

# Merging SeaWiFS and MODIS/Aqua Ocean Color Data in North and Equatorial Atlantic Using Weighted Averaging and Objective Analysis

Claire Pottier, Véronique Garçon, Gilles Larnicol, Joël Sudre, Philippe Schaeffer, and Pierre-Yves Le Traon

**Abstract**—Two approaches of ocean color data merging were tested and compared in the North and Equatorial Atlantic Basin: the weighted averaging and the objective analysis. The datasets used were the daily level-3 binned data of chlorophyll-*a* from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and Moderate Resolution Imaging Spectroradiometer on the Aqua satellite over the year 2003, which is the first common full year of operation. Since they represent input for both approaches, matchups between the satellite and the *in situ* data from the SeaWiFS Bio-optical Archive and Storage System and the Atlantic Meridional Transect were first studied to compute a spatial map of the root mean-square error and of the bias. Because of the log distribution of the chlorophyll fields, each approach was applied to untransformed and log-transformed values. The application of the weighted averaging to log-transformed values does not show significant differences in comparison to its application to untransformed values. This is not the case, however, for the objective analysis that gives better results when applied to log-transformed values. Both approaches give combined chlorophyll data of equivalent quality, although the objective analysis could be improved with a better statistical characterization of noise and signal covariance. The main advantage of the objective analysis is its ability to interpolate in space (and time) by taking into account the characteristic scales of chlorophyll-*a*. As a result, the spatial coverage of the combined data is at least twice as large in the case of objective analysis than weighted averaging.

**Index Terms**—Biology, marine vegetation, sea surface, signal analysis.

## I. INTRODUCTION

THE USE of satellite image data to investigate oceanic processes has become essential for oceanographic research and monitoring. The synoptic and global data play a fundamental role, since conventional platforms cannot adequately cover the vast and rapidly varying ocean at proper time and

space scales. Since the first ocean color mission, which is the Coastal Zone Color Scanner (CZCS), demonstrated the ability to observe globally the abundance and distribution of oceanic phytoplankton from space, the data have been used in many studies to better understand the role of the ocean in biogeochemical cycles, particularly as the biological carbon pump.

Chlorophyll-*a* (the main pigment in phytoplankton) distribution and variability can be determined from ocean color remote sensing. First, the normalized water-leaving radiance spectrum is derived from the calibrated top-of-atmosphere radiance measured by the satellite sensor at visible and near-infrared wavelengths by removing the contribution from the atmosphere through a process known as atmospheric correction [1]. Normalization refers to the correction for bidirectional effects (viewing angle dependence and effects of seawater anisotropy [2]). The spectrum of the normalized water-leaving radiance can subsequently be converted to chlorophyll-*a* concentration with biooptical algorithms, such as empirical band ratio formulations used for operational products [3].

Ten satellites having an ocean color sensor on board are now on orbit for regional or global scale applications. The main characteristics of the two studied here, which are the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) on OrbView-2 (NASA, USA) since August 01, 1997 and the Moderate Resolution Imaging Spectroradiometer (MODIS) on Aqua (NASA, USA) since May 04, 2002, are presented in Table I. Any individual ocean color mission is limited in ocean coverage by its swath width and gaps caused by clouds or sun glint, which disable the extraction of ocean color [4]. As an example, SeaWiFS can provide about 15% coverage of the global ocean at 0.5° resolution in one day under climatological cloudiness and sun glint [4]. This daily coverage would increase to 25% by merging SeaWiFS, MODIS/Aqua, and MODIS/Terra (MODIS sensor on the Terra satellite, since December 18, 1999) datasets. In this way, besides the increase of the statistical confidence in the data, the identification and study of biological and physical phenomena (such as primary production and algae blooms for instance) and their variability are made easier by filled maps.

There are many difficulties with ocean color data merging because the sensors have varying designs and characteristics. Let us have a look at those from SeaWiFS and MODIS/Aqua. SeaWiFS is in descending sun-synchronous near-polar circular orbit with a 12:20-noon local equator crossing time, whereas MODIS/Aqua is in ascending orbit with a 1:30 PM equator crossing time (Table I). The MODIS scan-mirror assembly uses a continuously rotating double-sided scan mirror to scan

Manuscript received July 4, 2005; revised April 11, 2006. This work was supported in part by the FP6 European IP Marine Environment and Security for the European Area (MERSEA) funding and in part by the CIFRE/ANRT fellowship to C. Pottier.

C. Pottier is with the Collecte Localisation Satellites (CLS) Space Oceanography Division, 31520 Ramonville-Saint-Agne, France, and also with the LEGOS/Centre National de la Recherche Scientifique (CNRS), 31401 Toulouse Cedex 9, France (e-mail: claire.pottier@cls.fr; claire.pottier@notos.cst.cnes.fr).

V. Garçon and J. Sudre are with the LEGOS/CNRS, 31401 Toulouse Cedex 9, France (e-mail: veronique.garcon@cnes.fr; joel.sudre@cnes.fr).

G. Larnicol and P. Schaeffer are with CLS Space Oceanography Division, 31520 Ramonville-Saint-Agne, France (e-mail: gilles.larnicol@cls.fr; philippe.schaeffer@cls.fr).

P.-Y. Le Traon is with Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), 29280 Plouzané, France (e-mail: pierre.yves.le.traon@ifremer.fr).

Digital Object Identifier 10.1109/TGRS.2006.878441

TABLE I  
CHARACTERISTICS OF SEAWIFS AND MODIS/AQUA SENSORS

Sensor	SeaWiFS	MODIS/Aqua
Satellite	Orbview-2	Aqua
Launch date	08/01/1997	05/04/2002
Equator crossing time	12:20 noon	1:30 pm
Spectral bands center (nm)	8 bands:	36 bands, with
	413	412
	443	443
	490	488
	510	531
	555	551
	670	667 & 678
	765	748
	865	869
Data available since...	09/04/1997	11/29/2002
Swath width	2,806 km	2,330 km
Spatial resolution	1100 m (LAC) 4500 m (GAC)	1000 m
Tilt	-20°, 0°, +20°	No
Orbit	Descending	Ascending

LAC stands for Local-Area Coverage. GAC stands for Global-Area Coverage.

$\pm 55^\circ$  [5]; its viewing swath is 10 km along track at nadir and 2330 km cross track at  $\pm 55^\circ$ . SeaWiFS is equipped with an off-axis folded telescope coupled with a rotating half-angle mirror (which rotates at half speed of the folded telescope), and that is phase-synchronized (<http://oceancolor.gsfc.nasa.gov/SeaWiFS/SEASTAR/SPACECRAFT.html>). In addition to the polarization scrambler, this arrangement permits a minimum level of polarization to be achieved: SeaWiFS nominal polarization sensitivity is therefore 20 times less than the maximum polarization sensitivity of MODIS/Aqua, and a polarization correction is thus needed for MODIS/Aqua to improve the agreement between both sensors' water-leaving radiances [6]. Although the maximum scan angle of SeaWiFS is  $\pm 58.3^\circ$  at MODIS-like altitude, the SeaWiFS global-area coverage (GAC) swath is limited to 1502 km. Contrary to MODIS/Aqua, SeaWiFS tilts during each orbit in order to reduce the effects of sun glint from the sea (Table I; <http://oceancolor.gsfc.nasa.gov/SeaWiFS/SEASTAR/SPACECRAFT.html>). As a result, SeaWiFS' products show a zonal band without any data. Both sensors provide a global coverage of the earth in two days. MODIS has 36 spectral bands from which nine are used for ocean color (bands #8 to #16, the others are used for measuring land/cloud/aerosols properties, atmospheric, and cirrus-cloud water vapor, surface/cloud/atmospheric temperature, cloud top altitude, and ozone; <http://modis.gsfc.nasa.gov/about/specifications.php>). SeaWiFS has eight spectral bands: all are used for ocean color measurement. MODIS bands 1–2 have a nominal nadir resolution of 250 m; bands 3–7 have a nadir resolution of 500 m; and bands 8–36 have a nadir resolution of 1 km [5]. SeaWiFS resolution for all eight bands is 1.1 km at nadir (Table I). Since GAC data are subsampled every fourth pixel and every fourth line, SeaWiFS' resolution is 4.5 km. MODIS spectral-band centers dedicated to ocean color are not exactly the same as SeaWiFS' (Table I). Therefore, the empirical algorithm to compute chlorophyll-*a* concentration from the normalized water-leaving radiances is not exactly the same for both sensors. However, the MODIS OC3M algorithm is similar to the SeaWiFS OC4v4 algorithm ([\[nasa.gov/eval/oc.cgi\]\(http://seabass.gsfc.nasa.gov/eval/oc.cgi\)\). Since merger activities depend on the calibration and validation quality of data products, differences in standard products among missions need to be evaluated: It is not the object of this study, but the interested reader can refer to \[7\]–\[9\].](http://seabass.gsfc.</a></p>
</div>
<div data-bbox=)

Data-merging methods can start at the level of the water-leaving radiance or at the level of the derived products such as chlorophyll. Maritorena and Siegel [10] used normalized water-leaving radiances from SeaWiFS and MODIS/Aqua together in a semianalytical biooptical model (GSM01) to produce global retrievals of three biogeochemically relevant variables (chlorophyll, combined dissolved and detrital absorption coefficient, and particulate backscattering coefficient). Gregg and Conkright [11] used a blended analysis to combine *in situ* data and satellite CZCS (NASA) chlorophyll in order to construct a climatological seasonal representation of the global chlorophyll distribution. Within the NASA SIMBIOS program, several methods were tested from the simple binning of the daily chlorophyll from different sources mapped to a common grid to more sophisticated methods [12]–[14].

Our objective here within the Global Monitoring for Environment and Security (GMES) integrated project Marine Environment and Security for the European Area (MERSEA) is to provide an accurate and consistent stream of ocean color data at a resolution and extent of coverage compatible with operational forecasting of the marine environment. Because we want to obtain daily maps of chlorophyll, we chose here to directly combine the chlorophyll-concentration products of both SeaWiFS and MODIS/Aqua. In the next section, we will present the datasets used and the results of the comparison to the comprehensive archives of *in situ* chlorophyll data. We then present and compare two approaches over the North and Equatorial Atlantic region for the year 2003: the weighted averaging and the objective analysis first introduced in oceanography by Bretherton *et al.* [15] and frequently used in altimetry [16]. A performance assessment of each approach will finally be provided at the North Atlantic basin scale as well as at the biogeochemical-province scale.

