Reports of the International Ocean-Colour Coordinating Group

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

IOCCG Report Number 4, 2004

Guide to the Creation and Use of Ocean-Colour, Level-3, Binned Data Products

Edited by:
David Antoine (Laboratoire d'Océanographie de Villefranche, France)

Report of an IOCCG working group on ocean-colour, data binning issues, chaired by David Antoine and based on contributions from (in alphabetical order):

David Antoine,	Laboratoire d'Océanographie de Villefranche, France
Janet W. Campbell	University of New Hampshire, USA
Robert H. Evans	RSMAS, University of Miami, USA
Watson W. Gregg	NASA Goddard Space Flight Center, Greenbelt, MD, USA
Marlon R. Lewis	Dalhousie University, Halifax, Nova Scotia, Canada
André Morel	Laboratoire d'Océanographie de Villefranche, France
Cyril Moulin	Laboratoire des Sciences du Climat et de l'Environnement, France
Hiroshi Murakami	JAXA Earth Observation Research and Application Center, Japan

Series Editor: Venetia Stuart

Correct citation for this publication:

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

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

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

http://www.ioccg.org

Contents

Chapter 1

General introduction

David Antoine

One of the goals of launching a number of ocean-colour sensors aboard various satellites is to build a long-term, multi-sensor, multi-year, ocean-colour archive (see IOCCG, 1999; McClain, 1998). The derived chlorophyll concentrations (in time and space) can be used to resolve inter-annual-to-decadal changes in oceanic phytoplankton biomass in response to global environmental changes.

Numerous obstacles stand in the way of this goal, including the different characteristics of the various sensors, (*e.g.* number of channels, their band widths and band centres, variations in the noise characteristics, differences in viewing geometry) as well as the differing approaches taken for the calibration of each sensor and the types of algorithms used to derive the appropriate geophysical quantities. Further, the methods used to average and quality control the basic pixel information, as well as the temporal and spatial binning schemes used to provide fields of geophysical quantities, also come into play when attempting to merge ocean-colour data from different sensors.

The aim of the working group convened by the International Ocean-Colour Coordinating Group (IOCCG) was to examine issues related to the diversity in current "binning" schemes used for ocean-colour data, and to recommend one or more basic approaches that could be used across Agencies or projects, parallel to the current schemes. The goal of this report is thus to produce the best possible, large-scale representation of the current status of binning schemes. It is important to avoid introducing artefacts when merging different data sets as a result of incompatibilities either in the space and time scales or in the interpretation of the information carried by these "Level-3" products[1].

The terms of reference for the data-binning working group (proposed during the 6th IOCCG meeting in La Jolla, CA) are as follows:

  1. Summarize time and space binning schemes and their products.

---

[1]"Level-2" data products are the geophysical values for each pixel, derived from the Level-1A raw radiance counts, by applying the sensor calibration, atmospheric corrections, and bio-optical algorithms. "Level-3" products consist of temporal (*e.g.* weekly, monthly, annually) and spatial composites of Level-2 products. The data are stored in a representation of a global grid or field, whose grid cells or "bins," are typically of the order of 50-100 km^2.

2. Summarize the rationale/reason for each of the different approaches.
3. Assess the impact/consequences of having different schemes.
4. Recommend one or more approaches that each project should incorporate.

This report is based on the above structure, but a few additional items have been added to embrace a wider range of topics. For example, we have examined issues related to merging of ocean-colour data from different sensors, as well as the synergistic use of ocean colour with other satellite-derived observations, such as sea-surface temperature, cloudiness and available radiation (*e.g.* for modelling primary production). In addition, we have examined the assimilation of ocean-colour data into global circulation models (GCMs) with embedded biological segments.

The aim of the report is not to define the best way to generate Level-3 ocean-colour data products from Level-2 data, which is outside the scope of this study, but rather to list previously unconsidered issues that may impair the use, inter-comparison or merging of Level-3 ocean-colour data sets, as well as their application with other satellite-derived data sets.

The following basic questions have been addressed in this report:

1. What are the present binning schemes for ocean-colour data?
2. How should one correct for the bidirectional effects in water-leaving radiances to merge several data sets (either at Level-2 or Level-3) in the appropriate manner? At first glance, there are no comparable issues for the other geophysical products that are usually derived from ocean-colour observations (such as chlorophyll concentration, aerosol optical thickness *etc.*).
3. How should one produce appropriate mean quantities for given variables, so as to fill the Level-3 basic elements (the "bins" of the gridded products) with information relevant to the space and time scales of the variable in question?
4. What space and time scales are best suited to merge Level-3 ocean-colour data sets from different sensors? What common space and time scales could be used for an easy merging and inter-comparison of ocean-colour data sets from different sensors?
5. Same questions as in (4), but for the case where Level-3 ocean-colour data is used in conjunction with other satellite-derived information (*e.g.* SST, cloudiness), when both are used in conjunction with a model (*e.g.* assimilation into GCMs).
6. Should we think of more elaborate products than simple gridded products, the production of which would be based mostly on physical grounds?

At present, the diversity of the various binning schemes makes the merging of different ocean-colour data sets, or their use in conjunction with other satellite data sets, a difficult task. Differences in spatial and temporal scales, differences in techniques used for generating mean quantities, as well as the

lack of correction for bidirectional effects, are the main problem areas, which can result in practical difficulties as well as scientific errors.

Answering the above questions is, however, far from a trivial task because each property has its own spatial/temporal distribution (in a geographical as well as statistical sense). To bin the data, the best spatial/temporal scales may thus be specific to the property in question. For instance, bio-optical properties are often log-normally distributed in the ocean over certain scales (Campbell, 1995), whereas this may not be true for all properties, over all scales. In the end, placing all products on the same grid may not necessarily be the wisest choice, although certainly useful from a practical point of view.

The new generation of ocean-colour sensors allows finer-scale processes and distributions to be resolved (*e.g.* filaments < 1 km in width, over hundreds of kilometres in length). This also raises the question of which of the accessible scales are to be kept when generating Level-3 products, and which are to be abandoned because of the spatial/temporal averaging process. Again, no unique answer exists.

Chapter 2 of the report starts with an inventory of the current binning schemes used for ocean-colour data, as well as schemes employed with other satellite data used in conjunction with ocean colour (*e.g.* modelling of primary production). The proliferation of binning-schemes is immediately apparent. To go beyond building a simple catalogue, several other issues are also examined and illustrated, including those relevant to the bidirectionality of the ocean reflectance (Section 3.1), the generation of mean quantities for a given bin (Section 3.2), issues related to the diversity of spatial and temporal scales (Chapter 4), and those relevant to the use of binned data in primary production models (Chapter 5). We do not claim to have provided an exhaustive discussion on all aspects; rather we have attempted to illustrate the possible consequences of the binning diversity by means of a few simple examples.

In Chapter 6, there is some discussion about the development of products beyond the simple production of "gridded" products (the essence of Level-3 products). General conclusions and recommendations are provided in Chapter 7, and in Chapter 8, a standard binning scheme is proposed that could be used across Agencies, in addition to other mission-specific, binning schemes.

One has to keep in mind that the mandate of the working group did not include data merging, which raises another entire suite of issues that are not within the scope of this report. The issues addressed here are the ones that should be resolved before data merging can easily, and accurately, be performed.

Membership of the IOCCG working group:
- ❖ David Antoine, Chairman, Laboratoire d'Océanographie de Villefranche, France
- ❖ Janet W. Campbell, University of New Hampshire, USA
- ❖ Robert H. Evans, RSMAS, University of Miami, USA

❖ Watson W. Gregg, NASA GSFC, USA
❖ Marlon R. Lewis, Dalhousie University, Canada
❖ André Morel, Laboratoire d'Océanographie de Villefranche, France
❖ Cyril Moulin, Laboratoire des Sciences du Climat et de l'Environnement, France
❖ Hiroshi Murakami, JAXA EORC, Japan

As already mentioned in previous IOCCG reports, the opinions expressed hereafter are those of the authors. They have been reviewed by the IOCCG Committee, but in no way do they represent the policies of Space Agencies or other institutions that support IOCCG activities.

Inventory of binning schemes

David Antoine, Robert Evans, Cyril Moulin and Hiroshi Murakami

2.1   Ocean-colour data

In this section, we summarize the time and space binning schemes and their products for the various past, present or firmly planned ocean-colour missions that have already generated, or will soon generate, Level-3 binned products. Some existing missions are not mentioned here because they are not currently producing any Level-3 products. This section also summarizes the rationale for the different approaches. Additional technical details of the various binning schemes can be found by consulting the list of URLs provided in the Appendix.

2.1.1   CZCS (NASA, November 1978–June 1986)

2.1.1.1   Grid characteristics

Level-3 CZCS data are binned to a fixed, linear latitude-longitude (equal angle) grid with a dimension of 1024 (latitude) × 2048 (longitude) and a resolution of ~18.5 km at the Equator. The global composites are either daily, weekly, monthly or annual averages. The annual composite is based on 12 monthly composites (not 365 daily composites).

2.1.1.2   Product averaging

Products mapped to this grid include phytoplankton pigment concentration, the diffuse attenuation coefficient at 490 nm, the normalized water-leaving radiance at 440, 520 and 550 nm, and the aerosol radiance at 670 nm. Monthly and seasonal composites are un-weighted, arithmetic, averages of pigment concentration for all pixels containing valid data, from daily composite images. None are true means in the sense of representing an unbiased, central tendency for a given pixel over the specified time and space interval. For example, the values shown in high latitude seas in the annual composites do not include any winter values during periods of ice cover, or during the polar winter.

Climatologies have also been produced for each month. For a given month, they are based upon arithmetic averages of pigment concentration for all available monthly composites of the month in question, between 1979 and 1986.

2.1.2 POLDER-1 and POLDER-2 (CNES, October 1996–June 1997, and December 2002–October 2003)

2.1.2.1 Introduction

The first POLDER instrument was launched onboard the Japanese platform ADEOS in August 1996. Data acquisition stopped on 30 June 1997 because of a failure of the satellite's solar panel. Eight months of data (November 1996 - June 1997) were processed by CNES. POLDER is a multi-mission sensor with directional and polarized capabilities to study clouds, aerosols and both marine and terrestrial biospheres (Deschamps *et al.*, 1994). A second POLDER sensor was launched in December 2002 onboard ADEOS-2, and the mission ended on 25 October 2003 due to satellite problems. POLDER-2 was identical to its predecessor (it was actually the backup instrument) which will ensure consistency between the two data sets, even if improved algorithms are used (Nicolas and Deschamps, 2000).

Ocean colour was one of the most difficult aspects of POLDER-1 processing, primarily because of problems with the 443 nm band, as well as calibration uncertainties in the blue bands (POLDER does not have an onboard calibration device and the calibration relies on a set of vicarious calibration techniques). The ocean-colour algorithm uses three red and near-infrared spectral bands (670, 765 and 865 nm) for atmospheric correction. Directional measurements are used in atmospheric correction to detect absorbing aerosols, and to estimate the "above-water, directional marine reflectances" in the three blue-green spectral bands (443, 490 and 565 nm). These reflectances are then averaged to compute the under-water diffuse marine reflectance necessary to estimate the chlorophyll concentration. Note that the bidirectionality is accounted for during this averaging procedure according to the methods of Morel and Gentili (1996). Up until now, polarization measurements were not used in the ocean-colour processing.

2.1.2.2 Grid characteristics

For POLDER-1, Level-3 ocean-colour products were generated on a monthly basis, whereas both Level-3 and mapped products are available for POLDER-2. The major differences between the two products are in spatial resolution and data format. The Level-3 and mapped products are somewhat similar to the binned and mapped SeaWiFS Level-3 products, respectively. Both products are described in this report even though mapped products are not yet available for POLDER-1. It
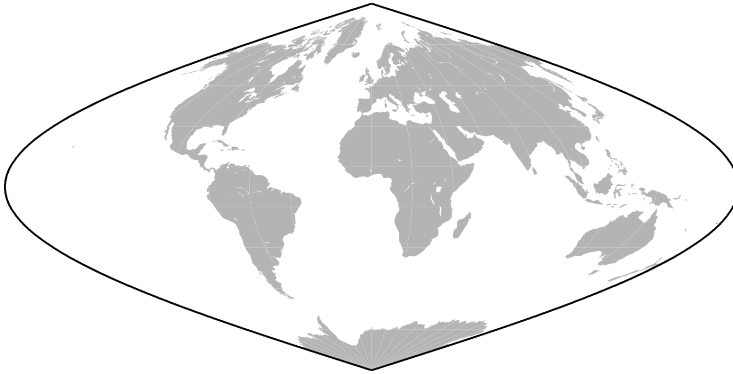
Figure 2.1   The POLDER projection: line (lin) is 1 to 3240 from top to bottom, column (col) is 1 to 6480 from left to right.

is, however, reasonable to believe that POLDER-1 data will be reprocessed, and will be made available to the community in the near future in both POLDER-2, Level-3 format, as well as the mapped format. A simplified version of these mapped products for POLDER-1 data can be downloaded from the POLDER web site (http://polder.cnes.fr/) in the "Scientific Products" menu.

Level-3 products

Level-3 ocean-colour products from POLDER are composed of one binary file per month containing global maps of several aerosol and marine variables for three, 10-day periods (decades), and for the month. The first and second decades extend from days 1 to 10 and 11 to 20, respectively. The third decade extends from day 21 to the end of the calendar month. The latter is thus derived from a variable number of days.

This product is projected on a "sinusoidal equal area" grid (Sanson-Flamsted; see Figure 2.1) at the full POLDER resolution (18 pixels per degree; 3240 rows of 6480 pixels) described below. Note that this grid is also used for Level-1 and -2 POLDER products. No spatial averaging is employed to produce the Level-3 products. The step is constant along a meridian with a resolution of 1/18 degrees. Thus, there are $180 \times 18 = 3240$ lines from pole to pole. Along a parallel, the step is chosen to have a resolution as constant as possible. Thus, the number of pixels from 180°W to 180°E is chosen to equal $2 \times$ NINT[3240 cos(latitude)] where NINT stands for nearest integer.

Note that the coordinates of the neighbours of a given pixel (lin, col) are *not* necessarily given by (lin ±1, col ±1). It is essential to account for the deformation of the projection with longitude. The following equations yield the latitude and longitude of a pixel given by its (lin, col) coordinates in the POLDER reference grid:

$$\text{lat} = 90 - \frac{\text{lin} - 0.5}{18}$$
$$N_i = \text{NINT}[3240 \ \cos(\text{lat})] \tag{2.1}$$
$$\text{lon} = \frac{180}{N_i}(\text{col} - 3240.5)$$

The following equations yield the lin±1, col±1 coordinates in the POLDER reference grid for a pixel of given latitude and longitude:

$$\text{lin} = \text{NINT}[18(90 - \text{lat}) + 0.5]$$
$$N_i = \text{NINT}\left[3240 \sin\left(\frac{\text{lin} - 0.5}{18}\right)\right] \tag{2.2}$$
$$\text{col} = \text{NINT}\left[3240.5 + \frac{N_i}{180}\text{lon}\right]$$

Mapped products

POLDER mapped ocean-colour products consist of a main directory that contains one HDF-EOS file per time period (3 decades and a month) and per variable. This main directory also contains a subdirectory in which one HDF-EOS file of chlorophyll concentration is available for each day of the month. The HDF-EOS format was chosen for compatibility with NASA products (*e.g.* SeaWiFS, MODIS).

These maps are projected on a "cylindrical" (or regular latitude-longitude) grid at a lowered resolution of 12 pixels per degree (2160 rows of 4320 pixels). As for the file format, this spatial resolution has also been chosen for compatibility with NASA products. The temporal averaging for the decades and the month is the same as for Level-3 products, and the spatial averaging is performed using a bi-linear interpolation.

2.1.2.3   Product averaging

For all variables, except chlorophyll, the 10-day and monthly means are obtained by an arithmetic average of the variable itself. For chlorophyll and pigment indices, the logarithm of the variable is used for averaging (Moulin *et al.*, 2001).

For all variables including chlorophyll, each daily value is weighted by the quality index of the single Level-2 retrieval, so that all Level-2 pixels used to compute the average do not have the same contribution to the 10-day or monthly mean. The quality index depends mostly on the atmospheric correction. It decreases with optical depth, but it also accounts for the coherence of the aerosol properties retrieved from the various viewing angles available for a given pixel (Nicolas and Deschamps, 2000).

Major evolutions from POLDER-1 to POLDER-2 are: (1) The mean marine diffuse reflectance is given at all wavelengths (*i.e.* 443, 490, 565 and 670 nm)

instead of being provided only at 565 nm; (2) Two pigment indices will be available for POLDER-2: the NDPI and the "SeaWiFS-like" chlorophyll already used for POLDER-1; (3) Three Inherent Optical Properties ($b_{b,565}$, a_{443} and a_{490}; Loisel and Stramski, 2000) will be computed. Note that for all ocean-colour variables, the standard deviation and other statistical information (minimum, maximum), is only available for the monthly mean (not for decades).

2.1.3   SeaWiFS (NASA, September 1997–present)

2.1.3.1   Introduction

This section describes the equal-area gridding scheme proposed by the RSMAS Remote Sensing Group for the binned, sea surface temperature fields, produced by the AVHRR Pathfinder project, an approach which is being adopted for SeaWiFS binned ocean-colour products. The gridding scheme is based on that adopted by the International Satellite Cloud Climatology Project (ISSCP).

Motivation for the need for an equal area grid for SeaWiFS, or other oceanographic products, can be found in a paper by Rossow and Gardner (1984), so it will not be further discussed here. Furthermore, this section describes only the design of the proposed equal-area grid, and does not discuss other related topics such as rules for spatially or temporally combining observations into the equal-area bins (see Section 3).

2.1.3.2   Grid characteristics

The gridding scheme proposed consists of rectangular bins or tiles, arranged in zonal rows. A compromise between data processing and storage capabilities, on the one hand, and the potential geophysical applications of satellite data, on the other, suggests that a suitable minimum bin size would be approximately 8-10 km on a side.

In the scheme proposed here, the tiles are approximately 9.28 km on a side. This size (9.28 km) was chosen because (a) it has approximately the desired minimum resolution, and (b) it results in 2160 zonal rows of tiles from pole to pole (*i.e.* 1080 in each hemisphere). This particular number of rows (2160) has some advantages which will be discussed in more detail below. Because the total number of rows is even, the bins will never straddle the Equator (*i.e.* there will be an equal number of rows above and below the Equator). This avoids possible situations where the Coriolis factor is zero, a characteristic that numerical modellers expect from any gridding scheme adopted.

The total number of ~9-km bins is 5,940,422. The bins or tiles are arranged in a series of zonal rows; the number of tiles per row varies. The rows immediately above and below the Equator have 4320 tiles. This number is derived
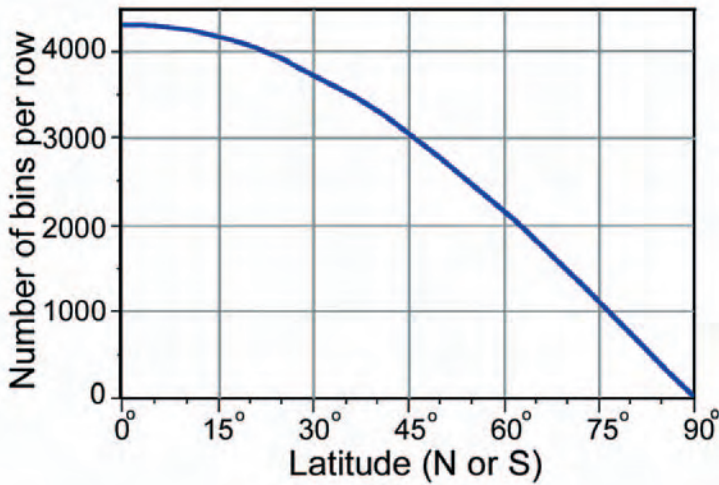
Figure 2.2 Number of 9.28 km tiles per zonal row as a function of latitude (North or South). The number of tiles is 4320 at the Equator, and decreases to 3 at the poles.

by dividing the perimeter of the Earth at the Equator by the standard tile size (*i.e.* $2\pi R_e/9.28$), where R_e is the Equatorial radius of the Earth ($R_e = 6{,}378.145$ km). The number of tiles per row decreases approximately as a cosine function as the rows get closer to each pole (there should be an adjustment for ellipticity of the Earth, as the Equatorial radius decreases progressively to the smaller polar radius, but this adjustment is not applied in the current implementation). At the poles, the number of tiles is always three. This special situation will be discussed in detail below. The number of tiles per row as a function of latitude is shown in Figure 2.2.

The number of bins in each zonal row is always an integer. To ensure an integer number of bins, the width of each bin (the size of a bin along a parallel, or x-length) must vary slightly from row to row. The bins, however, are always 9.28 km long along the meridians. That is, only one of the bin dimensions changes. The size of the bins at each zonal row is established in the following manner. First, a preliminary value for the number of tiles (N_p) at a given latitude (L) is computed as:

$$N_p = 2\pi r/X \qquad (2.3)$$

where X is the x-size of a bin at the Equator (9.28 km) and r is the radius of the circle produced by slicing the Earth with a plane parallel to the Equator at latitude L. The radius, r, can be calculated as:

$$r = R_e \cos(L) \qquad (2.4)$$

where R_e is the Equatorial radius of the Earth. If the fractional part of N_p is greater than, or equal to 0.5, then N_p is rounded up to the nearest integer (*i.e.* the final number of tiles would be the integer portion of N_p plus one), otherwise N_p is rounded down (the final number of tiles is the integer portion of N_p). Once the final integer number of tiles along a row is calculated, the x-size of the tiles must be adjusted. This is done by dividing the perimeter of the row ($2\pi r$) by the integer number of tiles. The result is the x-length of a tile (width) for a given row.

Because the x-length of the tiles is adjusted to ensure an integer number at each row, the "equal area" characteristics of this binning scheme are not rigorously preserved. However, variations in tile size are negligible throughout most of the globe, and only become relevant at very high latitudes, where there are fewer tiles per row, thus any adjustments are more noticeable. As soon as the number of tiles increases with distance from the poles, the difference between tile sizes rapidly becomes unnoticeable. To provide an idea of the magnitude of the fluctuations in tile size, the worst possible case occurs when half a tile remains "uncovered" after filling a zonal row with an integer number of tiles. Once a row has 100 bins (approximately 16 rows, or 148 km from the poles), the worst possible difference between the actual tile x-length and the standard x-length is of the order of 0.5% (*i.e.* half a tile's length redistributed among about 100 tiles). For a tile of about 9 km a side, this represents a difference in the x-length of about 45 m. Through a similar calculation, a row with 50 bins (about 80 km away from the poles) has a 1% variation with respect to the standard bin size.

The gridding scheme described here has an extremely useful feature: the number of 9.28 km tiles in each hemisphere (1080) is divisible by many numbers (*e.g.* 2, 3, 4, 5, 6), therefore it is extremely easy to generate an integer number of rows at many useful spatial resolutions. For instance, 12 rows of 9.28 km tiles can be combined to generate zonal bands of approximately one degree (one degree of latitude is equal to 111.12 km; 12 bins would form a band 111.20 km wide). Another example is the use of 30 rows to generate zonal bands of approximately 2.5° (a typical output resolution of atmospheric circulation models).

2.1.3.3 The poles

Both the North and South poles are special cases in the gridding scheme presented here. The pole areas are always covered by three tiles, shaped like pie sectors. While the meridional size of the polar bins (the y-length) will be the usual 9.28 km, the length of the bins along the arc of the sectors will be slightly larger. Ignoring sphericity, the area encompassed by the last row of tiles is πX^2, where $X = 9.28$ km. If we express the area of the circle as a rectangle of height X, the remaining dimension is πX. If we divide the perimeter by three (to yield three tiles), each tile will have dimensions X by $\pi X/3$ (approximately

$1.05X$). That is, the bases of the triangular polar tiles are about 5% larger than the x-length of the Equatorial tiles.

