowing poor quality data to exist longer than necessary. The IOCCG Working Group on Data Merging cannot provide a solution to the problem, although it can provide important information on identification of the problem, and its impact on data merging. The seriousness of the data set stability problem is enormous on data merging, but it is every bit as large on other areas of scientific pursuit, particularly trend analysis. This is another issue that could be addressed by an IOCCG working group.

## 8.3 More Research in Ocean-Colour Data Merging

More research into merging schemes will have to be undertaken, starting from existing attempts and leading to other unexplored approaches. A number of ocean-colour merger techniques have been proposed in Chapter 4. Many of these alternative methods address different problems concerned with merging multi-sensor, multi-year, multi-spectral, and multi-resolution data of varying calibration/ validation accuracy. As the merger techniques are implemented and tested, combinations of the methods can also be applied. Different merger algorithms should ideally be executed and optimized within the same processing environment so that detailed inter-comparisons and validations are fully correspondent. The research on merger methods will have to be accompanied by development of a comprehensive operational data processing, validation, and inter-comparison environment for quick testing of merger algorithm performance and subsequent algorithm modifications. Inter-comparison of merging techniques will certainly bring significant added knowledge, foster much thinking, and lead to gradual improvements in the ocean-colour merger technology. One of the first questions would be how to inter-compare the performance of merging schemes that can work very differently and even merge different products, such as chlorophyll and $L_{wN}$ (see Chapter 6). Another question would be the definition of the datasets for the effective and comprehensive merger algorithm inter-comparisons. The final goal is operational distribution of merged ocean-colour time series together with relevant ancillary information and documentation. One or a couple of merging methods might be selected, but the door should be left open to other algorithms, which will prompt further research on the topic.

## 8.4 Merging Method Intercomparison

Although more research is required to address data merging methods, there is also a need to inter-compare results. Given the international character of ocean-colour data merging efforts, it is conceivable that the IOCCG could coordinate such intercomparisons by soliciting interested investigators and providing notice and leadership for meetings. The intercomparison participants should agree in advance about the rules of the process, relying heavily on the metrics for success described in Chapter 6 of this report. A data set version freeze can be established to avoid the problems discussed above with respect to data set stability. Finally, a timeline for proceeding should be set with the agreement of all participants.

# Chapter 9

# Conclusions and Recommendations

The objective of this report is to chart the course leading to high quality, public archives of merged ocean-colour data. A detailed list of recommendations, with reference to the Chapters from which they were derived, is provided below.

## 9.1 Level-3 Data Access

Level-3 daily data should be produced by all ocean-colour missions, since this is the data set most frequently used by ocean-colour merging activities. The Level-3 daily data should be in standard format (preferably identical), within common dimensions and grid, and follow the recommendations of IOCCG Report No. 4 (IOCCG, 2004) (see Chapters 3 and 5).

## 9.2 *In Situ* Data Access

*In situ* data are the primary source for validation of merged data. Merged data *in situ* validation requirements are more stringent than individual mission data, because data are needed at spatial overlap points for merged data (see Chapter 6). Links to international data sets are provided on the IOCCG website and will be updated and expanded as and when new information is provided. Links to 11 data sets are currently available (see Chapter 8) and the IOCCG Data Merging Working Group encourages all interested researchers to submit information about new data sets to the IOCCG webmaster in a timely fashion.

## 9.3 Merge Common Products

Merging efforts should emphasize chlorophyll, the primary derived geophysical product of ocean-colour sensors. There is also interest in merging other common Level-3 products among two or more sensors, including, but not limited to, normalized water-leaving radiances, PAR, suspended sediments and CDOM. (see Chapter 7).

## 9.4 Knowledge of Data Performance and Sensor Characteristics

Knowledge of ocean-colour mission data is more important than quality. All the merging methodologies evaluated have some level of error-correction capability, and some have a bias-correction capability. Knowledge of sensor characteristics and data performance is required to promote research into data merging and to help identify the best methods. Such knowledge includes, but is not limited to:

- Level-3 products, definitions, coverage, time spans, and
- mission comparison with *in situ* data, including RMS and bias (log-transformed base-10 for chlorophyll, pigment, absorption).