## II. DATA

We used SeaWiFS and MODIS/Aqua daily level-3 binned data of chlorophyll concentration obtained from the NASA/Goddard Earth Science (GES)/Distributed Active Archive Center (DAAC). The datasets are issued from the reprocessing version 4 (July 2002, <http://oceancolor.gsfc.nasa.gov/REPROCESSING/SeaWiFS/R4/>) for SeaWiFS and version 1 (February 2005, <http://oceancolor.gsfc.nasa.gov/REPROCESSING/Aqua/R1/>) for MODIS/Aqua. The data were all considered whatever their flags are. These data are both on an equal area grid, in which the resolution is  $1/12^\circ$  and  $1/24^\circ$  at the Equator, respectively. For practical considerations, we resampled them, for this paper, on a regular grid of  $0.1^\circ \times 0.1^\circ$  (i.e.,  $11.1 \times 11.1$  km at the Equator) for SeaWiFS and  $0.05^\circ \times 0.05^\circ$  (i.e.,  $5.6 \times 5.6$  km at the Equator) for MODIS/Aqua. Fig. 1 shows an example of the location of a SeaWiFS and a MODIS/Aqua chlorophyll pixel at longitude *i* and latitude *j*). These resolutions were selected because they stand as a compromise between not being far from the original

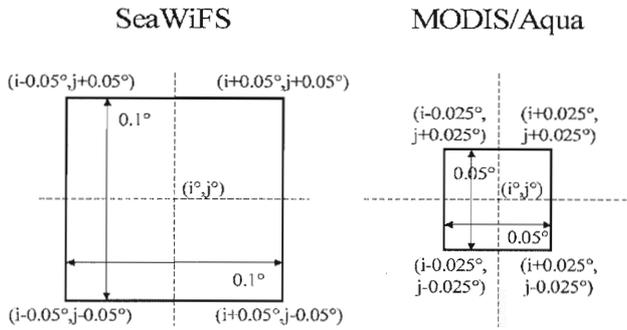


Fig. 1. Example of the location of SeaWiFS' and MODIS/Aqua's chlorophyll pixel at longitude  $i$  and latitude  $j$ .

level-3 resolutions and saving disk space. In addition, they conserve the ratio between the resolutions of level-3 binned data of both sensors. We will focus our study on the year 2003, which is the first common full year of operation of both SeaWiFS and MODIS/Aqua sensors. Our area of interest is the North and Equatorial Atlantic region, comprised between  $98.5^\circ$  W– $20^\circ$  E and  $20^\circ$  S– $70^\circ$  N. The distribution of untransformed and log-transformed SeaWiFS chlorophyll daily data (example shown: August 13, 2003) in the global ocean and in the North Atlantic basin (Fig. 2) confirms the lognormal distribution of ocean chlorophyll found by Campbell [17]: the logarithm of chlorophyll concentrations has a normal distribution. We will thus apply the merging algorithms on untransformed and log-transformed values.

*In situ* chlorophyll measurements collected from ships or oceanic platforms are usually used to validate original SeaWiFS, MODIS/Aqua, or even the combined chlorophyll datasets: These coincident in time and coverage observations between *in situ* and satellite measurements are called matchups. This allows computing errors and bias between each dataset. Two *in situ* datasets are used here. The first one was obtained from the NASA SeaWiFS Bio-optical Archive and Storage System (SeaBASS) [18] (Gregg, personal communication). Data are collected using a number of instrument packages (such as profilers and handheld instruments) on a variety of platforms (including ships and moorings). In the studied region, it includes 704 fluorometrically/spectrophotometrically derived chlorophyll-concentration values (milligrams per cubic meter) at depths of 0–5 m for the selected 2003-year period. The second set contains the Atlantic Meridional Transect (AMT) [19] underway data, which are obtained through the British Oceanographic Data Centre (BODC). The AMT Program undertakes biological, chemical, and physical oceanographic research during both annual passages of the R/V James Clark Ross between Plymouth (U.K.) and the Falkland Islands. Underway instruments continuously take measurements of the sea surface data (e.g., salinity, temperature, attenuation, chlorophyll, and nutrients) by measuring continuously pumped surface seawater. The database includes 62 613 fluorometrically derived chlorophyll-concentration values (milligrams per cubic meter) from through flow over two one-month periods: from May 12, 2003 to June 17, 2003 (AMT12 cruise) and from September 11, 2003 to October 13, 2003 (AMT13 cruise). This led to 63 317 *in situ* measurements. *In situ* data that were coincident (occurring

within the same day) and collocated (occurring within a single SeaWiFS or MODIS/Aqua level-3 pixel of chlorophyll) were averaged, following [20]. This led to 2754 *in situ* values on a  $0.1^\circ$  grid and 5144 on a  $0.05^\circ$  grid. After considering coincident, collocated averages, and cloud-free SeaWiFS and MODIS/Aqua data, the final result was 893 comparison matchup points for SeaWiFS and 1476 for MODIS/Aqua over 2003 in the North Atlantic basin. The locations of the matchups for both sensors are shown in Fig. 3. To perform a regional analysis, we subdivided the basin into 19 ocean domains, which are the biogeochemical provinces defined in [21] (Fig. 3). Statistical analyses were performed both globally and within these provinces.

The coefficient of determination  $r^2$  from the correlation analysis indicates the covariance between the satellite dataset and the *in situ* observations. For both sensors, the open ocean has higher  $r^2$  than the coastal provinces, as shown in Fig. 4. Actually, satellite chlorophyll concentrations less than  $1 \text{ mg/m}^3$ , which are generally found in the open ocean, have a better agreement with *in situ* data than chlorophyll concentrations higher than  $1 \text{ mg/m}^3$  (generally found in the coastal provinces).

For untransformed values, the error measurement is estimated by the root-mean-square (rms) error, which is defined as

$$\text{rms}_d(\text{mg/m}^3) = \sqrt{\frac{1}{N} \sum (S - I)^2} \quad (1)$$

and the bias is estimated by the averaged difference (AD) defined as

$$\text{AD}_d(\text{mg/m}^3) = \frac{1}{N} \sum (S - I) \quad (2)$$

where  $S$  indicates satellite (SeaWiFS or MODIS/Aqua) chlorophyll concentration,  $I$  indicates *in situ* chlorophyll concentration, and  $N$  is the number of samples.

For log-transformed values,  $\text{rms}_l$  and  $\text{AD}_l$  are computed by replacing  $S$  and  $I$  by  $\log(S)$  and  $\log(I)$ , respectively ("log" is the logarithm to base 10). Fig. 5 shows the values found per biogeochemical province and for the Atlantic basin. The rms error and the bias computed on chlorophyll are relatively weak in the open ocean (the rms error is less than  $0.09 \text{ mg/m}^3$  in the subtropical gyres and less than  $0.7 \text{ mg/m}^3$  everywhere else; the bias is less than  $0.04 \text{ mg/m}^3$  in the subtropical gyres and less than  $0.4 \text{ mg/m}^3$  everywhere else) and higher in the coastal provinces (the rms error and the bias are greater than  $1 \text{ mg/m}^3$ ). SeaWiFS is performing slightly better than MODIS/Aqua at the Atlantic basin scale. There are differences between the atmospheric correction schemes and biooptical algorithms applied to MODIS and SeaWiFS data. The much higher uncertainties affecting MODIS/Aqua's chlorophyll- $a$  with respect to SeaWiFS' in eastern U.S. coastal waters suggest that these differences might be highlighted in the presence of some optically complex waters and/or continental aerosols. In general, MODIS/Aqua's water-leaving radiances in the blue channels are lower than SeaWiFS' equivalents and field measurements [22]. This has been confirmed in coastal waters [23]. Eventually, it is likely that the determination of chlorophyll- $a$  is affected in different ways by the uncertainties associated with the respective radiometric products.

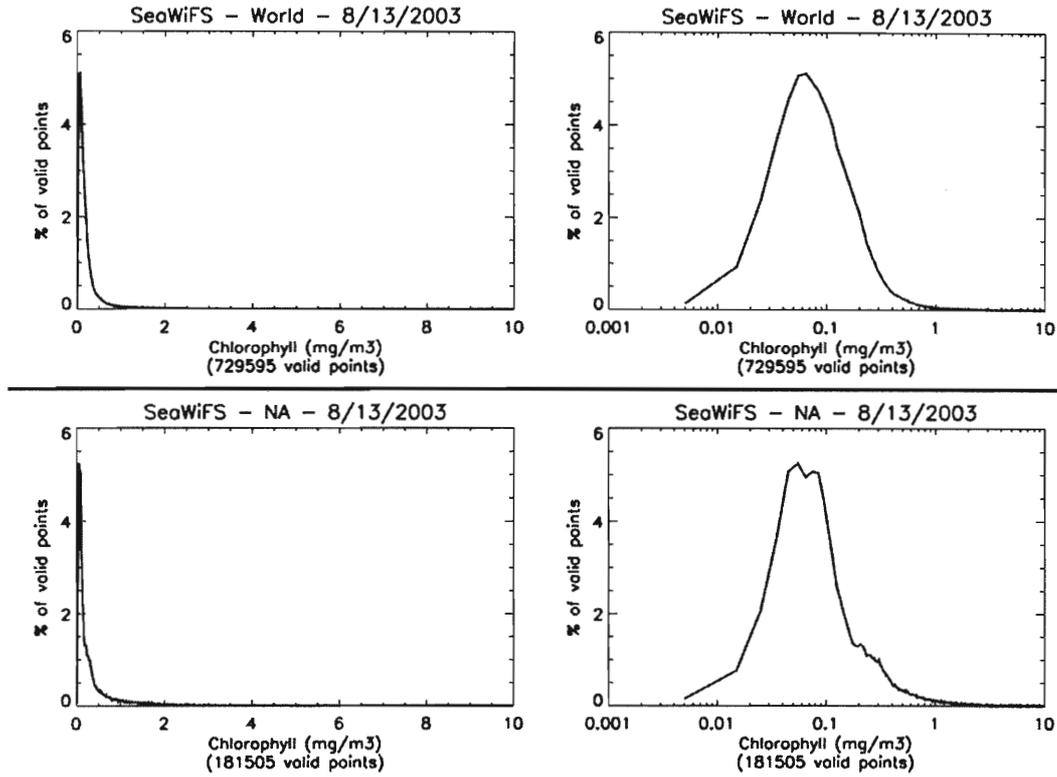


Fig. 2. Distribution of (left) untransformed and (right) log-transformed SeaWiFS chlorophyll daily data for August 13, 2003 in (top) the global ocean and (bottom) in the North and Equatorial Atlantic.

For our merging methods, a map of the measurement error and/or a map of the bias between *in situ* and satellite data are needed. We thus used the values given in Fig. 5. For the biogeochemical provinces for which there was no matchup, the rms error and the bias given were those from the nearest province. Nearest province means the closest geographically and the closest biogeochemically. These provinces associations are thus made following the biome (polar, westerlies, trade winds, and coastal; [21]) to which they belong (cf. Table II). In order to avoid sharp gradients on rms error and bias at boundaries between the biogeochemical provinces, the rms error and bias fields were smoothed at 100 km.

Results will be presented here for two particular days. First, August 13, 2003, for which the spatial coverage is remarkably good for both sensors (Fig. 6, right panel). Indeed, it is 23% for SeaWiFS and 20% for MODIS/Aqua. The second day chosen, which is April 1, 2003, is quite particular: In addition to the sparse spatial coverage of SeaWiFS (12%), MODIS/Aqua presents only one track (4%; Fig. 6, left panel).

### III. WEIGHTED AVERAGING

#### A. Description

The first approach used is the weighted averaging. From the map of chlorophyll concentrations for a given day and the map of the measurement errors for each sensor [i.e., the rms error computed between *in situ* measurements and each sensor data (see Section II)], we obtain the map of the combined chlorophyll for this day and the map of the associated errors.

The value of a combined chlorophyll pixel equals the weighted averaging of the chlorophyll pixels of each sensor and at the same location. The weights are the confidence we have in each sensor in comparison to the other, which are computed from the rms errors. Since SeaWiFS and MODIS/Aqua have not the same spatial resolution, two choices are available: The resolution of the combined data is either the lowest one ( $0.1^\circ$ : Algorithm WA1) or the highest one ( $0.05^\circ$ : Algorithm WA2).

1) *Algorithm WA1*: Let us begin with the untransformed values. The first step of Algorithm WA1 [see Fig. 7(a)] consists of oversampling MODIS/Aqua's chlorophyll pixels to bring them to SeaWiFS' resolution. Considering the location of the chlorophyll pixels of each sensor (Fig. 1), MODIS/Aqua's candidate  $M$  is obtained from nine original chlorophyll pixels with the following formula:

$$\begin{aligned}
 M &= \frac{1}{4}M_5 + \frac{1}{8}(M_2 + M_4 + M_6 + M_8) \\
 &\quad + \frac{1}{16}(M_1 + M_3 + M_7 + M_9) \\
 &= \sum_{i=1}^9 \lambda_i M_i
 \end{aligned} \tag{3}$$

with  $\sum_{i=1}^9 \lambda_i = 1$ .