2.1.3.4 Product averaging

Internally, the binning scheme carries the variables sum$X$, sumX^2, N (number of observation), and "Quality", where quality is variable dependent. For simple arithmetic averaging, the first three variables are simply summed according to binning rules that define how data of differing quality are to be treated. At present, only data of the highest available quality are retained, while the remainder are discarded. In a more advanced scheme, where weights are available, the data could be summed according to the relative weight of a given quality level. This pixel combining scheme is employed for summation in both space and time, allowing the space and time resolution fields to be reduced in resolution as needed. The final step of providing observation counts, pixel means and standard deviations, takes place in the mapping stage.

2.1.4   OCTS (NASDA, November 1996–June 1997)

2.1.4.1 Grid characteristics

The OCTS Level-3 binning algorithm and format is almost the same as that for SeaWiFS, except for the use of weekly time bins instead of 8-day time bins.

2.1.4.2 Product averaging

OCTS uses the logarithmic summation for the following reasons:
1. It follows the format initially planned for the SeaWiFS project[2].

$$\overline{X}_{\mathrm{mle}} = \exp(m + s^2/2)$$

   where m and s are maximum likelihood estimators for the sample mean and sample variance of $\ln(X)$ (see Campbell *et al.* (1995) for details).
2. The term "s" has been deleted because "s" enlarges spike noise from input Level-2 data.
3. The log-scheme reduces the influence of the spike noise.
    Variables are listed in Table 2.1.

There are other OCTS products through a collaboration of the SeaWiFS and SIMBIOS Projects and NASDA (JAXA). They have been processed by SeaWiFS-like algorithms except for calibration factors, and are written in a similar format to

---

[2]This was recommended at first by Campbell *et al.* (1995), but the SeaWiFS project eventually reverted to the use of the arithmetic average

Table 2.1   Contents of OCTS Level-3 ocean products

Product			Variable		Units
Level-2	**Level-3 Binned**	**Level-3 Mapped**	**Name**		
OC1	OC Daily Weekly[†] Monthly[‡]	OCL	1)	nLw412	$mW\ cm^{-2}\ sr^{-1}\ \mu m^{-1}$
			2)	nLw443	
			3)	nLw490	
			4)	nLw520	
			5)	nLw565	
			6)	La670	
			7)	La765	
			8)	La865	
			9)	$\tau_a 865$	Non-dimensional
		None	10)	EPSI 670/865	Non-dimensional
OC2		OCP	1)	CZCS PIG	$mg\ m^{-3}$
		OCC	2)	CHLO	$mg\ m^{-3}$
		OCK	3)	K490	m^{-1}
		None	4)	Integrated CHL	$mg\ m^{-2}$
SST[§]	SST	SST	SST		kelvin

†– calendar week   ‡– calendar month   §– daytime only

SeaWiFS Level-3 files. They do not include PAR because OCTS has no piecewise linear gain, and it saturates over clouds.

2.1.5   MODIS-Terra and MODIS-Aqua (NASA, December 1999–present, and May 2002–present)

The scheme for the production of Level-3 binned products is similar to that described in Section 2.1.3 for SeaWiFS. The only difference is that the basic bin size is 4.62 km (*i.e.* four times as many bins). In addition, the mapped products (equal angle grid) are provided at three different spatial resolutions: 4.88 km, 39 km and 1 degree, and at daily, weekly (8-day), monthly and yearly temporal increments. Unlike SeaWiFS, where product quality is defined only for chlorophyll, each MODIS product has a product specific quality field that permits quality definition to be tailored to the product's specific requirements.

2.1.6   GLI (NASDA, December 2002–October 2003)

The GLI sensor, which was onboard the ADEOS-2 platform, is an optical sensor that observes solar light reflected from the Earth's surface including land, ocean and clouds, or infrared radiation. Products include global and frequent

Table 2.2 Contents of GLI Level-3 ocean products

Level-2 Path-type Product	Variable Name	Level-3 bin Product	Level-3 Sta Product	Units
NL	1) nLw380	NW Daily 8 days Monthly[†]	NW	$mW\ cm^{-2}\ sr^{-1}\ \mu m^{-1}$
	2) nLw400			
	3) nLw412			
	4) nLw443			
	5) nLw460			
	6) nLw490			
	7) nLw520			
	8) nLw545			
	9) nLw565			
	10) nLw625			
	11)nLw666			
	12) nLw680			
	13) nLw710			
	1) aerosol albedo	LA	LA	Non-dimensional
	2) La865			$mW\ cm^{-2}\ sr^{-1}\ \mu m^{-1}$
	3) nLw_678			
	4) nLw_865			
	5) Tau_ 865			Non-dimensional
	6) angstrom_520			Non-dimensional
	7) PAR	TBD	TBD	$Einstein\ m^{-2}\ d^{-1}$
	8) 12 flags	None	None	32-bit quality flag
CS	CHLA	CS	CHLA	$mg\ m^{-3}$
	CDOM		CDOM	m^{-1}
	K490		K490	m^{-1}
	SS		SS	$g\ m^{-3}$
	12_ flags	None	None	32-bit quality flag
ST	SST	ST[‡]	SST_ all	kelvin
			SST_Day Night	
	flag	None	None	16-bit quality flag

†– calendar month   ‡– descending and ascending data stored separately

estimates of physical data such as surface temperature, vegetation distribution, and ice distribution. GLI is the follow-on mission to OCTS, and has greater precision (12-bit digitization), and more observation targets. GLI has 23 channels in the visible and near-infrared regions (VNIR, 380-865nm), 6 channels in the short-wavelength infrared region (SWIR, 1050-2210 nm), and 7 channels in the middle and thermal infrared region (MTIR, 3.7-12 m) for its multi-spectral observations. The spatial resolution is about 1-km at nadir, and six channels in the VNIR and SWIR have a resolution of 250 m. The swath width of GLI is about 1,600 km, achieved by rotating the scanning mirror mechanically. Its tilting angle is about +20 degrees or -20 degrees from the progressing direction.

2.1.6.1 Grid characteristics and product averaging

The GLI Level-3 binning algorithm and format is somewhat changed from that of OCTS and SeaWiFS. The characteristics are as follows:

1. A log-average scheme was used for the at-launch version, the same as that used for OCTS.
2. Variables are stored as Scientific Data (SD) in the Hierarchical Data Format (HDF), and not as V data.
3. OCTS Level-3 binned data included flag information from Level-2 pixels for each bin. However, GLI Level-3 data only stores a set of statistics for the flag bits in order to reduce the data volume.

Variables for GLI Level-2 and 3 products are listed in Table 2.2.

2.1.7 MERIS (ESA, March 2002–present)

2.1.7.1 Introduction

The MEdium Resolution Imaging Spectrometer (MERIS) was launched onboard ESA's Envisat satellite on 1 March, 2002. It is a 68.50 field-of-view push-broom imaging spectrometer that measures the Earth's reflected solar radiance in 15 spectral bands, programmable in width and position, in the 400 to 900 nm spectral range. MERIS has a dual spatial resolution, namely 300 and 1200 m and allows global coverage of the Earth in 3 days.

The primary objective of the MERIS mission is the measurement of ocean colour. The operational Level-1b and Level-2 products directly relevant to ocean studies include water-leaving reflectances in 13 of the 15 spectral bands, an algal pigment index derived from traditional band-ratio methods (optimized for Case 1 waters) as well as coastal zone products including an algal pigment index, suspended matter concentration, and yellow substance absorption, all three being derived by inversion techniques.

Though not included in the early suite of MERIS of products, Level-3 products should be released soon. These products are briefly described in the next section.

2.1.7.2 Grid characteristics and Product averaging

To start with, the coverage of these products would be global with a spatial resolution of about 9.3 km (5 minutes of arc), and their temporal resolution would be weekly. The binning methodology followed for the generation of ocean-colour product(s) would be largely inspired by the binning scheme used by SeaWiFS.

2.1.7.3 Data products

The preliminary Level-3 geophysical candidates currently under investigation are the algal pigment index derived for Case 1 waters, the MERIS Global Vegetation Index (MGVI), the total columnar water vapour concentration, the aerosol optical thickness (at the 865 nm band) and its spectral dependency, and the desert dust presence (based on the results of an absorbing aerosol flag). Additional products could be considered at a later date.

2.1.8   Summary tables

Table 2.3   Temporal averaging

Sensor	Temporal scales
CZCS	1 day, calendar month, calendar year
OCTS	1 day, 7 days (starting 3 November, 1996), calendar month
POLDER 1 and 2	1 day 1st decade of the month 2nd decade of the month Last part of the month (8, 9, 10, 11 days) Calendar month
SeaWiFS	1 day, 8 days (starting 1 January, 45 x 8 d periods + 5 or 6 days) Calendar month, calendar year
MODIS	1 day, 1 week, 3-week, calendar month
GLI	1 day, 8 days (starting 1 January, 45 x 8 d periods + 5 or 6 days) Calendar month
MERIS	8 days (starting 1 January, 45 x 8 d periods + 5 or 6 days)

Table 2.4 Spatial averaging: binned products (equal area projection)

Sensor	Resolution	N° Pixels/tiles/bin
CZCS	No binned products	
SeaWiFS, OCTS, GLI and MERIS	1/12° latitude 1/12° longitude at Equator Approximate size	2160 pole to pole 4320 at Equator, 3 at Poles 9.28 × 9.28 km
POLDER 1 and 2	1/18° latitude 1/18° longitude at Equator Approximate size	3240 pole to pole 6480 at Equator. 6.18 × 6.18 km
MODIS	1/24° latitude 1/24° long. at Equator Approximate size	4320 pole to pole 8640 at Equator 4.63 × 4.63 km

Table 2.5 Spatial averaging: mapped products (equal angle projection)

Sensor	N° Pixels/tiles/bin
CZCS	2048 pixels / 1024 lines, approximate size 18.5 km at the Equator
SeaWiFS, OCTS, GLI and MERIS	4096 pixels / 2048 lines, approximate size 9.25 km at the Equator
POLDER 1 and 2	No mapped products
MODIS	4096 pixels / 2048 lines, approximate size 9.25 km at the Equator 8192 pixels / 4096 lines, approximate size 4.6 km at the Equator 360 pixels / 180 lines, size 1-degree

Table 2.6 Calculation of mean quantities in a given bin.

Sensor	Methods
CZCS, SeaWiFS, MODIS and MERIS	Arithmetic mean
OCTS	Log mean
GLI	Log mean
POLDER 1 and 2	Log mean for Chlorophyll Arithmetic mean for other variables

2.2 Other satellite data

In this section, we summarize the time and space binning schemes and their products for existing "auxiliary" products that are often used in conjunction with ocean-colour data in models. This section is the counterpart of Section 2.1, which deals with ocean colour. It must be mentioned that the auxiliary data used when processing the ocean-colour, top-of-atmosphere observations are not dealt with here; it relates only to other data that are used in conjunction with ocean colour in modelling applications.

2.2.1 Sea-surface temperature (SST)

The binning schemes for temperature observations from the AVHRR pathfinder, OCTS and the two MODIS sensors, are identical to those adopted for ocean colour and have already been described in Section 2.1, with the exception of the averaging scheme for OCTS, which uses an arithmetic mean for temperature as opposed to a log mean for ocean colour.

The two other temperature products commonly used in conjunction with ocean colour are the Levitus climatology (Levitus, 1982) and Reynolds analyses (Reynolds, 1988; Reynolds, 1993; Reynolds and Marsico, 1993, Reynolds and Smith, 1994), both provided on the same 1-degree by 1-degree latitude/longitude grid (Table 2.7). The Levitus climatology comprises mean monthly fields generated from *in situ* measurements only, whereas Reynolds' analyses generate weekly (7-day) blended fields for years from 1982 to present, which combine *in situ* and satellite SSTs (AVHRR).

2.2.2 PAR

We have identified six Photosynthetically Available Radiation (PAR) data sets in production and available for distribution and use (see Table 2.8). All utilize an equal-angle projection, but spatial and temporal resolutions vary. Also, there are often limitations in the data set, such as ocean only, or domain limitations. In the case of MODIS and MERIS, PAR is only instantaneous and therefore is not useful for radiation budget analyses or primary production calculations (only usable for interpretation of the signal due to natural phytoplankton fluorescence).

2.2.3 Cloud cover

Cloud fraction data sets are more numerous than PAR, and our listing may be only partial. Data set binning characteristics are also more varied than PAR, al-

Table 2.7 Sea Surface Temperature (SST) data sets.

Data Set	Spatial Resolution	Temporal Resolution	Projection	Inclusive Dates
AVHRR	9 km 18 km 54 km	8-day Monthly	Equal-angle	2000-2003
OCTS	9 km	Daily Weekly Monthly	Equal-angle	Nov. 1996 to Jun. 1997
GLI	9 km	Daily 8-day Monthly	Equal-angle	Apr. 2003 to Oct. 2003
MODIS-binned	4.63 km	Daily Weekly Monthly Yearly	Equal-area	Feb. 2000 to present May 2002 to present
MODIS-mapped	4.88 km 39 km 1-degree	Daily Weekly Monthly Yearly	Equal-angle	Feb. 2000 to present May 2002 to present
LEVITUS	1-degree 0.5 E to 0.5 W and 89.5 S to 89.5 N	Monthly	Equal-angle	Climatology
Reynolds	1-degree 0.5 E to 0.5 W and 89.5 S to 89.5 N	Weekly	Equal-angle	Jan. 1982 to present

though again most are binned to an equal-angle projection. Spatial and temporal resolutions vary, but 1-degree and monthly bins are common (Table 2.9).

2.2.4 Products derived from altimetry

Generating continuous gridded fields from the measurements of space-borne altimeters raises significantly different issues compared to other measurements such as ocean colour or SST. Indeed, altimeter data are in essence along-track data, which need to be interpolated in order to produce continuous fields.

This is usually performed through global, sub-optimal space/time objective analyses (*e.g.*, Le Traon *et al.*, 1998; Ducet *et al.* 2000), where spatial and temporal scales are derived beforehand from historical data. These scales are intended to represent typical scales of sea level anomaly features (linked to the internal Rossby radius, varying with latitude from about 80 to 400 km), as well

Table 2.8 Photosynthetically available radiation (PAR) data sets.

Data Set	Spatial Resolution	Temporal Resolution	Projection	Inclusive Dates	Notes
ISLSCP	1-degree	3-hourly Monthly	Equal-angle	Jan. 1987 to Dec. 1988	
SeaWiFS	9 km	Weekly Monthly	Equal-angle	Sep. 1997 to present	Ocean only
MODIS	4 km	Daily Weekly Monthly	Equal-angle	Nov. 2000 to present	Ocean only Instantaneous
SSI	0.5-degree	2-hourly Daily	Equal-angle	Jan. 1991 to Dec. 1993	
ESSW	2.5-degree	Daily	Equal-angle	Jul. 1983 to Aug. 1994	Only to 60° lat.
PAR Project	2.5-degree	3-hourly Daily Monthly	Equal-angle	Jul. 1983 to Dec. 1993	

as temporal correlation scales that also vary with latitude (from about 10 to 40 days). From these analyses, gridded fields are eventually produced on regular latitude/longitude grids.

PODAAC data sets are also provided on the same two grid patterns: 1) 0.5 degree grid in latitude and longitude, with gaps where no data are present; 2) 1.0 degree grid, with all gaps filled in by interpolation. Sea Surface Heights (SSH) are given in millimetres, for October 1992 to December 2001 (later grids available online), averaged over 10 days, referenced to the 1993-2001 mean, with tidal and inverted barometer effects removed and all instrument and path corrections applied (Table 2.10).

Table 2.9   Cloud cover data sets.

Data Set	Spatial Resolution	Temporal Resolution	Projection	Inclusive Dates
ISSCP	2.5-degree	3-hourly Monthly	Equal-angle	Jul. 1983 to Dec. 1999
DAO	2-degree	Monthly	Equal-angle	Mar. 1980 to Nov. 1993
ISLSCP	1-degree	Monthly	Equal-angle	Jan. 1987 to Dec. 1988
SSI	0.5-degree	3-hourly	Equal-angle	Jan. 1991 to Dec. 1993
NCEP/NCAR	Variable	Daily reanalysis	T62 Gaussian	Jan. 1948 to present
MODIS	1-degree	Daily	Equal-angle	Nov. 2000 to present
DMSP	1-degree 2.5-degree	Monthly	Equal-angle	Jul. 1987 to present[†]
TOVS	1-degree	Monthly	Equal-angle	Jan. 1985 to Dec. 1992
PATMOS	1-degree	daily	Equal-area	Jul. 1981 to Jun. 2001
ECMWF	2.5-degree	6-hourly	Equal-angle	Jan. 1957 to present

†– Ocean only

Table 2.10   Delayed time-gridded altimetry products.

Data Set	Spatial Resolution	Temporal Resolution	Projection	Inclusive Dates
SLA (AVISO) (Jason-1 or T/P)	1/3° by 1/3°	6 weeks	Mercator	Oct. 1992 to Nov. 2002[†]
SSHA (PODAAC) (T/P)	0.5° by 0.5° 1° by 1°	10 days 10 days	Equal angle	Oct.1992 to Dec. 2001[‡]

†– Produced every 7 days ‡– Produced every 5 days

Chapter 3

Issues underlying generation of a binned quantity

André Morel, Janet Campbell and Hiroshi Murakami

Issues linked to the bidirectionality of the ocean's reflectance are considered here because they concern the primary product that should be merged across missions, *i.e.* the normalized water-leaving radiances. There are no similar issues for other products (such as chlorophyll), which are derived from the spectrum of the water-leaving radiances. In Section 3.2 we examine the issues linked to the generation of a relevant mean quantity for a given bin in a gridded product, assuming that geometrical artefacts are absent from the Level-2 products.

3.1 Bidirectionality of the ocean's reflectance

In this section, we will identify operations required at the Level-2 processing of the data, to permit production of compatible Level-3 products that can be easily and usefully merged between missions. Note that the discussion about merging data at Level-2, or at Level-3, is not of concern here; irrespective of the choice made, the Level-3 products should nevertheless be compatible with other Level-3 products, *i.e.* the normalized water-leaving radiances (nL_w) have to be fully normalized and the time/space scales compatible. For this reason, this section will be dedicated to the issue of normalization, which has to be considered in parallel, (or even prior to), the time/space binning issues.

The bidirectional structure of the upward field formed by the radiances leaving the ocean's surface is of crucial importance when interpreting remotely-sensed, ocean-colour data. The adjective "bidirectional" is appropriate, as the magnitude of any radiance emerging from the ocean depends on its own direction, as well as on the direction of the incident radiation (*e.g.* on the sun's direction, if the sun were in a black sky). In other words, the radiances leaving a uniform ocean, seen under various angles and differing illumination conditions, are not constant.

These characteristics combine with the way ocean-colour sensors view the Earth, under variable solar angles (as a function of latitude), and with varying viewing angles and azimuth differences between the solar plane and the mea-
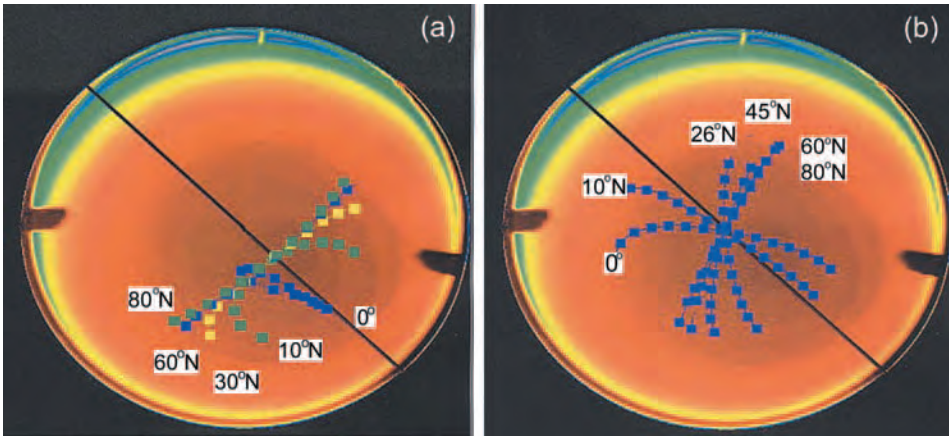
Figure 3.1   Variations with latitude of the SeaWiFS (left) and MODIS (right) scan line geometry. These images were taken by Ken Voss' "RADS" radiance camera, and show the hemispheric radiance distribution just below the sea surface. Black line represents the solar plane. Image courtesy of Ken Voss.

surement plane. This is illustrated in Figure 3.1, where the typical scan line geometry of SeaWiFS and MODIS are displayed for comparison. The range of geometric conditions in these scan lines is large enough for the bidirectionality of the ocean's reflectance to be taken into account.

As a consequence, the consistency of the spectral radiance data, derived from space observation (even under the assumption of a perfect atmospheric correction), is not necessarily ensured, if these geometrical effects are not removed. Coherence is, however, a prerequisite when comparing various pixels (*e.g.* observed within a swath or along the orbit, or viewed during successive days, or by different sensors). Merging the radiance data captured by several instruments operating simultaneously (the goal of the SIMBIOS project, for instance) requires that such an internal consistency be achieved.

The concept of spectral "normalized water-leaving radiance", denoted $[L_w(\lambda)]_N$ (λ is the wavelength), or nL_w, originally introduced by Gordon and Clark (1981), was an attempt at fulfilling this requirement. The underlying rationale was to produce, from actual water-leaving radiances, normalized quantities which would be comparable. With this aim, the actual radiances (measured at sea, or derived from satellite observation) must, ideally, be made independent from the illumination conditions at the instant of measurement. These conditions are essentially determined by the sun's position and the atmospheric diffuse transmittance. Therefore, these environmental influences can tentatively be removed by forming R_{rs} as follows:

$$[L_w(\lambda)]_N = [L_w(0^+, \theta, \phi, \lambda)/E_d(0^+, \lambda)]F_0(\lambda) = R_{rs}(\lambda)F_0(\lambda) \qquad (3.1)$$

where the measured radiance (0^+ is just above the surface, θ is the zenith angle

specifying the emerging radiance, and ϕ is the azimuth difference between the vertical planes containing the sun and the radiance) is divided by the actual irradiance at the sea level, $E_d(0^+)$, and then multiplied by the solar irradiance at the top of the atmosphere, F_0, at the mean sun-earth distance (d_0). The quantity R_{rs}, directly related to $[L_w]_N$, is called the "remote sensing reflectance". The above normalization can be applied equally to field measurements ("sea-truth" validation), or to marine radiances derived from satellite data, after atmospheric correction.

The same $[L_w]_N$ quantity can be interpreted in another way: the normalized water-leaving radiance is a hypothetical radiance which would be measured, if the sun were at zenith, in the absence of atmosphere, and when the Earth is at its mean distance from the Sun. In spite of this interpretation (and in spite of its name), the normalized water-leaving radiance is still a quantity that depends on geometry, since the ocean is not a lambertian reflector; indeed, the initial quantity $L_w(0^+, \theta, \phi)$ introduced in Equation 3.1, was obtained under specific angular conditions (θ, ϕ), and when the sun-zenith angle (θ_s) was not zero, and has thus taken a particular value.

