

# Regional Objective Analysis for Merging High-Resolution MERIS, MODIS/Aqua, and SeaWiFS Chlorophyll-*a* Data From 1998 to 2008 on the European Atlantic Shelf

Bertrand Saulquin, Francis Gohin, and René Garrello

**Abstract**—In this paper, we define the method that is used to merge high-resolution multisensor chlorophyll-*a* (chl-*a*) data on the Ireland–Biscay–Iberia Regional Ocean Observing System area from 1998 to the present at a resolution of 1.1 km. The method is based on geostatistics and is known as kriging. The merged variable is the daily anomaly of chl-*a*, with the anomaly being defined as the difference between the daily image and the mean historical field for the considered day. For each day, the continuous anomaly image is generated using the kriging method, and the mean historical field is then added to obtain the cloudless field of chl-*a*. The initial satellite chl-*a* data set used in the merging procedure is derived from the daily level-2 water leaving radiances of three ocean color sensors: the Sea-Viewing Wide Field of View Sensor on the Orbview platform, the Moderate Resolution Imaging Spectroradiometer on the Aqua platform, and the Medium Resolution Imaging Spectrometer Instrument on the ENVISAT platform. The chl-*a* concentration is obtained using a specific algorithm developed by Ifremer, known as OC5 product. Before merging, each satellite-derived chl-*a* product has been compared to *in situ* data and has been validated using a matchup data set. After this validation against *in situ* data, intercomparisons between the satellite data sets have been performed. As the chl-*a* anomaly variability depends, in this region, on the season and the distance from the shore, local space-time semivariograms have been calculated to estimate the spatiotemporal dependence or covariance of the chlorophyll anomalies. The semivariograms, used in the estimation of the kriged anomaly, are defined by their nuggets (noise), their spatial and temporal range (maximum distance for a nonnull covariance between the anomalies), and their sill (maximum variance). The spatial range of the semi-

variogram has been approximated locally on a regular grid. The nuggets and the sills have been deduced from the square of the mean of the chl-*a* concentration (the climatological reference) as we have observed a classical proportionality effect between the square of the chl-*a* mean, the variance of the distribution, and the parameters of the semivariograms. Compared to each original product, the analysis shows a complete coverage and differences with the *in situ* data that are statistically equivalent to those observed with the initial satellite data set. The merged product offers a number of applications for environmental monitoring such as the monitoring of the eutrophication risk required by the Water Framework Directive of the European Union.

**Index Terms**—Chlorophyll-*a* (chl-*a*), kriging, optimal interpolation, remote sensing.

## I. INTRODUCTION

OCEAN color satellite sensors provide a unique view of large-scale ocean biological processes. Chlorophyll-*a* (chl-*a*, which is the main pigment in phytoplankton) is a good indicator of primary production, which is the base of the food chain. Its concentration can be determined from the observations of space-based ocean color sensors. For ocean color purposes, sensors measure top-of-atmosphere radiances at different wavelengths in the visible spectrum between 400 and 700 nm. The contribution from the atmosphere is first removed from the top-of-atmosphere radiance through a process known as atmospheric correction [1] to obtain the water-leaving radiance ( $L_w$ ). The  $L_w$ s are then normalized (i.e., expressed in standard solar conditions (sun at zenith) in the absence of the atmosphere) and corrected from the bidirectional effects (viewing angle dependence and effects of seawater anisotropy [2]) to obtain the normalized water-leaving radiances ( $nL_{ws}$ ). From an optical perspective, in addition to pure water, the observed  $nL_{ws}$  spectrum is constrained by the concentration of the following three components [3]: pigments, such as chl-*a*, dissolved yellow substances (gelbstoff), and inorganic suspended matters. To retrieve the chl-*a* concentration from  $nL_{ws}$ , some analytical or empirical algorithms are used. In open sea and clear waters (defined as Case 1 waters), the algorithms are now considered as accurate. In coastal areas where the concentration of the suspended matters and yellow substances

Manuscript received April 16, 2009; revised November 9, 2009 and April 7, 2010. This work was supported in part by the European COastal sea OPerational observing and forecasting system (ECOOP) under Contract No. 36355, which is funded by the European Commission's Sixth Framework Programme, under the priority Sustainable Development, Global Change and Ecosystems, and in part by the MarCoast project, which is a Global Monitoring for Environment and Security Service Element funded by the European Space Agency.

B. Saulquin was with Ifremer, 29280 Plouzane, France. He is now with ACRI-ST, 06904 Sophia Antipolis, France (e-mail: bertrand.saulquin@acri-st.fr).

F. Gohin is with Ifremer, 29280 Plouzane, France (e-mail: francis.gohin@ifremer.fr).

R. Garrello is with Telecom Bretagne, Technopôle Brest-Iroise-CS 83818, 29238 Brest Cedex 3, France (e-mail: rene.garell@telecom-bretagne.eu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2010.2052813

may be important (defined as Case 2 waters), the inversion that is used to retrieve the component concentrations from  $nL_{ws}$  is more difficult. One of the most important issues in Case 2 waters comes from atmospheric correction failures that may lead to negative  $nL_{ws}$ . Ifremer has defined algorithms dedicated to Case 2 waters [4], which have been applied for more than ten years by Ifremer, based on simple but robust lookup tables that provide empirical correspondence between  $nL_{ws}$  and chl-*a* concentration. Systematic comparisons are made between the satellite-derived chl-*a* and the *in situ* data from cruises or, in a more quantitative way, from Ifremer's or partner's coastal monitoring networks. The user community of the OC5 products keeps on growing, and these satellite-derived chl-*a* products, particularly robust on the European turbid areas, will complement the data set collected from the conventional *in situ* monitoring networks for the monitoring requested by the new Water Framework Directive (WFD) [5]. The WFD aims at obtaining cleaner waters for Europe, and to that purpose, several physical and chemical parameters have been retained to evaluate the quality of the coastal environment. One of those parameters is the chl-*a* concentration as it is related to the phytoplankton abundance and the eutrophication risk.

The main limitation in the use of remote sensing ocean color data is that  $nL_{ws}$  can only be derived from cloud-free observations, which makes difficult the monitoring of environmental processes from day to day. To facilitate the end-user understanding [6, Benefits of Merging], we propose to merge the daily multisensor data to define a uniform product. The spatial coverage of the merged product, called the Ireland–Biscay–Iberia Regional Ocean Observing System (IBIROOS) area (<http://www.ibi-roos.eu/>), encompasses the West Europe between  $-12^\circ$  E to  $13^\circ$  W and  $36^\circ$  N to  $60^\circ$  N. The concerned period spans from 1998 to the present, and the resolution of the merged product is  $1.1 * 1.1 \text{ km}^2$ . The input data used in the merging procedure are the OC5 chl-*a* concentrations derived from  $nL_{ws}$  of the following three sensors: the Sea-viewing Wide Field of View Sensor (SeaWiFS), the Moderate Resolution Imaging Spectroradiometer (MODIS), and the MEdium Resolution Imaging Spectrometer Instrument (MERIS). Each sensor is embedded on a platform describing a polar orbit. Their main characteristics are described in Table I. The swath width is about 2800 km for SeaWiFS and 2300 km for MODIS, with a local area coverage (LAC) resolution of, respectively,  $1.1 * 1.1$  and  $1.0 \text{ km}^2$  at nadir. For MERIS, the swath width is about 1150 km at nadir, and the resolution of the reduced resolution (RR) product is  $1.04 * 1.16 \text{ km}^2$ . In terms of spatial coverage and by using the data that are available at Ifremer, the mean daily coverage for the LAC resolution was  $\sim 24\%$  for MODIS and  $\sim 22\%$  for SeaWiFS for 2004 on the IBI-ROOS area. This coverage could be extended to  $\sim 31\%$  by simply binning the two chl-*a* daily products. For 2008, the respective coverages were  $\sim 17\%$  for MERIS,  $\sim 22\%$  for MODIS, and  $\sim 28\%$  for the simple binned product.