The  $\lambda_i$  coefficients represent the fraction area of each MODIS/Aqua's original chlorophyll pixel covered by the MODIS/Aqua's candidate. Of course, only the existing data points are considered.

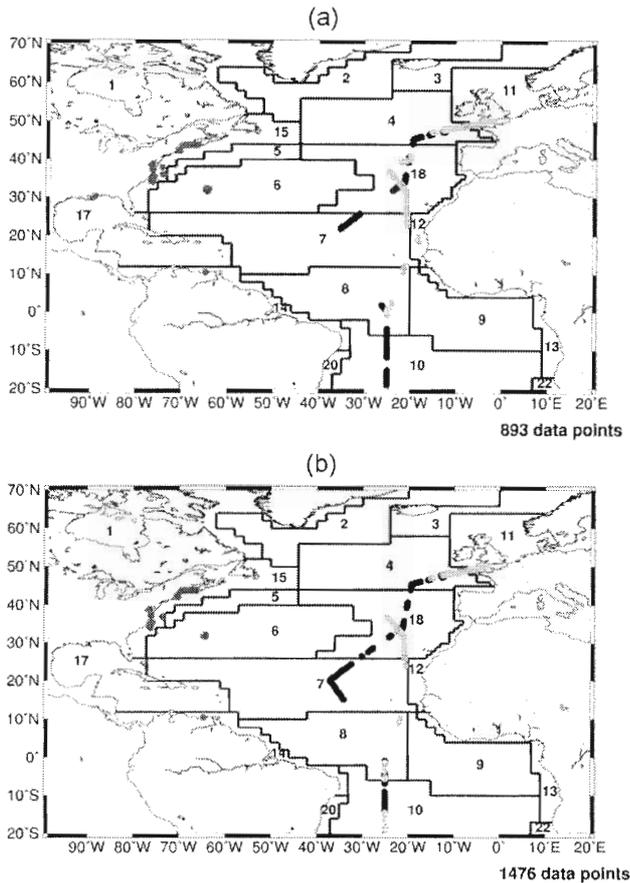


Fig. 3. Location of the matchups for (a) SeaWiFS and (b) MODIS/Aqua with SeaBASS's (in midgray), AMT12 cruise's (in black), and AMT13 cruise's (in pale gray) data. The biogeochemical provinces defined and numbered following [21] are delimited by gray lines (cf. Table II).

To compute the associated error EM, we can consider the distribution with an  $M$  mean and an EM standard deviation as a linear combination of the nine distributions of an  $M_i$  mean and an  $EM_i$  standard deviation (i.e., the error associated with the  $M_i$  chlorophyll pixel, which is the rms error associated with the point and is computed using matchups with *in situ* data in Section II). Since no specific rule exists to compute the standard deviation of a linear combination of lognormally distributed variables, the error (i.e., standard deviation) EM of  $M$  must be computed in a basic way as follows:

$$EM = \sqrt{\sum_{n=1}^N X_n^2} \quad \text{with } X_n = \sum_{i=1}^9 \lambda_i Y_{i,n}.$$

First, we build nine lognormal distributions  $Y_i$  ( $i = \{1, \dots, 9\}$ ) of  $N$  points (after several tests, we chose  $N = 1000$ ), with an  $M_i$  mean and an  $EM_i$  standard deviation; then, we compute their linear combination  $X_n$  (the weights are the  $\lambda_i$ ); the error EM of  $M$  is the standard deviation of  $X_n$ . The lack of several values out of the nine brings so little difference that we considered it had no real impact on the EM value.

The second and last step of Algorithm WA1 is the combination of SeaWiFS and MODIS/Aqua's candidates. The com-

bined chlorophyll pixel  $C$  is computed using the following:

$$C = \left(1 - \frac{\%ES}{\%EM + \%ES}\right) S + \left(1 - \frac{\%EM}{\%EM + \%ES}\right) M \quad (4)$$

where  $\%EX = EX/X$ . Actually, the weights of  $S$  and  $M$  represent the percentage of confidence we have in the corresponding chlorophyll pixel in comparison to the other. The associated error EC is computed in the same way as in step 1, i.e., it is considered as the standard deviation of a linear combination of two lognormal distributions of means  $S$  and  $M$  and standard deviations ES and EM.

To obtain MODIS/Aqua's candidate for log-transformed values, the " $X$ " values are replaced by the " $\log X$ " values. The geometric mean is used. Therefore, (3) is replaced by

$$M = 10^m \quad \text{with } m = \sum_{i=1}^9 \lambda_i \log M_i. \quad (5)$$

This time, let  $EM_i$  be the error associated with  $\log(M_i)$ . Since  $\log(M_i)$  has a normal distribution, we have

$$EM = \sqrt{\sum_{i=1}^9 (\lambda_i EM_i)^2}. \quad (6)$$

To compute the combined chlorophyll pixel, (4) is replaced by the following:

$$C = 10^{m'} \quad (7)$$

with

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

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

2) *Algorithm WA2*: Let us begin with the untransformed values. The first step of Algorithm WA2 [see Fig. 7(b)] consists of bringing down SeaWiFS' chlorophyll pixels to the resolution of MODIS/Aqua's. Considering the location of the chlorophyll pixels of each sensor (Fig. 1), SeaWiFS' candidates are obtained as the following:

$$\begin{aligned} S1' &= \frac{1}{4}(S1 + S2 + S4 + S5) \\ S2' &= \frac{1}{2}(S2 + S5) \\ S3' &= \frac{1}{4}(S2 + S3 + S5 + S6) \\ S4' &= \frac{1}{2}(S4 + S5) \\ S5' &= S5 \\ S6' &= \frac{1}{2}(S5 + S6) \\ S7' &= \frac{1}{4}(S4 + S5 + S7 + S8) \\ S8' &= \frac{1}{2}(S5 + S8) \\ S9' &= \frac{1}{4}(S5 + S6 + S8 + S9). \end{aligned} \quad (8)$$

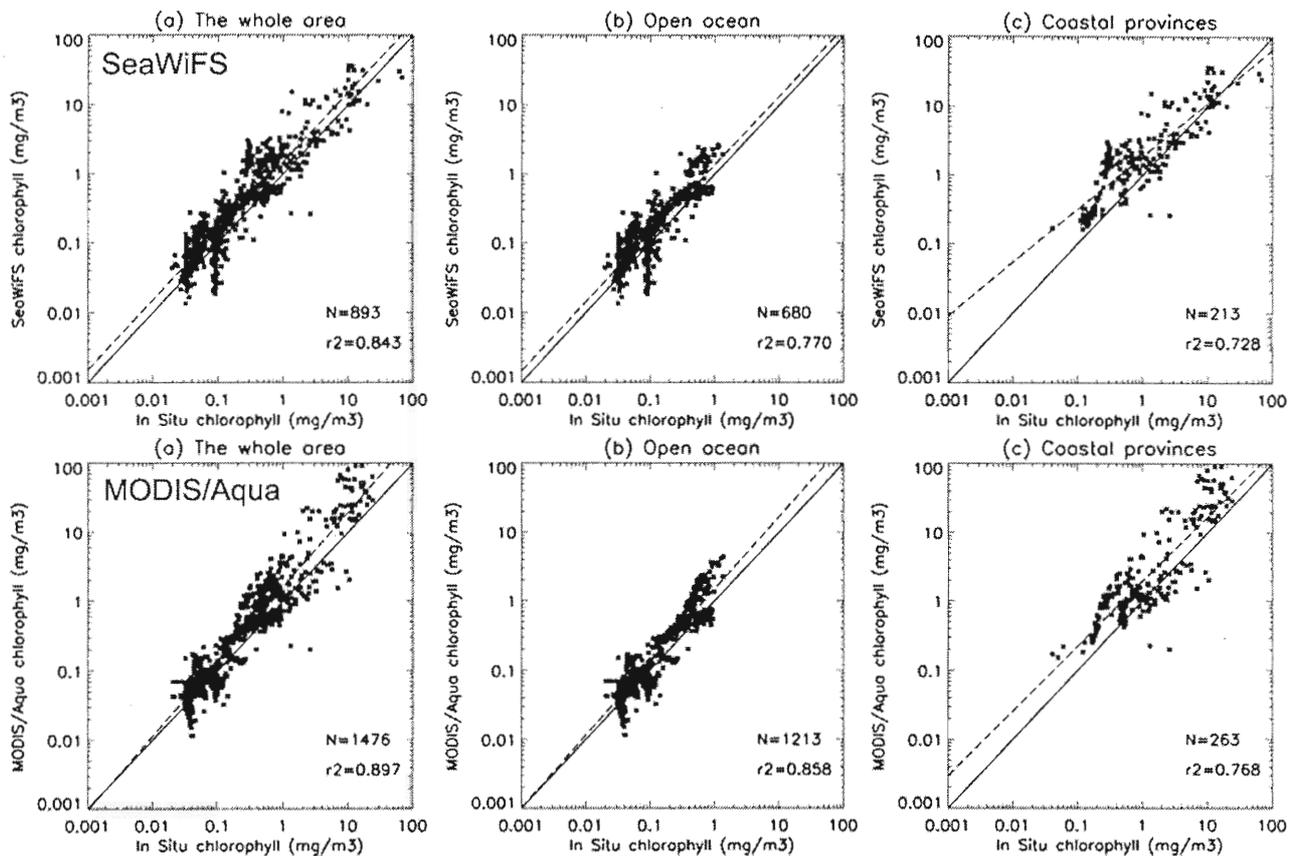


Fig. 4. Scatterplots of (top) *in situ*/SeaWiFS and (bottom) *in situ*/MODIS/Aqua chlorophyll data into (left) global ocean, (middle) open ocean (provinces No. 4, 5, 6, 7, 8, 10, and 18), and (right) coastal area (provinces No. 11, 12, 14, and 15).  $N$  is the number of points used for the computation;  $r^2$  is the coefficient of determination. The dark line is the ideal regression line; the dotted one is the real regression line.

The associated errors are obtained by the same way (just replace “S” by “ES”). The combination of SeaWiFS’ and MODIS/Aqua’s candidates is the same as for Algorithm WA1.

As far as log-transformed values are concerned, (8) are derived using the same techniques as those used in Algorithm WA1; the combination is then similar.

## B. Results

Algorithms WA1 and WA2 were applied on the North Atlantic basin for the whole year 2003. Both algorithms were used on untransformed and log-transformed values. For each of the four cases, matchups between the combined and *in situ* data were examined, and rms errors and biases were computed on untransformed and log-transformed values. In each case, the values of the rms error and of the bias range between those of SeaWiFS and MODIS/Aqua used alone. RMS errors and biases computed on chlorophyll and on log chlorophyll show that for Algorithm WA1 (combined data at SeaWiFS’ resolution) and log-transformed chlorophyll, the results are better than when using untransformed data. The results are opposite for Algorithm WA2 (combined data at MODIS/Aqua’s resolution) (Table III). They also indicate that Algorithm WA2 gives better results than Algorithm WA1. Therefore, the application of Algorithm WA2 on untransformed values gives the lowest rms errors and biases.

The application on untransformed values comprises computations of many lognormal distributions, therefore needing a large amount of CPU time. Therefore, the application on log-transformed values is the fastest one, with duration up to 10 min with a 700-MHz processor for both algorithms for one daily map on the North Atlantic basin. The application on untransformed values is quite slower: 10 to 20 min (Algorithm WA1) and 20 to 30 min (Algorithm WA2).