There are no minimum requirements on data set quality before merging research begins, just knowledge (as defined in Chapter 5).

## 9.5 Comprehensive Merged Data Set Evaluation Criteria

A thorough, comprehensive set of merger evaluation criteria are needed to evaluate merged products in a consistent and objective fashion. These include:

- A requirement for *in situ* comparison, especially at satellite data overlaps, including RMS, $r^2$, mean ratio (and bias).
- Frequency distribution with specified thresholds.
- Global and local averages and standard deviations with specified but qualitative thresholds.
- Local estimates that clearly represent major ocean biogeochemical features, such as mid-ocean gyres, equatorial upwelling regions, and high latitude seasonal blooms.
- Merged products should not show discontinuities caused by changes in the number of available data sources.
- No discontinuities using visual inspection, second order derivatives and gradients.
- Improvements in sampling frequency should be quantified including percent grid points covered by season and region daily, and increase in spatial coverage
- Users' preference

(see Chapter 6).

## 9.6 Method Intercomparison (Round Robin)

The best way to reach consensus on merging methods is via a sponsored, supervised, inter-comparison effort. An inter-comparison exercise requires agreement on validation methods, which should be derived from the recommendations of this report, as well as a common *in situ* data set, which should be assembled in advance.

Considering that a comprehensive merged archive requires data from international missions, it is probable that the IOCCG could sponsor and supervise such an inter-comparison exercise (see Chapter 8).

## 9.7 IOCCG Working Group on Data Set Stability

Since knowledge of ocean-colour data set quality and performance is necessary for evaluation of merging methods, it is necessary to have stable data sets. Constant reprocessing events cripple the ability of merging scientists to reach consensus on methodologies. Yet reprocessing is essential to improve data set quality. These conflicting interests severely impact the production of merged archives and yet clearly provide improved individual mission data sets. The IOCCG Data Merging Working Group is unable to define how to balance these competing interests, which should be assessed by a new working group comprised of members of the data user community, the mission data community, as well as the merging community (see Chapter 8).

## 9.8 Temporal Frequency and Spatial Resolution of Merged Data Set

Monthly one-degree data sets can be achieved with nearly full coverage by any one of the global ocean-colour missions. Merging data, however, improves coverage for shorter temporal frequencies. It also provides the potential for using the multiple observations for filling data gaps in high spatial resolution Level-3 data sets. At its best, merged data products serve both objectives in time and space if the highest possible resolutions are sought. At current mission capabilities, this means daily 1-km data. Spatial resolution can, however, be backed off to 9-km for more complete coverage, but the daily requirement still holds (see Chapter 1).

## 9.9 Source Data Defined

A new requirement for merged data sets is that source data must be defined (sensor and processing version) and that some means of identifying the source on a Level-3 grid basis, point-by-point, be created and distributed. This can be done using flags, quality levels, separate data sets or any other means that satisfies the requirement.

# References


Abbott, M.R., Brink, K.H., Booth, C.R., Blasco, D., Swenson, M.S., Davis, C.O. and Codispoti, L.A. (1995). Scales of variability of bio-optical properties as observed from near-surface drifters. *Journal of Geophysical Research* 100: 13,345-13,367.

Antoine, D., Morel, A., Gordon, H.R., Banzon, V.F. and Evans, R.H. (2005). Bridging ocean color observations of the 1980s and 2000s in search of long-term trends. *Journal of Geophysical Research* 110: C06009, doi: 101029/ 2004JC002620.

Bailey, S.W., McClain, C.R., Werdell, P.J. and Schieber, B.D. (2000). Normalized water- leaving radiance and chlorophyll a match-up analyses. *NASA Technical Memorandum 206892*, SeaWiFS Postlaunch Technical Report Series, Eds S.B. Hooker, E.R. Firestone, vol. 10, ch. 7, pp.45-52.