Several studies have already been done to merge multisensor data to provide complete fields of chl-*a*. Maritorena and Siegel [7] used  $nL_{ws}$  from SeaWiFS and MODIS/Aqua together through the semianalytical bio-optical model (GSM01) to pro-

TABLE I  
SEAWIFS, MODIS, AND MERIS SENSOR CHARACTERISTICS

Sensor	SeaWiFS	MODIS	MERIS
platform	Orbview-2	Aqua	ENVISAT
Launch-date	08/01/1997	05/04/2002	01/03/2002
Number of bands	8	36	15
Spectral band center used for ocean color (nm)	413, 443, 490, 510, 555, 670, 765, 865	412, 443, 488, 531, 551, 667, 678, 748, 869	412.5, 442.5, 490, 510, 560, 620, 665, 681.25, 705, 775, 865
Data since	09/04/1997	08/01/2002	15/06/2002
Swath width	2806 km	2300 km	1150 km
Resolution	$1.1 \text{ km}^2$ (LAC)	$1 \text{ km}^2$ (LAC)	$1.04 * 1.16 \text{ km}^2$ (RR)
orbit	descending	ascending	descending
Equator crossing time	12:20	13:30	10:00

duce global fields of chl-*a*. Within the National Aeronautics and Space Administration (NASA) Sensor Intercomparison for Marine Biological and Interdisciplinary Ocean Studies program, several methods [8], [9] were tested from simple binning of daily chl-*a* derived from different sources to more sophisticated methods. Within the integrated project Marine Environment and Security for the European Area, Pottier *et al.* [10] directly combined the chl-*a* concentrations of both SeaWiFS and MODIS/Aqua, using an optimal interpolation method, and produced global fields at a resolution of  $0.1^\circ * 0.1^\circ$ .

The method that we discuss in detail here is based on the geostatistical formalism of kriging [11]. This method has been frequently used in oceanography to merge sea surface temperature fields [12], [13] or altimetry data [14]. While some methods merge  $nL_{ws}$  or reflectances before retrieving the chl-*a* concentration [15]–[17], we chose to directly merge the chl-*a* anomalies which have been previously calculated for each sensor from a climatology defined over the period 1998–2008. In our approach, the chl-*a* anomaly, rather than the chl-*a*, was preferred for the merging as its spatial distribution is more isotropic (uniform in all directions), especially at the shores where the chl-*a* concentration depends of local conditions such as bathymetry, river mouths, etc. The strong space-time covariance of the chl-*a* anomaly is evaluated using the semivariograms. This covariance is then used in the kriging system to estimate the anomaly at any time and location. Conversely, to Pottier *et al.* [10] who worked in log-transformed data, we have preferred to use the data in natural units without transformation. Although our data set of chl-*a* concentration shows the classical log-normal distribution [18] and although the log-transformed data have better stationarity properties, we have preferred to work on the natural values. This was done with the purpose of avoiding biases in the return to natural units after merging. As the exponential of the mean of the log-transformed data is not the mean of the original data, the exponential of the interpolated data is not an unbiased estimator of the searched quantity.



Fig. 1. Map of the selected network stations.

## II. *IN SITU* AND SATELLITE DATA

### A. *In Situ* Data

We selected 14 stations from the monitoring networks of Ifremer and its partners to perform direct comparisons between the satellite and *in situ* chl-*a* over the period 1998–2008. The locations of the stations are shown in Fig. 1. Within French waters, stations were obtained from the following two main networks: the REPHY<sup>1</sup> phytoplankton network and the SOMLIT<sup>2</sup> network managed by INSU.<sup>3</sup> The Stonehaven station in the North Sea, provided by Rodney Forster from CEFAS (U.K.), and the station D2 AZTI, provided by AZTI-Tecnalia within the “Variaciones” project (supported by the Department of Agriculture, Fisheries and Food of the Basque Government), were added to the data set as they provide useful long time series in the North and South of the studied area. All of these stations were selected for their capacity to represent specific regional water conditions along the shores of the southern English Channel, the Bay of Biscay, and the Mediterranean Sea. To perform direct comparisons between the network stations and the satellite chl-*a*, some of the *in situ* locations were artificially shifted from the shore to avoid land contamination

<sup>1</sup>REPHY: Réseau de surveillance du phytoplancton et des phytotoxines.

<sup>2</sup>SOMLIT: Service d’Observation du Milieu Littoral.

<sup>3</sup>INSU: Institut National des Sciences de l’Univers.

TABLE II  
SELECTED NETWORK STATIONS, THEIR SAMPLING FREQUENCY,  
AND THEIR ASSOCIATED SHIFTS FOR COMPARISON  
WITH THE SATELLITE DATA

Name of the station	Network	Sampling frequency	Shift (km)
Stonehaven	CFAS smart buoy	daily	
Dunkerque_1	REPHY <sup>(1)</sup> /SRN <sup>(2)</sup>	biweekly	2
Dunkerque_3	REPHY <sup>(1)</sup> /SRN <sup>(2)</sup>	biweekly	
Boulogne_1	REPHY <sup>(1)</sup> /SRN <sup>(2)</sup>	biweekly	2
Boulogne_3	REPHY <sup>(1)</sup> /SRN <sup>(2)</sup>	biweekly	
Cabourg	REPHY <sup>(1)</sup>	biweekly	2
Cancale	REPHY <sup>(1)</sup>	biweekly	
St Cast	REPHY <sup>(1)</sup>	biweekly	2.23
Roscoff_Astan	SOMLIT <sup>(3)</sup>	biweekly	
Roscoff_Estacade	SOMLIT <sup>(3)</sup>	biweekly	3.6
Men du	REPHY <sup>(1)</sup>	biweekly	
Men er roue	REPHY <sup>(1)</sup>	biweekly	1
Ouest Loscolo	REPHY <sup>(1)</sup>	biweekly	
Boyard	REPHY <sup>(1)</sup>	biweekly	3.17
Men er Roue	REPHY <sup>(1)</sup>	biweekly	1
Ouest Loscolo	REPHY <sup>(1)</sup>	biweekly	
Boyard	REPHY <sup>(1)</sup>	biweekly	3.17
D2_azti	AZTI monitoring station.	monthly	
Banyuls	SOMLIT <sup>(3)</sup>	biweekly	3
Marseille Frioul	SOMLIT <sup>(3)</sup>	biweekly	2.8

on the satellite data. The shifts are detailed in Table II. In the same manner, all of the cruise data that are available at Ifremer were used to perform direct comparisons with the satellite data over the same period.

### B. *Satellite* Data

All of the OC5 products that are available at Ifremer cover the IBI-ROOS area. The SeaWiFS OC5 chl-*a* products are available on a daily basis from January 01, 1998, to December 31, 2004, the MODIS OC5 products are available from October 01, 2002, to the present, and the MERIS OC5 products are available from October 01, 2002, to the present. A few cloudless MERIS images are also available for the period January 01, 2003, to October 01, 2006. We directly download at Ifremer the level 2 products from the OceanColor ftp facilities for SeaWiFS and MODIS (<http://oceancolor.gsfc.nasa.gov/>). For MERIS, the level 2 RR products are downloaded from the European Space Agency (ESA) rolling archive of the Kiruna ground station (<https://oa-ks.eo.esa.int/ra/>). For MODIS, pixels flagged [19] as CLDICE and HILT were rejected. For SeaWiFS, pixels flagged [19] as HIGLINT, CLDICE, HILT, STRAYLIGHT, and DARKPIXEL were rejected. For MERIS, pixels flagged [20] as CLOUD and HIGLINT were discarded.  $nL_{ws}$  is then projected on a regular grid of  $1.1 * 1.1 \text{ km}^2$ , and a dedicated Lookup table (LUT) per sensor is applied to  $nL_{ws}$  to obtain

the OC5 chl-*a* products. For each sensor, the daily chl-*a* satellite data are available on the Nausicaa MarCoast Web server (<http://www.ifremer.fr/nausicaa/marcoast2/index.htm>) in jpg and netcdf formats.

### III. THEORETICAL MODEL OF THE MERGING PROCEDURE

#### A. Formulation of the Chl-*a* Anomaly

Chl-*a* anomalies were calculated from a monthly climatology computed over the period 1998–2008 using both MODIS and SeaWiFS OC5 chl-*a* data. This monthly climatology is a mean that is calculated from all of the daily images of a considered month in the period. These chl-*a* values were first filtered using a simple threshold that is equal to  $1.5 * \text{standard deviation}$  to remove extreme and doubtful observations. The chl-*a* climatological monthly fields used in the procedure are available on the Ifremer ftp facilities (<ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/ocean-color/climato/monthly>). Finally, the climatological fields that are used to derive the daily anomalies are calculated by applying a linear interpolation to the monthly fields, with weights related to the position of the day between two months.