As an example, Fig. 8 shows the result of the application of Algorithm WA2 on the untransformed values for April 1, 2003 (day for which the spatial coverage is the worst, left panel) and August 13, 2003 (day for which the spatial coverage is the best, right panel). The spatial coverage of the combined data is 15% and 34%, which is an improvement of 3% and 11%, respectively, compared to the initial SeaWiFS coverage.

## IV. OBJECTIVE ANALYSIS

The second approach studied is based on optimal interpolation used in oceanography by Bretherton *et al.* [15]. This methodology has been already successfully applied to satellite altimetric data in various applications [16].

Its aim is to determine the value of a field  $\theta$  at a point in space and time given various measurements of the field unevenly spread over space and time  $\Phi_{\text{obs},i}$  (with  $i \in [1, N]$ ). It computes an interpolated grid-point value as a weighted-linear

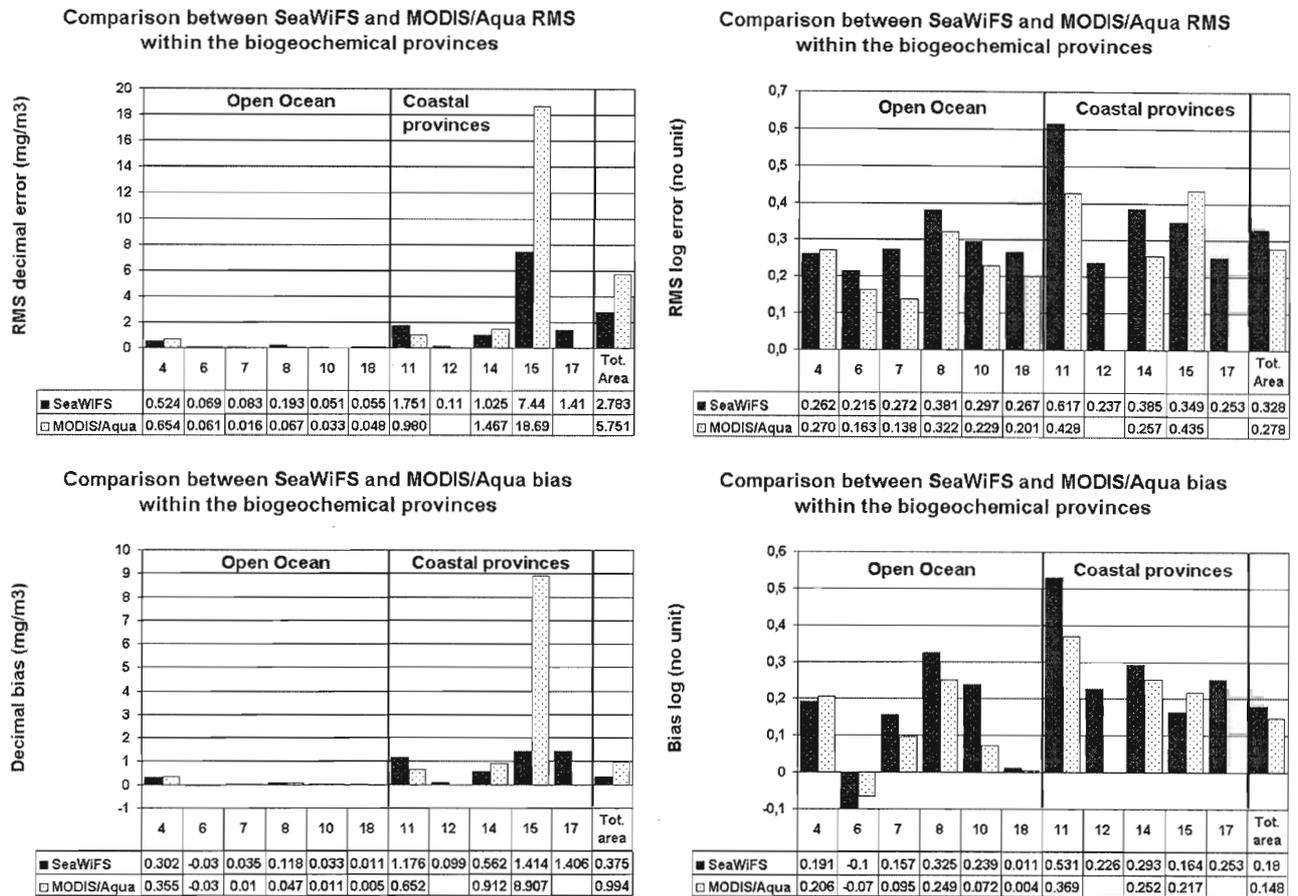


Fig. 5. (Top) RMS error computed on (left) chlorophyll and (right) log chlorophyll between *in situ* data and SeaWiFS (black) and MODIS/Aqua (dotted) chlorophyll datasets for the 11 provinces and the whole basin. (Bottom) Biases computed on (left) chlorophyll and (right) log chlorophyll.

TABLE II  
BIOGEOCHEMICAL PROVINCES IN THE NORTH AND EQUATORIAL ATLANTIC BASIN [21]. IN THE CLASSIFICATION OF THE ECOLOGICAL GEOGRAPHY OF THE SEA SUGGESTED BY LONGHURST [21], FOUR BIOMES APPEAR: THE COASTAL BIOME, THE POLAR BIOME, THE WESTERLIES BIOME, AND THE TRADE-WIND BIOME

Province #	Biome	Province	Association with province #
1	Polar	Boreal Polar Province	15
2	Polar	Atlantic Arctic Province	4
3	Polar	Atlantic Subarctic Province	4
4	Westerlies	North Atlantic Drift Province	x
5	Westerlies	Gulf Stream Province	x
6	Westerlies	North Atlantic Subtropical Gyral Province (West)	x
7	Trades	North Atlantic Tropical Gyral Province	x
8	Trades	Western Tropical Atlantic Province	x
9	Trades	Eastern Tropical Atlantic Province	8
10	Trades	South Atlantic Gyral Province	x
11	Coastal	NE Atlantic Shelves Province	x
12	Coastal	Canary Coastal Province	11 (MODIS/Aqua) – x (SeaWiFS)
13	Coastal	Guinea Current Coastal Province	11 (MODIS/Aqua) – 12 (SeaWiFS)
14	Coastal	Guianas Coastal Province	x
15	Coastal	NW Atlantic Shelves Province	x
17	Trades	Caribbean Province	14
18	Westerlies	North Atlantic Subtropical Gyral Province (East)	x
20	Coastal	Brazil Current Coastal Province	14
22	Coastal	Benguela Current Coastal Province	11 (MODIS/Aqua) – 12 (SeaWiFS)

The last column shows the association chosen between biogeochemical provinces where there are matchups (indicated by “x”) and where there is none.

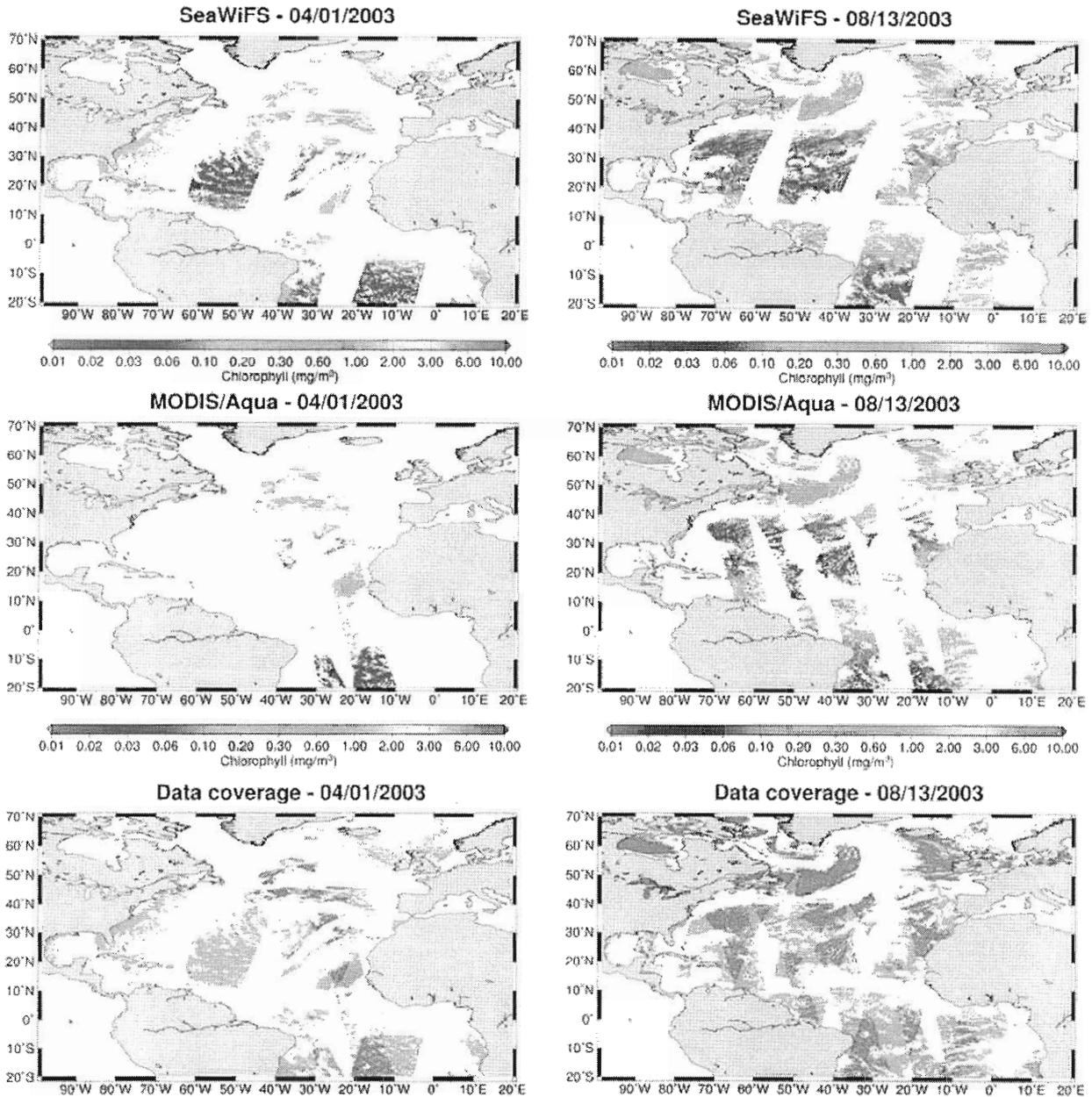


Fig. 6. (Top) SeaWiFS, (middle) MODIS/Aqua, and (bottom) a map of the data coverage of the two sensors for (left) April 1, 2003 and (right) August 13, 2003. Clear gray is for values lower than  $0.01 \text{ mg/m}^3$ , and dark gray is for values greater than  $10 \text{ mg/m}^3$ . On the maps of the data coverage, white is for no sensor, blue is for MODIS/Aqua only, green is for SeaWiFS only, and red is for both sensors.

combination of observations. The weights are optimal and take into account the spatial distribution of observations relative to one another (spatial correlation), the instrument errors and biases (that we have computed from the matchups in Section II), the variance of the field, etc. Two observed values at corresponding space/time coordinates do not have to be identical, and an interpolated variable of a field does not have to be identical with an observed value at corresponding space/time coordinates. Therefore, this method compensates for sensor-to-sensor differences in instrument design and characteristics, calibration peculiarities, and data processing.