Any slant water-leaving radiance, $L_w(0^+)$, before normalization, can be expressed as:

$$L_w(0^+, \theta_s, \tau_a, w, \theta, \phi, \lambda, \text{IOP}) = E_d(0^+, \theta_s, \tau_a, \lambda) \, \mathfrak{R}(\theta', w) \frac{f(\theta_s, \tau_a, w, \text{IOP})}{Q(\theta_s, \tau_a, w, \theta', \phi, \text{IOP})} \frac{b_b}{a} \quad (3.2)$$

This radiance, $L_w(0^+)$, at a given wavelength λ, is primarily governed by the inherent optical properties of the water body (IOP), but is also dependent upon the geometrical configuration (θ and ϕ), the illumination conditions (summarized by θ_s and τ_a, the aerosol optical thickness), and on the sea state (via the wind speed, w). The illumination conditions also determine the spectral downward irradiance, E_d (above the surface, at 0^+). Three parameters appear in Equation 3.2, namely Q, f, and $\mathfrak{R}$, which are examined below. The parameter Q (sr) is the ratio of upward irradiance (at null depth and just below the surface, denoted 0^-) to the slant (θ', ϕ) upward radiance, also at 0^- (see Equation 3.3).

$$Q = E_u(0^-)/L_u(0^-, \theta', \phi) \quad (3.3)$$

Note that if L_u was constant, Q would equal π sr, irrespective of the angles, θ' and ϕ. This is not generally the case, and Q is dependent upon the angular arrangement and on environmental factors. In addition, it is dependent upon the IOP of the water body (mainly the volume scattering function of the water). The variability of Q is illustrated in Figure 3.2 (reproduced from Morel *et al.*, 2002). The dimensionless coefficient, f, relates the irradiance reflectance, R (the ratio of upward to downward irradiance, beneath the surface), to the ratio (b_b/a) of the two inherent optical properties: b_b, the backscattering coefficient, and a, the absorption coefficient of the water body as follows:

$$R = E_u(0^-)/E_d(0^-) = f(b_b/a) \tag{3.4}$$

The coefficient, f, thus depends on environmental conditions, and also on the IOP. For given IOPs, both f and Q increase when θ_s increases, so that their ratio, which is incorporated in Equation 3.2, varies less (although it may vary within the range 0.07-0.18).

Finally, the dimensionless factor, $\mathcal{R}$, which depends on θ' and on wind speed, merges all the reflection and refraction effects that occur when downward irradiance and upward radiance propagate through the (wavy) interface. For small θ' values ($< 25°$), it amounts to 0.53 and is rather insensitive to the wind speed; in contrast it decreases and depends on w, when θ' increases toward the critical angle. Tables providing $\mathcal{R}$ (w and θ' as entries) are available. If normalization is considered to be too complex a task, simply removing data where θ' is larger than about 25° could be a solution (admittedly, rather drastic).

Because of the remaining dependence upon the geometrical conditions, the normalized water-leaving radiances, as defined in Equation 3.1, are not directly comparable. One way of circumventing this problem is to assume that the pixel is viewed vertically (nadir viewing) with the sun at zenith. This is in line with the initial approach developed by Gordon and Clark (1981), and uses the same logic, although such a hypothetical observation ignores the existence of the specular reflection of the sun disk.

In practise, $[L_w]_N$ (obtained through Equation 3.1) is divided by the f/Q ratio (which corresponds to the actual geometry), then re-multiplied by the f/Q ratio corresponding to the hypothetical viewing conditions ($\theta_s = \theta' = 0$). The same kind of operation can be repeated relative to the factor $\mathcal{R}$.

The "exact" (superscript "ex") normalized water-leaving radiance can be produced from the above transformation, by using $\mathcal{R}_0$ (the value of $\mathcal{R}$ when $\theta_s = 0$, in the absence of wind), f_0 (the f-value when $\theta_s = 0$), and Q_0 (the value of Q when both θ_s and θ' are zero) as follows:

$$[L_w]_N^{ex} = [L_w]_N \frac{\mathcal{R}_0}{\mathcal{R}(\theta', w)} \frac{f_0(\tau_a, w, \text{IOP})}{Q_0(\tau_a, w, \text{IOP})} \left(\frac{f(\theta_s, \tau_a, w, \text{IOP})}{Q(\theta_s, \theta', \phi, \tau_a, w, \text{IOP})} \right)^{-1} \tag{3.5}$$

To use the above equation, one would require information about all parameters involved ($\mathcal{R}_0$, $\mathcal{R}$, f_0, Q_0, f and Q) and for all possible geometrical configurations, environmental conditions, and IOPs. From a practical point of view, look-up tables can be used, but the entries of such look-up tables deserve some comment.

1. The three angles (θ', θ_s, ϕ) are known for each pixel (using satellite and astronomical ephemerides).
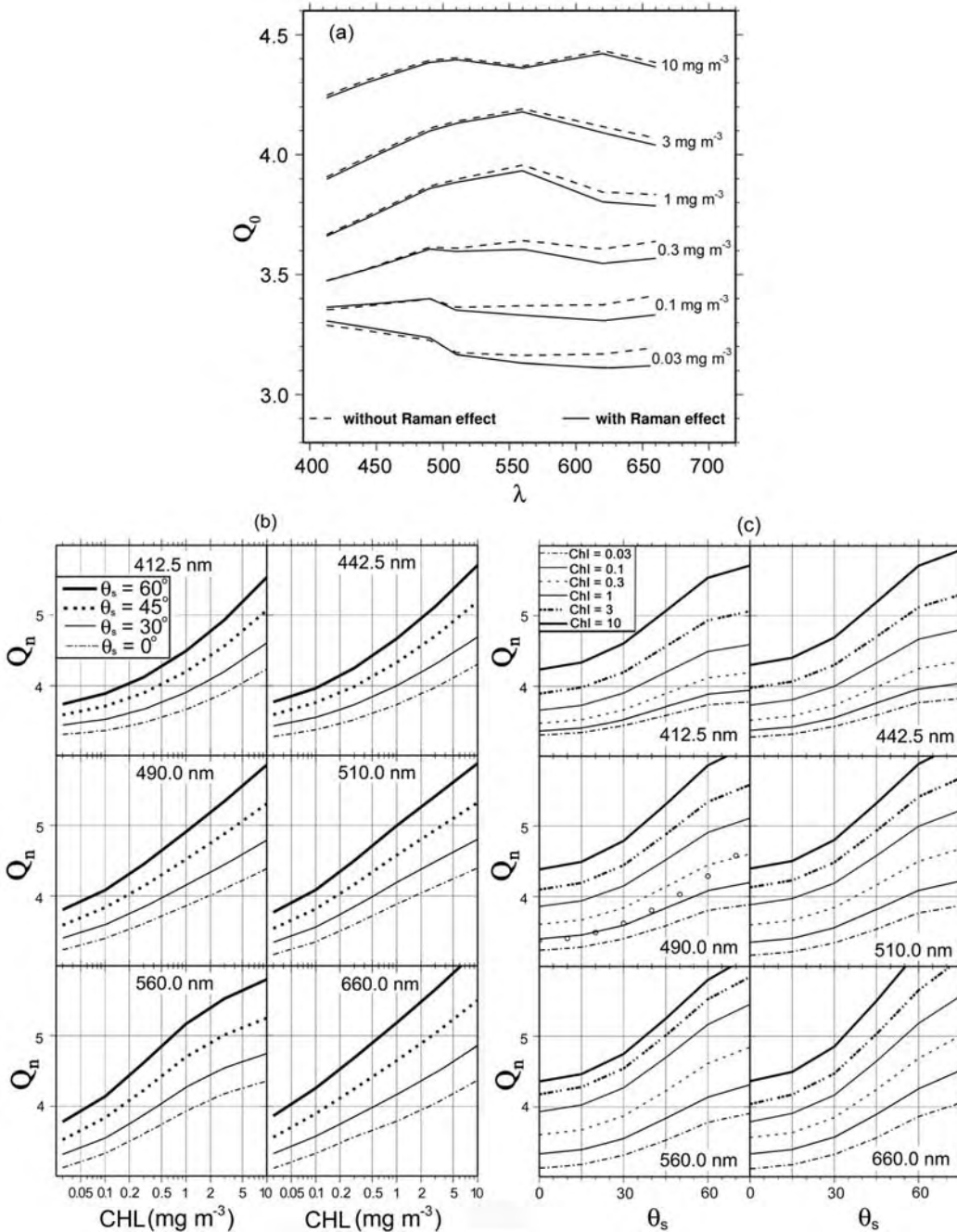
Figure 3.2 (a) Spectral values of the Q_0 function computed with and without Raman emissions for six chlorophyll (CHL) values. (b) Evolution of the Q_n function with increasing chlorophyll concentration for various wavelengths and solar zenith angles. (c) Q_n function of the solar zenith angle for various wavelengths and chlorophyll concentrations. Reproduced from Morel *et al.* (2002). Q_n represents the value of Q when $\theta' = 0$ (*i.e.* Q at nadir), and Q_0 represents the value of Q_n when $\theta_s = 0$ (*i.e.* Q at nadir and the sun at zenith).

2. The aerosol optical thickness (which has a weak influence) is a by-product of the atmospheric correction; the wind speed must be found through an independent means, as it strongly influences the $R(\theta')$ value for large angles (near the edges of the swath, for instance).
3. The IOPs of the section of ocean corresponding to the pixel are unknown, *a priori*.

Equation 3.5 cannot, therefore, be used directly, so an iterative process is required, which is briefly examined below. Note that two IOPs (b_b and a) already appear in Equation 3.2, and their ratio is thus explicitly involved in Equation 3.4 (the ratio $b_b/(a + b_b)$ can also be used in Equation 3.4, replacing f by f'). The dependence of f and Q on IOP is not restricted to the coefficients a and b_b, but also includes other IOPs, namely the volume scattering function (elastic scattering) and the Raman (inelastic) scattering coefficient.

At this point, a distinction should be made between Case 1 and Case 2 waters. In Case 1 waters, the spectral IOPs can be reasonably-well parameterized as a function of the chlorophyll concentration; the entries for the look-up tables are λ and chlorophyll. In such waters, the standard blue-green ratio algorithms are generally used to calculate chlorophyll concentration, *i.e.* ratios of $[L_w]_N$ (*sensu* Gordon and Clark, 1981) for two wavelengths, generally 443 or 490 nm (blue) and 555-560 nm (green). The f and Q parameters for a given chlorophyll concentration do not vary too widely when λ goes from 443 (or 490) to 555 nm, at least for a first approximation (when the angular and environmental conditions are fixed). The blue-to-green ratios, and thus the chlorophyll concentration retrievals, are not greatly influenced by the directional effects because ratios are taken (same compensation effects as for $\mathfrak{R}$). This chlorophyll value can be used as a starting point, and then the exact normalized water-leaving radiances can be produced. Double iterative procedures may also be envisaged (Morel and Gentili, 1996).

A remark must be added concerning the look-up tables. The quantities f and Q can be measured at sea with dedicated instruments, although very few determinations are currently available (but see Morel *et al.*, 1995, and Zibordi and Berthon, 2001). It is therefore necessary to rely on theory, and the look-up tables have to be constructed by solving the radiative transfer equation, under specified boundary conditions, and with the appropriate IOPs describing the water body as a function of the chlorophyll concentration (as far as Case 1 waters are concerned). This is the present situation, which requires the adoption of a bio-optical model.

For Case 2 waters, other optically significant substances are present (*e.g.* dissolved yellow substances, various sediments and particles), that vary independently of phytoplankton concentration. Consequently, their IOPs are less predictable, and uncoupled with chlorophyll concentration. Regarding their bidirectional properties, the dark "yellow substance-dominated" Case 2 waters

are similar to Case 1 waters with high chlorophyll concentrations. In contrast, the bright, turbid "sediment-dominated" Case 2 waters tend to produce a diffuse quasi-isotropic upward radiance field (Q approaching π). Currently, there is no simple and accurate solution to account for the bidirectional effects in such waters, notwithstanding the efforts of Loisel and Morel (2001). Conservative solutions could be as follows. The dark, yellow-substance rich waters are not detected as such, using the current algorithms. Instead, they are qualified as chlorophyll-rich waters, so that the corresponding bidirectionality correction for high chlorophyll is an acceptable approximation. For turbid Case 2 waters, which are easily identified during data processing through their abnormally high reflectance in the green domain, the situation is simplified. Indeed, the f/Q ratio is rather spectrally insensitive and varies weakly with geometry (as long as θ_s is below 60°; see Table 4 in Loisel and Morel, 2001), so that there is no need for a correction, except for the $\mathfrak{R}_0/\mathfrak{R}$ effect.

The production of normalized water-leaving radiances involves the adoption of the extra-terrestrial solar irradiance, $F_0(\lambda)$, and the computation of its value for the various spectral bands (position and width) which are available on the various sensors. Therefore, compatible $[L_w]_N$ values can be obtained only if the initial $F_0(\lambda)$ spectrum adopted by the various Agencies is the same. Recommendations in this regard have been provided by IOCCG in view of reaching an agreement.

Table 3.1 Summary table for the various radiance or reflectance products.

Sensor	Product	
SeaWiFS	$nL_w = [L_w]_N$ where $[L_w]_N = L_w/[\varepsilon_c \mu_s t_d(\theta_s)]$	"Normalized water-leaving radiance"
MODIS	$nL_w = [L_w]_N^{ex}$ According to Equation 3.5.	"Normalized water-leaving radiance"
OCTS and GLI	$nL_w = [L_w]_N$, same as for SeaWiFS and MODIS	
MOS	πnL_w	"Normalized reflectance"
POLDER 1 and 2	$\rho^+ = \pi nL_w/F_0$	"Directional marine reflectance"
	ρ^-	"Diffuse marine reflectance"
	$\rho^- = \dfrac{\rho^+ Q(chl,\lambda,\theta_s,\theta,\Lambda\phi)}{\pi\mathfrak{R}(\theta)}$	
	$\quad = E_u(0^-)/E_d(0^-),$	"Irradiance ratio"
MERIS	ρ_w	"Normalized surface reflectance"
	$\rho_w = \pi L_w(\theta_0,\theta,\phi)/E_d(0^+)$	
	$\quad = \pi[L_w]_N/F_0$	

3.2 Different ways of accumulating and averaging individual pixels into bins

3.2.1 Introduction

Binning algorithms employed by the various ocean-colour missions differ in several ways. For instance, they use different spatial and temporal bin sizes, calculate statistics of binned data in different ways, and use different schemes for weighting the data in each bin. In addition, the criteria used to screen data of lower quality, and even the definition of a "data day", may be different.

This section focuses on one aspect of using different binning algorithms, that is, the effect of different methods of averaging. Three methods of averaging data are described in Table 3.2, the two most common of which are the arithmetic average (AVG) and the geometric mean (GEO). The GEO is obtained by averaging the logarithm of the data, and then inverse-transforming the average logarithm. NASA currently uses AVG to create Level-3 SeaWiFS and MODIS products, whereas NASDA uses GEO for OCTS and GLI products. A third method, called the maximum likelihood estimator (MLE), was also used by NASDA to process CZCS data for distribution on a CD-ROM. The MLE was originally proposed by Campbell *et al.* (1995) for SeaWiFS, but was later rejected for reasons explained in Section 3.2.4. The MLE was intended to estimate the arithmetic mean (AVG) using saved statistics of ln(CHL). These saved statistics could also be used to estimate the mean of any power-law function involving chlorophyll. The MLE is the maximum likelihood estimator of the mean of a lognormal distribution, whereas GEO is an estimate of the median of a lognormal distribution (Campbell *et al.* 1995).

Table 3.2   Three averaging methods evaluated using SeaWiFS data.

Averaging Method	Equations	Used by
Arithmetic mean, "linear average"	$\text{AVG} = \frac{1}{n}\sum_{i=1}^{n} X_i$	SeaWiFS, MODIS
Geometric mean, "logarithmic average"	$\text{GEO} = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\ln X_i\right]$	OCTS, GLI
Maximum likelihood estimate of the mean of a lognormal distribution	$\text{MLE} = \exp\left[m + \frac{s^2}{2}\right]$ $m = \frac{1}{n}\sum_{i=1}^{n}\ln X_i$ $s^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln X_i - m)^2$	CZCS processed by NASDA

In Section 3.2.2, the three different methods are used to calculate monthly averaged chlorophyll concentrations, and the results are compared. These examples use SeaWiFS data, both at regional as well as at global scales. In Section 3.2.3, a similar comparison is made using OCTS data and in Section 3.2.4, a
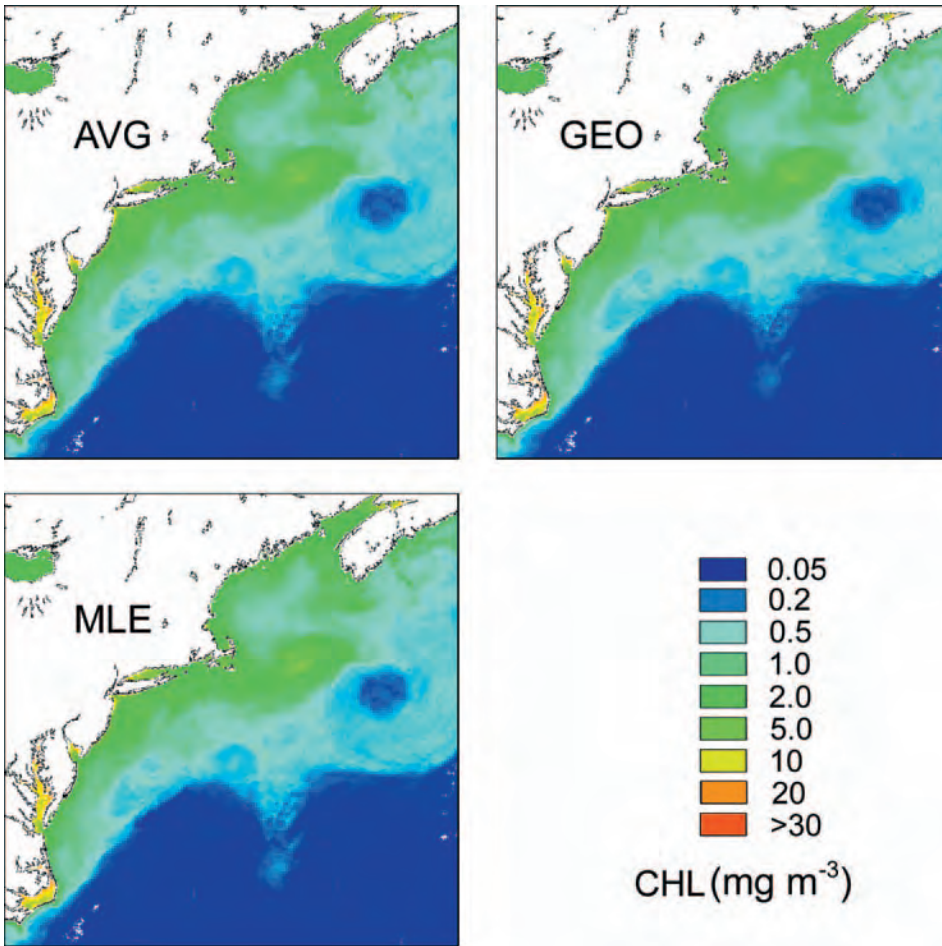
Figure 3.3 Regional monthly mean chlorophyll concentration values (CHL) calculated using the three different averaging methods. Differences are not apparent because of the logarithmic scale of the colour palette.

method is presented to resolve differences in the AVG and GEO averaging methods.

3.2.2   An example using SeaWiFS data

The three different averaging methods listed in Table 3.2 were applied to daily SeaWiFS chlorophyll data at regional and global scales. At the regional scale, high-resolution Level-1 data for passes over the western North Atlantic were processed to Level-2 using SeaDAS.v.4.3. Daily chlorophyll concentration data from 29 passes in July 2000 were remapped to the same conic projection, and then temporally binned to derive maps of the monthly average chlorophyll concentration. The resulting maps retained the original (1.1 km) spatial resolution of the Level-2 data. At the global scale, the analysis was based on SeaWiFS Level-2

GAC data for the month of March 2000. The GAC data have a nominal resolution of 4.4 km, achieved by sub-sampling every 4th pixel on every 4th scan line. The 16 orbital passes of GAC data each day were remapped to produce daily chlorophyll maps at the global scale. The remapped images had a spatial resolution of approximately 18 km at the Equator. These daily global maps were then binned over the month using the three averaging methods listed in Table 3.2. The regional-scale chlorophyll images derived from the three methods (Figure
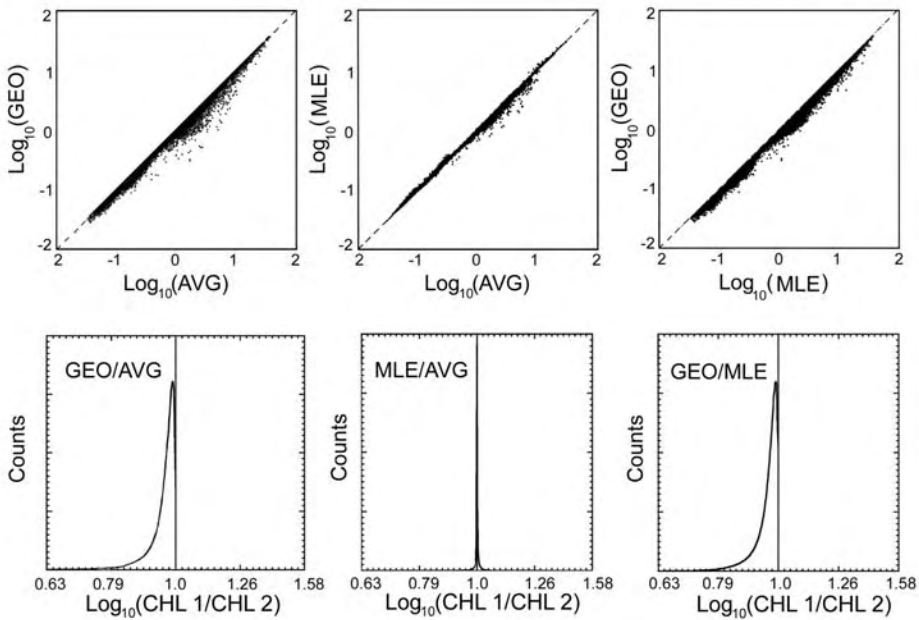
Figure 3.4   Scatter plots and histograms of ratios of CHL derived from the three averaging methods.

3.3) appear to be nearly identical when displayed with a log-scaled colour palette. However, scatter plots and histograms of ratios (Figure 3.4) show systematic differences, in the order of 20%. The geometric mean (GEO) was consistently lower than the arithmetic mean (AVG), whereas the MLE closely approximated the AVG.

Images of the ratios of the monthly mean chlorophyll values derived using the three averaging methods (Figure 3.5) show that the largest differences occur in dynamic frontal regions, such as along the northern boundary of the Gulf Stream, and in the vicinity of warm and cold eddies. What appears to be a deep southward meander of the Gulf Stream in Figure 3.3 is actually an artefact of the averaging. Inspection of the daily images reveals that a meander had formed at the beginning of the month, which later pinched off to form a cold-core ring that subsequently moved southward. The bins affected by this eddy had higher variance than bins in the surrounding waters. It is undoubtedly the variance within the binned data that gives rise to the differences in the averaging

methods.