#### B. Kriging Method for Interpolating the Anomaly

The parameter estimated by kriging is the chl-*a* anomaly ( $\bar{A}_{\text{Chl-}a}$ ). The theoretical scheme presented here is very similar to the technique applied by Bretherton *et al.* [21], which is well known in the oceanographic literature. For each day,  $\bar{A}_{\text{Chl-}a}$  is estimated at each pixel, centered on the position  $x_0$  and at time  $t_0$ , using  $n$  observed anomalies  $A_{\text{Chl-}a}(x_i, t_i)$  that are available in the space-time vicinity of  $(x_0, t_0)$ . The kriging estimator is said to be linear because the predicted value is a linear combination that may be written as

$$\bar{A}_{\text{Chl-}a}(x_0, t_0) = \sum_{i=1}^n \lambda_i * A_{\text{Chl-}a}(x_i, t_i). \quad (1)$$

Determining the weights  $\lambda_i$  in (1) is the main issue in optimal interpolation. For ensuring a well-balanced set of observations, in space and time, the anomaly data or control points were sought using a procedure which scans decreasing disks in space, starting with the largest disk for the current day, and then alternates between the previous and following days (Fig. 2; [22]).

The so-called ‘‘simple kriging assumption’’ sets that the mean value of the increments  $[A_{\text{Chl-}a}(x+d) - A_{\text{Chl-}a}(x)]$  is zero, and that the covariance (or the semivariogram of the increments) between  $A_{\text{Chl-}a}(x)$  and  $A_{\text{Chl-}a}(x+d)$  depends only on the spatiotemporal distance  $d$ . In geostatistics, it is assumed that, because the chl-*a* anomaly is a random process characterized by stationary properties, the variance of the estimation error (2) can be expressed using the semivariogram obtained through the structural analysis

$$V_K = \text{Var} \left[ \bar{A}_{\text{Chl-}a}(x_0, t_0) - \sum_{i=1}^n \lambda_i * A_{\text{Chl-}a}(x_i, t_i) \right]. \quad (2)$$

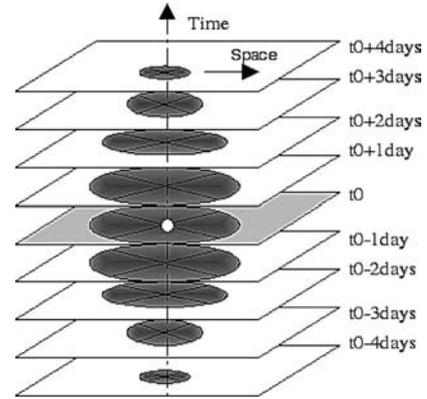


Fig. 2. Method for seeking the control points in space and time.

The experimental semivariogram (or the structure function) is obtained using the following equation, which is applied to  $n(dh, dt)$  points separated from distance  $dh$  and time  $dt$ :

$$\gamma(dh, dt) = (1/2 * n(dh, dt)) \times \sum_{i=1}^{n(h)} [A_{\text{Chl-}a}(x+dh, dt) - A_{\text{Chl-}a}(x)]^2. \quad (3)$$

The semivariogram can be directly expressed from the covariance when this latter exists

$$\gamma(dh, dt) = \sigma^2 - C(dh, dt). \quad (4)$$

In the simple kriging hypothesis, where the mean of the increments is zero and the semivariogram exists, the kriging variance can be developed as

$$V_K = - \sum_{i=1}^n \sum_{j=1}^n \lambda_i * \lambda_j \gamma(x_i, t_i; x_j, t_j) + 2 \sum_{i=1}^n \lambda_i \gamma(x_i, t_i; x_0, t_0). \quad (5)$$

By assuming that the sum of the coefficients is equal to one (ensuring no bias),  $\lambda_i$  is the solution of the kriging system if they minimize the variance  $V_K$  (5). Once  $\bar{A}_{\text{Chl-}a}(1)$  has been estimated, its value is added to the climatological mean of chl-*a* at location  $(x_0, t_0)$  to obtain the final chl-*a*.

#### C. Local Semivariograms

The improvement that we propose here to the standard kriging procedure is based on the definition of locally and per-month covariances (or semivariograms) to estimate  $A_{\text{Chl-}a}$ , as the anomaly is generally more variable in nutrient-rich coastal waters than in the open sea. This regional approach leading to local semivariograms is very similar to the method developed in [23] for estimating the sea surface temperature by merging the multisensor data sets. To study the space-time covariance of  $A_{\text{Chl-}a}$ , the IBI-ROOS area has been divided into subregions of 10 000 km<sup>2</sup>, and some extra subregions were added near the shores. The sampling grid is shown in Fig. 3. For each subregion, an experimental semivariogram (3) has been calculated per month over the period 2002–2004 using the SeaWiFS data. We considered that the chl-*a* anomaly was not structurally dependent on the direction, i.e., its distribution was isotropic.

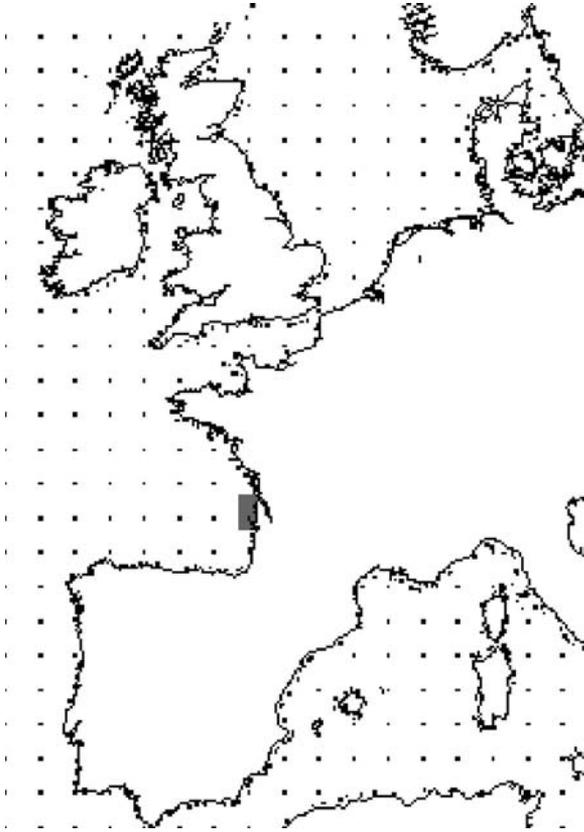


Fig. 3. Sampling grid. The dots represent the center of the  $100 \times 100 \text{ km}^2$  subregions. The selected region on which the variogram of Fig. 4 has been calculated is shown in gray.

The spatial anisotropy of the chl-*a* anomaly, if it exists, really depends on local conditions in space and time (river floods, circulation, etc.), and therefore, it would be particularly difficult to determine the persistent anisotropic features in the anomaly on such a large scale. In a practical sense, it also avoids the calculation and the use of multiple directional semivariograms that would have been heavy to handle operationally, moreover without clear and significant improvement in the method.

#### D. Automatic Adjustment of a Spherical Variogram to the Experimental Local Semivariogram

The theoretical spherical semivariogram (6) is defined by its spatial and temporal nuggets, by its spatial and temporal range, and by its sill. The description of these parameters and all of those used in our optimal interpolation procedure is provided in Table III.

$$\begin{aligned} \gamma(dh, dt) &= \text{sill} * (1.5 * d - 0.5 * d^3) + px + pt & \text{for } d < 1 \\ \gamma(dh, dt) &= \text{sill} + px + pt & \text{for } d > 1 \end{aligned} \quad (6)$$

with  $d = \sqrt{(dh/\text{spatial\_range})^2 + (dt/\text{temporal\_range})^2}$ .

For each subregion, a theoretical semivariogram of spherical type (6) has been adjusted on the experimental semivariogram (3) using a least square minimization procedure (Fig. 4).

TABLE III  
LIST OF SYMBOLS AND ABBREVIATIONS USED IN THE KRIGING PROCEDURE AND THEIR DESCRIPTION

Symbols & abbreviations	Description
$\hat{A}_{chl-a}(x_0, t_0)$	Estimator of $A_{chl-a}$ by kriging at the position $(x_0, t_0)$
$V_K$	Variance of the estimator
$\gamma(dh, dt)$	Semi-variogram of $A_{chl-a}$ calculated experimentally (3) for space (dh) and time (dt) lags. A theoretical semi-variogram (6) is adjusted on the experimental semi-variogram for the variance error estimation (5)
dh	Spatial distance in kilometers
dt	Temporal distance in days
Spatial range	Maximum spatial distance at which the correlation between the anomalies is not null
Temporal range	Maximum temporal distance at which the correlation between the anomalies is not null.
d	Spatio-temporal distance
sill	Maximum variance of $A_{chl-a}$
px	Spatial nugget: variance of the anomaly in space due to instrumental noise and micro-structures (at the pixel level)
pt	Temporal nugget: variance of the anomaly in time due to instrumental noise in time and day to day variations in the atmospheric corrections at the same location.