#### A. Description

The estimated chlorophyll field is computed as follows:

$$\theta_{\text{est}}(x) = \sum_{i=1}^N \sum_{j=1}^N A_{ij}^{-1} C_{xj} \Phi_{\text{obs}^i} \quad (9)$$

with  $\Phi_{\text{obs}^i} = \Phi_i + \varepsilon_i$ , which is the chlorophyll value measured by the sensor,  $\Phi_i$  is the *true* value of chlorophyll,  $\varepsilon_i$  is the measurement error, and  $A_{ij}$  is the covariance matrix between the observations:

$$A_{ij} = \langle \Phi_{\text{obs}^i} \Phi_{\text{obs}^j} \rangle = \langle \Phi_i \Phi_j \rangle + \langle \varepsilon_i \varepsilon_j \rangle \quad (10)$$

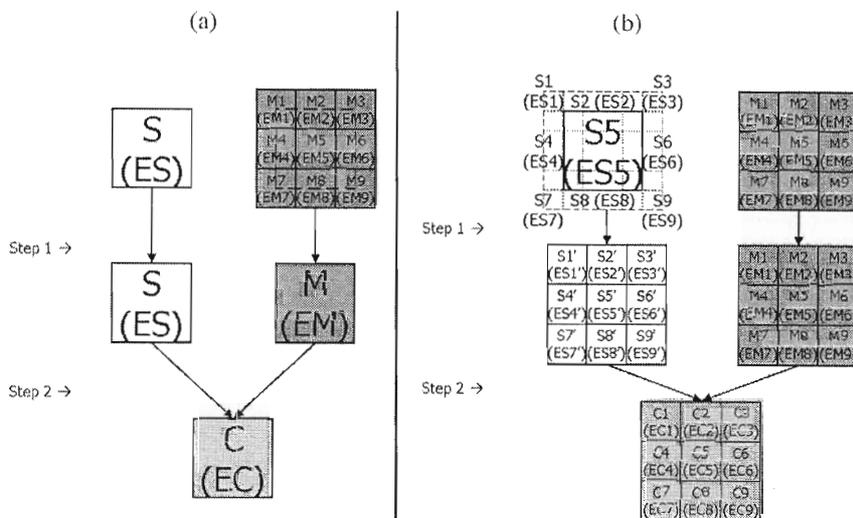


Fig. 7. (a) Description of the first algorithm of weighted averaging.  $X$  and  $EX$  are the pixel's chlorophyll concentration and the associated error, respectively (in milligrams per cubic meter with untransformed values or in no unit with log-transformed values).  $X$  is "S" for SeaWiFS, "M" for MODIS/Aqua, and "C" for the combined data. (b) Description of the second algorithm of weighted averaging: Same notations as for (a).

TABLE III  
COEFFICIENTS OF DETERMINATION  $r^2$ , RMS ERRORS, AND BIASES BETWEEN *In Situ* AND COMBINED CHLOROPHYLL OBTAINED BY WEIGHTED AVERAGING (WA1 AND WA2), AND BY OBJECTIVE ANALYSIS IN THE NORTH ATLANTIC BASIN

	N	Stat. on chlorophyll (mg/m <sup>3</sup> )			Stat. on log chlorophyll (no unit)		
		$r^2$	RMS	Bias	$r^2$	RMS	Bias
SeaWiFS	893	0.53	2.78	0.38	0.84	0.33	0.18
MODIS/Aqua	2249	0.63	5.75	0.99	0.88	0.28	0.15
Algorithm Chlorophyll	1288	0.47	5.48	0.83	0.88	0.29	0.16
WA1 Log chlorophyll	1288	0.47	5.22	0.76	0.88	0.29	0.16
Algorithm Chlorophyll	2268	0.47	3.76	0.53	0.86	0.30	0.18
WA2 Log chlorophyll	2268	0.47	3.96	0.55	0.86	0.30	0.17
Objective Chlorophyll	2261	0.58	3.58	0.60	0.85	0.34	0.21
analysis Log chlorophyll	2283	0.63	3.70	0.57	0.86	0.32	0.20

Statistics for SeaWiFS and MODIS/Aqua alone are reported in top two lines. N is the number of points used for the computation.

and  $C_{xj}$  is the covariance vector between the observations and the point to be estimated as follows:

$$C_{xj} = \langle \theta(x) \Phi_{\text{obs}^j} \rangle = \langle \theta(x) \Phi_j \rangle. \quad (11)$$

The variance of the error associated with the estimated chlorophyll field is given by

$$e^2 = C_{xx} - \sum_{i=1}^N \sum_{j=1}^N C_{xi} C_{xj} A_{ij}^{-1}. \quad (12)$$

The objective analysis used in [16] decomposes the errors into white noise and biases by expressing  $\langle \varepsilon_i, \varepsilon_j \rangle$  in the following forms:

- 1)  $\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{ij} b^2$  for points  $(i, j)$  from different sensors;
- 2)  $\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{ij} b^2 + E$  for points  $(i, j)$  from the same sensor.

$b^2$  is the variance of the measurement noise (i.e., the square of the rms errors computed in Section II), and  $E$  is the variance of the bias (i.e., the square of the mean ADs in Section II).

Since there are a high number of measurements, the interpolation is done with observations included in an "influence

bubble," whose size is defined by the zonal  $R_x$  and meridional  $R_y$  correlation scales.

### B. Pretreatment

1) *Data*: The practical requirement for the use of objective analysis is that a "first guess" field of the signal should exist, which the algorithm corrects by interpolating the input signal. The input data of the algorithm  $\Phi_{\text{obs}^i}$  are chlorophyll anomalies, i.e., the chlorophyll field minus this "first-guess." Different options are possible for this background field: it can be the previous-day chlorophyll map, the mean of the last-week daily chlorophyll maps, or the monthly climatology. Reynolds and Smith [24] have already addressed this problem of the choice of the background field for the application of objective analysis on sea-surface temperature (SST). A monthly climatology and the previous-day map for SST are both solutions with advantages and drawbacks. The preceding analysis has the advantage to be more accurate than the climatology. However, the climatology is more homogeneous in terms of statistics. This led us to choose the climatology solution. First, because the procedure to estimate the correlation function is more robust and easier to

develop, and second, because it is wiser to use a first guess with a coverage as extended as possible.

This monthly mean field is computed by using a one-year Fourier filtering on the 1997–2003 SeaWiFS monthly data. Indeed, this method only takes into account the general yearly trend of the signal and not the values of outliers that might occur once. The climatology for April and August is shown in Fig. 9.

All the following treatments have been computed on untransformed anomalies (chlorophyll minus monthly climatology) and log-transformed anomalies (logarithm to the base 10 of the chlorophyll minus logarithm to the base 10 of the monthly climatology).

2) *Variance*: The temporal variance was computed on untransformed and log-transformed chlorophyll values, as shown in Fig. 10. To avoid marked differences between neighboring points, which are inadequate for the objective analysis, a 50-km median filter was applied to both fields. Fig. 10 shows that the variance is relatively weak in the low-productivity oligotrophic regions such as the subtropical gyres and high everywhere else on the continental margins, in the subpolar gyre, near the Amazon Plume, near the upwelling of Mauritania, and in the equatorial zone. The variance of the untransformed values [Fig. 10(a)] exhibits sharper gradients than the variance of the log-transformed values [Fig. 10(b)].

3) *Correlation Function*: The empirical space–time correlation function has been first computed on untransformed anomalies in each point of a  $2^\circ \times 2^\circ$  grid in the North Atlantic basin using (13). Only the values included in a subdomain centered around each location studied are considered. The computation is done only if at least 50% of the subdomain is filled.

$$\text{CORR}_{\text{emp}}(\Delta\text{long}, \Delta\text{lat}, \Delta t) = \frac{\sum_{\text{map}=1}^{\text{NumberOfMaps}} \frac{\sum_{\text{pair}=1}^{\text{NumberOfPairs}} \text{PROD}(\text{map}, \text{pair}, \Delta\text{long}, \Delta\text{lat}, \Delta t)}{\text{VAR}(\text{subdomain}) \times \text{NumberOfPairs}}}{\text{NumberOfMaps}} \quad (13)$$

where  $\Delta\text{long} = \text{long}2 - \text{long}1$ ,  $\Delta\text{lat} = \text{lat}2 - \text{lat}1$ ,  $\text{pair} = (\text{long}1, \text{lat}1, 0) \leftrightarrow (\text{long}2, \text{lat}2, \Delta t)$ ; *NumberOfPairs* is the number of pairs of values included in the subdomain; and *NumberOfMaps* is the number of maps of chlorophyll in the time series used. The PROD product is given by

$$\begin{aligned} \text{PROD}(\text{map}, \text{pair}, \Delta\text{long}, \Delta\text{lat}, \Delta t) \\ = [\text{chl}(\text{map}, \text{long}1, \text{lat}1, 0) - \text{mean}(\text{subdomain})] \\ \times [\text{chl}(\text{map}, \text{long}2, \text{lat}2, \Delta t) - \text{mean}(\text{subdomain})] \end{aligned}$$

where chl stands for chlorophyll concentration and the variance of the subdomain, shown at the bottom of the page, where *Nlong*, *NLat*, and *NTps* are the numbers of points in longi-

tude, latitude, and time of the subdomain, respectively, and *NumberOfPoints* is the number of the points included in the subdomain.

The choice of the size of the subdomain for such computation is important. A small subdomain gives a very local information. A larger area gives a smoother information but a more accurate one because the computation is performed with more pairs. To have a first idea of the subdomain size, we use space–time correlation scales found in previous studies. Using the semivariogram approach from geostatistics on Level 3 daily standard mapped images (reprocessing 2) for the year 1998, Doney *et al.* [25] found for the global ocean a maximum zonal correlation scale of 250 km and a maximum meridional correlation scale of 200 km. The dimensions  $250 \times 200$  km of this subdomain are kept constant all over the North Atlantic, and we compute our empirical correlation scales over this subdomain size. Uz and Yoder [26] found low correlation between pairs of images as short as one-day apart. This allows us to work only in two dimensions (longitude, latitude) and then obtain a space correlation function. We also computed the correlation function with a time scale of five days but the results were not convincing. The correlation coefficients have been computed per month and for the whole year to examine a possible seasonal variation of the correlation function.

To model the empirical correlation function, two functions were studied.

First, let us examine the exponential model used by Kwiatkowska and Fargion [12]. The basis function is the following:

$$\text{CORR}_{\text{exp}}(r) = A_1 \exp(A_2 r) + A_3 \quad (14)$$

where  $r = \sqrt{(x/R_x)^2 + (y/R_y)^2}$ ;  $R_x$  and  $R_y$  are the zonal and meridional correlation radii (first zero crossing of  $\text{CORR}_{\text{exp}}$  with the zonal and meridional axes, respectively), and  $A_1$ ,  $A_2$ , and  $A_3$  are the coefficients to determine. The function must complete the following conditions: strictly decreasing,  $\text{CORR}_{\text{exp}}(0) = 1$ , and  $\text{CORR}_{\text{exp}}(1) = 0$  (i.e., null correlation at the first zero crossing). Therefore, we obtain

$$\text{CORR}_{\text{exp}}(r) = (1 - \alpha) \left( \frac{\alpha}{\alpha - 1} \right)^r + \alpha \quad (15)$$

with  $\alpha < 0$ . The smaller is the  $\alpha$ , the more sharp pointed is the function.