We examined this explicitly by plotting ratios against the standard deviation within each bin (Figure 3.6). The MLE/AVG ratio lies within a fan of points centred around 1, indicating that the differences tend to be unbiased, and that the relative difference increases as the intra-bin variance increases. The GEO/AVG ratio lies within a fan having more or less the same spreading pattern, but with a downward trend. The GEO/MLE ratio follows the deterministic curve:

$$\text{GEO/MLE} = \exp\left[-\frac{s^2}{2}\right] \qquad (3.6)$$

where s is the standard deviation of ln(CHL), as defined in Table 3.2.
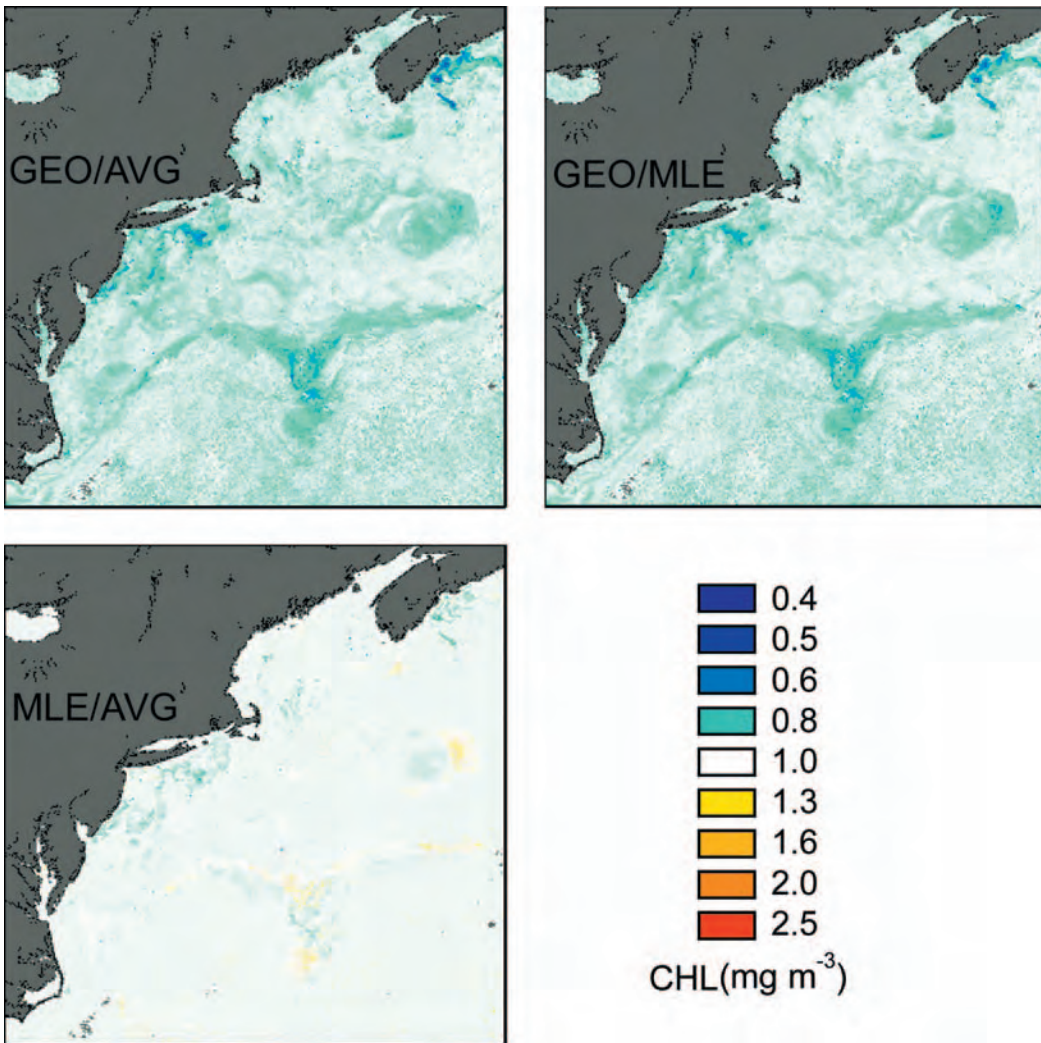

Figure 3.5 Ratios of the monthly mean CHL calculated using the three different averaging methods.
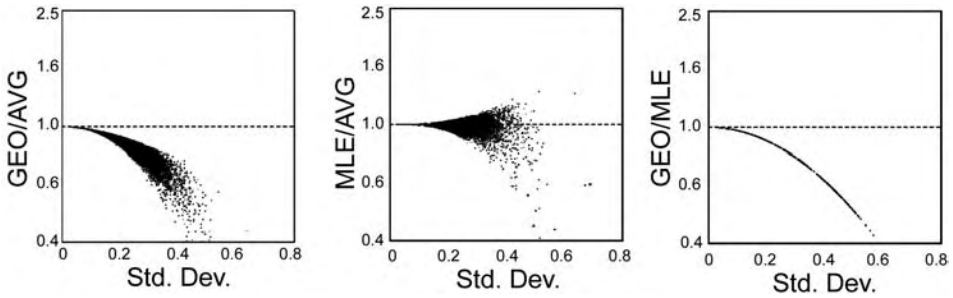
Figure 3.6 Ratios of the monthly mean CHL derived using different averaging methods, plotted against the standard deviation of CHL in the monthly sample.

In most respects, the results obtained for global-scale chlorophyll averages were similar to those obtained at regional scales. Chlorophyll maps (Figure 3.7) were indistinguishable, while maps of the ratios (Figure 3.8) showed similar levels of agreement between methods. Again, the AVG and MLE were most similar, and the GEO was systematically less than the other two. The scatter plots and histograms of ratios were essentially the same as those shown in Figure 3.4. Again, the greatest differences were associated with bins having high variances, as was found at regional scales.

In the global-scale comparisons, large differences between averaging methods appeared in the central gyres. It was somewhat surprising to find such large differences in these regions, which are generally considered to be highly stable environments. In some cases, the differences were larger than those found in dynamic sub-polar or tropical regions. Since the GEO/MLE ratio is strictly a function of the variance of ln(CHL) within the bins, maps of this ratio depict the relative intra-bin variance. Apparently, the variance of ln(CHL) in the central gyres is comparable to, or greater than, that found in regions with higher average CHL.

3.2.3   Another example using OCTS data

Here we compare the arithmetic mean, AVG, (also called the "linear" average) with the geometric mean, GEO, (also called "logarithmic average") using 31 days of OCTS data collected in May 1997. Daily Level-3 Binned Map (L3BM) chlorophyll-a (CHLA) data were used. In each case, the data in a 0.0879-degree grid were binned to a 0.25-degree grid, and averaged using the two different methods (Figure 3.9). The results were then further averaged over time to produce monthly Level-3 products.

Mean chlorophyll values derived using the geometric mean (GEO) log-scheme are smaller than the arithmetic average (AVG) by about 4%, on average (Figure 3.10). The difference is, however, as much as a factor of two in some areas.
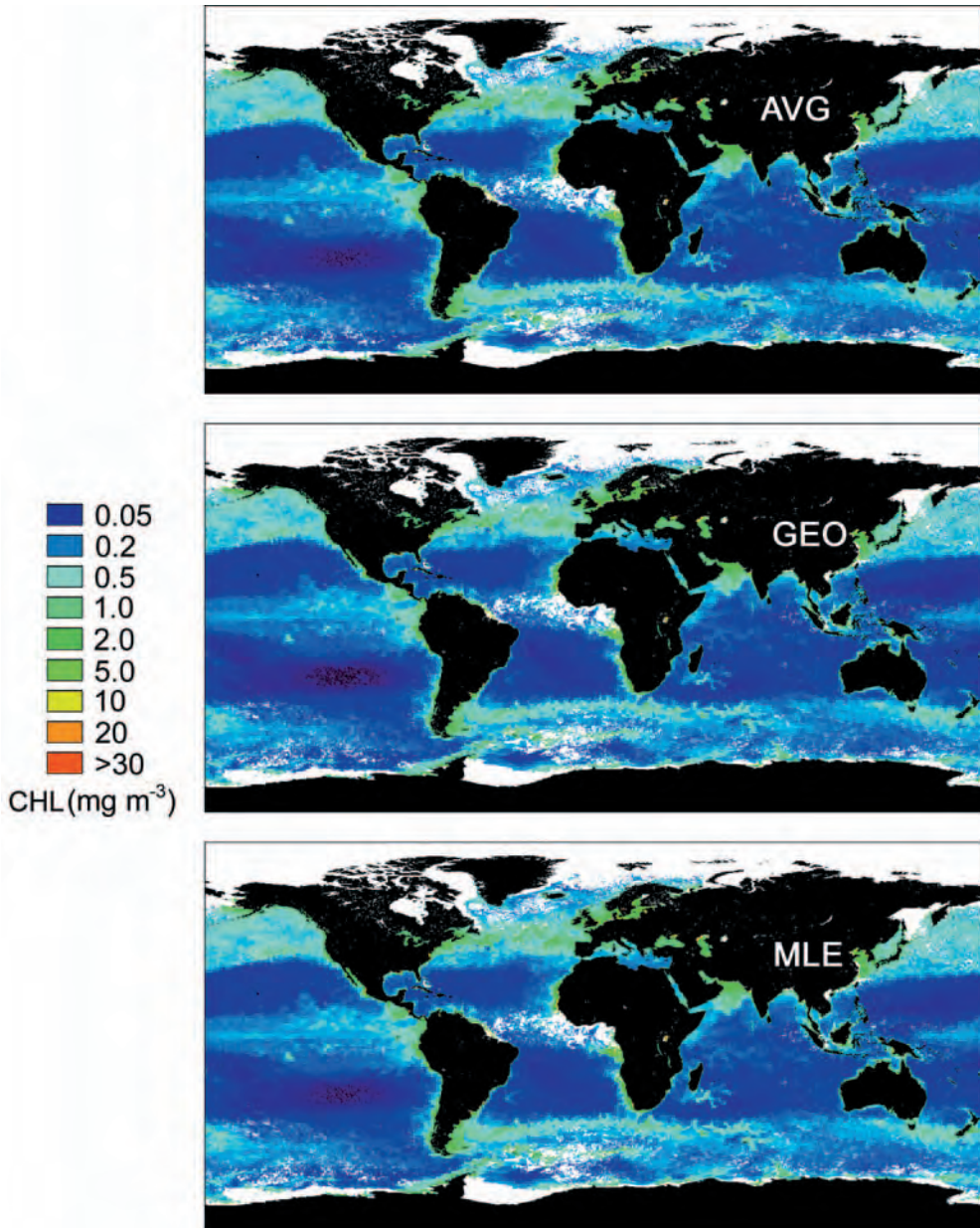
Figure 3.7 Global monthly mean chlorophyll concentration calculated using the three different averaging methods. Differences are not apparent because of the logarithmic scale of the colour palette.
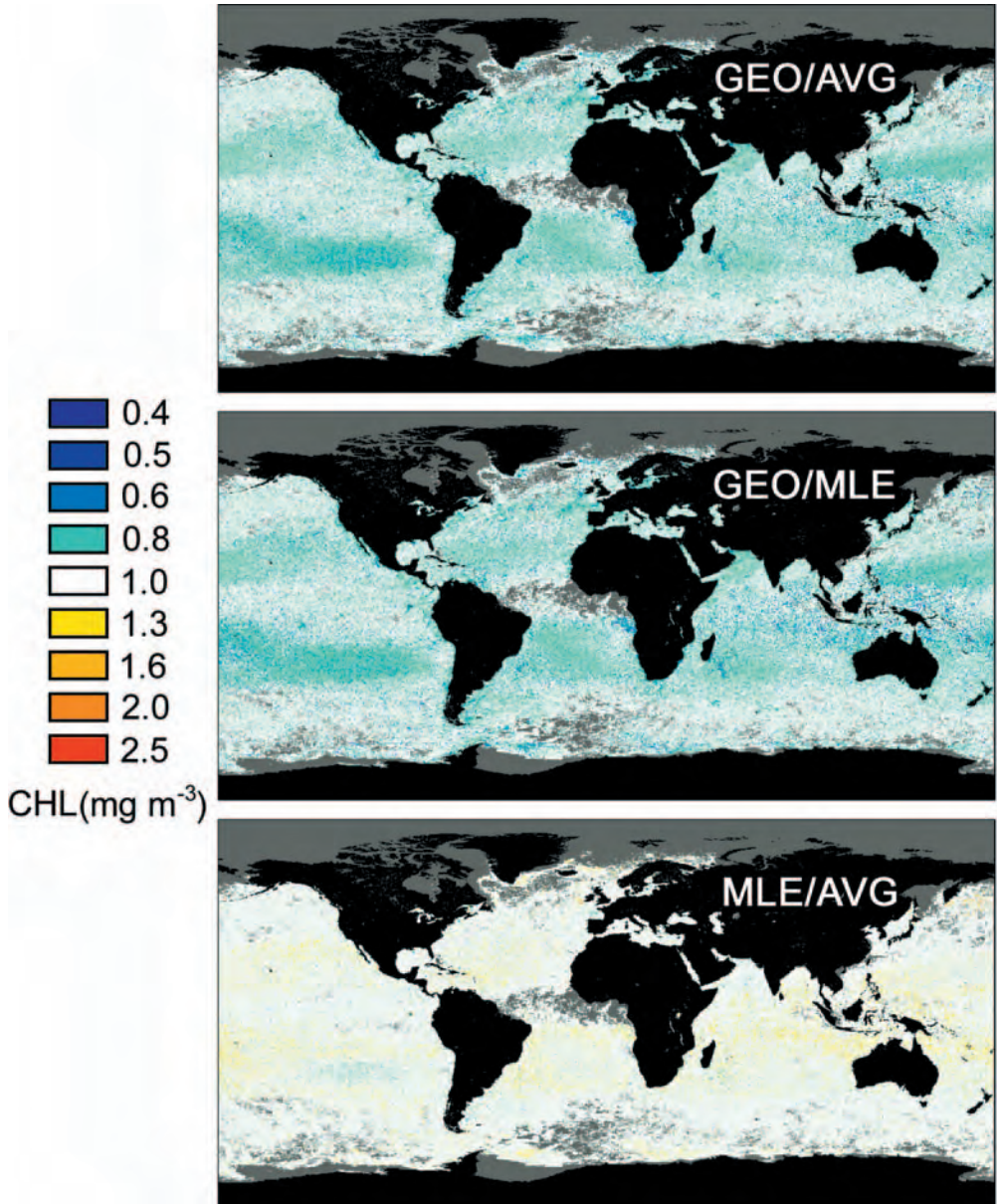
Figure 3.8 Ratios of the global monthly mean CHL calculated using the three different averaging methods.

The largest differences appear in areas of high chlorophyll variance (along the coastal margins and the sub arctic front). Upon closer inspection (Figure 3.11) it is apparent that the linear-scheme is more sensitive to spike noise, which may be caused by the near cloud contamination in the original daily data.
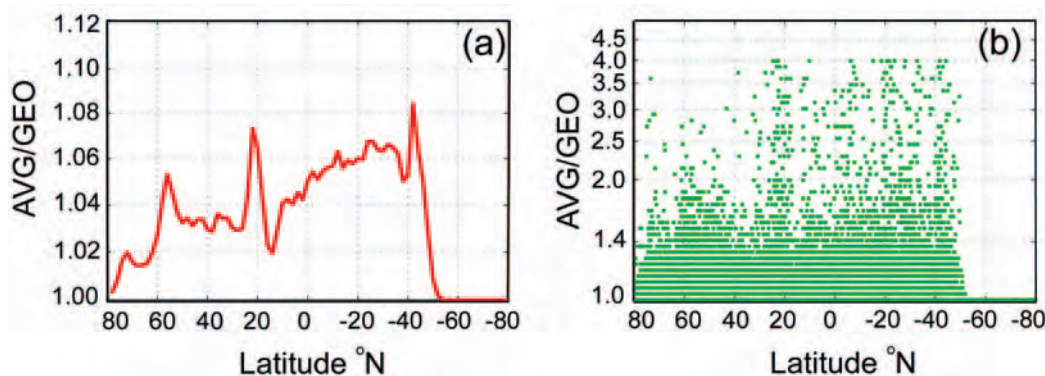

Figure 3.9 AVG/GEO ratios of monthly mean chlorophyll concentration. (a) Zonal average and (b) plots of each 0.25-degree grid.

The GLI ocean group has decided that the log-scheme will be used for at-launch binning codes, and will be changed to the linear-scheme after improvement of cloud screening in the atmospheric correction algorithm.

Another example is provided in Figure 3.12, which shows chlorophyll comparisons obtained using different binning schemes (AVG and GEO) and different algorithms (NASDA v4 versus Goddard algorithms) with the same OCTS data (May 1997), as well as comparisons with SeaWiFS data (May 1999). Differences caused by using different binning schemes are almost the same as differences caused by using different algorithms.

3.2.4 Discussion

The results of this analysis are consistent with those of Campbell *et al.* (1995), who compared the same averaging methods using CZCS data and moored fluorometer data (Medieros and Wirick, 1992). Collectively, these results are consistent with the assumption of an underlying log-normal distribution within the averaging bin. Use of the log-normal as a model for the spatial and temporal variability of chlorophyll has been supported both empirically and theoretically (Campbell, 1995). As long as the data being binned are approximately log-normally distributed, or the variance is relatively low, this assumption works well. Thus, in the SeaWiFS pre-launch technical memorandum describing Level-3 products, Campbell *et al.* (1995) recommended that sums and sum of squares of log-transformed data be saved. Given these saved statistics, they prescribed how the mean and variance of log-transformed data could be used to estimate
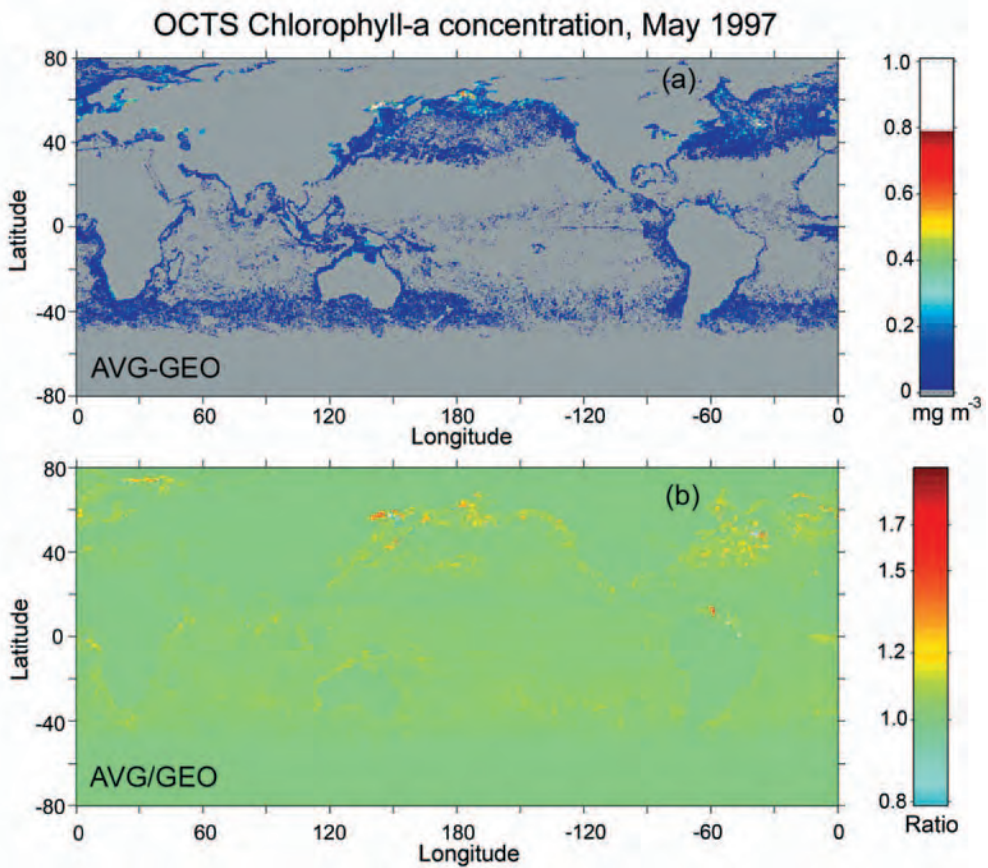
Figure 3.10 Distribution of the (a) difference and (b) ratio between the arithmetic average (AVG) and the geometric average (GEO) of monthly mean chlorophyll-a concentration in May 1997, measured by the OCTS sensor.

the mean, median, mode and variance of the variable itself, within the space-time bin.

Subsequent tests on simulated SeaWiFS data revealed that the MLE is sensitive to outliers. A thorough analysis of this situation (Campbell, unpublished) confirmed that the MLE is problematical when the data being binned are bimodal, specifically when predominantly high-chlorophyll values are averaged with a few low chlorophyll values. Although it was believed that such situations would be rare in the natural environment, it is not unusual to have this occur in satellite images contaminated by undetected clouds (low chlorophyll pixels often fringe the cloud mask). Thus, after the technical memorandum describing SeaWiFS Level-3 binning algorithms was published, the SeaWiFS Project decided to save the sums and sum of squares of the data (not the log-transformed data), and use the simple arithmetic average to create mapped products.

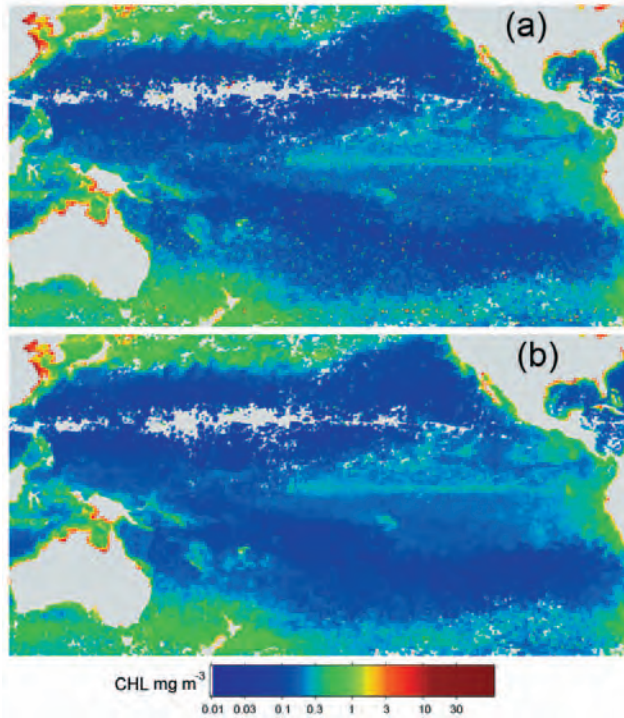Based on the recommendations of Campbell *et al.* (1995), NASDA used the

Figure 3.11 Monthly chlorophyll-a concentration averaged using (a) the linear-scheme (AVG) and, (b) the log-scheme (GEO). Data collected by the OCTS sensor in May 1997.

MLE to generate CZCS data products that were distributed on a CD-ROM. They had planned to use the MLE to generate Level-3 OCTS data, but later concluded that both the MLE and AVG were susceptible to errors due to spikey data, the result being anomalously high chlorophyll values. NASDA currently uses the GEO to create Level-3 products. As we have seen, the GEO (log-scene) averages are consistently lower than the averages produced by the other two methods. In the final analysis, the decision as to which average is best will depend on how the data are used. If the intra-bin variance has been saved, then the log-normal approximation can be used to convert AVG to GEO or visa versa. If the mean and variance of ln(CHL) have been saved, then

$$\text{AVG} \approx \text{GEO} \exp\left[\frac{s^2}{2}\right] \tag{3.7}$$

where s^2 is the variance of ln(CHL). If the mean and variance of CHL have been saved, then

$$\text{GEO} \approx \frac{\text{AVG}}{\sqrt{\text{CV}^2 + 1}} \tag{3.8}$$

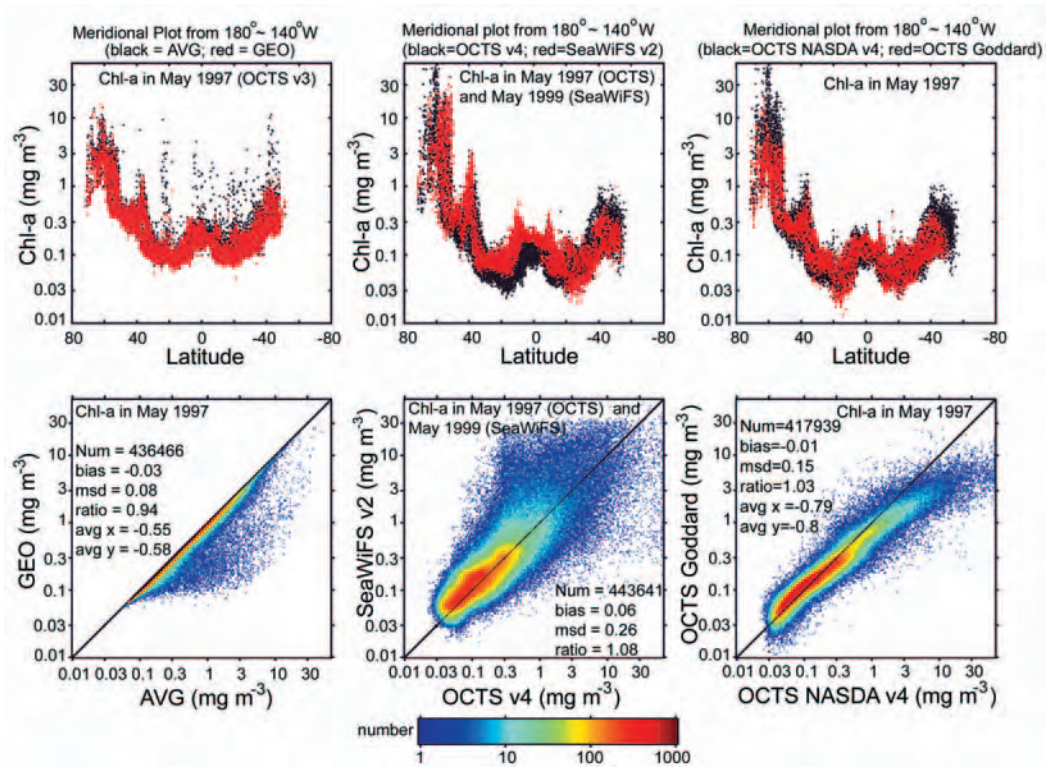where CV is the ratio of the standard deviation to the mean of CHL. Both of these

Figure 3.12 Comparisons of various OCTS-derived Level-3 data sets

equations are based on the assumption that the AVG is approximately equal to the MLE, an assumption that is supported by the results presented in Section 3.2.2.

