

Consistent merging of satellite ocean color data sets using a bio-optical model

Stéphane Maritorena^{a,*}, David A. Siegel^{a,b}

^a*Institute for Computational Earth System Science, University of California, Santa Barbara, Santa Barbara, CA 93106-3060, USA*

^b*Department of Geography, University of California, Santa Barbara, Santa Barbara, CA 93106-3060, USA*

Received 19 April 2004; received in revised form 26 August 2004; accepted 29 August 2004

Abstract

While many (and more on the way) ocean color satellite sensors presently provide routine observations of ocean biological processes, limited concrete effort has taken place to demonstrate how these data can be used together in any systematic way. One obvious way is to merge these data streams together to provide robust merged climate data records with measurable uncertainty bounds. Here, we present and implement a formalism for merging global satellite ocean color data streams to produce uniform data products. Normalized water-leaving radiances ($L_{wN}(\lambda)$) from SeaWiFS and MODIS are used together in a semianalytical ocean color merging model to produce global retrievals of 3 biogeochemically relevant variables (chlorophyll, combined dissolved and detrital absorption coefficient, particulate backscattering coefficient). The model-based merging approach has various benefits over techniques that blend end products, such as chlorophyll concentrations; (1) merging at the level of water-leaving radiance ensures simultaneity and consistency of the retrievals, (2) it works with single or multiple data sources regardless of their specific bands, (3) it exploits band redundancies and band differences, (4) it can account for the uncertainties of the incoming $L_{wN}(\lambda)$ data streams and, (5) it provides confidence intervals for the derived products. These features are illustrated through several examples of ocean color data merging using SeaWiFS and MODIS Terra and Aqua $L_{wN}(\lambda)$ imagery. Compared to each of the original data source, the products derived from the merging procedure show enhanced global daily coverage and lower uncertainties in the retrieved variables.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Ocean color; Data merging; Bio-optical model; SeaWiFS; MODIS

1. Introduction

There are presently (August 2004) eight ocean color satellite missions in polar orbit deployed by the various international space agencies with many more to follow in the coming years (http://www.ioccg.org/sensors_ioccg.html). At this point, there is little or no coordination between these missions, and each of them provides its individual data stream to its own user community. Consequently, a huge amount of separated data streams is collected without being fully exploited by the ocean color community. This has also resulted in some confusion as data users sometimes wonder which ocean color data set to use. The object of this

manuscript is not to solve these problems per se but rather to present and implement a formalism by which the near-simultaneous observation of ocean color with multiple satellite sensors can be used to improve climate data records of ocean color quantities. From our view, the answers to the above issues will come by measuring improvements in ocean color climate data records developed from multiple missions.

There are many ways where the apparent redundancy in coincident ocean color missions can be exploited. The most obvious way is through improvements in the space/time coverage by the merging of different data sets (e.g., see [Gregg et al., 1998](#); [Gregg & Woodward, 1998](#)). Any single ocean color mission has a limited daily coverage resulting mostly from gaps between the swaths, sun glint and cloud cover. Merged data coverage will increase simply due to the

* Corresponding author. Tel.: +1 8058934308.

E-mail address: stephane@icess.ucsb.edu (S. Maritorena).

combination of patchy and uneven daily coverage from sensors viewing the ocean at slightly different times and geometries. Inasmuch as sensors have different orbits and characteristics, the location and time of the areas they cover daily differ as are the locations affected by sun glint. The merging of ocean color data from different sensors is, thus, a way to achieve better global daily coverage.

The merging of ocean color data sets has other potential benefits, which can ultimately result in improved and more diverse data products with lower uncertainties (Siegel, 1998). Merging allows data from multiple sensors to be used to derive unique sets of products and, in essence, is the method for creating unified ocean color time series (so-called climate data records). Data merging may be the path by which space agencies transition from a mission-centric to a product-centric support of ocean color missions. In the midterm to long term, we believe that ocean color merging activities will provide scientific quality data sets for a wide variety of users (research, operations, education, etc.) involved in various ocean color-related applications.

The successful merging of satellite ocean color data must overcome problems related to the fact that sensors differ in design, calibrations, algorithms and accuracies. A good merging approach should be able to account for differences in the individual data sets involved (as long as these differences do not represent an outstanding bias between sensors). On the other hand, the merging procedure should not create biases, discontinuities or artifacts in the products it generates. While the merging of data from multiple sources has already been implemented for SST (e.g., see Reynolds & Smith, 1994), satellite altimetry (e.g., see Le Traon & Ogor, 1998) or clouds (Rossow & Schiffer, 1991), this remains a new topic in ocean color science. The development of methodologies for ocean color data merging was one of the components of the NASA SIMBIOS program (McClain et al., 2002). Several merging methods have been tested within the SIMBIOS frame from the simple binning of the daily data products from different sources mapped to a common grid to more sophisticated methods (Conkright & Gregg, 2003; Fargion & McClain, 2003; Kwiatkowska & Fargion, 2002, 2003). The approach in common with all of these methods is that they use final data products, such as the chlorophyll concentration (Chl), from different mission data sets to generate merged data products.

The ocean color data merging method presented here does not combine final determinations of chlorophyll concentration from remotely sensed data to produce a unified chlorophyll value. Rather, we collect available normalized water-leaving radiances from one or more sensors and use them all in the inversion of a semianalytical ocean color model to retrieve several variables simultaneously. This model-based approach has several benefits compared to techniques that blend end products from individual missions, including (1) ensuring the simultaneity and consistency of the retrievals by using a single bio-optical algorithm, (2) working with single or multiple data

sources regardless of their specific bands, (3) exploiting spectral band redundancies and band differences, (4) accounting for uncertainties in the input $L_{wN}(\lambda)$ data and (5) providing confidence intervals for the retrieved products.

In the following, the semianalytical model-based ocean color merging approach is presented, as well as some examples of unified products using SeaWiFS and MODIS data. Some of the main characteristics of the merged products are also documented. Lastly, discussion is presented, illustrating the potential of this approach and the necessary future steps to achieve high-quality unified ocean color data products from multiple sensors.

2. Methods and data

2.1. The GSM01 semianalytical bio-optical model

The core of the merging procedure is based on the semianalytical ocean color model described by Garver and Siegel (1997) and updated by Maritorena et al. (2002), thereafter referred to as the GSM01 model. The GSM01 model assumes 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., see Gordon et al., 1988) or