The second model is based on an inverse function ( $A_1$ ,  $A_2$ ,  $A_3$ , and  $A_4$  are coefficients to determine)

$$\text{CORR}_{\text{exp}}(r) = \frac{A_1}{A_2 r + A_3} + A_4 \quad (16)$$

$$\text{VAR}(\text{subdomain}) = \frac{\sum_{\text{long}=1}^{\text{NLong}} \sum_{\text{lat}=1}^{\text{NLat}} \sum_{\text{tps}=1}^{\text{NTps}} [\text{chl}(\text{map}, \text{long}, \text{lat}, \text{tps}) - \text{mean}(\text{subdomain})]^2}{\text{NumberOfPts}}$$

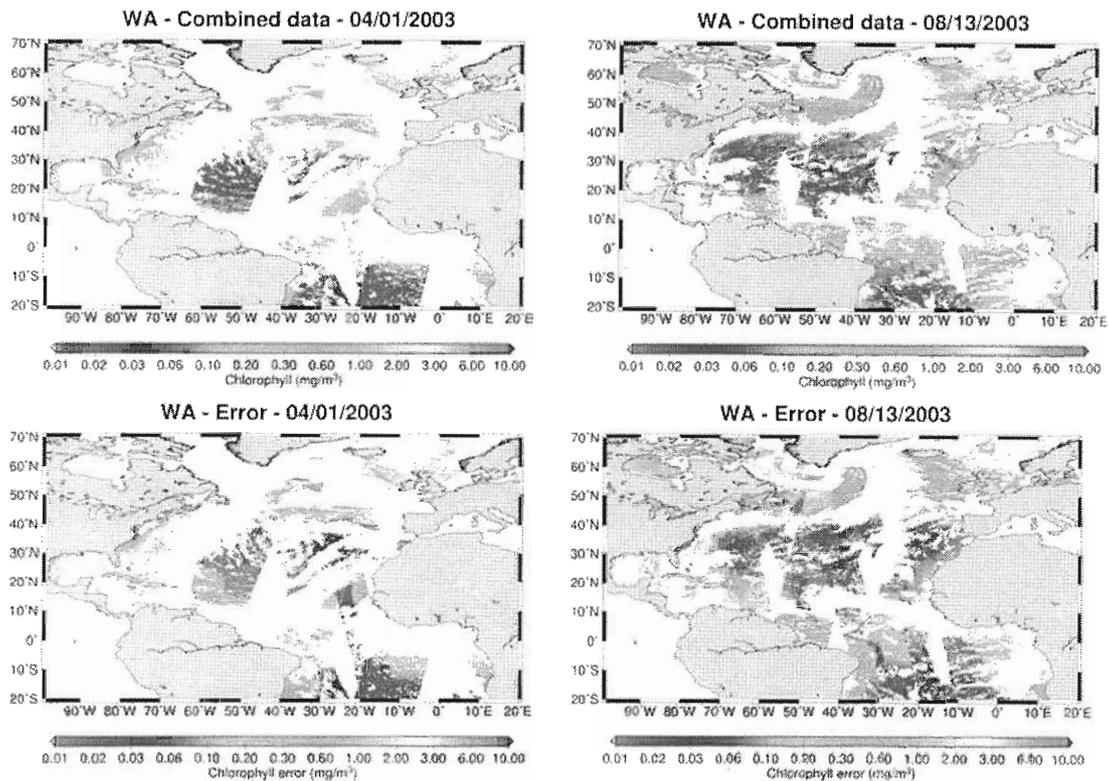


Fig. 8. Weighted averaging. Results of the application of Algorithm WA2 to untransformed values for (left) April 1, 2003 and (right) August 13, 2003. (Top) Map of the combined data. (Bottom) Map of the associated error. (Clear gray) Values lower than  $0.01 \text{ mg/m}^3$ . (Dark gray) Values greater than  $10 \text{ mg/m}^3$ .

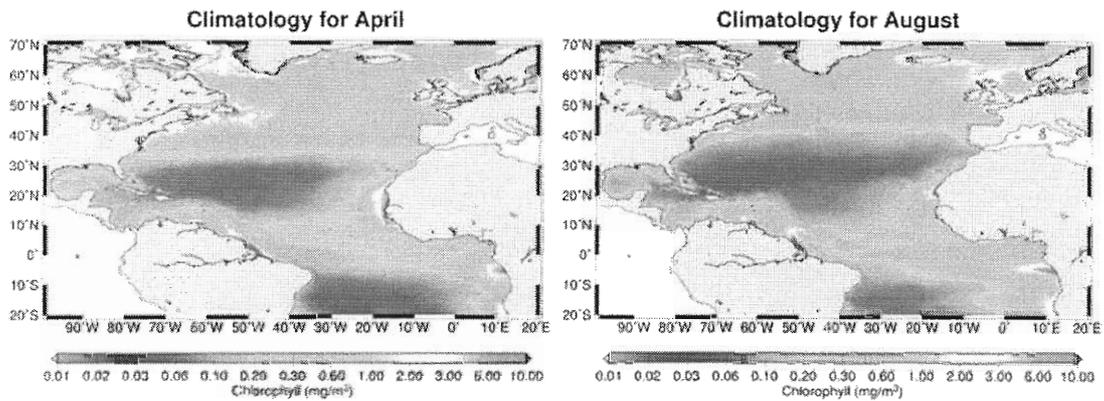


Fig. 9. Climatology for (left) April and (right) August.

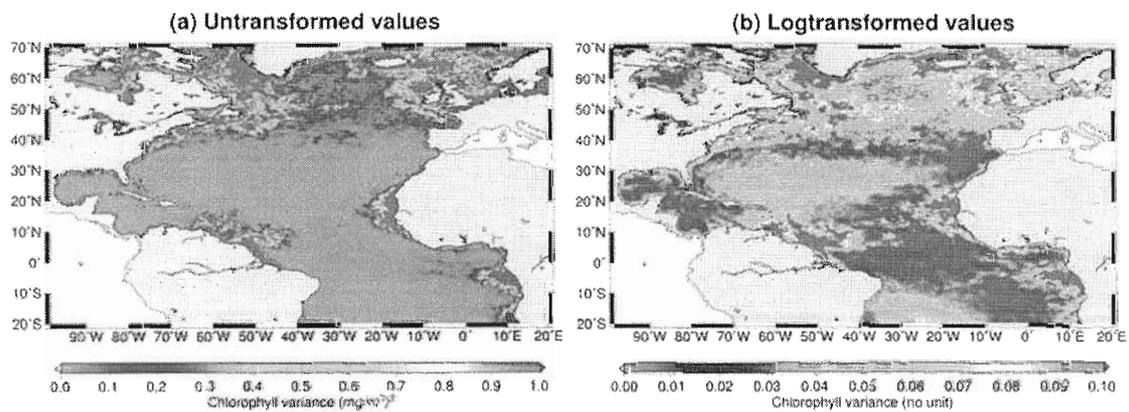


Fig. 10. Temporal variance of SeaWiFS chlorophyll for the year 2003, which is computed on (left) untransformed and (right) log-transformed values. Dark gray is for values greater than  $1 \text{ (milligrams per cubic meter)}^2$  and  $0.1$ , respectively.

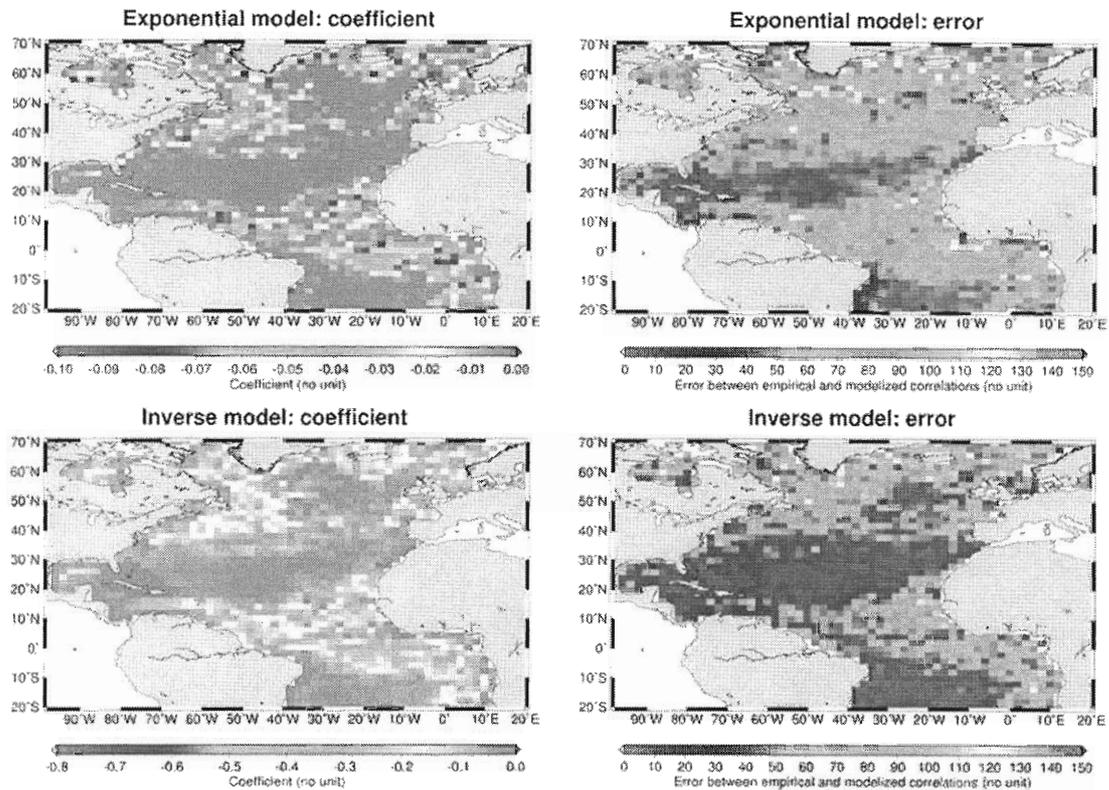


Fig. 11. (Top) Spatial distribution of (left) the coefficients  $\alpha$  of the exponential model in the North Atlantic basin and (right) the resulting absolute error from the empirical model. (Bottom) Same for the inverse model.