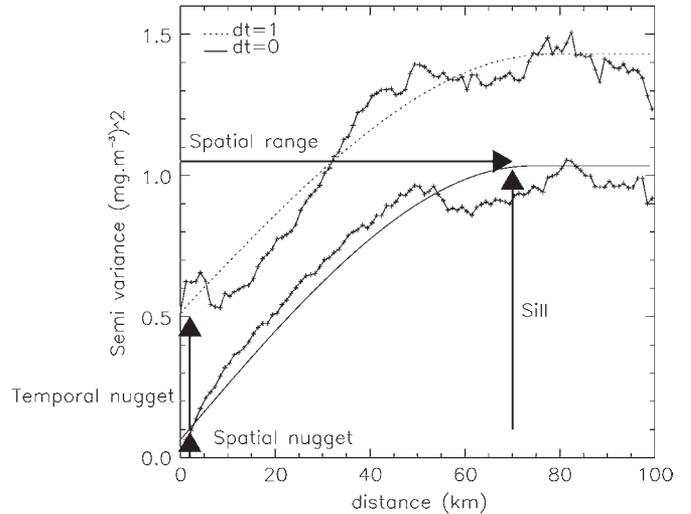


Fig. 4. (Crosses) Experimental spatial semivariogram and (line) the spherical semivariogram adjusted automatically for May and the selected subregion (area in gray; Fig. 3).

#### E. Generalization of the Theoretical Variogram Parameters to the Whole Area

The parameters of the local theoretical semivariograms, previously estimated using a least square minimization procedure (Fig. 4), were correlated to the square of the climatological monthly mean of chl-*a*. Fig. 5 shows, for April and each subregion of the grid shown in Fig. 3, the correlations between the sill, the nuggets, and the square of the mean of chl-*a*. We observe for April a correlation coefficient  $r^2$  of 0.74, 0.79,

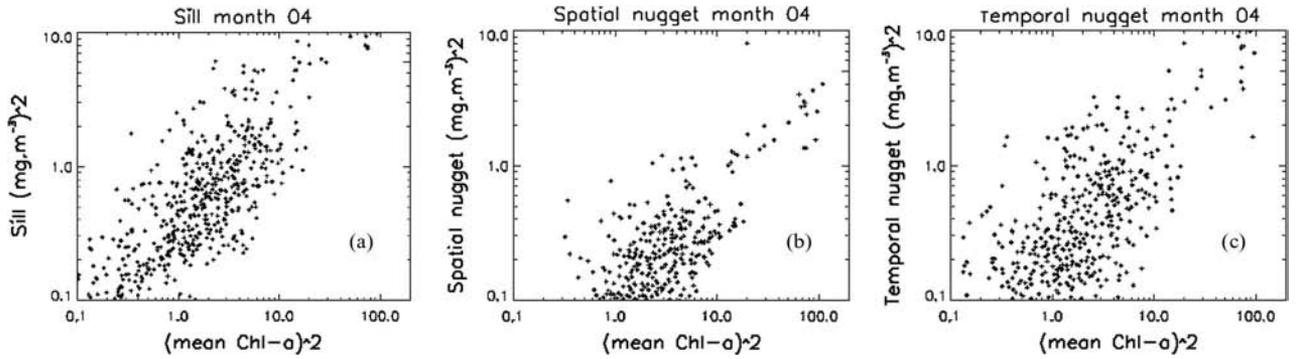


Fig. 5. Scatterplots between the variogram parameters and the climatological mean square values of chl-*a*. (a) Sills, (b) spatial nuggets, and (c) temporal nuggets for April.

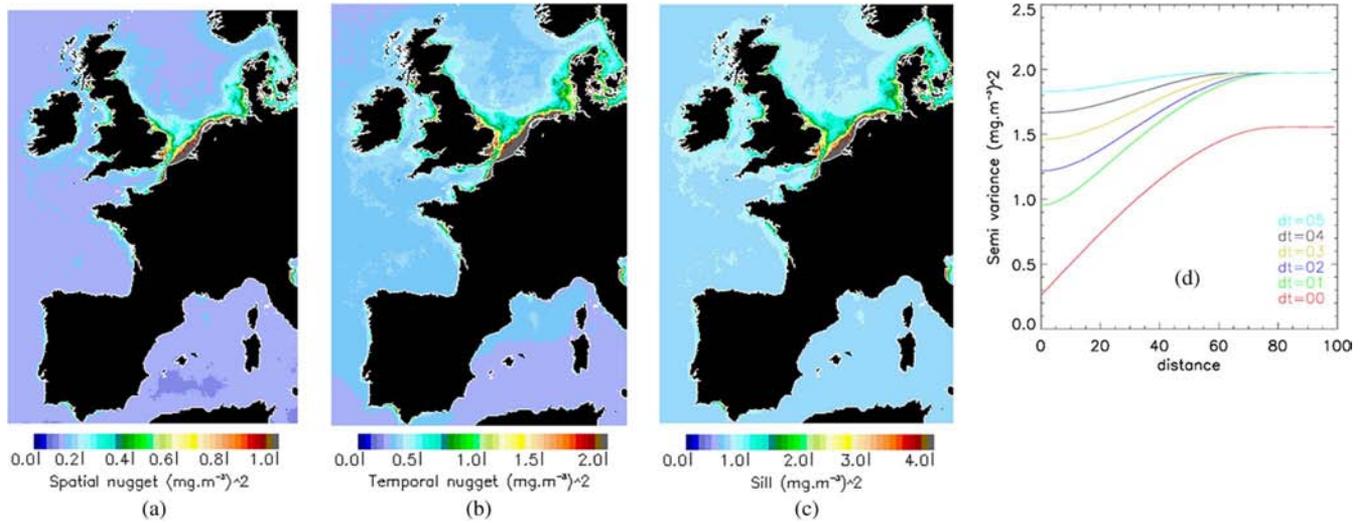


Fig. 6. Maps of the semivariogram parameters for April. (a) Spatial nuggets. (b) Temporal nuggets. (c) Sills. (d) Resulting spherical theoretical semivariogram at the same location as Fig. 4.

TABLE IV  
DEFAULT PARAMETERS OF THE SPHERICAL SEMIVARIOGRAM  
FOR DECEMBER AND JANUARY

Parameter	Value
Sill	$0.7 + 0.2 * [\text{chl-}a]^2$
Spatial nugget	$0.05 + 0.05 * [\text{chl-}a]^2$
Temporal nugget	$0.1 + 0.11 * [\text{chl-}a]^2$
Spatial range	80
Temporal range	10

and 0.71, respectively, for the sill and the spatial and temporal nuggets. For each month of the year, with the exceptions of January and February where the primary production is very weak here, we observe good correlations between these three parameters and the square of the mean of chl-*a*. Therefore, monthly continuous maps for the sill and the spatial and temporal nuggets were derived (Fig. 6) from the climatological

monthly fields using a linear regression. For December and January where the correlations were lower than 0.5, we used default parameters (Table IV). To determine the spatial ranges, which are not correlated with the climatological situation but with the local conditions, we directly interpolated the local spatial ranges using an interpolation procedure provided by the IDL toolbox to create continuous maps of spatial ranges on the area. For the temporal ranges, the regional temporal semivariograms were too noisy to justify the local characteristics. We then used a global semivariogram that is calculated over the whole area to determine this parameter. The estimated temporal range is ten days over the whole area.

## IV. RESULTS

### A. Validation of the Satellite-Derived Chl-*a* Concentration With In Situ Data

The satellite OC5 products are compared to the *in situ* data set, comprising selected stations and cruise data (Section II-A). To that purpose, we have produced “matchup” data pairs of satellite and *in situ* chl-*a* concentrations collocated in space (same pixel) and obtained during the same day. Fig. 7 shows the scatterplots and the statistics of the point-to-point comparisons. The coefficient of correlation ( $r$ ) has been calculated using log-transformed values. For the SeaWiFS OC5 product, the