Binder B.J. and DuRand, M.D. (2002). Diel cycles in surface waters of the equatorial Pacific. *Deep-Sea Research I*, 49: 2601-2617.

Bowers, D.G., Boudjelas, S. and Harker, E.L. (1998). The distribution of fine suspended sediments in the surface waters of the Irish Sea and its relation to tidal stirring. *International Journal of Remote Sensing* 19: 2789-2805.

Bretherton, F., Davis, R. and Fandry, C. (1976). A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Research* 23: 559-582.

Bricaud, A., Morel, A. and Prieur, L. (1981). Absorption by dissolved organic-matter of the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, 26: 43-53.

Campbell, J.W. (1995). The Lognormal Distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research* 100(C7): 13,237-13,254.

Carder K.L., Chen, F.R., Cannizzaro, J.P., Campbell, J.W. and Mitchell, B.G. (2004). Performance of the MODIS semi-analytical ocean color algorithm for chlorophyll-a. *Advances in Space Research* 33: 1152-1159.

Chang, G.C., Dickey, T.D., Schofield, O.M., Weidemann, A.D., Boss, E., Pegau, W.S., Moline, M.A. and Glen, S.M. (2002). Nearshore physical processes and bio-optical properties in the New York Bight. *Journal of Geophysical Research* 107: 3133, 10.1029/2001JC001018.

Claustre, H., Bricaud, A., Babin, M., Bruyant, F., Guillou, L., Le Gall, F., Marie, D. and Partensky, F. (2002). Diel variations in Prochlorococcus optical properties. *Limnology and Oceanography* 47: 1637-1647.

Claustre, H. and Maritorena, S. (2003). The many shades of ocean blue. *Science* 302: 1514-1515.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods.* Cambridge University Press, Cambridge, UK, pp. 204.

Daley, R. (1991). *Atmospheric Data Analysis.* Cambridge Univ. Press, Cambridge, UK, pp. 457.

Denman, K., Okubo, A. and Platt, T. (1977). The chlorophyll fluctuation spectrum in the sea. *Limnology and Oceanography*, 22: 1033-1038.

Doney, S.C., Najjar, R.M. and Stewart, S.E. (1995). Photochemistry, mixing, and diurnal cycles in the upper ocean. *Journal of Marine Research* 53: 341-369.

Doney, S.C., Glover, D.M., McCue, S.J. and Fuentes, M. (2003). Mesoscale variability of Sea-viewing Wide Field-of-view Sensor (SeaWiFS) satellite ocean color: Global patterns and spatial scales. *Journal of Geophysical Research* 108: 3024, 10.1029/2001JC000843.

Eplee, R.E., Robinson, W.D., Bailey, S.W., Clark, D.K., Werdell, P.J., Wang, M., Barnes, R.A. and McClain, C.R. (2001). Calibration of SeaWiFS. II. Vicarious techniques. *Applied Optics* 40: 6701-6718.

Eppley, R.W. (1972). Temperature and phytoplankton growth in the sea. *Fisheries Bulletin* 70: 1063-1085.

Field, C.B., Behrenfeld, M.J., Randerson, J.T. and Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281: 237-240.

Franz, B. A. and Kim, Y. (2001). A comparative study and intercalibration between OSMI and SeaWiFS. *Eos Trans. American Geophysical Union*, Fall Meeting, San Francisco, USA, 82(47): F661 (OS42D-09).

Gardner, W.D., Gundersen, J.S., Richardson, M.J. and Walsh, I.D. (1999). The role of seasonal and diel changes in mixed-layer depth on carbon and chlorophyll distributions in the Arabian Sea. *Deep-Sea Research I*, 46: 1833-1858.

Garver, S.A. and Siegel, D.A (1997). Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation .1. Time series from the Sargasso Sea. *Journal of Geophysical Research* 102: 18,607-18,625.

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J.W., Smith, R.C., Baker, K.S. and Clark, D.K. (1988). A semianalytic radiance model of ocean color. *Journal of Geophysical Research* 93: 10,909-10,924.