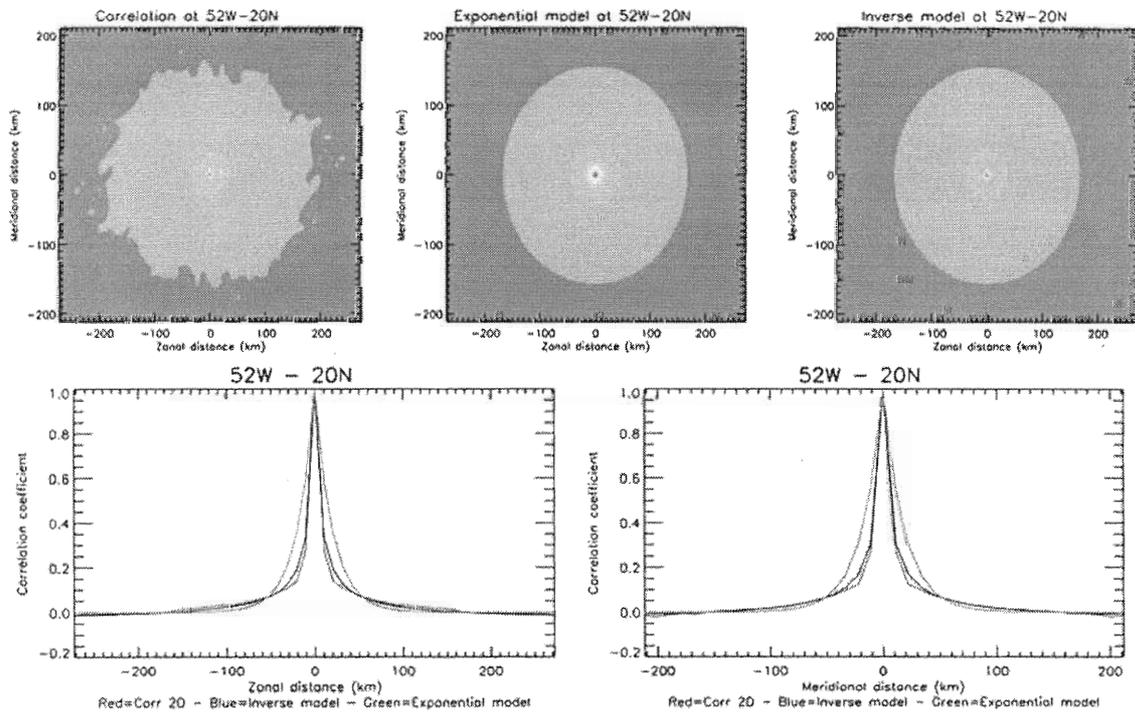


Fig. 12. Spatial correlation coefficients of SeaWiFS anomalies for the year 2003 at the grid point 52° W-20° N (Province 6). (Top) Value of the correlation coefficients versus zonal and meridional distances (in kilometers) for (left) the empirical function, (middle) the exponential model, and (right) the inverse model. Color scale: Red is for the value 1.0; dark blue is the zero field. (Bottom) Section for (left) a null meridional distance and (right) a null zonal distance; the empirical function is in red, the exponential model in green, and the inverse model in blue.

which gives the following, considering the above constraints:

$$\text{CORR}_{\text{exp}}(r) = \beta + \frac{\beta(1-\beta)}{\beta-r}. \quad (17)$$

Here, again, the smaller is the  $\beta$ , the more sharp pointed is the function.

The associated spectrum of both correlation models is a positive number for each wavenumber. Therefore, they are both nonnegative definite forms, which are a requisite [15].

At each point of the  $2^\circ \times 2^\circ$  grid,  $\alpha$  and  $\beta$  were computed so that the model correlation function gives the best fit to the empirical one. Fig. 11 shows the spatial repartition of these coefficients in the North Atlantic basin. For both models, they depend obviously on the biogeochemical provinces: In the subtropical gyres, they are close to zero, and the correlation model is very sharp pointed (i.e., the points are not well correlated between each other).

In the equatorial zone, the coefficients are higher in absolute value, and the correlation model is wider (i.e., the points are better correlated with each other). A monthly analysis reveals an interesting seasonal variation: The correlation scale is larger in summer (i.e., wide correlation model) than in winter (i.e., sharp-pointed correlation model).

Fig. 12 shows sections of the two-dimensional (2-D) correlation function at the grid point  $52^\circ \text{W} - 20^\circ \text{N}$ . Obviously, the inverse model fits better the empirical function than the exponential one. This is confirmed by the maps of the absolute error between the models and the empirical function in Fig. 11.

The same computation has been carried out on log-transformed anomalies, giving the same results.

4) *Correlation Scales*: Zonal (respectively meridional) correlation scales are given by the first zero crossing of the correlation function computed above with the zonal (respectively meridional) axis. They define the size of the influence bubble.

Fig. 13(a) displays the zonal correlation scales for the whole year 2003. They clearly depend on latitude and not on the biogeochemical-province partitioning. A study per month points out that they do not show any seasonal variation. Zonal correlation scales  $R_x$  were thus modeled with the following function:  $R_x = -0.03 \text{ lat}^2 + 220 \text{ km}$ . Thus, the coefficient of determination  $r^2$  is 0.914, showing the excellent fit of the modeled function.

Fig. 13(b) shows the meridional correlation scales for the year 2003. They depend neither on the latitude nor on the biogeochemical-province repartition. No seasonal variation emerges. The scales are close to 150 km all over the basin, and a constant value (150 km) has been used.

The time correlation scale computed (not shown) is about one day everywhere and until five days at few locations in the subtropical gyres, confirming the results in [26]. Therefore, the value of the time correlation scale has been set here to one day.

### C. Results

The objective analysis was applied on the North Atlantic basin for the whole year 2003. It was used on untransformed and log-transformed anomalies. The computation was made in order to have a combined data with a resolution of  $0.1^\circ \times 0.1^\circ$ .

Combined data obtained by the application of the objective analysis to untransformed anomalies (not shown here) are worse than those obtained by the application to log-transformed anomalies. Indeed, unrealistic values can be obtained at locations where the variance between points within the influence bubble is highly different. By this way, the covariance matrix is numerically ill conditioned, and its inversion gives a wrong result. This does not occur with lognormal anomalies because such differences do not exist in this case. Moreover, this method is based on the Gauss–Markov theorem, which is optimal only if applied to Gaussian variables. Therefore, objective analysis must be applied to log-transformed anomalies.

As for the weighted averaging, matchups between the combined and *in situ* data were examined for both cases, and rms error and bias were computed on untransformed and log-transformed chlorophyll. Coefficients of determination, rms errors, and biases computed on anomalies and log anomalies (Table III) confirm that better results are obtained by applying the objective analysis to log-transformed than to untransformed anomalies.

Naturally, for operational purposes, it is impossible to use all the valid points (up to 2000) in the influence bubble. Therefore, to decrease the CPU time, another parameter was added: the number of observations in the influence bubble considered for the computation at one given location. The CPU time increases of about 30 min with a 50-point increment. Different values were tested, and a value of 150 points was chosen. Indeed, selecting more points in the influence bubble does not change any more the chlorophyll value. The CPU time is about 1 h and 30 min for a daily map on a 700-MHz processor, depending clearly on the season (because of the number of points on the original grids).

As an example, Fig. 14 shows the result of the application of the objective analysis on log-transformed anomalies for April 1, 2003 (day for which the spatial coverage is the worst, left panel) and August 13, 2003 (day for which the spatial coverage is the best, right panel). The spatial coverage of the combined data is 50% and 71%, which is an improvement of 38% and 48%, respectively, compared to the initial SeaWiFS coverage.

## V. DISCUSSION AND CONCLUSION

The objective here was to present and test two approaches of ocean color data merging for operational purposes. Both were applied on the North and Equatorial Atlantic basins for the combination of SeaWiFS and MODIS/Aqua for the year 2003. Input values were untransformed and log-transformed chlorophyll-concentration values.

The first approach is an error-weighted averaging. Even considered as basic, it improves the spatial coverage by taking into account only the existing satellite values. It consumes low CPU time. Its main limitation is the discontinuity in the accuracy of the combined data, as shown in Fig. 8. If there is only one observed value, the accuracy of the combined value is the accuracy of this sensor's value. If there are two estimates (i.e., an observed value for both sensors), the accuracy of the combined data is improved. Its application on the North

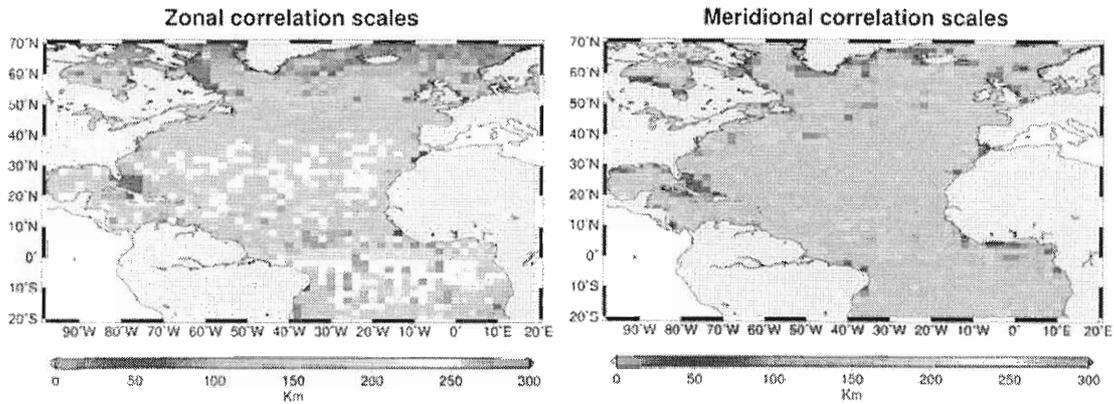


Fig. 13. Spatial repartition in the North Atlantic basin of (left) the zonal and (right) the meridional correlation scales.

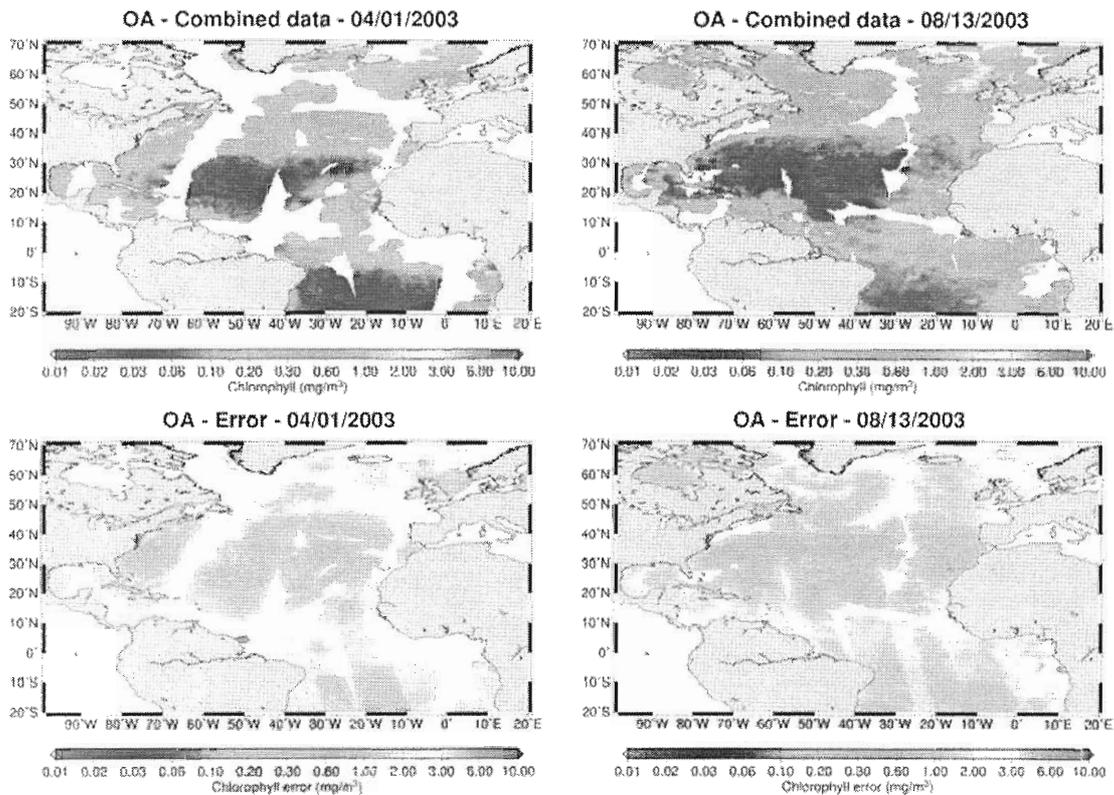


Fig. 14. Objective analysis. Results of the application to logtransformed values for (left) April 1, 2003 and (right) August 13, 2003. (Top) Map of the combined data. (Bottom) Map of the associated error. Clear gray is for values lower than  $0.01 \text{ mg/m}^3$ , and dark gray is for values greater than  $10 \text{ mg/m}^3$ .

Atlantic basin shows that the quality of the combined data does not differ much between the applications on untransformed and log-transformed values.