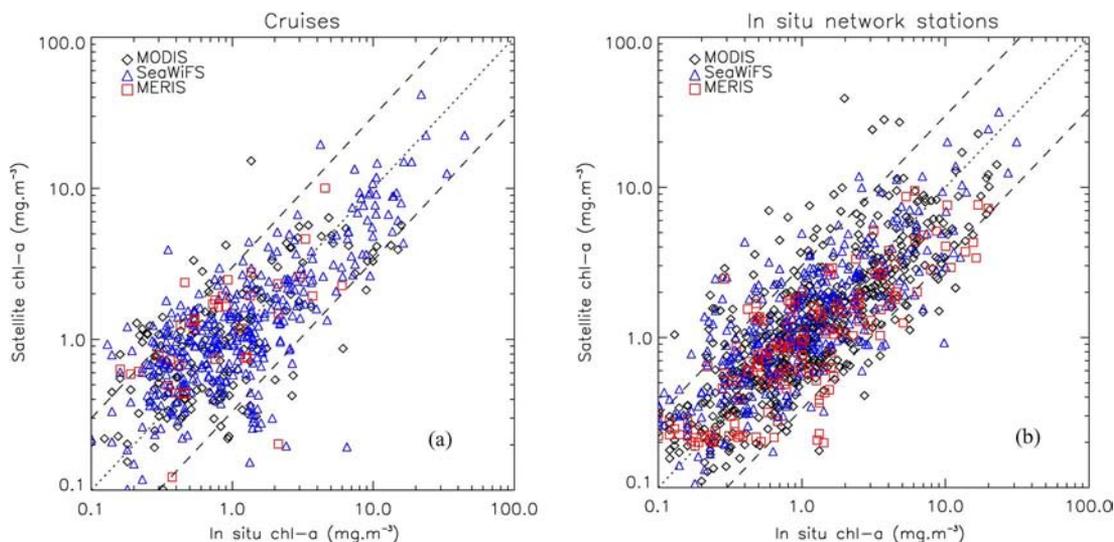


Fig. 7. Scatterplots of the SeaWiFS (from January 1998 to December 2004), MODIS (from August 2002 to December 2008), and MERIS (from October 2006 to December 2008) chl-*a* concentrations versus (a) the cruise data and (a) the coastal network data.

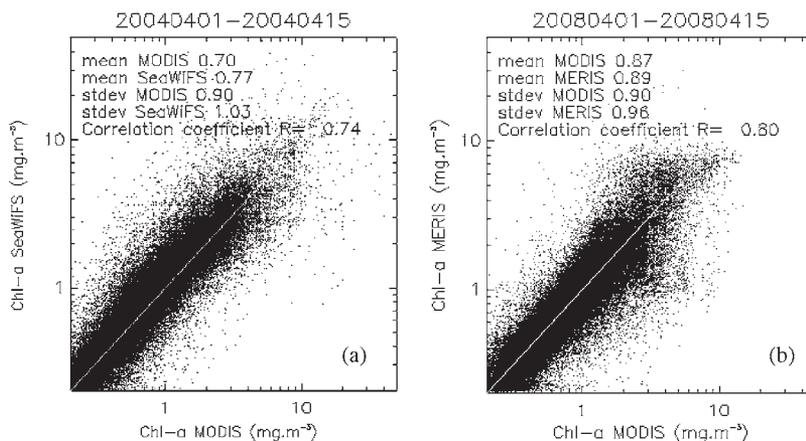


Fig. 8. Examples of the scatterplots (a) between SeaWiFS and MODIS OC5 chl-*a* products for the first two weeks of April 2004 and (b) between MERIS and MODIS for the first two weeks of April 2008.

observed bias is  $-0.31 \text{ mg} \cdot \text{m}^{-3}$  compared to the cruises data and  $0.04 \text{ mg} \cdot \text{m}^{-3}$  compared to the stations data. The correlation coefficient  $r$  is equal to 0.76 and 0.78 with the cruise and station data. For MODIS, the observed biases are  $-0.18 \text{ mg} \cdot \text{m}^{-3}$  with the cruise data and  $-0.09 \text{ mg} \cdot \text{m}^{-3}$  for the station data. The correlation coefficient is, respectively, equal to 0.67 and 0.79 with the cruise and station data sets. For MERIS, with only 36 matchups with the cruise data, the observed bias is  $0.37 \text{ mg} \cdot \text{m}^{-3}$ , with a coefficient of correlation of 0.59. MERIS OC5 showed a slight negative bias of  $-0.72 \text{ mg} \cdot \text{m}^{-3}$  with the station data, with a coefficient of correlation of 0.81.

### B. Intercomparisons Between the Satellite Data Sets

Before merging the different OC5 products, we compared carefully the satellite data sets over the same period. Comparisons between the OC5 products are performed for each month and over periods of two weeks by comparing the point-to-point values of the images. Fig. 8 shows the scatterplots between SeaWiFS and MODIS for the first two weeks of April 2004 and between MODIS and MERIS for the first two weeks of

April 2008. In April 2004, the observed bias of SeaWiFS compared to MODIS was  $0.07 \text{ mg} \cdot \text{m}^{-3}$ . In April 2008, MODIS appears to be lower than MERIS, with a mean difference of  $0.02 \text{ mg} \cdot \text{m}^{-3}$ . Globally, the comparisons between the MODIS and the SeaWiFS OC5 products show a very good agreement for all the levels of chl-*a*, with a slight negative bias for MODIS (Table V). In the same manner, MERIS showed a slight positive bias compared to MODIS. At this stage, it is difficult to evaluate which product over or under-estimates the chl-*a* concentration as the difference between the satellite products and the *in situ* data is on the same order of magnitude. For that reason, no further intercalibration procedure between the OC5 products was developed.

### C. Interpolating the Satellite Chlorophyll Concentration by Kriging

The kriging procedure is applied to each pixel of the image grid to estimate the daily chl-*a* field. For each interpolated day from 1998 to the present, the procedure takes into account the available images five days before and after the day they are being processed. For each point of the image, the parameters of

TABLE V  
STATISTICAL RESULTS OF THE INTERCOMPARISONS BETWEEN SEAWIFS AND MODIS OC5 Chl-*a* PRODUCTS FOR 2004 AND BETWEEN MERIS AND MODIS OC5 Chl-*a* PRODUCTS FOR 2008. FOR EVERY MONTH, ONLY THE FIRST TWO WEEKS ARE CONSIDERED. THE BIAS IS THE MEAN DIFFERENCE SEAWIFS-MODIS AND MERIS-MODIS FOR THE SELECTED PIXELS (IN MILLIGRAMS PER CUBIC METER)

Month	Bias	R	Month	Bias	R	Month	Bias	R
1	0.09/-0.04	0.84/0.60	5	0.11/0.20	0.61/0.75	9	0.10/0.09	0.78/0.80
2	0.09/0.01	0.69/0.75	6	0.04/0.10	0.73/0.77	10	0.06/0.04	0.76/0.77
3	0.11/0.08	0.74/0.75	7	0.04/0.16	0.80/0.78	11	0.06/0.01	0.82/0.70
4	0.07/0.10	0.74/0.80	8	0.06/0.08	0.77/0.84	12	0.04/0.07	0.86/0.65

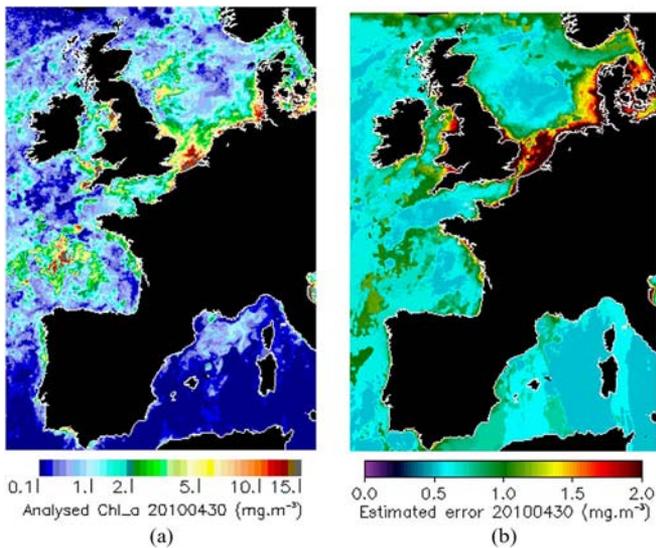


Fig. 9. (a) Analysis chl-*a* product for April 07, 2009. (b) Estimated error.

the variogram are extracted from the monthly parameter maps (Fig. 6). For a specific day, the parameter values are interpolated linearly using these monthly maps and are weighted regarding the situation of the day within the month. For each point to estimate, 50 observations are collected, and chl-*a* anomalies that are greater than  $1.5 \times \text{stdev}$  are filtered. A specific procedure written and optimized in C, based on the GNU GSL scientific library <http://www.gnu.org/software/gsl/>, has been developed to process a daily chl-*a* image of  $2401 \times 1467$  pixels in 12 min using a 2-GHz processor. Fig. 9(a) shows an example of the result of the interpolation procedure for April 30, 2010, and the estimated error associated. For that example, six MODIS and six MERIS available images were used.