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

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. The absorption and backscattering coefficients are inherent optical properties (IOP) which spectra can be partitioned into subcomponents. The absorption coefficient is composed of the seawater absorption, $a_w(\lambda)$, the phytoplankton absorption, $a_{ph}(\lambda)$, and the combined absorption of dissolved and detrital particulate, $a_{cdm}(\lambda)$ (considered together as a single term because of their similar spectral shapes; Carder et al., 1991; Nelson & Siegel, 2002; Nelson et al., 1998). The backscattering coefficient is partitioned into terms due to seawater, $b_{bw}(\lambda)$, and suspended particulates, $b_{bp}(\lambda)$. The nonwater absorption and scattering terms are parameterized as a known shape with an unknown magnitude,

$$a_{ph}(\lambda) = \text{Chl } a_{ph}^*(\lambda) \quad (2)$$

$$a_{cdm}(\lambda) = a_{cdm}(\lambda_0) \exp(-S(\lambda - \lambda_0)) \quad (3)$$

$$b_{bp}(\lambda) = b_{bp}(\lambda_0) \left(\frac{\lambda}{\lambda_0} \right)^{-\eta} \quad (4)$$

where $a_{ph}^*(\lambda)$ is the chlorophyll a specific absorption coefficient, S is the spectral decay constant for CDM absorption (Bricaud et al., 1981), η is the power law

exponent for particulate backscattering coefficient, and λ_o is a scaling wavelength (443 nm). For $a_{ph}(\lambda)$, $a_{cdm}(\lambda)$ and $b_{bp}(\lambda)$, the unknown magnitudes are the chlorophyll *a* concentration Chl, the CDM absorption coefficient ($a_{cdm}(443)$) and the particulate backscatter coefficient ($b_{bp}(443)$), respectively. In the application of Eqs. (1)–(4), $a_w(\lambda)$, $b_{bw}(\lambda)$, $F_o(\lambda)$, n_w , t and g are taken from the literature, whereas the values of η , S , $a_{ph}^*(\lambda)$ were determined by “tuning” the model against a large in situ data set (Maritorena et al., 2002). The unknown values in Eqs. (1)–(4), Chl, $b_{bp}(\lambda_o)$ and $a_{cdm}(\lambda_o)$, are retrieved by applying a nonlinear least squares technique to fit Eq. (1) to $L_{wN}(\lambda)$ data collected at four or more wavelengths. With in situ global $L_{wN}(\lambda)$ data, the performance of GSM01 Chl retrievals are comparable to those of SeaWiFS’s OC4v4 (Maritorena et al., 2002). The GSM01 model has been used successfully with SeaWiFS data (e.g., see Behrenfeld et al., submitted for publication; Chomko et al., 2003; Siegel et al., 2002; Siegel et al., submitted for publication). The model was initially optimized for the SeaWiFS bands

so $a_{ph}^*(\lambda)$ values for the 531 and 551 nm bands were determined by linear interpolation in order to use GSM01 with MODIS (the other MODIS bands are considered similar to SeaWiFS bands).

2.2. Application of the GSM01 semianalytical model to data merging

The application of the GSM01 model to the merging of ocean color 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 in an array while a separate array with the relevant waveband information is also created. The combined array of $L_{wN}(\lambda)$ observations is then used in the model inversion procedure to produce merged data products (Fig. 1). When different sensors have the same spectral $L_{wN}(\lambda)$ observations, these data are used individually, “as is”, without any averaging or other transformation.

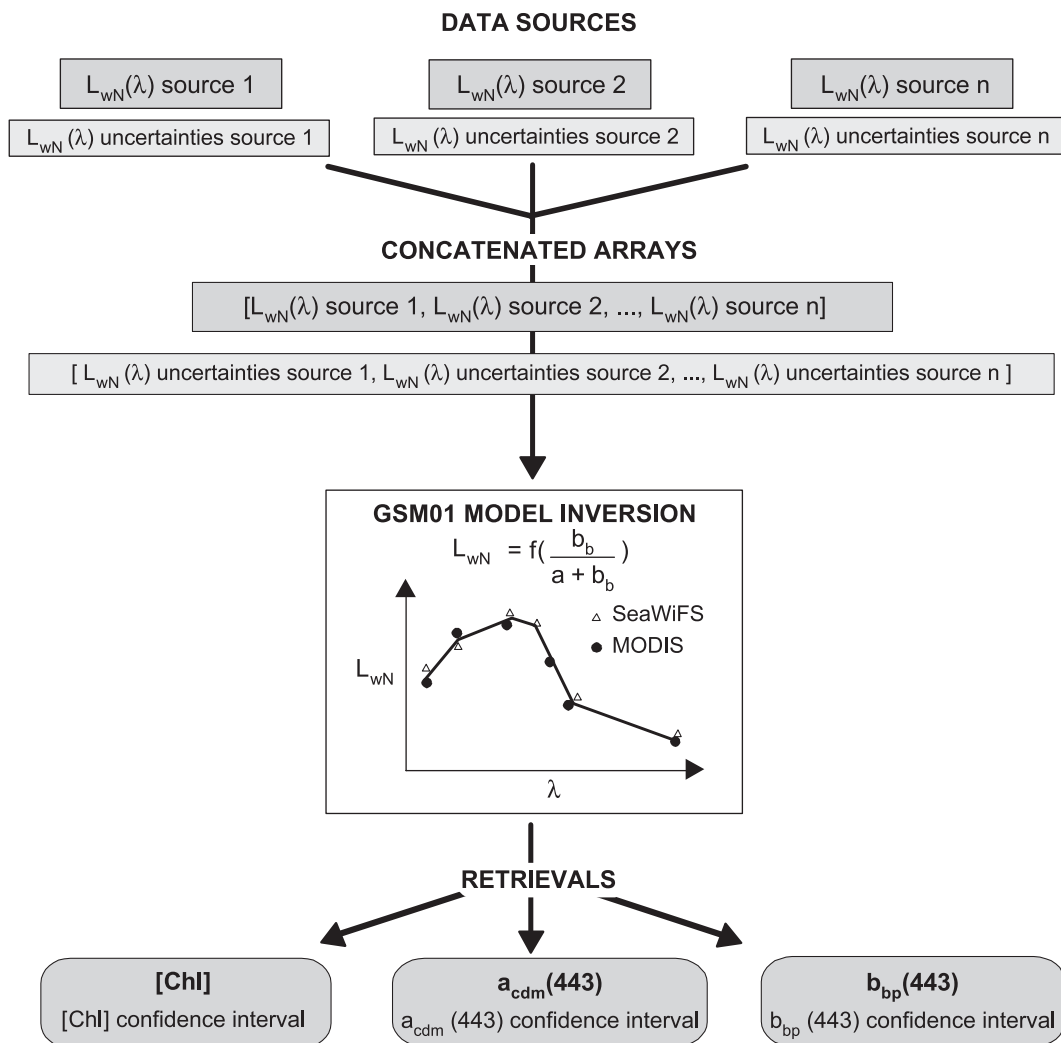


Fig. 1. Schematic of the input data and output products of the GSM01 merging model.

When data sources have different bands, the merging procedure takes advantage of the improved spectral resolution that this affords. When multiple sensors are available, a larger number of observations are provided to the GSM01 inversion procedure. This increases the number of statistical degrees of freedom in the inversion problem, which contributes greatly to reduce the confidence intervals of the final data products. The direct use of all available $L_{wN}(\lambda)$ data in the merging procedure assumes that all sources have similar or close uncertainty levels, and that none of the data sources contains significant bias. The

present merging procedure can also differentially weight individual $L_{wN}(\lambda)$ observations to ensure that the best observations are given a higher weight in the fitting procedure that generates the retrievals. For reasons explained later, this feature cannot be rigorously implemented for satellite data at this time, but the potential of this approach can be demonstrated with simulated data (see Results and discussion).