The second approach is an objective analysis. Given a first guess and statistical parameters computed directly on satellite data, this approach interpolates values where there is none. As a result, the spatial coverage of the combined data is improved compared to the error-weighted averaging. Its application on untransformed anomalies sometimes gives poor results, especially because of the lognormal distribution and the time dynamics of the chlorophyll. Therefore, the objective analysis must be applied to log-transformed data. The results shown

here constitute a preliminary application of this method on chlorophyll-*a* data. It thus needs to be improved: tuning, better characterization of signal and noise, correlation function, selection of the observations considered for the computation, etc.

Assessing the quality of the combined data is quite difficult. The best way to do it is to compare the combined data to *in situ* measurements, as we made here. Table III shows that the rms errors and biases of the error-weighted averaging are slightly smaller than for the objective analysis. But, since the matchups are very sparse and rare in space and time, an estimation of the global quality (extreme values, mean, and standard deviation) of the combined data can also be computed.

The CPU time is much quicker for the weighted averaging than for the objective analysis because of the difference of the space coverage between both approaches. Nevertheless, the CPU time is quite low, and both techniques can be applied for operational purposes.

In conclusion, we have tested and compared two approaches for ocean color data merging of SeaWiFS and MODIS/Aqua in the North Atlantic basin for the year 2003: weighted averaging and objective analysis. The main advantage of the objective analysis is its ability to interpolate in space (and time) by taking into account the statistical properties of chlorophyll-*a*. Moreover, matchups between *in situ* and combined data show that both approaches yield about the same rms errors and biases, although the spatial coverage is at least twice as large in the case of the objective analysis as the weighted averaging. It also seems that, due to the nature of the chlorophyll distribution, it is better to use mergers on log-transformed values.

#### ACKNOWLEDGMENT

The authors would like to thank Orbimage for the SeaWiFS mission, the OceanColorWeb support, and the SeaBASS and the AMT archivers (especially G. Moncoiffé for her efficient help) of chlorophyll, and the contributors to these databases. The authors would also like to thank W. Gregg for giving access to the recent SeaBASS datasets and for the helpful comments. They would also like to thank P. Mazzega (LEGOS), S. Ruiz (CLS), and F. Mélin (JRC, Ispra, Italy) for fruitful discussions, as well as the three reviewers for the very constructive comments that improved this paper.

#### REFERENCES

- [1] H. R. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm," *Appl. Opt.*, vol. 33, no. 3, pp. 443–452, Jan. 1994.
- [2] A. Morel, D. Antoine, and B. Gentili, "Bidirectional reflectance of oceanic waters: Accounting for Raman emission and varying particle phase function," *Appl. Opt.*, vol. 41, pp. 6289–6306, Oct. 30, 2002.
- [3] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res.*, vol. 103C, no. C11, pp. 24937–24953, Oct. 1998.
- [4] W. W. Gregg, W. E. Esais, G. C. Feldman, R. Frouin, S. B. Hooker, C. R. McClain, and R. H. Woodward, "Coverage opportunities for global ocean color in a multimission era," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1620–1627, Sep. 1998.
- [5] J. Xiong, G. Toller, V. Chiang, J. Sun, J. Esposito, and W. Barnes, *MODIS Level 1B Algorithm Theoretical Basis Document*, (2005). [Online]. Available: [http://www.mcst.ssai.biz/mcstweb/info/online\\_doc.html](http://www.mcst.ssai.biz/mcstweb/info/online_doc.html)
- [6] G. Meister, E. J. Kwiatkowska, B. A. Franz, F. S. Patt, G. C. Feldman, and C. R. McClain, "Moderate-Resolution Imaging Spectroradiometer ocean color polarization correction," *Appl. Opt.*, vol. 44, no. 26, pp. 5524–5535, Sep. 2005.
- [7] C. R. McClain, W. Esais, G. Feldman, R. Frouin, W. Gregg, and S. Hooker, "The proposal for the NASA sensor intercalibration and merger for biological and interdisciplinary oceanic studies (SIMBIOS) program," NASA Goddard Space Flight Center, Greenbelt, MD, NASA TM-2002-210008, 2002.
- [8] E. J. Kwiatkowska, "Comparisons of daily global ocean color data sets: MODIS-Terra/Aqua and SeaWiFS," in *MODIS Validation, Data Merger, and Other Activities Accomplished by SIMBIOS Project: 2002–2003*, NASA Goddard Space Flight Center, Greenbelt, MD, NASA-TM-2003.
- [9] B. A. Franz, "A long-term intercomparison of oceanic optical property retrievals from MODIS-Terra and SeaWiFS," in *MODIS Validation, Data Merger, and Other Activities Accomplished by SIMBIOS Project: 2002–2003*, NASA Goddard Space Flight Center, Greenbelt, MD, NASA-TM-2003.
- [10] S. Maritorena and D. A. Siegel, "Consistent merging of satellite ocean color data sets using a bio-optical model," *Remote Sens. Environ.*, vol. 94, no. 4, pp. 429–440, Feb. 2005.
- [11] W. W. Gregg and M. E. Conkright, "Global seasonal climatologies of ocean chlorophyll: Blending *in situ* and satellite data for the coastal zone color scanner era," *J. Geophys. Res.*, vol. 106, no. C2, pp. 2499–2516, Feb. 2001.
- [12] E. J. Kwiatkowska and G. S. Fargion, "Merger of ocean color data from multiple satellite missions within the SIMBIOS project," in *Proc. SPIE Symp.—Remote Sensing Atmosphere, Ocean, Environment and Space*, 2002, pp. 168–182. Invited Paper.
- [13] ———, "Merger of ocean color data from multiple satellite missions within the NASA SIMBIOS project office," in *Proc. 5th IEEE Int. Conf. Inf. Fusion*, 2002, vol. 1, pp. 291–298.
- [14] ———, "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 Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2844–2860, Dec. 2003.
- [15] F. P. Bretherton, R. E. Davis, and C. B. Fandry, "A technique for objective analysis and design of oceanographic experiments applied to MODE-73," *Deep-Sea Res.*, vol. 23, no. 7, pp. 559–582, Jul. 1976.
- [16] P.-Y. Le Traon, F. Nadal, and N. Ducet, "An improved mapping method of multisatellite altimeter data," *J. Atmos. Ocean. Technol.*, vol. 15, no. 2, pp. 522–534, Apr. 1998.
- [17] J. W. Campbell, "The lognormal distribution as a model for bio-optical variability in the sea," *J. Geophys. Res.*, vol. 100, no. C7, pp. 13237–13254, Jul. 1995.
- [18] P. J. Werdell and S. W. Bailey, "The SeaWiFS bio-optical archive and storage system (SeaBASS): Current architecture and implementation—Introduction," NASA Goddard Space Flight Center, Greenbelt, MD, NASA Tech. Memo 2002-211617, 2002.
- [19] D. B. Robins and J. Aiken, "The Atlantic Meridional Transect: An oceanographic research programme to investigate physical, chemical, biological and optical variables of the Atlantic Ocean," *Underwater Technol.*, vol. 21, no. 4, pp. 8–14, 1996.
- [20] W. W. Gregg and N. W. Casey, "Global and regional evaluation of the SeaWiFS chlorophyll data set," *Remote Sens. Environ.*, vol. 93, no. 4, pp. 463–479, Dec. 2004.
- [21] A. Longhurst, *Ecological Geography of the Sea*. San Diego, CA: Academic, 1998, p. 398.
- [22] B. A. Franz, P. J. Werdell, G. Meister, S. W. Bailey, R. E. Eplee, G. C. Feldman, E. Kwiatkowska, C. R. McClain, F. S. Patt, and D. Thomas, "The continuity of ocean color measurements from SeaWiFS to MODIS," *Proc. SPIE*, vol. 5882, pp. 1–13, 2005.
- [23] G. Zibordi, F. Mélin, and J. F. Berthon, "Comparison of SeaWiFS, MODIS and MERIS radiometric products at a coastal site," *Geophys. Res. Lett.*, vol. 33, no. 6, L06617, DOI: 10.1029/2006GL025778, 2006.
- [24] R. W. Reynolds and T. M. Smith, "Improved global sea surface temperature analyses," *J. Clim.*, vol. 7, no. 6, pp. 929–948, Jun. 1994.
- [25] S. C. Doney, D. M. Glover, S. J. McCue, and M. Fuentes, "Mesoscale variability of Sea-viewing Wide Field-of-view Sensor (SeaWiFS) satellite ocean color: Global patterns and spatial scales," *J. Geophys. Res.*, vol. 108, no. C2, 3024, DOI: 10.2912/2001JC000843, 2003.
- [26] B. M. Uz and J. A. Yoder, "High frequency and mesoscale variability in SeaWiFS chlorophyll imagery and its relation to other remotely sensed oceanographic variables," *Deep-Sea Res. II*, vol. 51, no. 10/11, pp. 1001–1017, May/June 2004.



**Claire Pottier** received the engineering degree from the Ecole Supérieure d'Electronique de l'Ouest, Angers, France, in 2003, and the diploma in signal and image processing from the University of Rennes, Rennes, France. She is currently working toward the Ph.D. degree at the University of Toulouse, Toulouse, France.

Since 2003, she has been working with Collecte Localisation Satellites (CLS) and LEGOS, Toulouse. Her research interests are in the areas of signal and image processing adapted to oceanographic data.



**Véronique Garçon** received the Ph.D. thesis degree in energy and environmental pollution from the University of Paris VII, Paris, France, in 1981, and the Habilitation à Diriger les Recherches in oceanography from Toulouse University, Toulouse, France, in 1995.

She is currently a Centre National de la Recherche Scientifique (CNRS) Senior Scientist and Head of the Group Physical Dynamics/Marine Biogeochemistry with the LEGOS Laboratory, Toulouse, France. Her research interests are in the areas of marine biogeochemical cycles and dynamics of ecosystems in response to climate changes. This implies activities in remote sensing, coupled physical/biogeochemical modeling, analysis of *in situ* data, data assimilation, and experimental work at sea.



**Gilles Larnicol** received the engineering degree from the Ecole Nationale Supérieure de Techniques Avancées, Paris, France, in 1993, and the Ph.D. degree from the University of Bretagne Occidentale, Brest, France, in the field of ocean data analyses, in 1998.

Since 1998, he has been working with Collecte Localisation Satellites (CLS) Space Oceanography Division, France. He has experiences both in operational and research oceanography. He is currently the Head of the Oceanography Department of the Space Oceanography Division.



**Joël Sudre** received the Master's degree from the University of Méditerranée, Aix-Marseille II, France.

He is currently a CNRS Engineer with the LEGOS Laboratory, Toulouse, France, working on satellite data analysis and image processing for oceanographic studies.



**Philippe Schaeffer** received the engineering degree from the Institut National des Sciences Appliquées (ENSAIS), Strasbourg, France, which is a French superior engineering school, in 1991.

He then worked six years with the French National Research Council (CNRS) as a Research Engineer in the domains of orbitography, geodynamics, and satellite altimetry. Since 1997, he has been a Research Engineer with the CLS Space Oceanography Unit, Ramonville-Saint-Agne, France.



**Pierre-Yves Le Traon** graduated the French Civil Engineering School of Saint Etienne in 1984. He received the Master thesis degree in dynamical and coastal oceanography of Brest, in 1985, and the Ph.D. degree in physical oceanography from Toulouse University, Toulouse, France, in 1990.

He is formerly the Vice-Director of the CLS Space Oceanography Division. He is now Program Director for operational oceanography systems with the Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Plouzané, France. His fields of interest include operational oceanography, data assimilation, and *in situ* and satellite ocean observing systems.