#### D. Validation of the Multisensor Analysis With In Situ Data

The interpolated fields were compared to the *in situ* data using the same selected stations and cruises over the 1998–2008 period. Fig. 10 shows the scatterplots between the analyzed product and the *in situ* data. The number of matchups with the cruise data has increased from 561 (for SeaWiFS, MODIS, and MERIS) to 1447 and from 1086 to 3970 for the station data. The observed biases of the analysis compared to the cruise data are  $-0.17$  and  $-0.26 \text{ mg} \cdot \text{m}^{-3}$  for the station data. This negative bias with the station data is more important compared

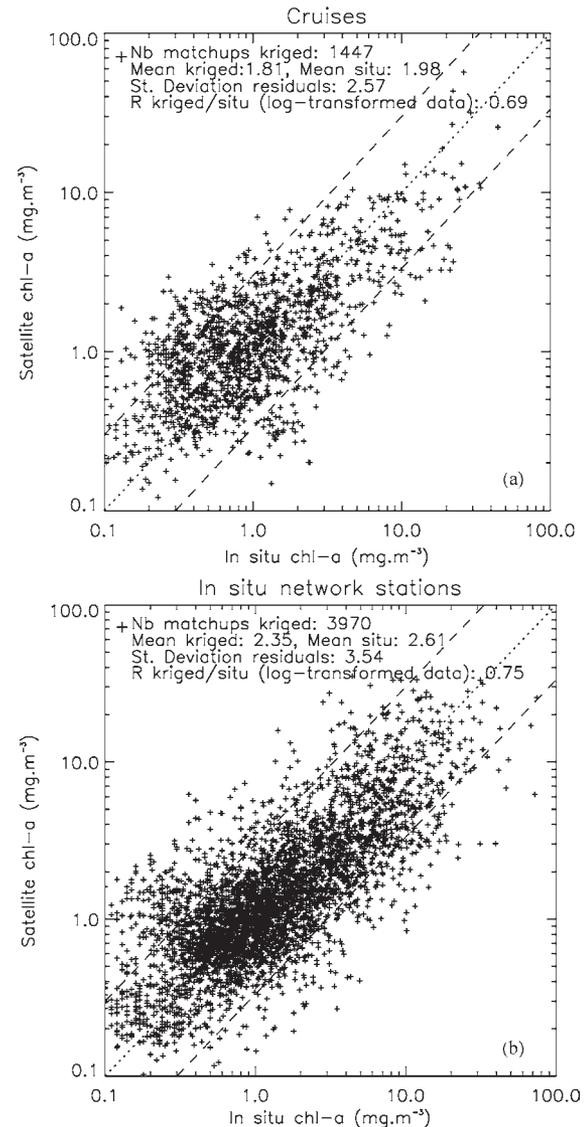


Fig. 10. Scatterplots of the multisensor chl-*a* analysis versus (a) the cruise data and (b) the coastal network data for the period 1998–2008.

to the direct satellite/*in situ* comparison (cf., Section IV-A). This might be explained by the fact that these stations are localized at the shore, and the kriging procedure collects more data offshore, which can contribute to this slight underestimation, considering the general gradient of increasing chlorophyll shorewards. However, this bias is almost negligible. Globally,

the method does not introduce any bias, which was expected as it has been designed to verify this condition. The multisensor analysis also shows a standard deviation with the *in situ* data, which is equivalent to that obtained from direct matchups based on the original satellite data. The coefficient of correlation between the analyzed product and the cruise data is 0.69 and 0.75 between the analysis and the station data.

### E. Using the Analysis for the WFD

The percentile 90 (P90) value of a data set is defined as the level where 90% of the observations are equal to or less than its value. Therefore, this can be used as a good indicator of high chl-*a* concentration events, and the P90 of chl-*a* is now used to characterize the eutrophication status [4] of the coastal water bodies in the scope of the new WFD. Most of the selected stations in this paper contribute to the monitoring network of the WFD. In most of the cases, only one station with a biweekly sampling frequency is used to estimate the eutrophication status of a water body, which may be larger than 1000 km<sup>2</sup> in case of large bays. Therefore, the use of the satellite data, with increased spatial and temporal coverage, is very promising.

Fig. 11 shows the comparisons at the stations of the mean and P90 values calculated over the productive period (March to October) using the *in situ* and satellite data. We observe a good general agreement between the satellite and *in situ* data (in mean and P90), with a slight underestimation of the satellite products. To evaluate the satellite potential for the WFD, we calculate the statistics in mean and P90 for the 14 selected stations. The biases shown in Fig. 11 are the averages of the mean and P90 biases at the 14 stations. For SeaWiFS, from 1998 to 2004, during the productive period [Fig. 11(a)], the observed bias of the satellite P90 compared to the *in situ* data is  $-1.68 \text{ mg} \cdot \text{m}^{-3}$ . For the same period, using the analyzed data [Fig. 11(b)], the bias is  $-1.44 \text{ mg} \cdot \text{m}^{-3}$ . The coefficients of correlation were, respectively, 0.96 between the SeaWiFS and the *in situ* P90 and 0.94 between the analysis and the *in situ* data. For MODIS [Fig. 11(c)], from 2003 to 2008, the bias value is  $-2 \text{ mg} \cdot \text{m}^{-3}$  in P90 and  $-1.42 \text{ mg} \cdot \text{m}^{-3}$  for the analyzed data over the same period [Fig. 11(d)]. The coefficient of correlation values are 0.96 and 0.97, respectively, for the MODIS and the analyzed product. Over the whole period [Fig. 11(e)], the analyzed product shows a slight negative bias of  $-0.94 \text{ mg} \cdot \text{m}^{-3}$  and a coefficient of correlation of 0.96 for the P90. Fig. 12 shows the ratio between the P90 and the mean at the stations. This ratio value, obtained using the *in situ* data for the period 1998–2004, is 2.16, compared to 2.09 when using the SeaWiFS data [Fig. 12(a)]. For the period 2003–2008, the ratio value is 2.04 with the *in situ* data and 1.95 with the MODIS data [Fig. 12(b)]. For the whole period [Fig. 12(c)], the ratio value is 2.09 with the *in situ* and 2.16 with the analysis. This good agreement in mean and P90 between the analysis and the *in situ* data enhances the interest of using the analysis for the application of the WFD. Its high temporal frequency will give a better estimation of the mean and P90 of chl-*a*, through the productive period, than the original satellite imagery or even than the *in situ* monitoring. This better estimation of the mean of each pixel included in the coastal water bodies, associated

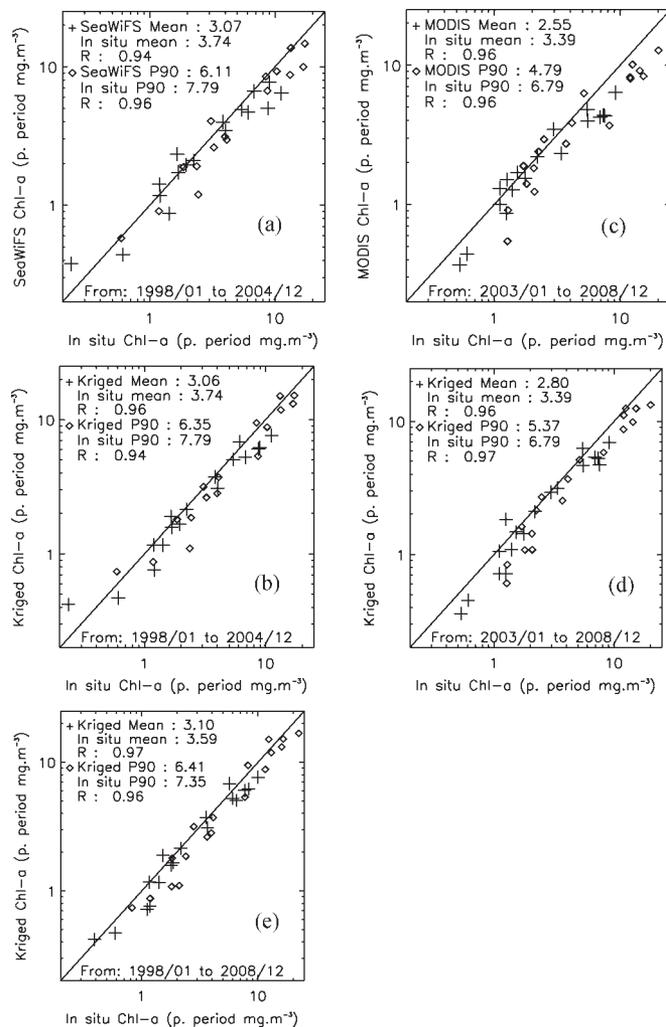


Fig. 11. Satellite mean and P90 at the coastal stations. For each station, all the *in situ* and satellite data that are available during the productive period (March to October) are considered. (a) SeaWiFS versus *in situ* (1998–2004). (b) Analysis versus *in situ* (1998–2004). (c) MODIS versus *in situ* (2003–2008). (d) Analysis versus *in situ* (2003–2008). (e) Analysis versus *in situ* (1998–2008).

with this unexpected but interesting correlation between the mean and the P90, will help in improving the assessment of the eutrophication status of the water bodies.