Issues linked to the diversity of spatio-temporal grids

Robert Evans and David Antoine

4.1 Example 1: Definition of a "data-day"

Satellites orbit the earth continuously, irrespective of what happens on the ground in terms of sun rise and sun set, and in terms of abrupt changes in dates. Accordingly, when generating a daily field of a satellite-derived quantity, issues arise regarding whether or not a given pixel, in a given scan line, of a given orbit, is to be assigned to the daily field of day N or day $N + 1$. The simplest and most unambiguous solution is to start collection of data for day N at 00:00:00 UTC and end it at 23:59:59 UTC. Large gaps or discontinuities in the data would, however, occur with this solution, so other approaches have been adopted for ocean-colour missions such as SeaWiFS, based on previous experiences with NOAA's AVHRR mission. The following paragraph is extracted from a manuscript by Kilpatrick *et al.* (2001) entitled "Overview of the NOAA/NASA AVHRR pathfinder algorithm for sea surface temperature and associated matchup database". The authors summarize the issue surrounding a "data-day" for the Pathfinder project. Podestá (1995) has also published a detailed description of a "data-day".

"The basic products generated by the AVHRR Pathfinder Oceans project are global daily fields of sea surface temperature. To construct these fields a consistent definition of a data-day was adopted as described in considerable detail by Podestá (1995). The definition of the data-day encompasses both a spatial and temporal component. If either a temporal (*e.g.* a 24 h period) or spatial definition (*e.g.* full global coverage) were to be used alone, the resulting daily product would have large gaps or discontinuities at the time or space boundary due to satellite orbital mechanics. For the Pathfinder project, the spatial component defines the start and end time of a given data-day and is defined as the time at which the satellite orbit track crosses the 180° meridian nearest to the Equator. The start and end of the data-day therefore changes as a function of the satellite orbit. This strict spatial definition results in large temporal discontinuities at

the meridian boundary as the orbit precesses. To eliminate these discontinuities, the temporal component includes any data taken within 2 hours before or 2 hours after the meridian crossing. A data-day therefore may represent data taken over a 24 to 28 hour period."

4.2 Example 2: Projecting the same data set onto grids of different spatial resolutions

A common operation performed on Level-3 fields is their simple comparison on a pixel-by-pixel basis. For example, this is often performed when a new ocean-colour mission has generated its initial Level-3 fields. These are then cross-checked against similar fields produced by other validated missions. This kind of operation requires that identical grids are used to project the individual pixel information into each of the bins of the Level-3 fields, which, in practise, is rarely the case. Assessing the impact of using different Level-3 fields on the results of various oceanographic studies also necessitates homogeneity between the spatio-temporal scales of these fields (see Bricaud *et al.*, 2002). A simple exercise was performed to illustrate the possible consequences of heterogeneity in the Level-3 fields. Firstly, a large set of SeaWiFS, Level-2 GAC images (collected over the Mediterranean Sea, in 1999) was pooled. Next, three different sets of monthly fields were generated from the data, each of which was based on the use of an equal-area projection. The three projections were the ones used by the SeaWiFS mission (resolution 1/12 of a degree, see Section 2.1.3,) and by the POLDER mission (resolution 1/18 and 1/9 of a degree, see Section 2.1.2). The arithmetic average of all SeaWiFS pixels falling into a given bin of these equal-area grids was computed for periods of one month.

The next step consisted of re-projecting the resulting monthly fields onto a unique rectangular grid of 2048 lines, each with 4096 pixels, with a constant increment in latitude and longitude. This grid is similar to the "standard mapped image" (SMI) format used by the SeaWiFS mission, except that the resolution is higher. Thus, from an initial set of Level-2 pixels, we produced three different Level-3 maps; differences between these maps are only due to the three different pathways followed from the initial data set.

In the end, a pixel-by-pixel comparison of the different maps was performed and histograms of the ratio were produced. The overall process is shown in Figure 4.1, and results are displayed in Figure 4.2.

The histogram in Figure 4.2 shows that the successive projections described above, when applied to a given set of individual pixels, produce noise in the final product, typically of the order of ±15% (*i.e.*, the standard deviation of the distribution shown in Figure 4.2). Errors larger than ±50% also occur, yet they only represent about 10% of the total number of comparison points. In addition,
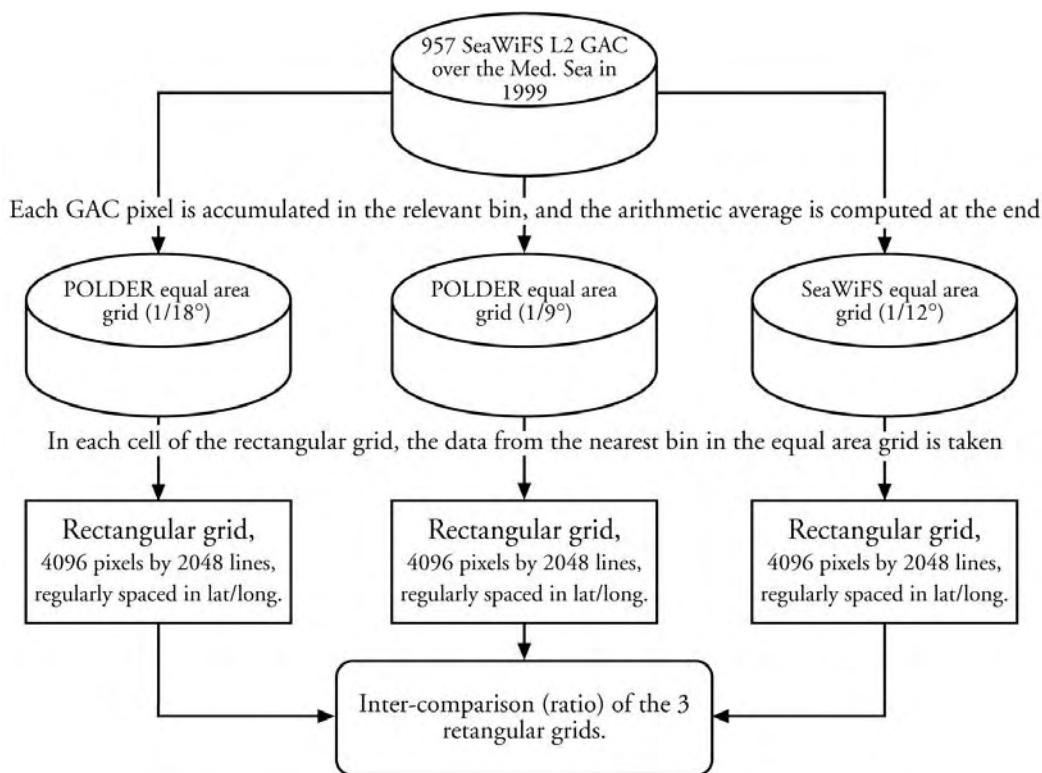
Figure 4.1 Overall process for generating three different monthly, Level-3, chlorophyll fields based on three equal-area projections of differing spatial resolution, and on reorganization of these fields into a unique rectangular grid (see text).

these successive operations do not introduce any bias. Some of this variance may, however, have been introduced when the three different resolution equal area bin files ($1/18°$, $1/12°$ and $1/9°$) are reprojected onto the equal angle 4096 x 2048 SMI (Standard Mapped Image) maps. Different rows from the each of the equal area bin files are dropped during reprojection in order to make them fit into the power of two grid of the SMI map resolutions. It is possible that a "true" equal angle map projection, rather than the Standard Mapped Images currently in use, might provide less variance.

It should also be noted, that the results presented here were obtained for the Mediterranean Sea, which is characterized by relatively weak chlorophyll gradients, except in very localized areas such as the Alboran Sea. Repeating the same exercise for oceanic areas, characterized by a higher variance and sharper fronts, will no doubt result in the introduction of a greater noise when comparing Level-3 fields (the generation of which went through grids of different spatial resolutions). The same comments probably hold true for high latitude areas, where the distortion of grids makes successive use of different grids even more
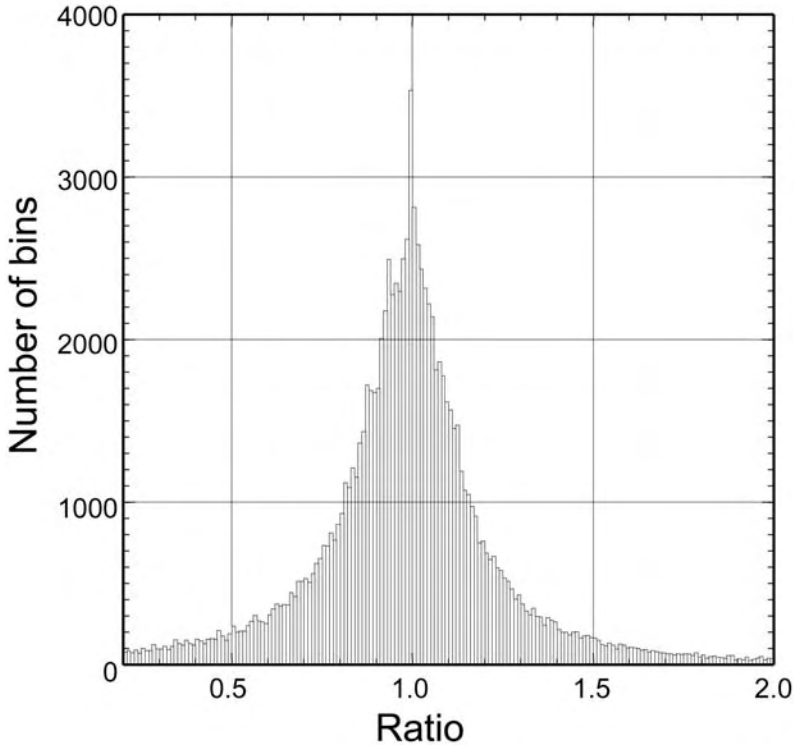
Figure 4.2   Histogram of the bin-by-bin ratio of chlorophyll concentrations, as initially binned following POLDER and SeaWiFS equal-area projections, with spatial resolutions of 1/18 and 1/12 of a degree, respectively.

problematic.

Another illustration is given below, dealing with the temporal resolution of the Level-3 fields (this subject overlaps somewhat with that in Section 5). This example is taken from a case study in the Mediterranean Sea carried out by Ras (1999), assessing the effect of different temporal resolutions in the Level-3 chlorophyll fields, when estimating primary production using the spectral light-photosynthesis model of Morel (1991). Figure 4.3 shows the difference between primary production (integrated over various regions of the Mediterranean Sea) based on 8-day chlorophyll composites, and that based on monthly chlorophyll composites (both using SeaWiFS data). In this example, the overall effect of using the 8-day fields is to produce a larger value in the end, yet the difference is small compared to the annual primary production estimate. The converse effect is produced for areas where the spatio-temporal changes are *a priori* larger and faster (Adriatic Sea, Alboran Sea, Gulf of Lions). It was expected that the effect would be small at the scale of the whole Mediterranean Sea, and possibly larger and more significant at the scale of regional seas.

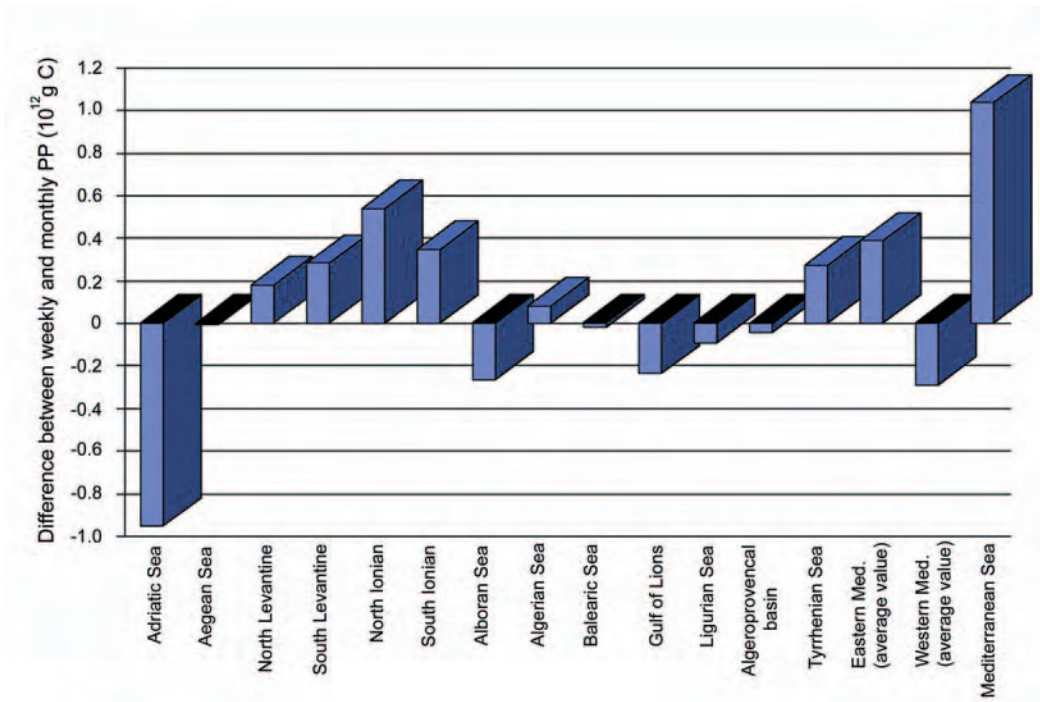A single example cannot be used to predict whether it is better to use 8-day

Figure 4.3   Absolute values of the difference between 8-day and monthly derived annual primary production values (10^{12} g C) for various Mediterranean basins. Reproduced from Ras (1999).

or monthly fields (not the goal here). It simply illustrates another aspect of the impact of using differing binning schemes in biogeochemical studies.

<div style="border:1px solid #000; display:inline-block; padding:10px;">

Chapter 5

</div>

Issues linked to the use of binned data in models: An example using primary production

Janet Campbell

The degree of commonality/discrepancy between satellite ocean-colour data sets and other satellite-derived data sets, in terms of space/time scales, was explored in Chapter 2. In this chapter, we provide examples of the impact of this discrepancy when the data are used in models, and include an example of modelling basin-scale primary production. Modelling of primary production is one of the most common, synergistic uses of ocean-colour data with other satellite data. The assimilation of ocean-colour data into ocean general circulation models (GCM's) could also have been treated here, but it was considered more relevant for the next chapter, about future directions.

5.1 Introduction

We consider here the effect of using binned data to derive primary production (PP). It is common practice to apply primary productivity algorithms to monthly-averaged input fields (Behrenfeld and Falkowski, 1997; Behrenfeld *et al.* 2001). In the processing of MODIS data, weekly averages are used as input to primary productivity algorithms. An alternative to using binned data is to apply the algorithms to Level-2 data, and then average the resulting daily primary production estimates. In this section, we compare these various approaches using two primary productivity algorithms.

One algorithm is the vertically generalized production model (VGPM) of Behrenfeld and Falkowski (1997). According to this algorithm, daily euphotic zone primary production (mg C m^{-2}) is given by:

$$PP = 0.66125 \, D_{irr} \, P^B_{opt} \, \frac{E_0}{(E_0 + 4.1)} \, C_{sat} \, Z_{eu} \qquad (5.1)$$

where D_{irr} is the day length in hours, P^B_{opt} is the maximum rate of carbon fixation in the water column in units of mg C (mg Chl)$^{-1}$ h^{-1}, E_0 is the photosynthetically available radiation (PAR) in mol quanta m^{-2} d^{-1}, C_{sat} is the surface chlorophyll

concentration in mg Chl m^{-3}, and Z_{eu} is the depth of the euphotic zone, or 1% light level, in meters, which is a function of C_{sat} (Morel and Berthon, 1989). According to this algorithm, P^B_{opt} is a 7th order polynomial function of SST.

The other algorithm is a modified version of the Howard-Yoder-Ryan algorithm used by MODIS to estimate mixed-layer primary production (Howard and Yoder, 1997; Campbell *et al.* 2002). We modified this algorithm to give euphotic primary production:

$$PP = 24\ P^B_{max}\ \frac{E_{bar}}{(E_s + E_{bar})}\ C_{sat}\ Z_{eu} \tag{5.2}$$

where P^B_{max} is the maximum photosynthetic rate in units of mg C (mg Chl)$^{-1}$ h^{-1}, E_{bar} is the average PAR in the euphotic zone, in W m^{-2}, and $E_s = P^B_{max}/\alpha$, and is the half-saturation irradiance in W m^{-2}. In this algorithm, the depth of the euphotic zone, Z_{eu}, is $4.6/K_{par}$, where K_{par} is a nonlinear function of C_{sat} (Nelson and Smith, 1991). This algorithm assumes an "average" P-E curve in which $P^B_{max} = \exp(0.09\ SST)$ is from Eppley (1972), and $\alpha = 0.11$ is assumed to be constant. The average euphotic PAR is $E_{bar} = 0.99\ E_0/4.6$, where E_0 is the surface PAR in W m^{-2}, averaged over a 24-hour period. Note that this formulation is based on the average PAR over 24 hours, and thus P^B_{max} is multiplied by 24 hours, whereas the VGPM algorithm scales P^B_{opt} by the day length, D_{irr}.
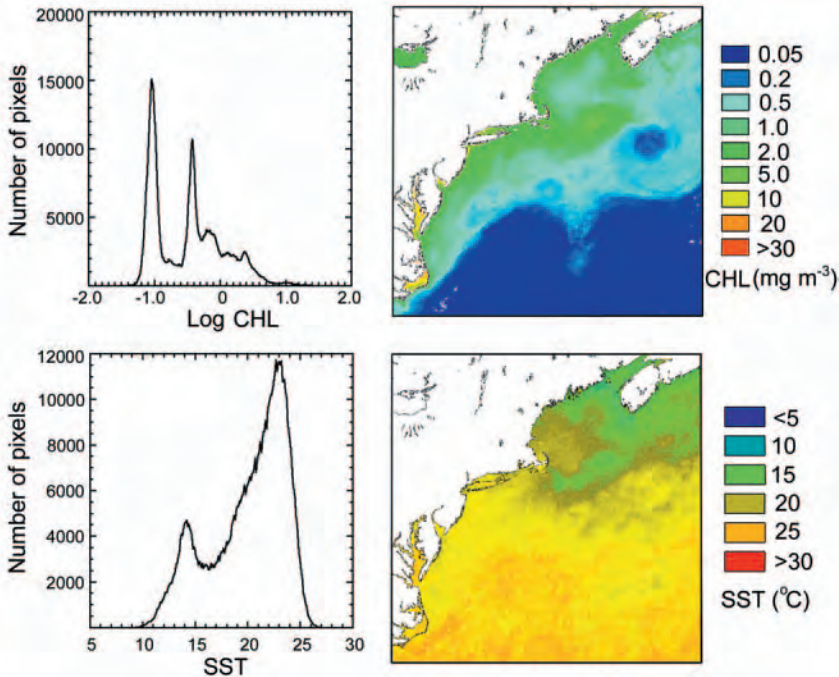

Figure 5.1 Monthly average CHL (upper) and SST (lower) used as input to derive PP$_m$. CHL was derived from SeaWiFS data for July 2000, and SST from AVHRR data for the same month.
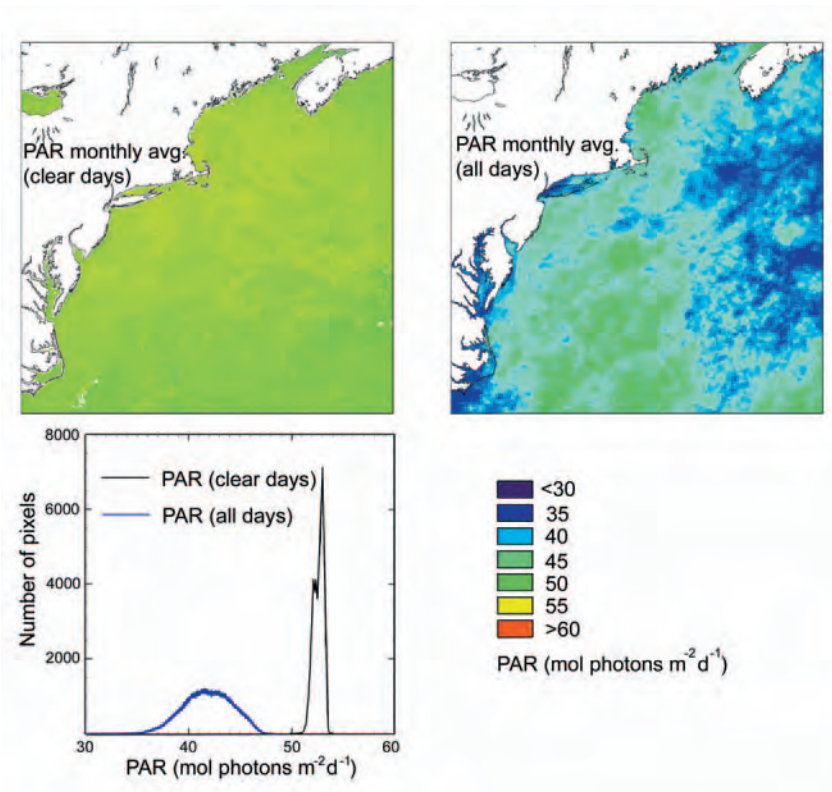
Figure 5.2 Monthly average PAR (units mol quanta m^{-2} d^{-1}) used as input to derive PP_m^* (left) and PP_m (right). The image on the left shows the average of all clear-sky PAR values used in deriving PP_d.

Both algorithms were applied to Level-2 data and the results averaged to estimate the monthly average daily production, which we denote PP_d. Next, the algorithms were applied to monthly average input fields, and the monthly average daily production was derived, which we denote PP_m. All averages were based on the arithmetic average (AVG). For C_{sat} we used the SeaWiFS chlorophyll data for July 2000 (as described in Section 3.2.2). PAR, in mol quanta m^{-2} d^{-1}, was derived from the Level-1 SeaWiFS data by the algorithm of Frouin *et al.* (2001), and converted to average daily PAR in W m^{-2} for the Howard-Yoder-Ryan (HYR) algorithm. Coincident SST data for the same dates were obtained from the AVHRR Pathfinder archive at the University of Rhode Island (J. Yoder and S. Schollaert, pers. comm.). Data from the overpass nearest in time to the SeaWiFS overpass were selected. All three input fields were mapped to the same projection, and the cloud masks were merged to form a cloud mask for the daily PP calculations.