The quantification of confidence intervals, c_i , of the retrieved products is another major characteristic of the merging procedure. The model provides confidence interval

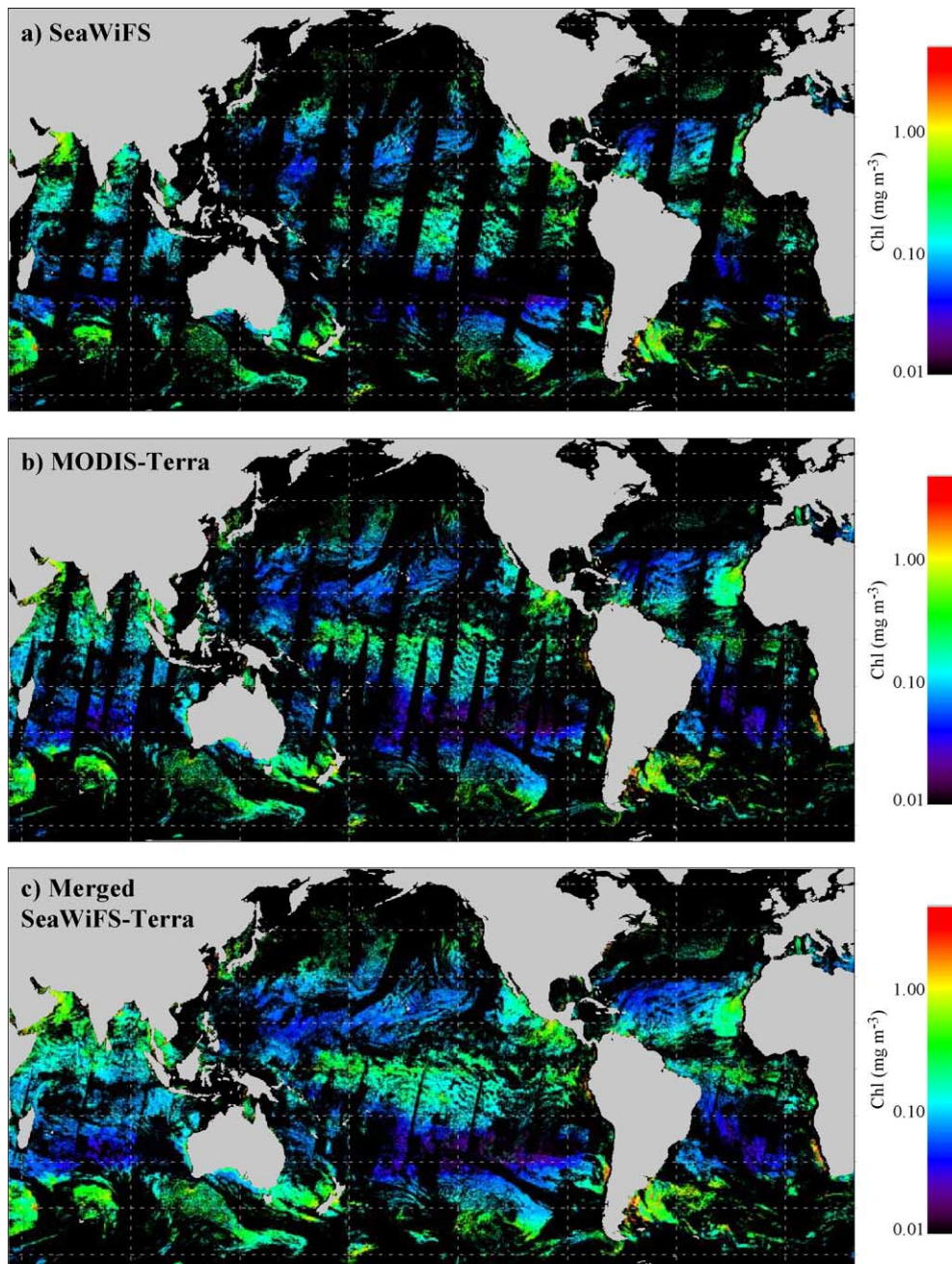


Fig. 2. Global chlorophyll maps derived with the GSM01 model from daily level 3 $L_{wN}(\lambda)$ data (Dec. 4, 2000), (a) using the SeaWiFS $L_{wN}(\lambda)$ data only, (b) using MODIS Terra $L_{wN}(\lambda)$ data only and (c) using the pooled $L_{wN}(\lambda)$ from SeaWiFS and MODIS Terra.

estimates for each of the retrievals using a linear approximation to the calculation of nonlinear regression inference regions (Bates & Watts, 1988). Using local linear approximation, the geometry of nonlinear least squares allows approximate inference parameters to be calculated. Similar to the linear case, if Y is a vector of p retrieved products, then the confidence interval of product Y_p , $ci(Y_p)$, can be expressed as

$$ci(Y_p) = se(Y_p)t(n-p; \alpha/2) \quad (5)$$

where $se(Y_p)$ is the standard error of the p th estimates of Y , and $t(n-p, \alpha/2)$ is the upper $\alpha/2$ quantile for Student's T -distribution with $n-p$ degrees of freedom (n —number of observations, i.e., input $L_{wN}(\lambda)$ and p —number of unknowns). The standard error of the parameter estimator is defined as

$$se(Y_p) = \text{RMS}^{\frac{1}{2}} \left\| \text{QR}(V)^{-1} \right\|_p^{\frac{1}{2}} \quad (6)$$

where RMS is the residual mean square (or variance estimate based on $n-p$ degrees of freedom) between observed and modeled $L_{wN}(\lambda)$ (using the retrieved unknowns, i.e., forward model) and $\left\| \text{QR}(V)^{-1} \right\|_p^{\frac{1}{2}}$ is the

length of the p th row of the inverted and QR decomposed matrix of partial derivatives, V , of the model (Eq. (1)) with respect to each unknown. In the linear approximation, this corresponds to the linear mapping of the projection of the residual vector onto the expectation plane to the parameter (i.e., unknowns) plane (see Bates & Watts, 1988 for details).

2.3. SeaWiFS and MODIS data processing

We have implemented the model-based merging approach using daily level 3 data from SeaWiFS (reprocessing #4, 9 km) and MODIS (collection #4, 4.6 km) for 18 dates between December 4, 2000, and March 22, 2003. Only the MODIS “best quality” data (i.e., quality 0) were used during these tests. Inasmuch as the SeaWiFS and MODIS $L_{wN}(\lambda)$ data products have different spatial resolution, it is necessary to adapt the MODIS data to the SeaWiFS resolution by averaging four 4.6 km bins into a 9 km bin and to have the two data sources set to a common binned grid. The data were processed between 65°N and 65°S. The results presented below cover the period where MODIS Terra $L_{wN}(\lambda)$ collection #4 products are validated (11/1/2000–3/19/2002). Some results outside this time