### F. Using the Analysis for Monitoring the Phytoplankton Blooms

The multisensor analysis, by filling the cloudy pixels, provides high-frequency retrievals of chl-*a*, which can contribute to a better monitoring of the phytoplankton biomass. In comparison, most of the *in situ* networks have a sampling period of at least 15 days. During such period, significant blooms of phytoplankton may occur and may be missed by the *in situ* monitoring. These short-time scale events will appear on the satellite data, except in the case of persistent clouds in the vicinity of the bloom, and will be identified using the multisensor analysis. Fig. 13 shows the comparisons between the satellite and *in situ* data for 2007 at some of the selected network stations. The comparison with the instrumented MAREL buoy, located along the Carnot dyke, in the Boulogne harbour is also shown [Fig. 13(d)] as it provides high-frequency chl-*a* data

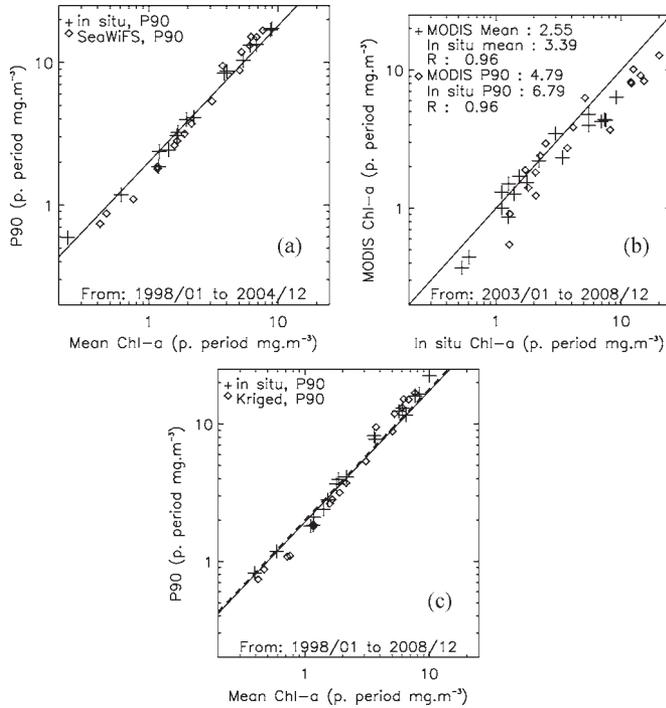


Fig. 12. Relation between the mean and the P90 at the stations. (a) SeaWiFS and *in situ*. (b) MODIS and *in situ*. (c) Analysis and *in situ*.

(derived from an embedded fluorimeter). The continuous lines represent the climatological mean calculated using all of the *in situ* data that are available at the stations for the period 1998–2008, and the dotted line shows the climatological P90 using the same data. For each network station, one retrieval per day is, in general, provided by the multisensor analysis. This frequency, compared to a biweekly or a monthly sampling frequency for the *in situ* data, improves the monitoring.

Fig. 13(a) shows that, during the period from April 15 to May 20, 2007, no satellite data were available at Boulogne from MODIS or MERIS, while in the same time, the analysis clearly showed the spring bloom (with a maximum of  $25 \text{ mg} \cdot \text{m}^{-3}$ , which was reached in April 17) followed by the decrease of the chl-*a* concentration caused by the lack of nutrients. Fig. 13(b) shows the results at the station of Cancale, characterized by very clear waters and low sediment concentrations. The results in 2007 appear to be less productive than the average year, and the *in situ* and satellite data (kriged or not) were in good agreement. The station of Cabourg [Fig. 13(c)] is located in the vicinity of the mouth of the Seine River, in a nutrient-rich area where the turbidity is relatively low and where the potential for production is very high. At Cabourg, the direct satellite data in 2007 did not catch all the high chl-*a* concentration events, while in the same time, the analyzed product catches some of them. The instrumented buoy of Carnot [Fig. 13(d)] provides high-frequency chl-*a* data, extrapolated from an empirical relationship between the fluorometry measurement and chl-*a*. These data clearly show the chl-*a* cycle in 2007. At the end of April, the instrumented buoy shows some concentrations of  $25 \text{ mg} \cdot \text{m}^{-3}$ , and no data are available from MERIS and MODIS at this location. The analysis shows the same maximum concentration. From May 15 to June 15, the instrumented buoy shows very low chl concentrations, not confirmed by the

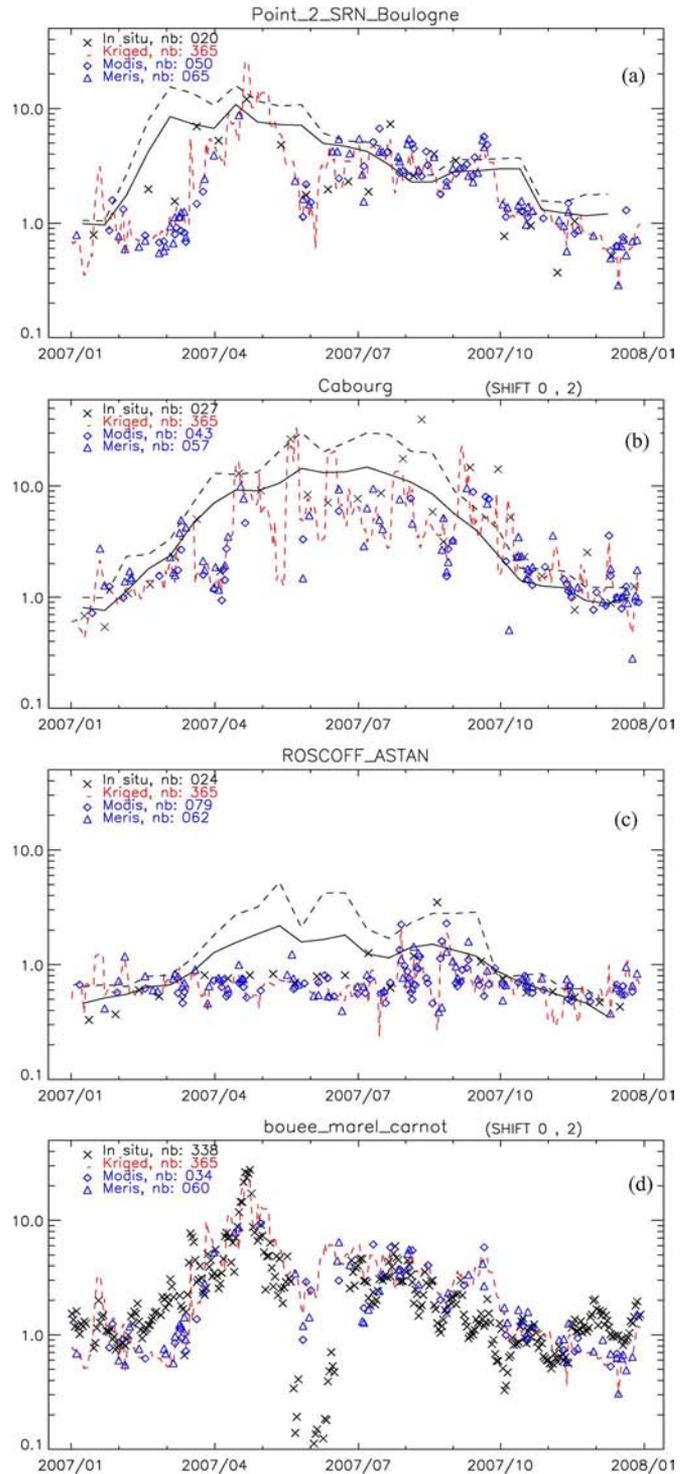


Fig. 13. Chl-*a* from satellite and *in situ* data at the network stations in 2007. The continuous line plotted for the first three stations represents the mean *in situ* chl-*a* seasonal curve calculated from all of the historical *in situ* data that are available at each station, and the dotted line shows the P90 *in situ* chl-*a* seasonal curve.

satellites or the analyzed product. We do believe that, in that period, the buoy encountered a maintenance failure.