The PAR algorithm does not employ a cloud mask since there is sunlight beneath clouds, and thus there are PAR values at all pixels in the scene. Their monthly average is more representative of the true radiant energy available for
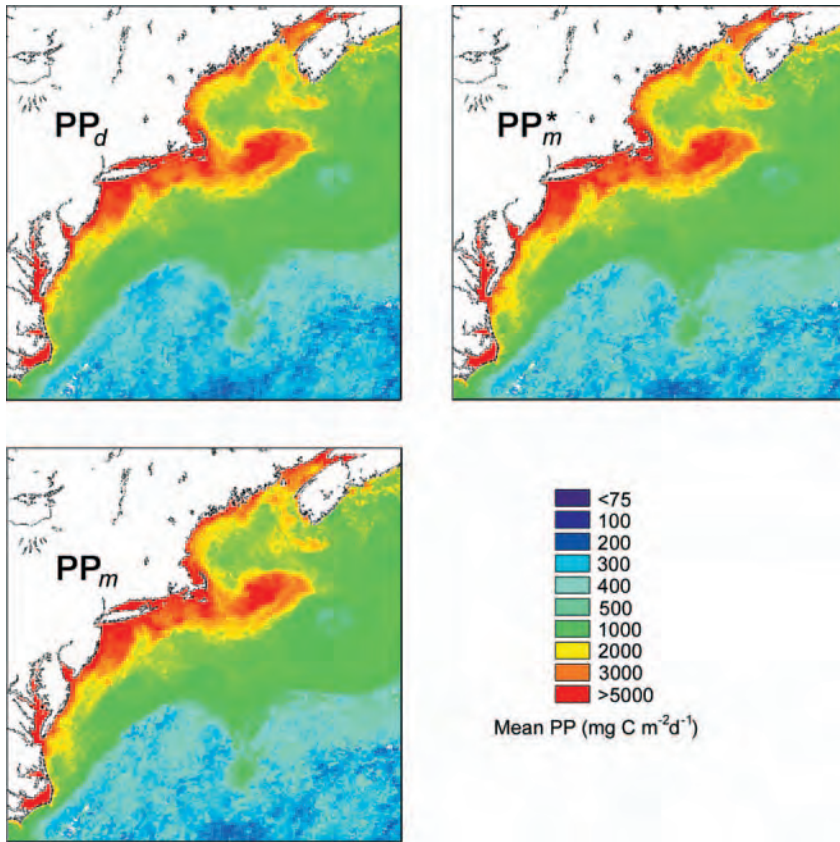
Figure 5.3   Results of deriving monthly mean primary productivity (mg C
m^{-2} d^{-1}) using the VGPM algorithm (Behrenfeld and Falkowski, 1997) by
three methods. For PP_d the algorithm is applied to daily data and then
averaged over the month. For PP_m, the algorithm is applied to monthly
averaged input fields of CHL, PAR, and SST. PP_m^* is the same as PP_m, except
that the average PAR is based on clear-sky pixels only.

photosynthesis than is the average of the clear-sky PAR values used in calculat-
ing PP_d. In other words, since daily PP calculations can only be made at cloud-
free locations, PP_d is biased to clear-sky conditions. To evaluate the effect of this
bias, PP was also derived using the monthly average PAR for cloud-free pixels
only. This result is denoted PP_m^*.

The monthly average input fields of C_{sat} and SST input are shown in Figure
5.1 along with histograms of these fields, and two PAR fields are shown in Figure
5.2. The average PAR used to calculate PP_m^* (left, based on cloud-free pixels only)
was substantially higher than the average PAR for all days (right) used in PP_m.
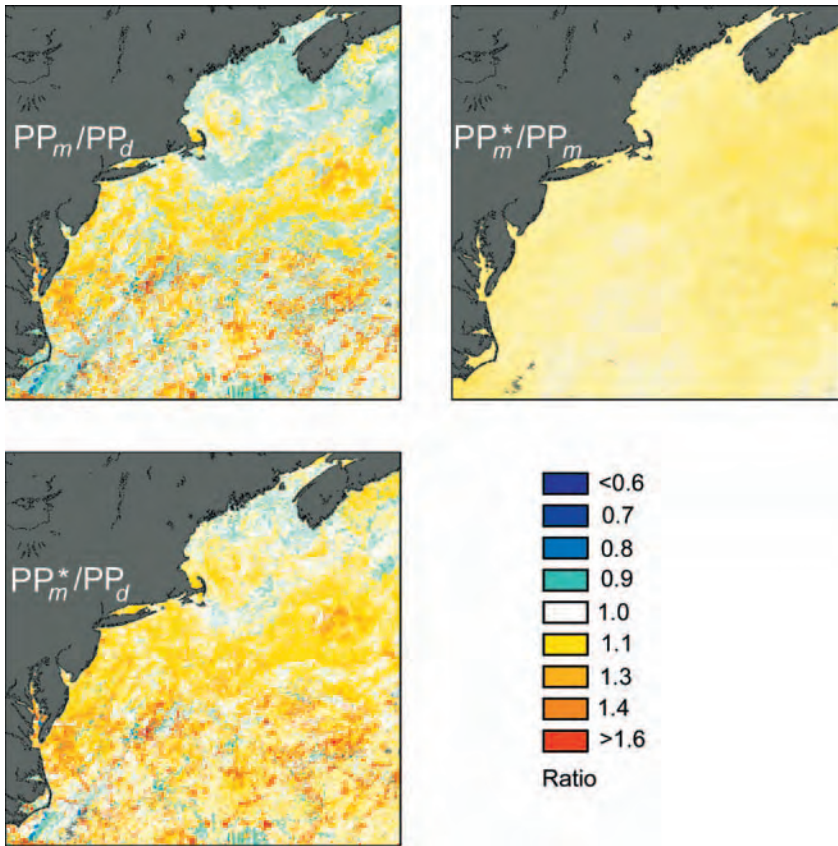
Figure 5.4 Ratios of the fields shown in Fig. 5.3.

5.2 Results

Results for the VGPM algorithm are shown in Figures 5.3 and 5.4. In Figure 5.3, the differences between PP_d and PP_m are most apparent in the open-ocean waters south of the shelf where $PP_d < PP_m$, whereas there is no apparent difference between PP_m and PP_m^* in this figure. The ratios shown in Figure 5.4 reveal that $PP_d < PP_m$ over most of the scene, although there are exceptions in the Gulf of Maine, Georges Bank, and off Cape Hatteras. The clear sky bias is negligible ($PP_m \approx PP_m^*$), and certainly does not explain the trends seen here. We discuss the importance of PAR later in the context of both algorithms. When the primary production is summed over the whole region for the month, the differences in methods are small: $\Sigma PP_d = 0.0264$ giga tonnes (GT); $\Sigma PP_m = 0.0272$ GT; and $\Sigma PP_m^* = 0.0278$ GT. The small differences in the aerial production are deceptive, however, because there are offsetting trends that mask more significant differences in a particular area.

Similar results for the HYR algorithm are shown in Figures 5.5 and 5.6. The
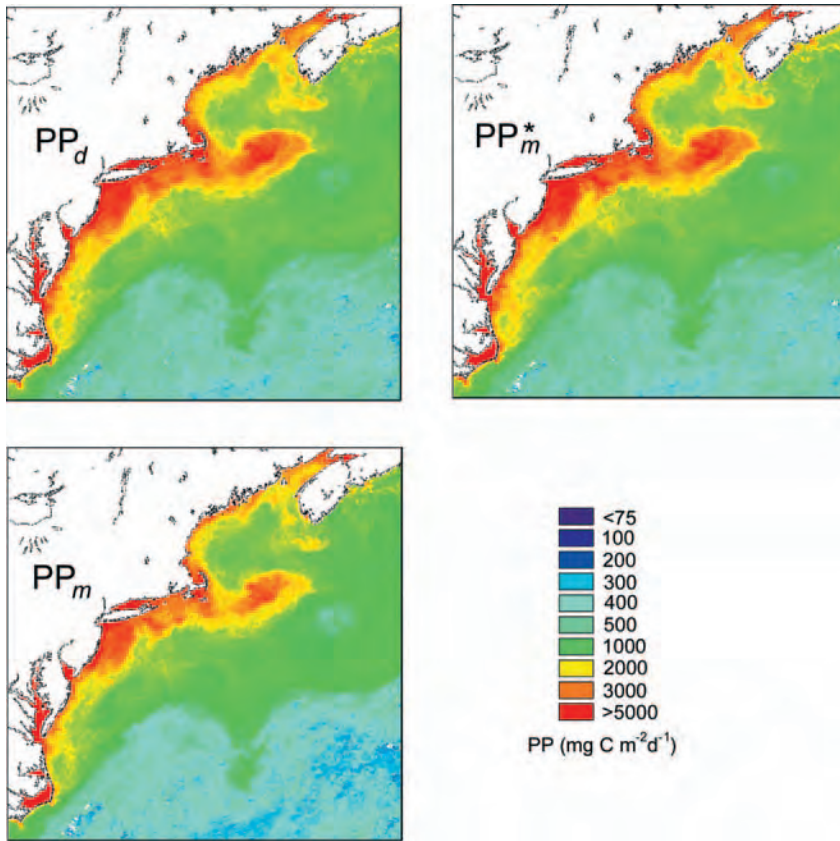
Figure 5.5   Results of deriving monthly mean primary productivity using the HYR algorithm (Howard and Yoder, 1997; Campbell *et al.* 2002) by the same three methods described in Figure 5.3.

colour scale used in Figure 5.5 tends to reduce the apparent difference between PP_d and PP_m. Ratios of PP shown in Figure 5.6, however, reveal that PP_m <PP_d <PP_m^* over the entire scene. These results indicate that the clear-sky bias is the main reason for the difference between PP_d and PP_m. Despite their systematic differences, there were relatively small differences in the total monthly production for the whole region: $\Sigma PP_d = 0.0285$ GT; $\Sigma PP_m = 0.0245$ GT; and $\Sigma PP_m^* = 0.0291$ GT. We examine the sensitivity of each algorithm to the input variables (SST, PAR, and CHL) in the following section.

5.3   Separating the effects of SST, PAR and CHL

In this section, we examine the three input fields to determine their relative contribution to the differences in PP_m and PP_d, but first we introduce new notation. Let <X> denote the monthly mean of X obtained by calculating X on each day, and then averaging over all days in a month. As described earlier,
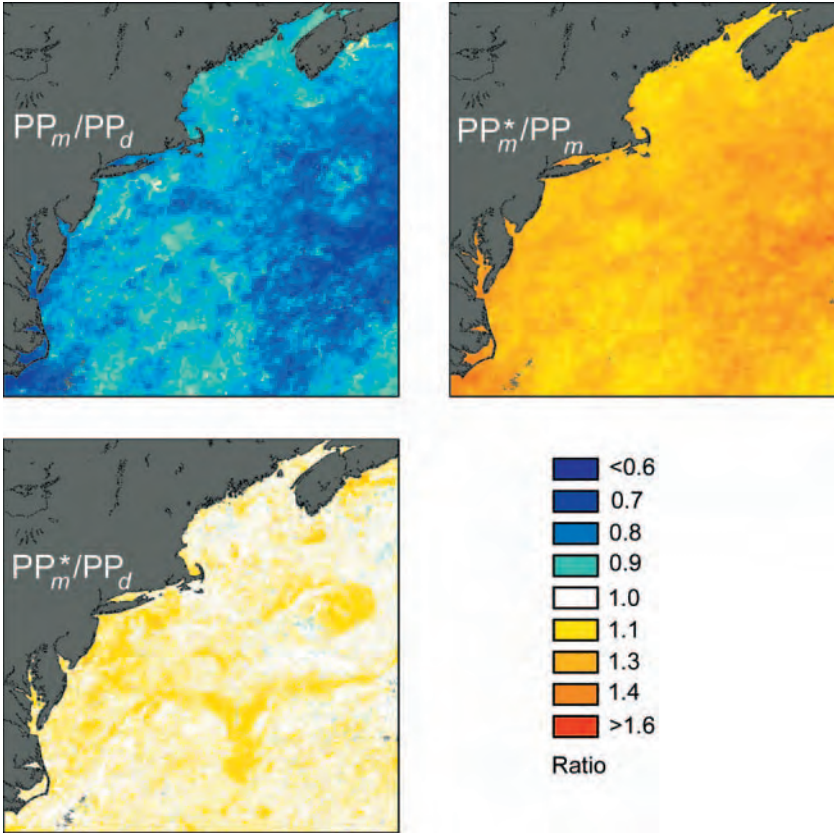
Figure 5.6   Ratios of the fields shown in Figure 5.5.

we found significant differences between <PP> = PP_d and PP_m which was estimated by substituting <SST>, <PAR>, and <CHL> into the primary productivity algorithms.

The VGPM model can be expressed in terms of three factors, each a function of a different satellite-derived input field:

$$PP = 0.66125 \, D_{irr} \, f_1(SST) \, f_2(PAR) \, f_3(CHL), \tag{5.3}$$

where $f_1(SST) = P_{opt}^B$ which is a seventh-order polynomial function of SST, $f_2(PAR) = E_0/(E_0 + 4.1)$, and $f_3(CHL) = C_{sat}Z_{eu}$. To the extent that variations in D_{irr}, SST, PAR, and CHL are statistically independent within a given month, the mean of PP is given by the product of the means of the various functions:

$$<PP> = 0.66125<D_{irr}> \, <f_1(SST)> \, <f_2(PAR)> \, <f_3(CHL)> \tag{5.4}$$

Thus, for the VGPM algorithm, the ratio between PP_m and PP_d is

$$\frac{PP_m}{PP_d} = \frac{f_1(<SST>)}{<f_1(SST)>} \frac{f_2(<PAR>)}{<f_2(PAR)>} \frac{f_3(<CHL>)}{<f_3(CHL)>} = \frac{f_{1m}}{f_{1d}} \frac{f_{2m}}{f_{2d}} \frac{f_{3m}}{f_{3d}} \tag{5.5}$$

Figure 5.7 Ratios of the three functions as defined in Equations 5.3 to 5.5 for the VGPM algorithm. f_1 is the SST-dependent factor; f_2 is the PAR-dependent factor, and f_3 is the CHL-dependent factor.

Plots of these three ratios (Figure 5.7) reveal that the major reason for the difference between PP_m and PP_d in the VGPM algorithm is the use of monthly average SST values in the computation of P^B_{opt}. Use of monthly means in a non-linear function will result in an over-estimation of the mean of the function if the second derivative is negative (*i.e.*, the function is concave downward), and an under-estimation of the mean if the second derivative is positive (*i.e.*, the function is concave upward) (Trela *et al.* 1995). In the case of the seventh-order polynomial defining P^B_{opt} as a function of SST, the second derivative changes sign several times over the range of SST values. The curvature is positive over the range of cooler SST values found in the Gulf of Maine and Georges Bank in July 2000, thus using <SST> underestimates <P^B_{opt}>. In the warmer waters of the shelf and slope waters, the function has a negative second derivative, resulting in an over-estimation of <P^B_{opt}>.

The f_2(PAR) ratio is < 1 throughout the region. This is a result of the clear-sky bias associated with <f_2(PAR)>. When only clear-sky values of PAR were used to estimate <PAR>, the ratio (not shown) is positive, since f_2(PAR) is con-

cave downward. In the VGPM algorithm, f_3(CHL) is a power-law function of CHL with an exponent between 0.6 and 0.9, and thus its second derivative is negative. As a result, the f_3(<CHL>) exceeds <f_3(CHL)> nearly everywhere.
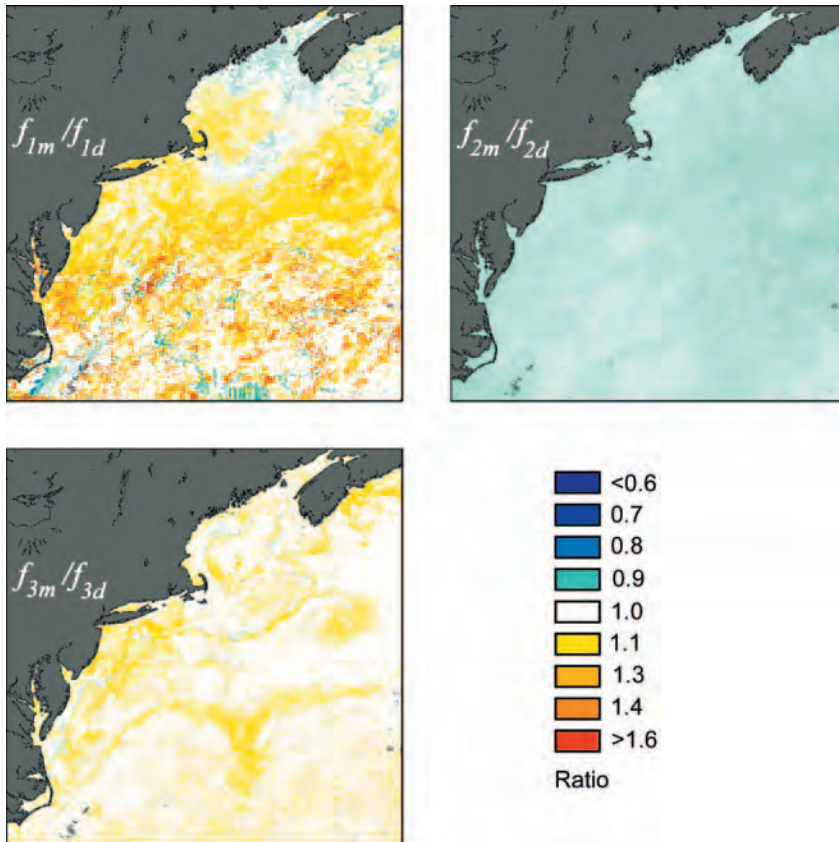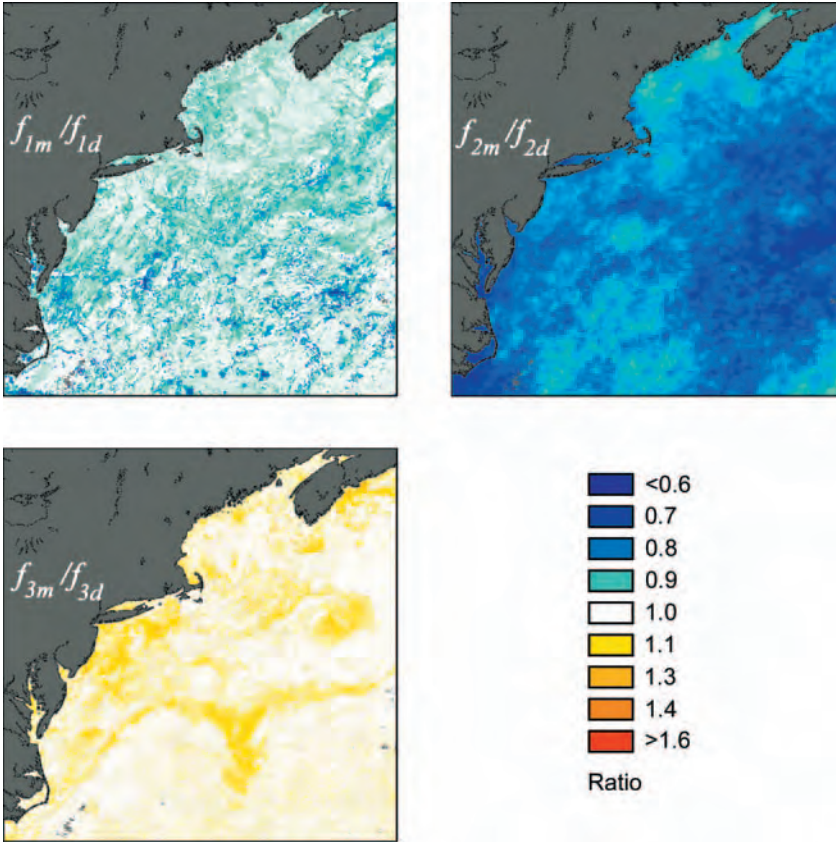

Figure 5.8 Ratios of the three functions as defined in Equation 5.6 for the HYR algorithm. f_1 is the SST-dependent factor; f_3 is the CHL-dependent factor, and f_2 is dependent on both PAR and SST.

In the case of the HYR algorithm, the effects of the three input fields cannot be separated as readily. However, it is possible to separate the algorithm into three factors that are comparable with those of the VGPM algorithm. Although the assumption of independence is not met, we approximate:

$$<PP> \approx 24 <f_1(SST)> <f_2(PAR, SST)> <f_3(CHL)> \qquad (5.6)$$

where $f_1(SST) = P_{max}^B = \exp(0.09\,SST)$, $f_2(PAR, SST) = E_{bar}/(E_s + E_{bar})$, and $f_3(CHL) = C_{sat}Z_{eu}$. Plots of the three associated ratios for the HYR algorithm are shown in Figure 5.8. The $f_1(SST)$ ratio is < 1 everywhere, consistent with the fact that its second derivative is positive. The $f_3(CHL)$ ratio is very similar to that of the VGPM algorithm, although the formula for Z_{eu} was somewhat different. The primary reason that $PP_d >PP_m$ for this algorithm is the clear-sky

bias which only affects $f_2(\text{PAR}, \text{SST})$. The ratio associated with $f_2(\text{PAR}, \text{SST})$ is < 1 throughout the image, and its pattern is similar to that of $\text{PP}_m / \text{PP}_d$.

5.4 Discussion

The fundamental question is, which method is better? Is it better to calculate PP on a daily basis, and then average over time? Or is it better to use temporally averaged input fields to calculate PP? Since most algorithms involve non-linear functions of the input fields, we would expect the two methods to differ. In general, one should evaluate a non-linear function at the highest spatial and temporal resolution possible, and then average the results to obtain its mean over time and space. However, when computing PP from daily Level-2 satellite data, a clear-sky bias is introduced since PP is only calculated where there are no clouds. By using monthly average input fields to calculate PP, there are much fewer missing data in general, and furthermore, the average PAR can be based on all data (not just cloud-free pixels).

One of the surprising results of this study was the insensitivity of the VGPM algorithm to the clear-sky bias. In the VGPM algorithm, PAR (E_0 in Equation 5.1) is used to scale P_{opt}^B, and its effect diminishes for values of PAR $\gg 4.1$ mol quanta m^{-2} d^{-1}. Based on the range of PAR during July 2000 (Figure 5.2), the VGPM scaling factor had a limited range ($0.82 < f_2 < 0.93$). The clear-sky bias had a much more pronounced effect on the HYR algorithm because of its sensitivity to PAR. In this algorithm, PAR (E_0 in Equation 5.2) is used to scale P_{max}^B, and its effect diminishes for values of $E_0 \gg P_{\text{max}}^B / \alpha$. For the range of temperatures and PAR involved in this study, the factor used to scale P_{max}^B in the HYR algorithm spanned a much greater dynamic range ($0.10 < f_2 < 0.60$).

The seventh-order polynomial used in the VGPM algorithm to relate P_{opt}^B to SST led to spatially heterogeneous results. The results can be explained in terms of the curvature, or second derivative, of this polynomial with respect to SST. In a temperature range where the curvature is positive (below 2°C, and between 6°C and 13°C), PP_d exceeded PP_m, and where the curvature was negative (between 2°C and 6°C, and above 13°C), PP_m exceeded PP_d. This could lead to large-scale artefacts in regional estimates of PP. For example, in cooler sub-polar regions, PP would be systematically under estimated, whereas in the warmer subtropical gyres, it would be over estimated.

Campbell *et al.* (1995) have recommended that the ratio of CHL to K_{490} be binned. The rationale is that many PP algorithms involve the product: $C_{\text{sat}} Z_{\text{eu}} = 4.6 C_{\text{sat}} K_{\text{par}}$, and the Level-2 variable K_{490} can serve as a surrogate for K_{par}. Thus, in effect, $<f_3(\text{CHL})>$ would be available for computing PP using monthly average (Level-3) data. Given the heterogeneous behaviour of P_{opt}^B in the VGPM algorithm, it would also be worthwhile binning P_{opt}^B if this algorithm

is used. Then monthly average PAR could be used together with $<P^B_{\text{opt}}>$ and $<C_{\text{sat}}\ Z_{\text{eu}}>$ since this algorithm is insensitive to PAR.

Returning to the question of which method is better, a compromise approach would be to use weekly averaged input fields. This is the method used to derive the MODIS "Level-4" primary productivity products. MODIS has the added advantage of having both SST and CHL measured simultaneously, thus avoiding errors associated with cloud masking. Weekly input fields from MODIS data provide nearly complete global coverage, and thus are ideal as input to PP algorithms.