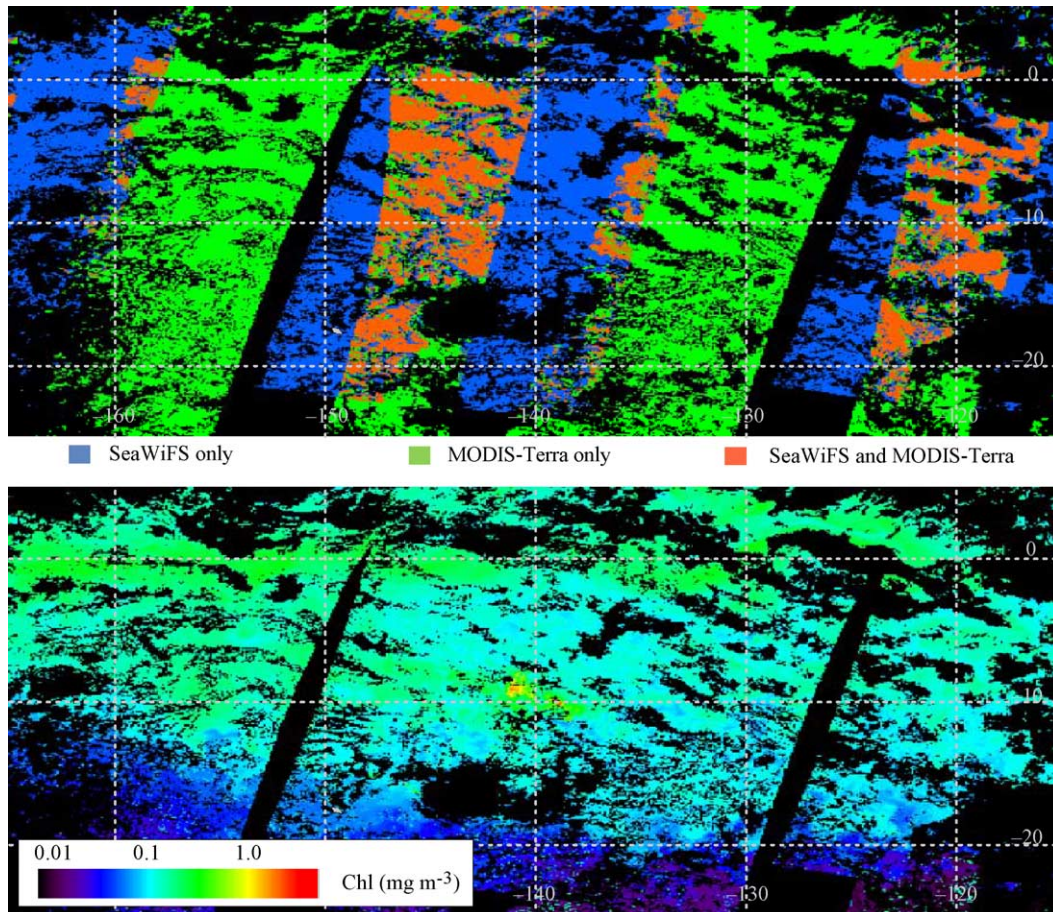


Fig. 3. Blowup of Fig. 2c for the subequatorial South Pacific region centered on the Marquesas Islands ($\sim 10^{\circ}\text{S}$, 140°W). The upper map shows the origin of the data used in the GSM01 model (blue—SeaWiFS $L_{wN}(\lambda)$ data only, green—MODIS Terra $L_{wN}(\lambda)$ data only, red—both SeaWiFS $L_{wN}(\lambda)$ and MODIS Terra $L_{wN}(\lambda)$ data). The lower map shows the very good consistency and continuity in the retrieved chlorophyll between regions having different data sources.

window are also presented to illustrate SeaWiFS/Terra/Aqua data merging and to discuss improvement in coverage when merging ocean color data.

3. Results and discussion

Global chlorophyll maps derived using the GSM01 model with SeaWiFS $L_{wN}(\lambda)$ data alone, MODIS Terra $L_{wN}(\lambda)$ data alone, and the pooled $L_{wN}(\lambda)$ from both sources are presented in Fig. 2. The products generated by the merging model show excellent consistency between the individual satellite data sets and when merged. No obvious discontinuities appear where the model switches from an area with a single data source to one where both data sets are merged. This is better demonstrated in a blow up of the subequatorial South Pacific region and centered on the Marquesas Islands (Fig. 3) which shows the very good continuity in the Chl product between regions having different data sources. The consistency in the Chl fields actually reflects the overall good agreement between the $L_{wN}(\lambda)$ data from both sensors for this particular date and location. The $a_{cdm}(443)$ and the $b_{bp}(443)$ images generated by merging MODIS Terra and SeaWiFS $L_{wN}(\lambda)$ data for the same date are presented in Fig. 4. The merged images clearly show the improvement in spatial coverage when

using these two data sources. Improvement in coverage is obvious when SeaWiFS and both MODIS Terra and MODIS Aqua data are used together (Fig. 5 from another date).

Daily coverage from an individual satellite sensor depends upon various factors such as its orbital characteristics, sun glint, cloud cover and season. The increased coverage that results from the use of multiple data sources is illustrated in Fig. 6 for the 18 dates we have used. Daily coverage jumps from ~12–15% of the ocean surface (in the 65°N to 65°S range) when SeaWiFS is used alone to ~26% when it is used with MODIS Terra. When MODIS Aqua data are used in the merging process along with SeaWiFS and Terra, the daily percentage coverage reaches ~35% of the ocean surface. These numbers agree well with those derived from a theoretical analysis conducted prior to the launches of SeaWiFS and MODIS (Gregg & Woodward, 1998) and with other studies (Fargion & McClain, 2003). With a three-sensor combination resulting in a daily coverage of 30–35%, composite images over a 4–5 days span should allow almost complete coverage of the world ocean which would be very useful to study ocean processes with high temporal resolution.

Because merged products use more quasi-independent observations (i.e., $L_{wN}(\lambda)$) in the nonlinear inversion of Eq. (1), their confidence intervals should be reduced compared

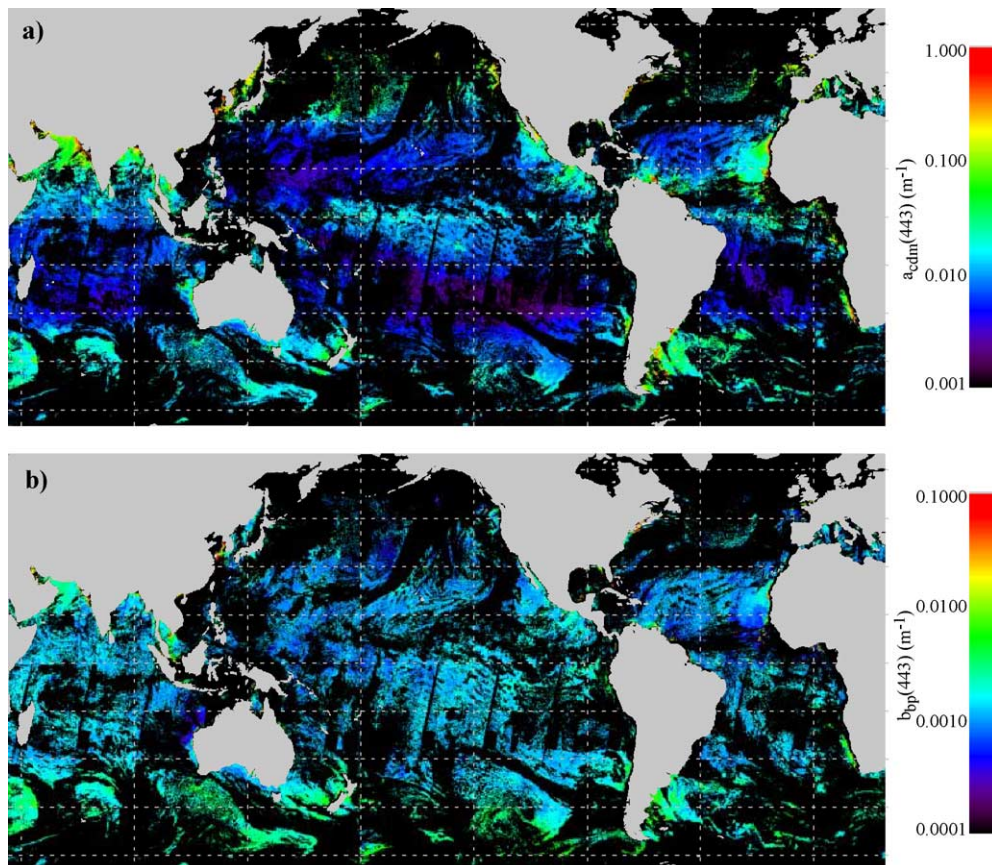


Fig. 4. (a) $a_{cdm}(443)$ and (b) $b_{bp}(443)$ images generated by the GSM01 model from Terra and SeaWiFS daily level 3 $L_{wN}(\lambda)$ data (Dec. 4, 2000).