Gregg, W.W. and Woodward, R.H. (1998). Improvements in high frequency ocean color observations: Combining data from SeaWiFS and MODIS. *IEEE Transactions on Geoscience and Remote Sensing* 36: 1350-1353.

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

Gregg, W.W., Conkright, M.E., O'Reilly, J.E., Patt, F.S., Wang, M., Yoder, J.A. and Casey, N.W. (2002). NOAA-NASA Coastal Zone Color Scanner reanalysis effort. *Applied Optics* 41: 1615-1628.

Gregg, W.W., Ginoux, P., Schopf, P.S., and Casey, N.W. (2003). Phytoplankton and Iron: Validation of a global three-dimensional ocean biogeochemical model. *Deep-Sea Research* II 50: 3143-3169.

Gregg, W.W. and Casey, N.W. (2004). Global and regional evaluation of the SeaWiFS chlorophyll data set. *Remote Sensing of Environment* 93: 463-479.

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

Ichoku, C., Remer, L.A., Eck, T.F. (2005). Quantitative evaluation and intercomparison of morning and afternoon Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol measurements from Terra and Aqua. *Journal of Geophysical Research* 110: D10S03, 10.1029/2004JD004987.

IOCCG (1999). *Status and Plans for Satellite Ocean-Colour Missions: Considerations for Complementary Missions.* Yoder, J.A. (ed.). Reports of the International Ocean- Colour Coordinating Group, No. 2, IOCCG, Dartmouth, Canada,. 43 pp.

IOCCG (2004). *Guide to the Creation and Use of Ocean-Colour, Level-3, Binned Data Products.* Antoine, D. (ed.). Reports of the International Ocean-Colour Coordinating Group, No. 4, IOCCG, Dartmouth, Canada, 88 pp.

Julian, P.R. and Thiebaux, H.J. (1975). On some properties of correlation functions used in optimum interpolation schemes. *Monthly Weather Review* 103: 605-616.

King, M.D., Kaufman, Y.J., Tanré, D. and Nakajima, T. (1999). Remote sensing of tropospheric aerosols from space: past, present and future. *Bulletin of the American Meteorological Society* 80: 2229-2259.

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(47): F675 (OS52A-0514).

Kwiatkowska, E.J. and Fargion, G.S. ( 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 Fargion, G.S. (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, 23-27 October 2002, Hangzhou, China.* Ocean Remote Sensing and Applications 4892: 168-182.

Kwiatkowska, E.J. and Fargion, G.S. (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 Transactions on Geoscience and Remote Sensing* 41: 2844-2860.

Kwiatkowska, E.J. (2003a). Cross-calibration of multi-sensor and multi-year datasets to create a consistent and calibrated global ocean color baseline. In: *MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003.* Eds, Fargion, G.S. and McClain C.R., NASA Technical Memorandum, Goddard Space Flight Center, Greenbelt, Maryland, pp. 51-61.

Kwiatkowska, E.J. (2003b). Local area application of data merger: enhancement of oceanic features in lower resolution imagery using higher resolution data. In: *MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003.* Eds, Fargion, G.S. and McClain, C.R., NASA Technical Memorandum, Goddard Space Flight Center, Greenbelt, Maryland, pp. 66-69.

Kwiatkowska, E.J. (2003c). Statistical objective analysis; spatial and temporal interpolation of multi-sensor ocean color data onto daily global binned coverage. In: *MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003.* Eds, Fargion, G.S. and McClain, C.R., NASA Technical Memorandum, Goddard Space Flight Center, Greenbelt, Maryland, pp. 61-66.

Le Traon, P.Y., Nadal, F. and Ducet, N. (1998). An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technologies* 15: 522-534.

Le Traon, P.Y. and Ogor, F. (1998). ERS-1/2 orbit improvement using TOPEX/ POSEIDON: The 2 cm challenge. *Journal of Geophysical Research* 103: 8045-8057.