Future Directions: What is beyond simple "binning/gridding"

Watson Gregg and Marlon Lewis

6.1 From binning to optimal interpolation

The objective of the approaches described above is to produce an ocean state or field that is as close as possible to reality. This approach should, in principle, take into consideration not only the observational errors, but the spatial and temporal error fields as well. Finally, it would be desirable to ensure that the resulting fields are dynamically consistent, taking into account all the available information - certainly the observational satellite data, but also model dynamics, *in situ* data, physical constraints, and climatology. This extends the options beyond simple binning schemes to include a variety of objective analysis methods, which reach their current state of the art in new four dimensional variational data assimilation schemes.

A relatively simple approach would be to weight the observations (or their transforms) based on their distance from the grid point, an approach that dates to the successive correction methods of Cressman (1959). Say we wish to create an analysis field at a specified location and at a given time $\mathbf{x}_a(j)$, from observations taken at different locations and at different times $\mathbf{y}(i)$. The analysis field in our context would be the Level-3 gridded data product, and the observations would be the remotely-observed normalized radiances (or perhaps some derivation of same, such as chlorophyll concentration, or attenuation length, or their statistical transforms). We might assume that the value or usefulness of the observations might diminish the farther one is away from the grid point j in space and time,

$$\mathbf{x}_a(j) = \frac{\sum_{i=1}^{n} w(i,j)\mathbf{y}(i)}{\sum_{i=1}^{n} w(i,j)}, \tag{6.1}$$

where the weights $w(i,j)$ must be determined.

In the simplest case (actually the one taken for the majority of the ocean-colour binning schemes described above), a "zone of influence" is defined in

space and time such that $w(i, j)$ is equal to one, if the observation is within the zone, and zero when it is not. For example, in the SeaWiFS case above, the rectangular tiles or bins of ~ 85 km^2 constitute the spatial zone of influence, whereas the various daily, weekly and monthly products define the temporal zone,

$$w(i, j) = \left(\frac{1, \ d(i, j) < R}{0, \ d(i, j) > R} \right), \tag{6.2}$$

where R represents the spatial or temporal extent of the bin, and $d(i, j)$ represents the location of the observation, both relative to the bin centre.

Fixing the dimensions of the zone of influence, and setting the weights implicitly defines the expectation of the statistical nature of the spatial and temporal scales of variability, specifically the autocorrelation function. The weighting, as applied currently to most ocean-colour data sets, is rather crude, first because the implied autocorrelation function is not realistic (sharp cut off at the scale of the zone), and second, because it assumes that the de-correlation scale is somehow related to changes in the area of the bin with latitude, as a function of the Earth's sphericity. Neither is realistic.

A marginally better approach would be to recognize a weighting that falls off as the "distance" of the observation from the grid point increases. Lewis *et al.* (1988) applied such a method to the gridding of the global Secchi disc climatology as,

$$w(i, j) = \max \left(0, \frac{R^2 - d_{i,j}^2}{R^2 + d_{i,j}^2} \right), \tag{6.3}$$

where the "value" or weight given to a given observation falls off as the inverse square of the distance from the grid point. A point falling directly on the grid is given full weight, and observations outside the defined zone of influence are not accumulated for the given bin. Other schemes (*e.g.* weighting like $\exp\left(-\frac{d_{i,j}^2}{R^2}\right)$) can be envisioned. For the Secchi disk climatology, all points falling in a given temporal bin (seasonal and annual) were given a temporal weight of one, and the above only applied to the spatial field. R was taken to be 2 degrees of both longitude and latitude and the data were gridded to a one-degree resolution.

This approach represents a slight improvement in that it more realistically approximates the shape of observed autocorrelation functions, although fixing R as was done, does not reflect the nature of ocean variability. A more accurate approach was taken by Petrie *et al.* (1999) who were interested in mapping the horizontal (x and y), vertical (z) and temporal (t) distributions of nutrients along the North Atlantic coast, given the historical database of observations. They derived a weighting as:

$$w(i,j) = \exp(-d_{i,j}) * \left(1 + d_{i,j} + \frac{d_{i,j}^2}{3} \right), \tag{6.4}$$

where d_{ij} is defined as a non-dimensional distance (pseudo-distance), given by,

$$d_{i,j} = \sqrt{x'^2 + y'^2 + z'^2 + t'^2}, \tag{6.5}$$

and the primed quantities represent the scaled variables, *e.g.* x' is the distance in the x direction divided by the scale in that direction. The x, y, z, and t variables represent the horizontal, vertical and time coordinates. The appropriate scale varies with location and seasonally; the autocorrelation scales must be determined *a priori*. They estimated that the (x, y, z, t) scales for the continental shelf varied from (40 km, 40 km, 15 m, 45 days) in winter for $z = 0 - 30$ m (the z scale increased to 25 m in deeper water), to (30 km, 30 km, 15 m, 30 days) for all other seasons. Such an approach may have some utility for the ocean-colour binning problem, but would require that appropriate decorrelation scales be determined for the global ocean throughout the seasons. A useful extension of these purely statistical approaches would be to combine the observations with models. Models can potentially bridge the satellite observations in space and time, and provide meaningful information that observations cannot. In practice, models are deficient in their representation of processes and interactions, and consequently their outputs stray from the observations. If they can be linked to the observations, then models can provide greatly enhanced understanding of biogeochemical cycling, by identifying the nature of the deficiencies and providing clues to improvement, as well as by nudging model variables toward realism. The general approach of data assimilation, as applied to the production of realistic fields, constrains both data and models, and allows for prediction of ocean biogeochemical processes in a hind cast, analysis, or forecast sense.

Assume a "background" state, which can be the results of a previous model run, persistence or climatology, $\mathbf{x}_b(j)$. We wish to derive the new analysis field, $\mathbf{x}_a(j)$ as above, subject to the assimilation or constraints of observational data, $\mathbf{y}(i)$, taken at places and times not necessarily coincident with the grid points, j,

$$\mathbf{x}_a(j) = \mathbf{x}_b(j) + \frac{\sum_{i=1}^n w(i,j)\{\mathbf{y}(i) - \mathbf{x}_b(i)\}}{\sum_{i=1}^n w(i,j)} \tag{6.6}$$

The $\mathbf{x}_b(i)$ is the "background" field interpolated to the observational points i. The weights, $w(i,j)$, are as defined above or determined by least-squares criteria or optimal interpolation (see below). Note that more weight can be given to the model (or climatology) by setting the weights less than one for $i = j$. Essentially, this is then a weighted average between the background and the observations.

Such an approach has been used for the production of "blended" fields, which couple satellite observations, *in situ* observations, and a relaxation to either persistence (*e.g.* the previous analysis) or climatology as the background field in the above (Reynolds and Smith, 1994; 1995). This has seen limited use for ocean-colour fields, but with new *in situ* observational platforms now available, could be a useful approach. For example, Gregg and Conkright (2001) have produced a blended ocean-colour product from the CZCS data. This analysis assumes that *in situ* data are valid and uses these data directly in the final product. The satellite chlorophyll data are inserted into the final field using Poisson's equation,

$$\nabla^2 C^b = \Psi, \qquad (6.7)$$

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

$$C_{ibc} = I, \qquad (6.8)$$

where the subscript $_{ibc}$ indicates internal boundary condition, and I is the *in situ* measured value of chlorophyll. Thus, *in situ* data appear un-adjusted in the final blended product. This method assimilates directly on the basis of spatial variability, as inferred from Poisson's equation. The data field is retained in the analyzed field, while the model retains its spatial variability, adjusted for bias by the data field.

Further improvements are possible. Issues with the above include the lack of dynamical constraints (*e.g.* does not have to respect physical laws), that it is not always clear how to specify the weights *a priori*, and that it is not always easy to treat poor observations (*e.g.* adjust weights). Therefore, a more suitable statistical approach is needed, based on some sort of least squares approach. In a general sense, this optimal interpolation can be given as:

$$\mathbf{x}_a(j) = \sum_{i=1}^{n} [\alpha(i,j)\{\mathbf{y}(i)\}], \qquad (6.9)$$

where the optimal weights, $\alpha(i,j)$, represent the least-squares estimates of the true weights. In this method, the weight matrix is chosen to minimize the expected error variance of the analyzed field (Daley, 1991). It differs from the spatial analysis method by allowing the covariance between the model and data to determine the error correlation length scale, and from the blended analysis by use of a statistical approach to defining the weights. The weight matrix now represents the error correlations, and is referred to as the error covariance matrix.

Extensions to the above approaches form the basis on which dynamical biogeochemical models and data (both satellite and *in situ*) can be merged to produce forecasts of ocean biogeochemical processes. These forecasts can, in turn,

provide a "background" against which significantly improved analysis fields can be produced that obey physical/biological constraints, and which avoid some sources of aliasing in the resulting product. For example, the adjoint data assimilation method differs considerably from the methods described above. In essence, this method iterates model parameters, boundary and initial conditions, and forcing functions to minimize a cost function in a least-squares sense. This cost function is a measure of the difference between model output and observations over a specified time and space interval. The model is run forward in time to evaluate the cost function, then an adjoint model, which is forced essentially by the deviations between the model and data, is run backward in time to evaluate the gradient of the cost function. An algorithm is applied that determines how the model should be adjusted to reduce the disparities. The adjustments are made and the entire procedure is run again until minimization results.

The advantage of the method is its inherent diagnosis of the model leading to explicit model improvement. The method has been utilized to considerable success in some biological oceanographic applications (*e.g.* McGillicuddy *et al.*, 1998). More advanced methods include the Kalman filter, which is a four-dimensional state-space approach to variational data assimilation. Building on previous efforts to reduce the state space of the Kalman filter (cf., Cane *et al.*, 1996) in linear reduced-gravity models, the so-called SEEK (Singular Evolutive Extended Kalman) filter has been applied to the tropical Pacific Ocean (Gourdeau *et al.*, 2000; Verron *et al.*, 1999). The approach to reducing the computational burden of the Kalman filter consists essentially in approximating the error covariance matrix by a singular low rank matrix, which leads to making corrections only in those directions for which the error is not naturally attenuated by the system. These directions evolve with time according to the model evolution, yielding to the adaptive nature of the filter. The filter is initialized by a method based on Empirical Orthogonal Functions issued from free runs of the model. This reduction in the error covariance matrix avoids the overwhelming burden of computing the temporal evolution of the prediction error with all the degrees of freedom of the full state vector. An excellent example of this approach, and an exceptionally clear description of the methodology, as applied to a physical/biogeochemical model and the SeaWiFS ocean-colour data set, is given in Natvik and Evensen (2003a, b).

6.2   Ocean-colour products for assimilation

Assimilation data differ from forcing data in that they affect the model biogeochemical distributions in an *a posteriori* and non-causal fashion. They improve model results by enforcing convergence and constraint to observed data distri-

butions. It is critical that data assimilation sources have the highest quality, since they will directly affect model results.

There are several satellite-derived biogeochemically-related output products for potential assimilation into models (Table 6.1). Although chlorophyll is the assimilation data set of primary importance, investigations using other MODIS, MERIS, or GLI data products can provide important information in the assimilation model for biogeochemical cycling. For example, MODIS may contain information on coccolithophores and possible cyanobacteria abundances (Table 6.1) that in the future could be used to refine the model distributions in an assimilation capacity. Furthermore, advances in remote sensing algorithms for the detection of other phytoplankton groups, may provide other opportunities for assimilation.

Table 6.1 Potential data sets for biogeochemical model assimilation, their purpose, and possible sources of data.

Variable	Purpose	Candidate Sensor
Chlorophyll concentration	Total phytoplankton distribution	MODIS/SeaWiFS
Coccolith concentration	Coccolithophore distribution	MODIS
Phycoerythrin	Cyanobacteria distribution	MODIS
CDOM absorption coeff.	CDOM distribution	MODIS/MERIS

6.3   Spatial and temporal resolution

A survey of global ocean carbon and biogeochemical models indicates that the state of the science is generally > 1 degree horizontal resolution (Table 6.2). Future improvements in the next 10 years are expected to require 0.1 degree resolution. Coastal models require much greater resolution, and are typically run at about 0.1 degree. However, future coastal models may be expected to require 0.01 degree (1 km) resolution. Requirements for spatial resolution of ocean-colour data products should meet, or exceed, the highest resolution models to provide maximum usefulness for data assimilation. This is presently about 0.1 degrees, based on coastal model requirements, and may be expected to reach 0.01 degrees in 5 to 10 years.

Daily or monthly assimilation requirements are typical for present efforts, although biogeochemical assimilation is fairly immature. However, these temporal frequencies must be considered the minimum requirement for ocean-colour data products. Monthly data can be quite biased, which can affect assimilation. Often a single observation represents an entire monthly mean for a given bin. If this observation occurs at the beginning or end of the month, or before or after a major bloom/recede event, it is a biased representation of the

Table 6.2 Representative sample of global ocean carbon cycle and biogeo-chemical models currently in use, and their horizontal grid structure and orientation.

Organisation	Resolution (lon. x lat., degrees)	Rectangular
Alfred Wegener Institute (AWI)	5 x 4, 2.5 x 2	yes
Commonwealth Science and Industrial Research Organization (CSIRO)	5.6 x 3.2	yes
Institute for Global Change Research (IGCR)	4 x 4	yes
Institut Pierre Simon Laplace (IPSL)	2 x 1.5 (0.5 Equator)	yes
Lawrence Livermore National Laboratory (LLNL)	4 x 2	yes
Massachusetts Institute of Technology (MIT)	2.8 x 2.8	yes
Max Planck Institut fuer Meterologie (MPIM)	5 x 5	yes
NASA/Global Modeling and Assimilation Office (GMAO)	1.25 x 0.67	yes
National Center for Atmospheric Research	(NCAR) 3.6 x 1.8 (0.8 Equator)	yes
Nansen Environmental and Remote Sensing Center (NERSC)	2 x 3.2 cos(lat)	no
Physics Institute University of Bern (PIUB)	Zonal basin average	yes
Princeton University /Geophsyical Fluid Dynamics Laboratory (GFDL)	3.75 x 4.5	yes
Southampton Oceanography Centre (SOC)	2.5 x 3.75	yes
University of Liege (UL)	3 x 3	yes

Data from Dutay *et al.* (2002)

monthly mean. Therefore daily products are required for assimilation. Future assimilation efforts may take advantage of higher temporal frequencies, such as 3 to 6 hourly, if data are available from merged products. This is despite the loss of coverage from such highly refined temporal products. But models are time-stepped, and can potentially utilize ocean-colour data in assimilation at these intervals. Such high frequency assimilation may be advantageous for models, as they will produce less initialization shock and more coherent model-data matches. Disadvantages are associated with stopping model runs more frequently and re-initialization, which will slow model execution. However, we should look toward model improvement as the ultimate goal.

Ocean-colour data product standardization and ease of use are important to

assimilation efforts, since assimilation is a complex effort and model execution is computer-intensive. Regularly-spaced grids, preferentially rectangular in orientation, are most useful for most model activities, and equal-angle grids are preferred. Data products should contain only one variable and "value-added" data formats, such as HDF, net-CDF, GRIB, T62, tend to require more effort by modellers to utilize data sets. Standard IEEE flat-file formats are useable by virtually all operating platforms at the present time, often requiring only simple modifications in executable scripts or software.

Chapter 7

Conclusions and recommendations for binning of ocean-colour data

David Antoine

The summary of ocean-colour data binning schemes presented in this report has highlighted their diversity, and the examples provided above have illustrated some of the consequences of this diversity for the outcome of various scientific studies. The working group has come up with a set of recommendations that could help reduce the ambiguities arising from using different binning schemes. They are summarized later in this section.

It is, however, clear that one unique scheme for binning ocean-colour data would most likely not satisfy all objectives. This is why a "minimum" set of recommendations was issued for the most common uses of ocean-colour data, irrespective of other specific schemes. Alternative solutions are not proposed for these specific applications, in particular for assimilation of ocean-colour data into biogeochemical models, simply because this field is rapidly evolving and demands temporal resolutions that are not currently available for ocean-colour missions. Using this "minimum" set of recommendations, would help to facilitate the generation of multi-mission products, one of the acknowledged goals of the ocean-colour community (*e.g.*, McClain, 1998). It should be borne in mind, though, that the comments and suggestions provided in this report are confined to our present state of knowledge of the physical and biological mechanisms that underlie ocean-colour time and space distributions.

A first recommendation to space Agencies would be to produce Level-3 products. While this statement might be obvious, it should not be taken for granted. At the time of writing, only two missions, SeaWiFS and MODIS, produce Level-3 products. POLDER-2 and GLI on ADEOS-2 also produced Level-3 products, but unfortunately their mission ended in October 2003. All other missions produce only Level-2 products.

The Level-3 products have proven to be extremely useful for research, applications and education. They will also become part of the operational oceanography scenario in the near future (*e.g.*, the MERCATOR Project, Bahurel *et al.*, 2002), when the temporal and spatial scales of this data will be amenable to this type of

application. Level-3 products are the only way to synthesize the large quantity of observations that the satellite sensors provide, and they also enhance their usefulness. It allows general features, as well as the mean temporal evolutions, to be described and quantified. Level-3 products have been used, for example, in biogeochemical modelling of primary production and the carbon cycle (Longhurst *et al.* 1995; Behrenfeld and Falkowski, 1997) and for the description of aerosol distributions (Wang *et al.*, 2000). In terms of the physical ocean, the same data have also served to quantify the role of phytoplankton in the modulation of the heating rate of the upper ocean layers, and in the modulation of the Earth's albedo at regional and global scales (Nakamoto *et al.* 2000; Frouin and Iacobellis, 2002), and also in describing circulation features and scales of phenomena (Machu *et al.*, 1999). In addition, verifying the long-term stability of ocean-colour sensors through analysis of time-series over major subtropical, oligotrophic, gyres cannot be achieved without Level-3 products. Level-3 products do have some limitations though, such as when focusing on very small scale and short term events, which are better examined with Level-2 data.

From a practical point of view, Level-3 products are the only products that can be readily stored and manipulated by users examining regional or global scale applications. Corresponding volumes of Level-2 data would be totally unmanageable. Users should nevertheless be aware of the possible limitations when using Level-3 data in highly non-linear models (see Chapters 5 and 6).

It is apparent that log-transforms of the data and the use of the geometrical mean (GEO) come into play when generating relevant ocean-colour-derived quantities through accumulation of individual pixel measurements into bins. One reason for this is the coherence of knowledge about the distribution of bio-optical variables in the ocean (Campbell, 1995), and the inherent ability of geometrical mean transformation to minimize the impact of outliers, such as incorrectly processed data, that may have slipped through the net of quality control procedures. The significance of the log-transform is, however, only verified when the spatial and temporal scales are large enough, *e.g.* a monthly mean over a 10 km × 10 km area. It may loose its importance when data from the same area are averaged over a single day, for instance.

It is therefore recommended that the simple arithmetic mean (AVG) be used for spatial scales below tens of kilometres, on a daily scale, whereas the geometric mean (GEO) should be used for all larger scales. If the intra-bin variance has been saved, then the log-normal approximation can be used to convert AVG to GEO, or vice versa. This reversibility should be conserved as far as possible (this is also true for grids). The difference between the AVG and GEO estimates may also be used as a quality indicator (large differences may indicate artefacts in the data distribution). A corollary is that any technique that reduces the noise in Level-2, will facilitate the production of unbiased Level-3 products, implying that quality control procedures should be as effective as possible. Another op-

tion is to use GEO for quality control, and AVG for the computation of the final mean quantities.

The bidirectional character of the light field emerging from the ocean should also be accounted for in the Level-2 processing schemes, so that a full normalization of the water-leaving radiance is achieved. This would ensure that the Level-3 products are free of geometric artefacts, and would allow, for example, a worthwhile comparison of average radiances in a bin near the polar circle with that in a bin near the Equator. This would also permit successful merging of Level-2 or Level-3 radiance fields from different missions.

Correcting for these geometrical effects is well documented (*e.g.*, Morel and Gentili, 1996; Morel *et al.*, 2002) and can be performed easily by relying on look-up tables of the relevant quantities. The look-up table approach will also allow any new knowledge to be incorporated into operational processing chains by simply updating the corresponding "auxiliary products" (*i.e.*, the look-up table files).

As far as temporal scales of ocean colour are concerned, the ease of use of multi-mission data sets and the commonality between missions are more important than mission characteristics (*e.g.*, orbit cycle). The minimum temporal scale that is essential, is the daily scale. From that, any other temporal accumulation of the data can be produced (weekly, 8-day, monthly, decadal, seasonal, annual or climatologies) either by Agencies who support this work, or by scientific and other users who have the necessary resources. That being said, the definition of a data-day still has to be standardised. It would be most advantageous if the definition adopted by the SeaWiFS project, based on previous experience with the AVHRR mission, were adopted as a standard (see Chapter 4).

To avoid leaving the decision up to the end-users, it is strongly recommended that Agencies produce and distribute weekly and monthly products, as standard products. The examples provided in this report advocate the production of these weekly gridded products. Inevitably, daily products have incomplete fields (because of cloudiness and sun glint), and are thus not optimal. It is further recommended that the basic weekly products be 7-day averages.

Ocean-colour data assimilation is somewhat different with respect to temporal scales, since the assimilation scenarios often require global products of, at least, daily temporal resolution, which are currently not attainable with present ocean-colour missions. It will remain unattainable until routine observations become available from a constellation of satellites, or from satellites in geostationary orbit. It is feasible that these types of missions may never exist. The only foreseeable way to produce daily fields with the least amount of missing data is to merge data from several missions. Although it is not the central topic discussed here, data merging should be facilitated if the recommendations proposed in this report are adopted.

As for the type of projection, a compromise has to be reached between respecting the spatial/statistical structure of the data (best achieved using the

equal-area binning schemes) and facilitating easy use and merging of the data (best achieved with equal-angle, regular latitude/ longitude grids used for "mapped products"). The equal-angle grid is recommended for historic reasons, as well as for its simplicity and commonality with other data sets. Bins (or tiles) are not recommended, although it may be possible to nest the ones selected for a particular mission into other missions, or vice versa. Interpolation methods do exist, but the level of noise they introduce is not easily quantifiable.

Space Agencies should be encouraged to support studies to improve the gridding/mapping methodologies. They should also endeavour to incorporate more physics in the process of generating large-scale fields of various properties. These studies may include an investigation into typical spatial scales of phenomena (eddies, fronts, jets, meso-scale to sub-meso-scale features *etc.*) as well as the time constants of the same phenomena (permanent, transient, seasonal, pulse-like *etc.*). The relationship between the features observed in the ocean-colour fields and other environmental variables (SST for instance) still need to be documented, as well as the physical mechanisms that produce the observed distributions. In the end, the mathematical methods used to generate basin-scale, Level-3 products have to respect the characteristics of the data.

It is imperative that all Space Agencies possess the capability to reprocess archived data. Reprocessing is essential to incorporate calibration adjustments and algorithm improvements. This is true for the production of Level-2 products, as well as for Level-3 products. The latter may benefit from new developments in the area of production of the best possible fields, for a variety of applications.

The last recommendation, related to the creation and use of Level-3 ocean-colour products, is of a practical nature. Space Agencies should provide the necessary tools for converting one product into another one, *e.g.* products based on two different grids. For instance, it is useful to go from the equal-area binned products to the equal-angle mapped products (this is possible for SeaWiFS products), or to change the spatial resolution of a given mapped or binned product. To limit the need for "in-house" code development, data formats should not be mission-specific, but should rather be basic and widely accepted and used. Data formats should also be enforced (*e.g.*, HDF, net-CDF, flat files).

Finally, it should be noted that some issues, which may also impair the easy use of Level-3 data from different missions, such as differences in the spectral band sets of the various missions, as well as differences in the algorithms used to generate the Level-2 products, have not been examined here. Another aspect that was not addressed, is whether it is best to average spatially the basic pixel information, and then stratify temporally the resulting spatially-averaged quantities, or vice versa.

Recommended "common" or "minimum" binning scheme

David Antoine

According to the terms of reference for the working group, we have developed a recommendation for a binning scheme that could be used across Agencies. This is a"minimum" binning scheme and is a practical expression of the more general recommendations provided in the previous section.