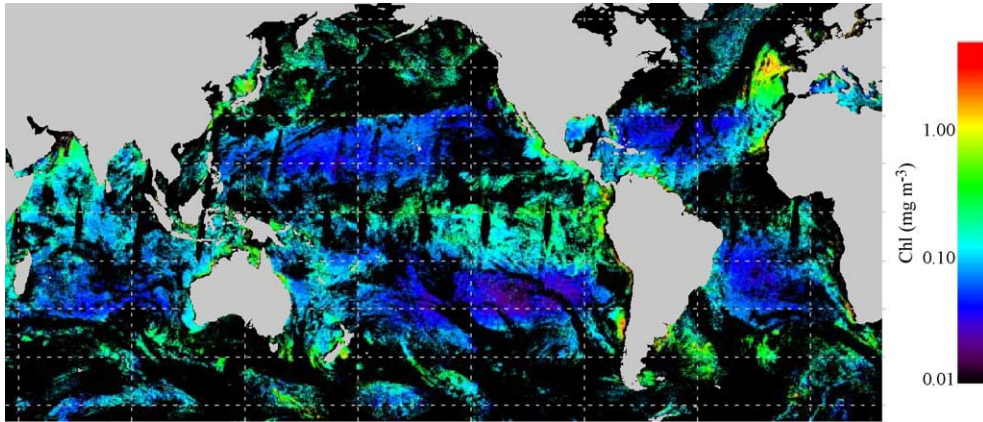


Fig. 5. Global chlorophyll map derived with the GSM01 model using daily level 3 $L_{wN}(\lambda)$ data from SeaWiFS, MODIS Terra and MODIS Aqua (March 21, 2003).

to those from a single data source. Fig. 7 shows, for each GSM01 product, the frequency distribution of the ratio of the confidence intervals of the merged product over the confidence intervals obtained using either SeaWiFS or MODIS $L_{wN}(\lambda)$ data alone. Overall, the confidence intervals tend to decrease when multiple sources are used. This is true for all three products generated by the GSM01 model. The reduction in the confidence intervals of the merged products is observed at all the dates we processed. In the worst case, among the 18 dates we have tested, at least 70% of the pixels that had both SeaWiFS and Terra data showed lower confidence intervals in the merged products. Confidence intervals of merged products show greater improvements when compared to SeaWiFS observations, suggesting that

MODIS Terra data are overall more consistent with the GSM01 bio-optical model than are SeaWiFS data. An examination of the spatial distribution of regions where the merged products confidence intervals, ci_{merged} , is higher than in products from individual sensors are frequently located next to gaps caused by sun glint or gaps between the swaths. They also appear on the eastern part of some swaths and seem to be mostly located in the Southern Hemisphere.

Clearly, there are instrumental issues for each individual mission which influence the data sets and thus the success of the data-merging approach. We hypothesize that increased confidence intervals in the merged products (i.e., $ci_{merged}/ci_{MODIS} > 1$) is likely to occur if there are large discrepancies in the source $L_{wN}(\lambda)$ data. A comparison of the input

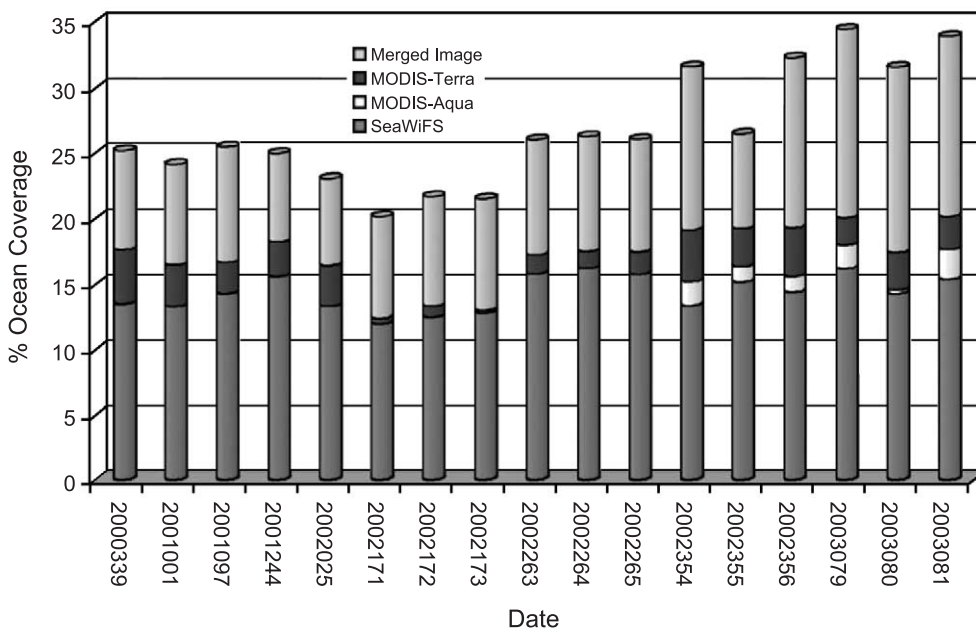


Fig. 6. Improvement in daily coverage from the merging of SeaWiFS, MODIS Terra and MODIS Aqua data for the 18 dates tested. MODIS Aqua data started on 2002354.