Léon, J.-F., Chazette, P., Dulac, F., Pelon, J., Flamant, C., Bonazzola, M., Foret, G., Alfaro, S.C., Cachier, H., Cautenet, S., Hamonou, E., Gaudichet, A., Gomes, L., Rajot, J.-L., Lavenu, F., Inamdar, S.R., Sarode, P.R. and Kadadevarmath, J.S. (2001). Large scale advection of continental aerosols during INDOEX. *Journal of Geophysical Research* 106: 28,427-28,439.

Mahadevan, A. and Campbell, J.W. (2002). Biogeochemical patchiness at the sea surface. *Geophysical Research Letters* 29: 1926, 10.1029/2001GL014116.

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 11(7): 674-693.

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

Maritorena, S. and Siegel, D.A. (2005). Consistent merging of satellite ocean color data sets using a bio-optical model. *Remote Sensing of Environment* 94: 429-440.

Mélin, F., Zibordi, G. and Berthon, J.-F. (2003). Assessment of SeaWiFS atmospheric and marine products for the Northern Adriatic Sea. *IEEE Transactions on Geosciences Remote Sensing* 41: 548-558.

Mélin, F. and Zibordi, G. (2005). Aerosol variability in the Po Valley analyzed from automated optical measurements. *Geophysical Research Letters* 32: L03810, 10.1029/2004GL021787.

Moore, G.F., Aiken, J., and Lavender, S.J. (1999). The atmospheric correction of water colour and the quantitative retrieval of suspended particulate matter in Case II waters: application to MERIS. *International Journal of Remote Sensing* 20: 1713-1733.

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

Neveux, J., Dupouy, C., Blanchot, J., Le Bouteiller, A., Landry, M.R. and Brown, S.L. (2003). Diel dynamics of chlorophylls in high-nutrient, low-chlorophyll waters of the equatorial Pacific (180°): Interactions of growth, grazing, physiological responses, and mixing. *Journal of Geophysical Research* 108: 8140, 10.1029/2000JC000747.

Núñez, J., Otazu, X., Fors, O., Prades, A., Palà, V. and Arbiol, R. (1999). Multi-resolution-based image fusion with additive wavelet decomposition. *IEEE Transactions Geoscience and Remote Sensing* 37(3): 1204-1211.

O'Dowd, C.D. (2002). On the spatial extent and evolution of coastal aerosol plumes. *Journal of Geophysical Research* 107: 8105, 10.1029/2001JD000422.

Oort, A.H.(1983). Global atmospheric circulation statistics, 1958-1973. NOAA Professional Paper 14, 180 pp.

Pinkerton, M.H., Lavender, S.J. and Aiken, J. (2003). Validation of SeaWiFS ocean color satellite data using a moored databuoy. *Journal of Geophysical Research* 108: 3133, doi:10.1029/2002JC001337.

Pottier, C., Garçon, V., Larnicol, G., Sudre, J., Schaeffer, P. and Le Traon, P.-Y. (2006). Merging SeaWiFS and MODIS/Aqua ocean color data in North and Equatorial Atlantic using weighted averaging and objective analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44 (11): 3436-3451. doi:10.1109/TGRS.2006.878441.

Reynolds, R.W. (1988). A real-time global sea surface temperature analysis. *Journal of Climate* 1:75-86.

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

Rossow, W.B. and Schiffer, R.A. (1991). ISCCP cloud data products. *Bulletin of the American Meteorological Society* 72: 2-20.

Siegel, D.A., Dickey, T.D., Washburn, L., Hamilton, M.K. and Mitchell, B.G. (1989). Optical determination of particulate abundance and production variations in the oligotrophic ocean. *Deep-Sea Research* I 36: 211-222.

Smirnov, A., Holben, B.N., Kaufman, Y.J., Dubovik, O., Eck, T.F., Slutsker, I., Pietras, C. and Halthore, R.N. (2002). Optical properties of atmospheric aerosol in maritime environments. *Journal of Atmospheric Sciences* 59: 501-523.

Stramski, D. and Reynolds, R.A. (1993). Diel variations in the optical properties of a marine diatom. *Limnology and Oceanography* 38: 1347-1364.