To take advantage of the experience gained with present missions, and to minimize potential new sources of error, the recommended scheme is essentially the one used by the SeaWiFS and MODIS missions, with specific modifications/features listed below. The scheme should:

❖ Be applied to the chlorophyll concentration[1], to the aerosol optical thickness in the NIR (usually around 865 nm), and to fully normalized water-leaving reflectances. Other products could be binned successfully using the same grid, if they exist as Level-2 products of a given mission. The normalized water-leaving reflectance is obtained by multiplying the normalized water-leaving radiance (Equation 3.5 in Section 3.1) by the ratio $(\pi/F_0(\lambda))$.

❖ Use an equal-area grid, with a spatial resolution of one twelfth of a degree in latitude and one twelfth of a degree in longitude at the Equator[2]. The poles are formed by three bins. Other characteristics have been described in detail in Section 2.1.3.

❖ Produce daily and monthly mean quantities, $\overline{X}$, according to the definition of the arithmetic mean in Table 3.2.

❖ Use the AVHRR-SeaWiFS definition of a "data-day".

---

[1]Irrespective of the various definitions of this product.

[2]Recent studies at the time of going to press suggest that "true" equal-angle map projections be used rather than SMI projections".

References

Bahurel, P., De Mey, P., Le Provost, C. and Le Traon, P.Y. (2002). A GODAE prototype system with applications – Example of the Mercator system. *"En route to GODAE" 2002 Proceedings*. Edited by CNES, France, pp. 137-142.

Behrenfeld, M.J. and Falkowski, P.G. (1977). Photosynthetic rates derived from satellite-based chlorophyll concentration, *Limnology and Oceanography* **42**: 1-20.

Behrenfeld, M.J., Randerson, J., McClain, C., Feldman, G., Los, S., Tucker, C., Falkowski, P., Field, C., Frouin, R., Esaias, W., Kolber, D. and Pollack, N. (2001). Temporal variations in the photosynthetic biosphere, *Science* **291**: 2,594-2,597.

Bricaud A., Bosc, E. and Antoine, D. (2002). Algal biomass and sea surface temperature in the Mediterranean basin: intercomparison of data from various satellite sensors, and implications for primary production estimates. *Remote Sensing of the Enviroment* **81**: 163-178.

Campbell, J.W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research* **100**: 12,237-13,254.

Campbell, J.W., Blaisdell, J.M. and Darzi, M. (1995). Level-3 SeaWiFS data products: Spatial and temporal binning algorithms. In: *SeaWiFS Technical Report Series*, Hooker, S.B., Firestone, E.R. and Acker J.G. (eds.). NASA Technical Memorandum 104566, Vol. 32, 73 pp.

Campbell, J.W., Antoine, D., Armstrong, R., Arrigo, K., Balch, W., Barber, R., Behrenfeld, M., Bidigare, R., Bishop, J., Carr, M.-E., Esaias, W., Falkowski, P., Hoepffner, N., Iverson, R., Kiefer, D., Lohrenz, S., Marra, J., Morel, A., Ryan, J., Vedernikov,V., Waters, K., Yentsch, C. and Yoder, J. (2002). Comparison of algorithms for estimating ocean primary productivity from surface chlorophyll, temperature, and irradiance. *Global Biogeochemical Cycles* **16(3)**: 10.1029 / 2001GB001444.

Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C. and Busalacchi, A. (1996). Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter *Journal of Geophysical Research* **10**: 22,599-22617.

Cressman, G.P. (1959). An operational objective analysis system. *Monthly Weather Review* **87**: 367-374.

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

Deschamps P.Y., Bréon, F.M., Leroy, M., Podaire, A., Bricaud, A., Buriez, J.C. and Seze, G. (1994). The POLDER Mission: Instrument characteristics and scientific objectives. *IEEE Transactions in Geosciences and Remote Sensing* **32**, 598-615.

Ducet, N., Le Traon, P-Y., and Reverdin, G. (2000). Global high resolution mapping of ocean circulation from Topex/Poseidon and ERS-1/2. *Journal of Geophysical Research* **105**, 19,477-19,498.

Dutay, J.-C., and 28 others (2002). Evaluation of ocean model ventilation with CFC-11: Comparison of 13 global ocean models. *Ocean Modelling* **4**: 89-120.

Eppley, R.W. (1972). Temperature and phytoplankton growth in the sea. *Fisheries Bulletin* **70**: 1,063-1,085.

Frouin, R., Franz, B. and Wang, M. (2001). Algorithm to estimate PAR from SeaWiFS data. Version 1.2 - Documentation. http://seadas.gsfc.nasa.gov/doc/msl12/seawifs_par_wfigs.pdf

Frouin R., and Iacobellis, S.F. (2002). Influence of phytoplankton on the global radiation budget. *Journal of Geophysical Research* **107(D19)**: 5-1 to 5-10.

Gordon, H.R. and Clark, D.K. (1981). Clear water radiances for atmospheric correction of Coastal Zone Colour Scanner imagery. *Applied Optics* **20**: 4,175-4,180.

Gourdeau, L., Verron, J., Delcroix, T., Busalacchi, A.J. and Murtugudde, R. (2000). Assimilation of TOPEX/Poseidon altimetric data in a primitive equation model of the tropical Pacific Ocean, during the 1992-1996 ENSO period. *Journal of Geophysical Research* **105**: 8,473-8,488.

Gregg, W.W. and Conkright M.E. (2001) Global seasonal climatologies of ocean chlorophyll: Blending *in situ* and satellite data for the Coastal Zone Color Scanner era. *Journal of Geophysical Research* **106(C2)**: 2,499-2,515.

Howard, K.L., and Yoder, J.A. (1997). Contribution of the subtropical ocean to global primary production. In: *Space Remote Sensing of the Sub-tropical Oceans*, edited by Liu, C.-T. Pergamon, New York. pp. 157-168.

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

Kilpatrick K.A., Podesta G.P. and Evans, R.H. (2001). Overview of the NOAA/NASA AVHRR pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research* **106**: 9,179-9,197.

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

Levitus, S. (1982). Climatological atlas of the world ocean, NOAA Professional paper 13, U.S. Government Printing Office, Washington, xv + 173 pp.

Lewis, M.R., Kuring, N. and Yentsch, C. (1988). Global patterns of ocean transparency: Implications for the new production of the open ocean. *Journal of Geophysical Research* **93**: 6,847-6,856.

Loisel, H. and Morel, A. (2001). Non-isotropy of the upward-radiance field in typical coastal case 2 waters. *International Journal of Remote Sensing* **22**: 275-295.

Loisel, H. and Stramski, D. (2000). Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. *Applied Optics* **39**: 3,001-3,011.

Longhurst A.R., Sathyendranath, S., Platt, T., and Caverhill, C. (1995). An estimate of global primary production from satellite radiometer data. *Journal of Plankton Research* **17**: 1,245-1,271.

Machu, E., Ferret, B., and Garçon, V.C. (1999). Phytoplankton distribution from SeaWiFS data in the subtropical convergence zone south of Africa: a wavelet analysis. *Geophysical Research Letters* **26**: 1,469-1,472.

McClain C.R. (1998). SIMBIOS background. In: *SIMBIOS Project 1998 Annual Report*, Eds. McClain, C.R. and Fargion, G. NASA/TM-1999-208645, NASA Goddard Space Flight Center, MD 20771, USA.

McGillicuddy, D.J., Lynch, D.R., Moore, A.M., Gentleman, W.C., Davis, C.S. and Meise, C.J. (1998). An adjoint data assimilation approach to diagnosis of physical and biological controls on *Pseudocalanus spp.* in the Gulf of Maine — Georges Bank region. *Fisheries Oceanography* **7**: 205-218.

Medeiros, W.H. and Wirick, C.D. (1992). SEEP II: Shelf Edge Exchange Processes II. Chlorophyll *a* fluorescence, temperature, and beam attenuation measurements from moored fluorometers. *BNL 47211 Informal Report*. Brookhaven National Laboratory, Upton, New York, 205 pp.

Morel, A. (1991). Light and marine photosynthesis: a spectral model with geochemical and climatological implications. *Progress in Oceanography* **26**: 263-306.

Morel, A., Antoine, D. and Gentili, B. (2002). Bidirectional reflectance of oceanic waters: Accounting for Raman emission and varying particle phase function. *Applied Optics* **41**: 6,289-6,306.

Morel, A. and Berthon, J.-F. (1989). Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. *Limnology and Oceanography* **34**: 1,545-1,562.

Morel, A. and Gentili, B. (1996). Diffuse reflectance of oceanic waters. III. Implication of the bidirectionality for the remote sensing problem. *Applied Optics* **35**: 4,850-4,862.

Morel, A., Voss, K.J. and Gentili, B. (1995). Bidirectional reflectance of oceanic

waters: a comparison of modeled and measured upward radiance fields, *Journal of Geophysical Research* **100**: 13,143-13,150.

Moulin C., Giraud, S., Poitou, J. and Bréon, F.-M. (2001). Spécifications de la chaîne de traitement Couleur de l'Océan de niveau 3 Classe 2, *CNES report n° P2-ST-364-3620-CEA*, Ed.0 Rev.0.

Nakamoto S., Prasanna Kumar, S., Oberhuber, J.M., Muneyama, K. and Frouin, R. (2000). Chlorophyll modulation of sea surface temperature in the Arabian Sea in a mixed-layer isopycnal general circulation model. *Geophysical Research Letters* **27**: 747-750.

Natvik, L.-J and Evensen, G. (2003a). Assimilation of ocean-colour data into a biochemical model of the North Atlantic. Part 1. Data assimilation experiments. *Journal of Marine Systems* **40-41**: 127-153.

Natvik, L.-J and Evensen, G. (2003b). Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 2. Statistical analysis. *Journal of Marine Systems* **40-41**: 155-169.

Nelson, D.M., and Smith, W.O. (1991). Sverdrup revisited: critical depths, maximum chlorophyll levels, and the control of Southern Ocean productivity by the irradiance-mixing regime. *Limnology and Oceanography* **36**: 1,650-1,661.

Nicolas J.M. and Deschamps, P.-Y. (2000). Document des algorithmes et produits de la chaîne de traitement de niveau 2 "Océan" de Classe 2, CNES report n°P2-SP-OZ-1128-LOA, Ed.1 Rev.1.

Petrie, B., Yeats, P. and Strain, P. (1999). Nitrate, silicate and phosphate atlas for the Scotian Shelf and the Gulf of Main. *Canadian Technical Report of Hydrography and Ocean Sciences* **203**: 99 pp.

Podestá, G.P. (1995). Case studies for SeaWiFS calibration and validation, Part 3, SeaWiFS global fields: What's in a day? *NASA Technical Memorandum* 10456, 27 pp.

Preisendorfer, R.W. (1976). *Hydrologic Optics. Vol. I. Introduction*, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratory, Honolulu, 218 pp.

Ras, J. (1999). Primary production modelling from ocean colour satellite data (SeaWiFS): a practical study on the Mediterranean Sea. *Dissertation in fulfilment of M.Sc in Oceanography*, University of Southampton, 51 pp.

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

Reynolds, R. W. (1993). Impact of Mount Pinatubo aerosols on satellite-derived Sea Surface Temperatures. *Journal of Climate* **6**: 768-774.

Reynolds, R.W. and Marsico, D.C. (1993). An improved real-time global sea surface temperature analysis. *Journal of Climate* **6**: 114-119.

Reynolds, R.W. and Smith, T.M. (1994). Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate* **7**: 75-86.

Reynolds, R.W. and. Smith, T.M (1995). A high-resolution global sea surface temperature climatology. *Journal of Climate* **8**: 1,571-1,583.

Rossow W. and Gardner, L. (1984). Selection of a map grid for data analysis and archival. *Journal of Climate and Applied Meteorology* **23**: 1,253-1,257.

Trela, P., Sathyendranath, S., Moore, R.M. and Kelley, D.E. (1995). Effect of the nonlinearity of the carbonate system on partial pressure of carbon dioxide in the oceans. *Journal of Geophysical Research* **100(C4)**: 6,829-6,844.

Verron, J., Gourdeau, L., Pham, D. T., Murtugudde, R. and Busalacchi, A.J. (1999). An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical Pacific Ocean: Method and validation. *Journal of Geophysical Research* **104**: 5,441-5,458.

Wang M., Bailey, S. and McClain, C.R. (2000). SeaWiFS provides unique global aerosol optical property data. *EOS Transactions American Geophysical Union* **81**: 197-202.

Zibordi G. and Berthon, J.-F. (2001). *In situ* relationships between the Q-factor and seawater optical properties in coastal regions. *Limnology and Oceanography* **46**: 1,130-1,140.

Acronyms and Abbreviations

ADEOS   Advanced Earth Observing Satellite

AOP   Apparent Optical Properties (*sensu* Preisendorfer, 1976).

AVG   Arithmetic Average

AVHRR   Advanced Very High Resolution Radiometer

AVISO   Archiving, Validation and Interpretation of Satellite Oceanographic data

BRDF   Bidirectional Reflectance Distribution Function

CDOM   Coloured Dissolved Organic Matter

CNES   Centre National d'Etudes Spatiales

CZCS   Coastal Zone Color Scanner

DAO   Data Assimilation Office (NASA)

DMSP   Defence Meteorological Satellite Program

ECMWF   European Centre for Medium Range Weather Forecast

EORC   Earth Observation Research and Application Centre

EOS   Earth Observing System

ESA   European Space Agency

ENVISAT   Environmental Satellite

EORC   Earth Observation Research Centre

ESSW   Earth System Science Workbench

GAC   Global Area Coverage

GCM   General Circulation Model

GEO   Geometric mean

GFDL   Geophysical Fluid Dynamics Laboratory

GLI   GLobal Imager

GSFC   Goddard Space Flight Center

GT   Giga tonnes

HDF-EOS   Hierarchical Data Format - Earth Observing System

HYR   Howard-Yoder-Ryan (algorithm)

IOCCG	International Ocean-Colour Coordinating Group
IOP	Inherent Optical Properties (*sensu* Preisendorfer, 1976).
IPSL	Institut Pierre Simon Laplace
ISCCP	International Satellite Cloud Climatology Project
ISLSCP	International Satellite Land Surface Climatology Project
JAXA	Japan Aerospace Exploration Agency
LOV	Laboratoire d'Océanographie de Villefranche, France
LSCE	Laboratoire des Sciences du Climat et de l'Environnement
MERIS	Medium Resolution Imaging Spectrometer
MLE	Maximum Likelihood Estimator
MODIS	MODerate Resolution Imaging Spectrometer.
MPI	Max Planck Institute
MTIR	Medium to Thermal Infrared
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency of Japan
NCEP/NCAR	National Center for Environmental Predictions / National Center for Atmospheric Research
NDPI	Normalized Difference Pigment Index
NIR	Near Infrared
NOAA	National Oceanic and Atmospheric Administration
NPOESS	National Polar-orbiting Operational Environmental Satellite System
OCTS	Ocean-Colour and Temperature Scanner
OI	Optimal Interpolation
PAR	Photosynthetically Available Radiation
PATMOS	AVHRR Pathfinder Atmosphere
POLDER	POLarization and Directionality of the Earth Reflectance
PODAAC	Physical Oceanography Distributed Active Archive Center
PP	Primary Production
RSMAS	Rosenstiel School of Marine and Atmospheric Sciences
SeaDAS	SeaWiFS Data Application Software
SeaWiFS	Sea-viewing Wide Field of view Sensor
SLA	Sea Level Anomaly
SMI	Standard Mapped Image
SSH	Sea Surface Height
SSHA	Sea Surface Height Anomaly

SSI	Surface Solar Irradiance
SST	Sea Surface Temperature
SWIR	Short wave infrared
TOA	Top of Atmosphere
TOVS	TIROS Operational Vertical Sounder
T/P	Topex / Poseïdon
VGPM	Vertically Generalized Production Model
VSF	Volume Scattering Function

Symbols and Units

$\alpha(\lambda)$	Absorption coefficient	m^{-1}
$b_b(\lambda)$	Backscattering coefficient	m^{-1}
CHL	Chlorophyll concentration	mg m^{-3}
C_{sat}	"Satellite" chlorophyll concentration	mg m^{-3}
D_{irr}	Day length	h
Δ_ϕ	Relative azimuth difference angle	degrees
E_{bar}	Average PAR in the euphotic zone	W m^{-2}
$E_d(\lambda)$	Downward irradiance	$\text{W m}^{-2}\mu^{-1}$
$E_u(\lambda)$	Upward irradiance	$\text{W m}^{-2}\mu^{-1}$
E_s	"Half saturation irradiance" $(= P^B_{\text{max}}/\alpha)$	W m^{-2}
E_0	Notation of PAR in the VGPM model	W m^{-2}
ε_c	Correction factor applied to $F_0(\lambda)$, and accounting for the changes in the Earth-sun distance. It is computed from the eccentricity of the Earth orbit, e = 0.0167, and from the day number D, as $\varepsilon_c \left(1 + e \cos\left(\frac{2\pi(D-3)}{365}\right)\right)^2$	dimensionless
$F_0(\lambda)$	Mean extraterrestrial spectral irradiance	$\text{W m}^{-2}\text{ nm}^{-1}$
K_{par}	Diffuse attenuation coefficient for the downwelling PAR	m^{-1}
K_{490}	Diffuse attenuation coefficient for downwelling irradiance at 490 nm	m^{-1}
$L_u(\lambda)$	Upward radiance	$\text{W m}^{-2}\text{ sr}^{-1}\mu^{-1}$
$L_w(\lambda, \theta_s, \theta', \Delta\phi)$	Water-leaving normalized radiance	$\text{mW cm}^{-2}\text{ sr}^{-1}\mu^{-1}$

$[L_w]_N(\lambda, \theta_s, \theta', \Delta\phi)$ or nL_w	Normalized water-leaving radiance	mW cm^{-2} sr$^{-1}\mu^{-1}$
λ	Wavelength	nm
P^B_{opt}	Maximum rate of carbon fixation in the water column	g C (g Chl)$^{-1}$ h^{-1}
P^B_{max}	Maximum photosynthetic rate per unit chlorophyll biomass	g C (g Chl)$^{-1}$ h^{-1}
PP_d	Monthly average daily primary production obtained from daily ocean-colour observations.	g C m^{-2}
PP_m	Monthly average daily primary production obtained from monthly composites of ocean-colour observations.	g C m^{-2}
PP^*_m	Monthly average daily primary production obtained from monthly composites of ocean-colour observations and clear-sky pixels only.	g C m^{-2}
$Q(\lambda, \theta_s, \theta', \Delta\phi)$	Q factor (*i.e.*, E_u/L_u) Subscript 0 when $\theta' = \Delta\phi = 0$	sr
θ_s	Solar zenith angle (cosine is μ_s)	degrees
θ_v	Satellite zenith angle (cosine is μ_v) ("viewing angle")	degrees
θ'	$\theta' = \sin^{-1}(\sin(\theta_v)/1.34)$	degrees
$R(\lambda, \theta_s)$	Measured reflectance just below sea surface, *i.e.* $[(E_u(0^-)/(E_d(0^+))*0.96)]$, with a solar zenith angle θ_s	dimensionless
R_e	Earth radius	km
R_{rs}	Remote sensing reflectance $(L_w(0^+, \theta, \phi, \lambda)/E_d(0^+, \lambda))$	sr^{-1}

$\Re(\theta')$	Geometrical factor, accounting for all refraction and reflection effects at the air-sea interface (Morel and Gentili, 1996) $\Re(\theta') = \left[\frac{(1-\overline{\rho})}{(1-\overline{r}R)} \frac{(1-\rho_F(\theta'))}{n^2} \right]$ (subscript 0 when $\theta' = 0$)	dimensionless
	where:	
	n is the refractive index of water	dimensionless
	$\rho_F(\theta)$ is the Fresnel reflection coefficient for incident angle θ	dimensionless
	$\overline{\rho}$ is the mean reflection coefficient for the downwelling irradiance at the sea surface	dimensionless
	$\overline{r}$ is the average reflection for upwelling irradiance at the water-air interface	dimensionless
$t_d(\lambda, \theta)$	Diffuse transmittance for angle θ $t_d(\lambda, \theta) = L_{TOA}(\lambda, \theta_s, \theta_v, \Delta\phi)/L_0 + (\lambda, \theta_s, \theta_v, \Delta\phi)$, *i.e.*, the ratio of the radiance at the TOA to the radiance at the sea level (0^+) often computed as $\exp\left(-\frac{\tau_r/2 + \tau_{o3}}{\cos(\theta)}\right)$	dimensionless
$\tau_r(\lambda)$	Rayleigh optical thickness	dimensionless
$\tau_a(\lambda)$	Aerosol optical thickness	dimensionless
$\tau_{o3}(\lambda)$	Ozone optical thickness	dimensionless
Z_{eu}	Euphotic zone, *i.e.*, that depth where PAR is reduced to 1% of its value just below the sea surface.	m

Appendix: URLs for various data sets listed in the report

Ocean-colour data

CZCS information
http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/OCDST/OB_Documentation.html

SeaWiFS project homepage
http://seawifs.gsfc.nasa.gov/SEAWIFS.html

MODIS ocean homepage
http://modis-ocean.gsfc.nasa.gov/

OCTS
http://kuroshio.eorc.jaxa.jp/ADEOS/Project/Octs.html

POLDER 1 and 2
http://polder.cnes.fr/

GLI
http://sharaku.eorc.jaxa.jp/GLI/index.html

SST data

AVHRR
http://podaac.jpl.nasa.gov/sst/sst_data.html#data_description

OCTS
http://kuroshio.eorc.jaxa.jp/ADEOS/SpecialData/Global_octs.html

GLI
http://sharaku.eorc.jaxa.jp/GLI/index.html

MODIS
http://modis-ocean.gsfc.nasa.gov/

LEVITUS
http://ingrid.ldeo.columbia.edu/SOURCES/.LEVITUS/

Reynolds
http://www.cdc.noaa.gov/

PAR data

ISLSCP - International Satellite Land Surface Climatology Project
http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/ISLSCP/DATASET_DOCUMENTS/MTH3HRAD.html

SeaWiFS - Sea-viewing Wide Field-of-view Sensor
DATA: http://daac.gsfc.nasa.gov/data/dataset/SEAWIFS/06_New_Products/index.html

MODIS - Moderate Resolution Imaging Spectroradiometer
http://acdisx.gsfc.nasa.gov/data/dataset/MODIS/03_Ocean/index.html

SSI - High-resolution (DX) Surface Solar Irradiance
http://daac.gsfc.nasa.gov/data/dataset/SSI/01_Data_Products/index.html

ESSW - Earth System Science Workbench, UC Santa Barbara
http://essw.bren.ucsb.edu/products/2001.shtml

PAR Project - University of Maryland
http://www.meto.umd.edu/~srb/par/cgi-bin/dataaccess.cgi

Cloud cover data

ISSCP - International Satellite Cloud Climatology Project
http://isccp.giss.nasa.gov/overStatPg.html

DAO - Data Assimilation Office (NASA)
http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/FTP_SITE/INT_DIS/readmes/assim54A_mo_subset

ISLSCP - International Satellite Land Surface Climatology Project
http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/ISLSCP/DATASET_DOCUMENTS/CLOUDS.html

SSI - High-resolution (DX) Surface Solar Irradiance
http://daac.gsfc.nasa.gov/data/dataset/SSI/01_Data_Products/index.html

NCEP/NCAR Reanalysis - National Center for Environmental Prediction/National Center
for Atmospheric Research
http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html

MODIS - Moderate Resolution Imaging Spectroradiometer
http://acdisx.gsfc.nasa.gov/data/dataset/MODIS/02_Atmosphere/02_Level_3/index.html

DMSP - Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager(SSM/I)
http://lwf.ncdc.noaa.gov/oa/satellite/ssmi/ssmiproducts.html

TOVS - TIROS Operational Vertical Sounder
http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/FTP_SITE/INT_DIS/readmes/tovs.html

PATMOS - AVHRR Pathfinder Atmosphere
http://www.saa.noaa.gov:8080/cocoon/nsaa/products/search

ECMWF - European Centre for Medium-range Weather Forecasting
http://www.ecmwf.int

Products derived from altimetry

http://podaac.jpl.nasa.gov/order/order_topex.html
and
http://www-aviso.cls.fr/