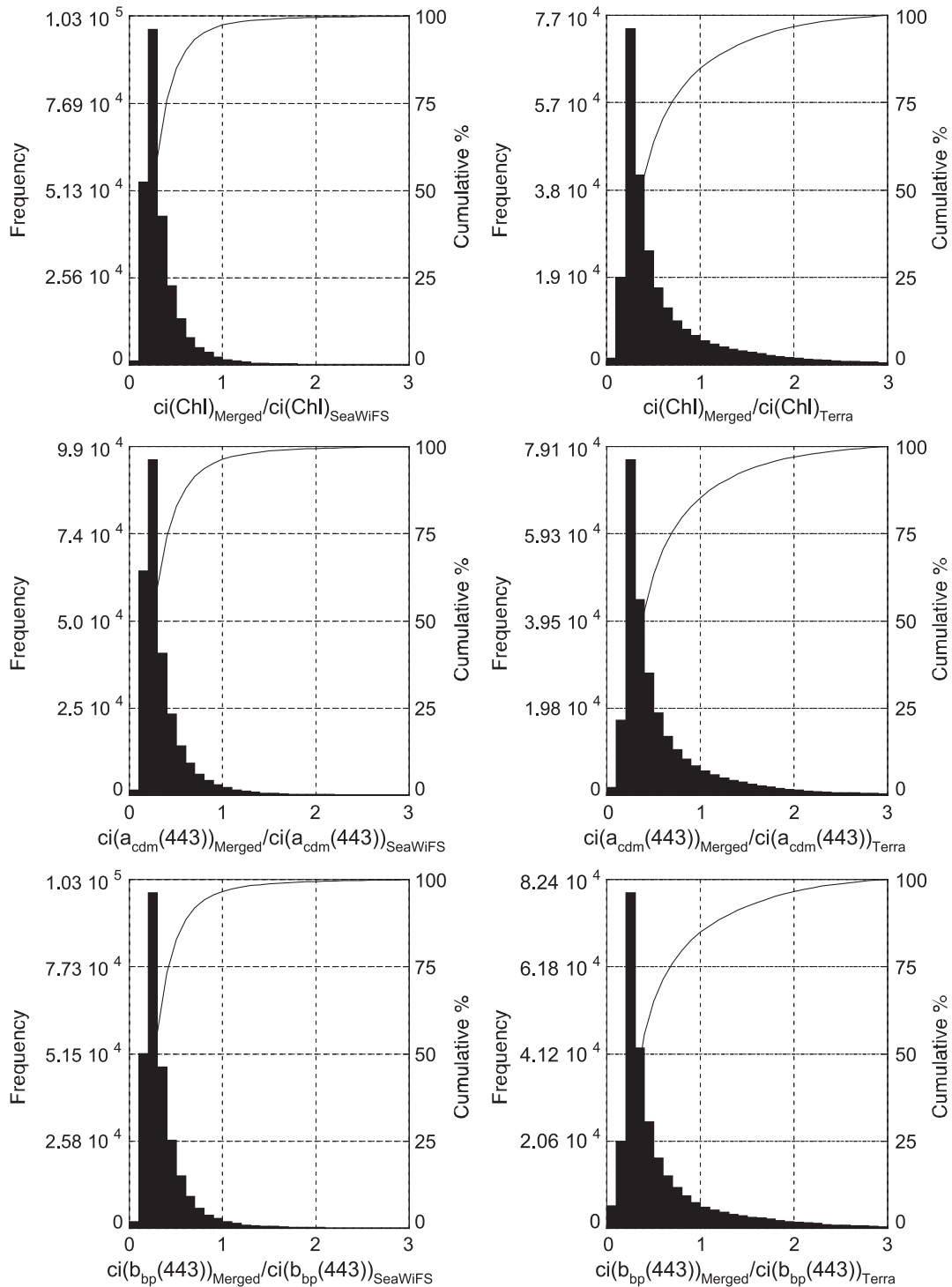


Fig. 7. Frequency distribution of the ratio of the confidence intervals of each GSM01 merged product over the confidence intervals obtained using either SeaWiFS or MODIS $L_{wN}(\lambda)$ data alone. Terra and SeaWiFS daily level 3 $L_{wN}(\lambda)$ data (Dec. 4, 2000).

SeaWiFS and MODIS Terra $L_{wN}(\lambda)$ data used in the merging is shown in Fig. 8. When $ci_{merged}/ci_{MODIS} < 1$, there is excellent agreement between the SeaWiFS and MODIS Terra $L_{wN}(\lambda)$ data sets (Fig. 8). However, when merged confidence intervals are poor ($ci_{merged}/ci_{MODIS} > 1$), there are clear discrepancies in the $L_{wN}(\lambda)$ data. The dependence of

merging algorithm performance on the quality of the input data is confirmed by the statistical analysis. Statistical comparisons become much worse for the pixels with higher confidence intervals in the merged Chl, and the scatter plots show clusters of data points away from the 1:1 line (Fig. 8). Differences in the $L_{wN}(\lambda)$ data can be caused by either a

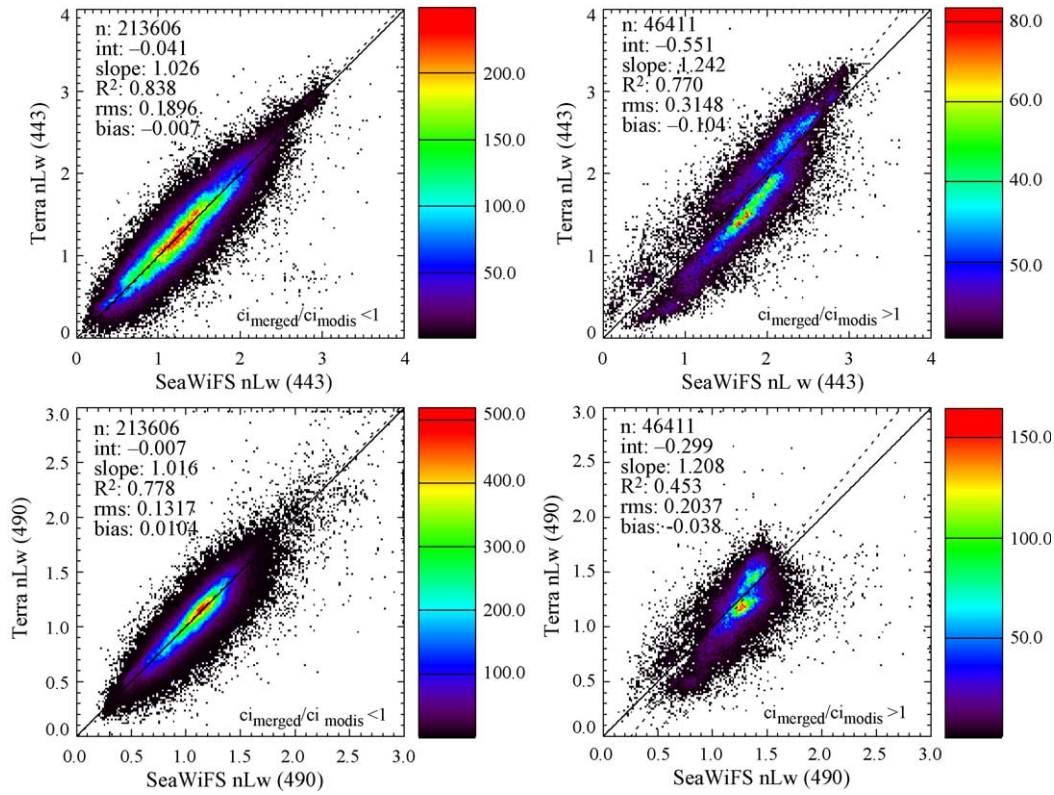


Fig. 8. Comparison the SeaWiFS and MODIS Terra $L_{wN}(\lambda)$ data at 443 and 490 nm for pixels with reduced confidence intervals in the Chl product (left panels) and for the pixels where the Chl confidence intervals are higher in the merged data set (right panels). Terra and SeaWiFS daily level 3 $L_{wN}(\lambda)$ data (Dec. 4, 2000).

lack of synopticity or other issues associated with algorithms, calibrations and other sensor characteristics (e.g., solar diffuser, sensitivity to polarization, see also Fargion and McClain (2003) for a SeaWiFS/MODIS Terra comparison). Detection of poor correspondence in the input $L_{wN}(\lambda)$ data streams may be useful for assessing data quality before the data merging is performed.

It is very unlikely that “simultaneous” observations of $L_{wN}(\lambda)$ from different sensors will ever “perfectly” match for the reasons alluded to in the above. For each sensor, uncertainties in $L_{wN}(\lambda)$ retrievals will always exist, and these uncertainties will differ between satellite sensors and will be variable in time and space, as well as spectrally. Furthermore, the quality of the $L_{wN}(\lambda)$ retrievals has a direct influence on the accuracy (and confidence intervals) of the products derived from them. Provided that the $L_{wN}(\lambda)$ uncertainties are known, this information can be used within the merging procedure. That is, input $L_{wN}(\lambda)$ data can be weighted by their uncertainty levels to ensure that the best data dominate in the fitting procedure. For example, uncertainties ($\sigma_i(\lambda_j)$) of input $L_{wN}(\lambda)$ can be accounted for in the least squares minimization (LSM) procedure as

$$LSM = \sum_{i=1}^{N_{sat}} \sum_{j=1}^{N_{\lambda_i}} \left[\frac{L_{wN-i}(\lambda_j)_{mod.} - L_{wN-i}(\lambda_j)_{meas.}}{\sigma_i(\lambda_j)} \right]^2 \quad (7)$$

where N_{sat} is the number of data source, and N_{λ} is the number of bands for each source. This function allows the best data to have a higher weight in the fitting procedure.