## V. DISCUSSION AND CONCLUSION

In this paper, we have shown that, through statistical comparisons, it is possible to provide a better temporal and spatial

coverage of the chl-*a* concentration by interpolating the satellite observations without degrading the statistical results compared to the *in situ* data set. This method will help in providing high-frequency retrievals for the monitoring of chl-*a*. We have based our interpolation scheme on realistic hypotheses that are based on regional and seasonal variograms, as the intensity of the phytoplankton blooms considerably varies in the considered area during the year, with a dominant shoreward gradient. The variogram parameters (sill and nuggets), used in the estimation of the spatiotemporal covariance of the studied variable (the anomaly of chl-*a*), have been related to the square of the climatological mean of chl-*a*. Finally, the monthly maps for these parameters have been proposed, which constitute our improvement to the initial kriging method. The variance map [Fig. 9(b)] of the merged chl-*a* is also proposed. This map gives a realistic idea of the quality of the interpolated product. However, this error does not take into account the uncertainty in the estimation of the satellite-derived chl-*a*, but it is based on the spatial and temporal variability of the satellite-derived data from day to day.

A first version of the multisensor analysis has been developed, and daily cloudless fields have been calculated at a resolution of  $1.1 \times 1.1 \text{ km}^2$  on the IBI-ROOS area for the period 1998–2008 using SeaWiFS, MODIS, and MERIS OC5 chl-*a* products. The data are available on the Previmar Web site ([http://www.previmar.org/previsions/production\\_primaire/](http://www.previmar.org/previsions/production_primaire/)) and on the new Nausicaa Web server (<http://www.ifremer.fr/nausicaa2/marcoast>). The data can also be directly downloaded from the Ifremer ftp facility (<ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/ocean-color/atlantic/EUR-L4-CHL-ATL-v01/>). We do believe that, in the near future, the daily data set provided by such analysis, associated with the confidence of end users in its quality, will be used in numerous applications. The first application will be the validation of biological models by providing a cloudless reliable data set of satellite-derived chl-*a*. The second application will be on the surveillance of the phytoplankton blooms and on the water quality monitoring in the scope of the new WFD. This paper was successful in producing reliable and cloudless daily fields of chl-*a*. In the forthcoming months, it will be extended to the suspended matters and the Kpar products as these analyses can be of great interest in forcing the light in operational models of coastal oceanography [24].

#### ACKNOWLEDGMENT

The authors would like to thank NASA, particularly the ocean color group, for the production and delivery of the level 2 data for the MODIS and SeaWiFS sensors; ESA for the production and delivery of the level 2 data for the MERIS sensor; and CEFAS and AZTI for the provision of the long *in situ* time series.

#### 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, no. 30, pp. 6289–6306, Oct. 30, 2002.
- [3] IOCCG Report 3 "Remote Sensing in Coastal, and Other Optically-Complex, Waters," 2000.
- [4] F. Gohin, B. Saulquin, H. Oger-Jeanneret, L. Lozac'h, L. Lampert, A. Lefebvre, P. Riou, and F. Bruchon, "Towards a better assessment of the ecological status of coastal waters using satellite-derived chlorophyll-*a* concentrations," *Remote Sens. Environ.*, vol. 112, no. 8, pp. 3329–3340, Aug. 2008.
- [5] F. Gohin, J. N. Druon, and L. Lampert, "A five channel chlorophyll algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters," *Int. J. Remote Sens.*, vol. 23, no. 8, pp. 1639–1661, Apr. 2002.
- [6] IOCCG Report 6 "Ocean-Colour Data Merging," 2007.
- [7] S. Maritorena and D. A. Siegel, "Consistent merging of satellite ocean color datasets using a bio-optical model," *Remote Sens. Environ.*, vol. 94, no. 4, pp. 429–440, Feb. 2005.
- [8] E. J. Kwiatkowska, "MODIS validation, data merger, and other activities accomplished by SIMBIOS project: 2002–2003," NASA Goddard Space Flight Center, Greenbelt, MD, NASA-TM-2003, 2003, Comparisons of daily global ocean color datasets: MODIS-Terra/Aqua and SeaWiFS.
- [9] 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 Sens. Atmos., Ocean, Environ. Space*, Oct. 2002, pp. 168–182, Invited Paper.
- [10] C. Pottier, V. Garçon, G. Larnicol, J. Sudre, P. Schaeffer, and P.-Y. Le Traon, "Merging SeaWiFS and MODIS/aqua ocean color data in North and Equatorial Atlantic using weighted averaging and objective analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3436–3451, Nov. 2006.
- [11] G. Matheron, *The Theory of Regionalized Variables and Its Applications*. Paris, France: Cahier du centre de Morphologie Mathématique de Fontainebleau, 1971, p. 211, No. 5.
- [12] Y. Faugere, P. Le Borgne, and H. Roquet, "Réalisation d'une climatologie mondiale de la température de surface de la mer," *La Météorologie*, vol. 35, pp. 24–35, Nov. 2001.
- [13] B. Saulquin and F. Gohin, Mean seasonal cycle and evolution of the sea surface temperature from satellite and *in-situ* data in the english channel for the period 1986–2006, to be published.
- [14] P.-Y. Le Traon, F. Nadal, and N. Ducet, "An improved mapping method of multi-satellite altimeter data," *J. Atmos. Ocean. Technol.*, vol. 15, no. 2, pp. 522–534, Apr. 1998.
- [15] F. Melin, G. Zibordi, and S. Djavidnia, "Merged series of normalized water leaving radiances obtained from multiple satellite missions for the Mediterranean Sea," *Adv. Space Res.*, vol. 43, no. 3, pp. 423–437, Feb. 2009.
- [16] A. Morel, Y. Huot, B. Gentili, P. J. Werdell, S. B. Hooker, and B. A. Franz, "Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach," *Remote Sens. Environ.*, vol. 111, no. 1, pp. 69–88, Nov. 2007.
- [17] 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*. Greenbelt, MD: NASA Goddard Space Flight Center, Sep. 2003, pp. 19–31, NASA-TM-2003.
- [18] J. W. Campbell, "The lognormal distribution as a model for bio-optical variability in the sea," *J. Geophys. Res.*, vol. 100, no. C7, pp. 13 237–13 254, Jul. 1995.
- [19] SeaWiFS Postlaunch Technical report series Nasa Tech. Memo., 2003-206892, vol. 22
- [20] MERIS Level 2 Detailed Processing Model. [Online]. Available: [http://earth.esa.int/pub/ESA\\_DOC/ENVISAT/MERIS/MERIS\\_DPML2\\_i7r2A\\_re-issued.pdf](http://earth.esa.int/pub/ESA_DOC/ENVISAT/MERIS/MERIS_DPML2_i7r2A_re-issued.pdf)
- [21] 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.
- [22] K. Laffont, Apport des données satellite «Couleur de l'eau» à la surveillance de la qualité des eaux côtières, rapport de l'option géostatistique de l'Ecole Nationale Supérieure des Mines de Paris, p. 80, 2005.
- [23] J. L. Høyer and J. She, "Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea," *J. Mar. Syst.*, vol. 65, no. 1–4, pp. 176–189, Mar. 2007.
- [24] M. Huret, F. Gohin, D. Delmas, and M. Lunven, "Use of SeaWiFS data for light availability and parameter estimation of a phytoplankton production model of the Bay of Biscay," *J. Mar. Syst.*, vol. 65, no. 1–4, pp. 509–531, Mar. 2007.



**Bertrand Saulquin** received the M.S. degree from the University of Rennes, France.

He is currently an Engineer with ACRI-ST, Sophia Antipolis, France, working on satellite data analysis and geostatistics in Global Monitoring for Environment and Security projects of the European Union (Aquamar and MarCoast2).



**Francis Gohin** received the B.S. degree in biology and geostatistics from the Paris School of Mines, Paris, France, in 1987.

He is currently with Ifremer, Plouzane, France. His experience mainly deals with the quantitative assessment of water quality parameters (sea surface temperature, chlorophyll, turbidity, and suspended particulate matters) in the spirit of the Global Monitoring for Environment and Security program of the European Union. He is involved in several national and European projects that are aimed at enhancing

the application of Earth observation data to environmental surveillance (European Coastal Sea Operational Observing and Forecasting System, Water Framework Directive, and MarCoast2, funded by the European Space Agency).



**René Garrello** received the Ph.D. degree in signal processing from the Institut National Polytechnique de Grenoble (INPG), Grenoble, France, in 1981 and the Habilitation (HDR; Habilitation to Supervise Research) in 1995.

He has been with the Ecole Nationale Supérieure des Télécommunications de Bretagne (Telecom Bretagne), Brest, France, since 1985. His main research interests include remote sensing, 2-D signal processing, and statistical and spectral analysis applied to ocean surface feature detection and

characterization.