Thiebaux, H.J. and Pedder, M.A. (1987). *Spatial Objective Analysis with Applications in Atmospheric Sciences.* Academic Press, New York.

Uz, B.M. and Yoder, J.A. (2004). High frequency and mesoscale variability in SeaWiFS chlorophyll imagery and its relation to other remotely sensed oceanographic variables. *Deep-Sea Research* II 51: 1001-1017.

Wang, M. and Franz, B.A. (2000). Comparing the ocean color measurements between MOS and SeaWiFS: A vicarious intercalibration approach for MOS. *IEEE Transactions on Geosciences Remote Sensing* 38(1): 184-197.

Wang, M., Isaacman, A., Franz, B.A. and McClain, C.R. (2002). Ocean color optical property data derived 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: 974-990.

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

Wiggert, J.D., Granata, T.C., Dickey, T.D. and Marra, J. (1999). A seasonal succession of physical/biological interaction mechanisms in the Sargasso Sea. *Journal of Marine Research* 57: 933-966.

Zibordi, G., Mélin, F., Hooker, S.B., D'Alimonte, D. and Holben, B.N. (2004). An autonomous above-water system for the validation of ocean color radiance data. *IEEE Transactions on Geosciences Remote Sensing* 42: 401-415.

Zibordi, G., Mélin, F. and Berthon, J.-F. (2006). A time series of above-water radiometric measurements for coastal water monitoring and remote sensing product validation. *IEEE Geoscience and Remote Sensing Letters* 3(1): 120-124, 10.1109/LGRS.2005.858486, 2006.

# Acronyms

| | |
|---|---|
| ADEOS | Advanced Earth Observing Satellite |
| AOP | Apparent Optical Property |
| ATSR | Along Track Scanning Radiometer |
| AVHRR | Advanced Very High Resolution Radiometer |
| BRDF | Bidirectional Reflectance Distribution Function |
| CDOM | Chromophoric Dissolved Organic Matter |
| CPU | Central Processing Unit |
| EOS | Earth Observing System |
| Envisat | Environmental Satellite |
| GAC | Global Area Coverage |
| GLI | Global Imager |
| GMES | Global Monitoring for Environment and Security |
| IBC | Internal Boundary Condition |
| IOP | Inherent Optical Property |
| LAC | Local Area Coverage |
| $L_{\mathrm{wN}}$ | Normalized water-leaving radiance |
| MERIS | Medium Resolution Imaging Spectrometer |
| METEOSAT | Meteorology Satellite |
| MISR | Multiangle Imaging SpectroRadiometer |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| NPP | NPOESS Preparatory Project |
| NPOESS | National Polar-orbiting Operational Environmental Satellite System |
| NRT | Near Real Time |
| OI | Optimal Interpolation |
| PAR | Photosynthetically Available Radiation |
| POLDER | Polarization and Directionality of the Earth's Reflectances |
| RMS | Root Mean Square |
| SeaWiFS | Sea-viewing Wide Field-of-view Sensor |
| TOA | Top of Atmosphere |
| TOMS | Total Ozone Mapping Spectrometer |

# Appendix

## Current ocean-colour Sensors

| SENSOR | AGENCY | SATELLITE | LAUNCH DATE | SWATH (km) | RESOLUTION (m) | NO. OF BANDS | SPECTRAL COVERAGE (nm) | ORBIT |
|---|---|---|---|---|---|---|---|---|
| MERIS | ESA (Europe) | ENVISAT (Europe) | 01/03/02 | 1150 | 300/1200 | 15 | 412-1050 | Polar |
| MMRS | CONAE (Argentina) | SAC-C (Argentina) | 21/11/00 | 360 | 175 | 5 | 480-1700 | Polar |
| MODIS-Aqua | NASA (USA) | Aqua (EOS-PM1) | 04/05/02 | 2330 | 1000 | 36 | 405-14,385 | Polar |
| MODES-Terra | NASA (USA) | Terra (USA) | 18/12/99 | 2330 | 1000 | 36 | 405-14,385 | Polar |
| OCM | ISRO (India) | IRS-P4 (India) | 26/05/99 | 1420 | 350 | 8 | 402-885 | Polar |
| OSMI | KARI (Korea) | KOMPSAT (Korea) | 20/12/99 | 800 | 850 | 6 | 400-900 | Polar |
| PARASOL | CNES (France) | Myriade Series | 18/12/04 | 2100 | 6000 | 9 | 443-1020 | Polar |
| SeaWIFS | NASA (USA) | OrbView-2 (USA) | 01/08/97 | 2806 | 1100 | 8 | 402-885 | Polar |