The application of Eq. (7) requires that the uncertainties of the input $L_{wN}(\lambda)$ are known for each band and each sensor. In principle, these uncertainties could be determined through matchup analyses comparing satellite and in situ radiances obtained independently within close space and time windows. These analyses are performed by NASA/GSFC (http://seabass.gsfc.nasa.gov/matchup_results.html) and can provide a good general assessment of the $L_{wN}(\lambda)$ uncertainty levels of some ocean color sensors. However, the current ocean color matchup data sets have some limitations mostly related to their limited size which does not allow an accurate assessment of seasonal or regional variability in the $L_{wN}(\lambda)$ uncertainties. It is not even possible to use the current matchup data sets to test the $L_{wN}(\lambda)$ weighting function inasmuch as SeaWiFS and Terra only have seven common matchup points with both $L_{wN}(\lambda)$ and Chl in situ measurements, while SeaWiFS and Aqua only have one common point. Inasmuch as our goal here is to show the potential of accounting for input data uncertainties, this can be achieved using simulated data.

A semianalytical model (Morel & Maritorena, 2001) was used to generate normalized water-leaving radiance at the five first SeaWiFS and MODIS bands from a range of Chl

values ($0.03\text{--}10\text{ mg m}^{-3}$). The model was slightly modified from its original version by including a correction for the Q factor (Morel & Gentili, 1996) in order to retrieve $L_{\text{wN}}(\lambda)$ rather than the original irradiance reflectance. Noise was added to one of the model-generated $L_{\text{wN}}(\lambda)$ spectrum by multiplying each band by an independent normally distributed random deviates (with a mean of 1.0 and a standard deviation of 0.3). An example of the amount of noise introduced at each waveband is illustrated in Fig. 9a. The noisy and nonnoisy $L_{\text{wN}}(\lambda)$ sets are then used in two different merging runs of the GSM01 model: one without the $L_{\text{wN}}(\lambda)$ uncertainties weighting and one with spectral $L_{\text{wN}}(\lambda)$ uncertainties proportional to the added noise. The use of the $L_{\text{wN}}(\lambda)$ weighting function results in two major improvements over the basic method; chlorophyll retrievals are more accurate and their associated confidence intervals are lower (Fig. 9b). The improvement when using $L_{\text{wN}}(\lambda)$ uncertainties is obvious when looking at the statistics of the regression between the inputted and modeled Chl values. For the example presented in Fig. 9, the slope, R^2 and RMS

error (in log space) are 1.324, 0.742 and 0.520, respectively, when the $L_{\text{wN}}(\lambda)$ uncertainties are not used in the model inversion, while the same statistical parameters are 0.993, 0.943 and 0.178 when the uncertainty weighting function is used. This demonstrates how data with different levels of quality can be efficiently used in data merging and the potential usefulness of this $L_{\text{wN}}(\lambda)$ weighting approach. However, a rigorous implementation of this weighting approach with real data requires a better characterization of the sensors than is currently available.

4. Conclusions and future directions

Our main objective here is the demonstration of a model-based merging approach. The performance of the GSM01 model and a comparison with the retrievals from the SeaWiFS operational algorithm (OC4) were presented in a previous paper using in situ data (Maritorena et al., 2002). It is out of the scope of the present paper to compare the GSM01 satellite retrievals with those of other algorithms as such a comparison, using SeaWiFS data, is documented elsewhere for Chl (Siegel et al., submitted for publication).

We have presented and implemented a simple formalism for merging global satellite ocean color data streams to produce uniform data products while accounting for uncertainty in the input and product data streams. Preliminary results with SeaWiFS and MODIS imagery demonstrate a decrease in the product confidence intervals and an apparent absence of discontinuities in the mapped product fields. These are two important criteria for the development of a successful and generally useful data merging procedure.

Several modifications in the merging procedure are currently being developed and should result in a better overall design of the model along with improved robustness and flexibility. These changes include a fully hyperspectral (or “bandless”) bio-optical model and updated parameterizations that will relax some simplifying assumptions that exist in the current version (e.g., photoadaptation and the modeling of $a_{\text{ph}}^*(\lambda; \text{Chl})$). Corrections for bidirectional reflection distribution function (BRDF) are not yet validated for SeaWiFS and MODIS data. Once an agreed upon and validated BRDF correction scheme is available for satellite ocean color data, it should be beneficial for data merging activities because of presumably improved $L_{\text{wN}}(\lambda)$ estimates. The merging approach presented here requires better sensor characterization than is available at present. A reasonably good knowledge of the $L_{\text{wN}}(\lambda)$ uncertainties for each sensor and their variations in time and space along with the identification of possible biases are key steps towards a fully efficient merging activity. Future tests will also need to assess whether the weighting of $L_{\text{wN}}(\lambda)$ can result in discontinuities in the merged products. This is not expected unless important biases exist in some bands. It is possible that biased $L_{\text{wN}}(\lambda)$ retrievals between sensors have a lesser effect for some other merging methods, but the opposite can also happen, depend-

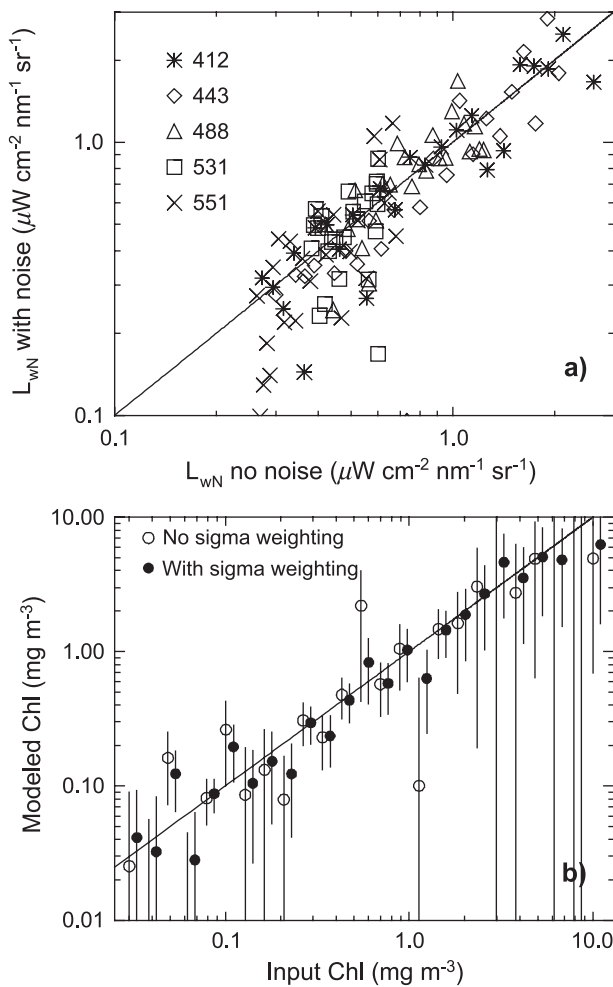


Fig. 9. (a) Comparison of the noisy and nonnoisy simulated $L_{\text{wN}}(\lambda)$ data for the MODIS bands. (b) Chl retrievals and associated confidence intervals when the MODIS noisy and nonnoisy simulated $L_{\text{wN}}(\lambda)$ data were used in GSM01 with a nonnoisy SeaWiFS simulated $L_{\text{wN}}(\lambda)$ set.

ing on the kind of biases and the wavebands affected. Lastly, a validation of the merged data products through matchup analyses and comparisons with other operational products is also to be performed in the future.

The ocean color community can take advantage of the current abundance of satellite ocean color data in various ways. Clearly, the development of merged data products with documented uncertainties is essential for both the users and the satellite agencies. This was one of the goals of the SIMBIOS program (McClain et al., 2002) and requires the cooperation among the satellite missions. The present method of merging individual satellite estimates of $L_{wN}(\lambda)$ is a start towards this goal. However, each satellite sensor can/will have its own atmospheric correction procedure with various levels of success. In the future, data merging could also be conducted using top of the atmosphere radiance determinations from individual missions. This could be a very powerful technique enabling a large increase in the number of degrees of freedom for constraining the eventual least squares problem (following Eq. (7) or similar) due to the different illumination and viewing geometries that need to be considered simultaneously. This approach will also insure consistency in data processing and satellite calibrations applied. A new generation of coupled ocean-atmospheric correction algorithms have been proposed (Chomko & Gordon, 2001; Chomko et al., 2003; Stamnes et al., 2003) which seem able to achieve these goals. It is hoped that the continual improvement of ocean color data processing algorithms will lead to the fulfillment of the vision of unified ocean color science products.

Acknowledgments

This work was supported by the NASA SIMBIOS Project (Contract #NAS5-00196) and NASA ReaSoN CAN (Contract #NNG04GE66G). We thank three anonymous referees for their helpful comments.

References

- Bates, D. M., & Watts, D. G. (1988). Nonlinear regression analysis and its applications. Wiley 365 pp.
- Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M., (2004). Global remote sensing of phytoplankton physiology and carbon-based productivity. *Global Bio-Geochem.* In Press.
- Bricaud, A., Morel, A., & Prieur, L. (1981). Absorption by dissolved organic matter in the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, 26, 43–53.
- Carder, K. L., Hawes, S. K., Baker, K. A., Smith, R. C., Steward, R. G., & Mitchell, B. G. (1991). Reflectance model for quantifying chlorophyll a in the presence of productivity degradation products. *Journal of Geophysical Research*, 96, 20599–20611.
- Chomko, R. M., & Gordon, H. R. (2001). Atmospheric correction of ocean color imagery: Test of the spectral optimization algorithm with SeaWiFS. *Applied Optics*, 40, 2973–2984.
- Chomko, R. M., Gordon, H. R., Maritorena, S., & Siegel, D. A. (2003). Simultaneous determination of oceanic and atmospheric parameters for ocean color imagery by spectral optimization: A validation. *Remote Sensing of Environment*, 84, 208–220.
- Conkright, M. E., & Gregg, W. W. (2003). Comparison of global chlorophyll climatologies: In situ, CZCS, blended in situ-CZCS and SeaWiFS. *International Journal of Remote Sensing*, 24(5), 969–991.
- Fargion, G. S., & McClain, C. R., (2003). MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002–2003. NASA TM-2003-212249, NASA/GSFC, Greenbelt, MD, USA, p54.
- Garver, S. A., & Siegel, D. A. (1997). Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: I. Time series from the Sargasso Sea. *Journal of Geophysical Research*, 102, 18607–18625.
- Gordon, H. R., Brown, O. B., Evans, R. H., Brown, J. W., Smith, R. C., Baker, K. S., et al. (1988). A semi-analytic radiance model of ocean color. *Journal of Geophysical Research*, 93, 10909–10924.
- Gregg, W. W., Esaias, W. E., Feldman, G. C., Frouin, R., Hooker, S. B., McClain, C. R., et al. (1998). Coverage opportunities for global ocean color in a multimission era. *IEEE Transactions On Geoscience and Remote Sensing*, 36, 1620–1627.
- Gregg, W. W., & Woodward, R. H. (1998). Improvements in coverage frequency of ocean color: Combining data from SeaWiFS and MODIS. *IEEE Transactions On Geoscience and Remote Sensing*, 36, 1350–1353.
- Kwiatkowska, E. J., & Fargion, G. S. (2002, Oct.). Merger of ocean color data from multiple satellite missions within the SIMBIOS project. Proc. SPIE Symp.—Remote Sensing of the Atmosphere, Ocean, Environment, and Space. *Ocean Remote Sensing and Applications, Hangzhou, China, vol. 4892.* (pp. 168–182).
- Kwiatkowska, E. J., & Fargion, G. S. (2003). Application of machine-learning techniques toward 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(12), 2844–2860.
- Le Traon, P. Y., & Ogor, F. (1998). ERS-1/2 orbit improvement using TOPEX/POSEIDON: The 2 cm challenge. *Journal of Geophysical Research*, 103, 8045–8057.
- Maritorena, S., Siegel, D. A., & Peterson, A. (2002). Optimization of a semi-analytical ocean color model for global scale applications. *Applied Optics*, 41(15), 2705–2714.
- McClain, C. R., Esaias, W., Feldman, G., Frouin, R., Gregg, W., & Hooker, S. (2002). The Proposal for the NASA sensor intercalibration and merger for biological and interdisciplinary oceanic studies (SIMBIOS) Program, 1995. 2002, NASA TM-2002-210008, NASA/GSFC, Greenbelt, MD, USA, p54.
- Morel, A., & Gentili, B. (1996). Diffuse reflectance of oceanic waters: 3. Implication of bidirectionality for the remote-sensing problem. *Applied Optics*, 35(24), 4850–4862.
- Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. *Journal of Geophysical Research*, 106(C4), 7163–7180.
- Nelson, N. B., & Siegel, D. A. (2002). Chromophoric DOM in the open ocean. In D. A. Hansell, & C. A. Carlson (Eds.), *Biogeochemistry of marine dissolved organic matter*. Academic Press.
- Nelson, N. B., Siegel, D. A., & Michaels, A. F. (1998). Seasonal dynamics of colored dissolved material in the Sargasso Sea. *Deep-Sea Research. Part 1. Oceanographic Research Papers*, 45(6), 931–957.
- Reynolds, R. W., & Smith, T. M. (1994). Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate*, 7, 929–948.
- Rossow, W. B., & Schiffer, R. A. (1991). ISCCP cloud data products. *Bulletin of the American Meteorological Society*, 72, 2–20.
- Siegel, D. A., (1998). Spectral data assimilation for merging satellite ocean color imagery. Extended abstract published on the CD proceedings from Ocean Optics XIV.
- Siegel, D. A., Maritorena, S., Nelson, N. B., Hansell, D. A., & Lorenzi-Kayser, M. (2002). Global ocean distribution and dynamics of colored

- dissolved and detrital organic materials. *Journal of Geophysical Research*, 107, 3228.
- Siegel, D. A., Maritorena, S., Nelson, N. R., & Behrenfeld, M. J., (2004). Independence and interdependences of global ocean optical properties viewed using satellite ocean color imagery. Submitted to *J. Geophys. Research*.
- Stamnes, K., Li, W., Yan, B., Eide, H., Barnard, A., Pegau, W. S., et al. (2003). Accurate and self-consistent ocean color algorithm: Simultaneous retrieval of aerosol optical properties and chlorophyll concentrations. *Applied Optics*, 42, 939–951.