## Scheduled ocean-colour Sensors

| SENSOR | AGENCY | SATELLITE | SCHEDULED LAUNCH | SWATH (km) | RESOLUTION (m) | NO.OF BANDS | SPECTRAL COVERAGE (nm) | ORBIT |
|---|---|---|---|---|---|---|---|---|
| GOCI | KARI/ KORDI | COMS-1 (Korea) | 2008 | 2500 | 500 | 8 | 400-865 | Geo-stationary |
| HES-CW | NOAA/ NESDIS | GOES-R (USA) | 2012 | 400 | 30-300 | 14 | 412-900 | Geo-stationary |
| OCM-II | ISRO (India) | IRS-P7 (India) | 2007 | 1400 | 1-4 km | 8 | 400-900 | Polar |
| OLCI | ESA (Europe) | GMES Sentinel-3 | 2012 | 1120 | < 300 | 15 | 400-900 | Polar |
| S-GLI | JAXA (Japan) | GCOM-C (Japan) | 2011 | 1150 | 250/1000 | 16 | 375-12,500 | Polar |
| VIIRS | NASA/ IPO | NPP | 2008 | 3000 | 370/740 | 22 | 402-11,800 | Polar |
| VIIRS | NASA/ IPO | NPOESS | 2012 | 3000 | 370/740 | 22 | 402-11,800 | Polar |

## Historical ocean-colour Sensors

| SENSOR | AGENCY | SATELLITE | OPERATING DATES | SWATH (km) | RESOLUTION (m) | NO. OF BANDS | SPECTRAL COVERAGE (nm) | ORBIT |
|---|---|---|---|---|---|---|---|---|
| CMODIS | CNSA (China) | Shen Zhou-3 (China) | 25/03/02 - 15/9/02 | 650-700 | 400 | 34 | 403-12,500 | Polar |
| COCTS | CNSA (China) | Hai-Yang-1 (China) | 15/05/02 - 1/04/04 | 1400 | 1100 | 10 | 402-12,500 | Polar |
| CZCS | NASA (USA) | Nimbus-7 (USA) | 24/10/78 – 22/06/86 | 1556 | 825 | 6 | 433-12,500 | Polar |
| CZI | CNSA (China) | Hai Yang-1 (China) | 15/05/02 – 1/12/03 | 500 | 250 | 4 | 420-890 | Polar |
| GLI | NASDA (Japan) | ADEOS-II (Japan) | 14/12/02 – 25/10/03 | 1600 | 250/1000 | 36 | 375-12,500 | Polar |
| MOS | DLR (Germany) | IRS P3 (India) | 21/03/96 – 31/5/2004 | 200 | 500 | 18 | 408-1600 | Polar |
| OCI | NEC (Japan) | ROCSAT-1 (Taiwan) | 27/01/99 – 16/6/04 | 690 | 825 | 6 | 433-12,500 | Polar |
| OCTS | NASDA (Japan) | ADEOS (Japan) | 17/08/96 – 1/07/97 | 1400 | 700 | 12 | 402-12,500 | Polar |
| POLDER-2 | CNES (France) | ADEOS-II (Japan) | 14/12/02 – 25/10/03 | 2400 | 6000 | 9 | 443-910 | Polar |
| POLDER | CNES (France) | ADEOS (Japan) | 17/08/96 – 1/07/97 | 2400 | 6000 | 9 | 443-910 | Polar |