imum of data with a minimum of scientist's exhaustion. Ergo, data can be readily shared between participants, which is a relatively straightforward consideration and meets the principles of the LOICZ Data and Information System Plan (Boudreau et al. 1996).

Data acquisition is carried out as retrieval from national data centers, submission as e-mail attachment, URL, hardcopy, etc. Data archival is got to work by WDC-MARE staff at a central place. Depending on skill, data access is offered through (1) a simple, GOOGLE™-like data search engine; (2) a sophisticated data mining tool *ART*; (3) a client/server data management tool *4th Dimension*®; (4) various PANGAEA-related data analysis and visualization software tools. As some data sets refer to existing publications, others are yet unpublished or non-public, all data sets were reconditioned under the same criteria to satisfy PANGAEA's formal RDBMS standard, the so-called database model (Fig. A.1.2), to ensure maximum quality, and to finally be disseminated at URL http://cmtt.pangaea.de/PangaVista?query=CMTT (cf. Dittert et al. 2001).

A.1.5 Conclusion

The amount of publications in natural sciences doubles about every 18 months, the estimated growth rate in related scientific data is even higher. This evolution is accelerated by technological progress as well as by growing public interest. Since printing of scientific data is economically no more acceptable, the context (scientific unit) gets lost more easily. Moreover, binding data standards are (if at all) poorly developed and rather confusingly established and global change science, however, requires a good availability of enormous amounts of analytical data (e.g., Alverson et al. 2001). Respecting the international WDC standard, WDC-MARE/PANGAEA accompanied CMTT during the final synthesis phase to ensure:

(a) Philosophy of consistent data sets
(b) Geo-coding of analytical data in a RDBMS environment
(c) Quality check of data according to existing meta-standards
(d) Mutual effect of the unit CMTT scientific community – WDC – Publisher
(e) Maximum of information exchange among participants

A.2 Introduction to SeaWiFS/MODIS Chlorophyll Data Products and Data Analysis Tools[2]

Chuanmin Hu, I.-I. Lin and Chun-Chi Lien

A.2.1 Introduction

Data products of chlorophyll-a concentration [CHL, $\mathrm{mg\,m^{-3}}$] in the surface layer of the global ocean have been obtained for the period of September 1997 to December 2002 from the Distributed Active Archive Center (DAAC) of the National Aeronautics and Space Administration (NASA) of the United States. These data are based on the measurements from the Sea-viewing Wide Field-of-View Sensor (SeawiFS, Hooker 1992; McClain et al. 1998) onboard the Orbview-II satellite, launched in August 1997 (property of Orbimage Corp., now GeoEye). The computer files of the data products and an analysis software tool are included in the distribution of this book. Other online analysis tools are also introduced here.

The software tool was developed using the Interactive Data Language (IDL). It allows a user to extract

C. Hu (✉)
Institute for Marine Remote Sensing (IMaRS), College of Marine Science, Univ. of South Florida, 140 7th Ave. South, St Petersburg, FL 33701, USA
e-mail: hu@marine.usf.edu

[2]This work was supported by NASA (grant NAG5-10738). SeaWiFS data are property of Orbimage Corp. (now GeoEye), and their use here is in accordance with the SeaWiFS Research Data Use Terms and Conditions Agreement of the NASA SeaWiFS project. Additional support was provided by the National Center for Ocean Research (NCOR) in Taipei, Taiwan. An unabridged version of this article is available online. (See Appendix C.6.)

time-series of CHL at any oceanic location(s), specified by latitude(s), and longitude(s) or by a predefined area. A brief description of the SeaWiFS CHL data and the data extraction tool are provided below. The purpose of this introduction is to provide a general overview of the existing global CHL data products to a general audience.

This service was provided in early 2000s, when online services for ocean color products were limited and global CHL data products from the MODerate resolution Imaging Spectroradiometer (MODIS) were not available. Since then, significant progress has been made in data-processing algorithms and online tools for ocean color. At the end of this introduction, we will also briefly describe the MODIS data products and online tools.

A.2.2 SeaWiFS CHL

Details of the SeaWiFS mission and its various aspects from sensor design, algorithm, and implementation, to calibration and validation, can be found in the literature as well as in the SeaWiFS Prelaunch and Postlaunch Technical Report Series (http://seawifs.gsfc.nasa.gov/). For clarity this background is briefly summarized here.

The SeaWiFS sensor is in a sun-synchronous polar-orbit, from where it measures light intensity (radiance) exiting from the top of the atmosphere (TOA) in eight spectral bands (wavelengths) centered at 412, 443, 490, 510, 555, 670, 765, and 865 nm, respectively. The sensor design (wavelength choice, signal-to-noise ratio, etc.) was optimized for ocean color measurements. SeaWiFS provides synoptic coverage of the global ocean near noon every 2 days at high resolution (1.1 km at nadir, i.e., for a location directly beneath the satellite).

The color (spectral reflectance or water-leaving radiance, see below) of the ocean is determined by the various water constituents. In waters where the major optical constituent is phytoplankton, the phytoplankton pigment for photosynthesis, namely chlorophyll-a, strongly absorbs the blue light (peaks around 440 nm). Therefore, in principle, the color measured from space can be used to infer CHL in marine waters. In practice, there are three general "steps" to convert the sensor signal to CHL.

First, the sensor signal is radiometrically and vicariously calibrated to a total radiance, $Lt(\lambda)$, where L is radiance and λ is wavelength (for brevity the wavelength dependency is omitted hereafter). This conversion includes (1) correction for various factors that affect the sensor signal such as temperature, (2) correction of sensor degradation, and (3) adjustment of the resulting signal to a modeled total radiance from concurrent field measurements obtained using a marine optical buoy (MOBY, Clark et al. 2001). The MOBY signal is propagated to the TOA according to an atmospheric correction model (Gordon and Wang 1994). This process is called a "vicarious calibration" (Gordon 1987).

Second, water-leaving radiance (L_w, radiance that exits from the ocean as measured by an imaginary sensor just above the ocean surface) is obtained with the same atmospheric correction model as used in the vicarious calibration (Gordon and Wang 1994; Ding and Gordon 1995). This correction is based on the conceptual relationship: $L_t = L_A + tL_w$, where L_A is radiance from the atmosphere that also includes the light reflected by the surface (Fresnel reflection). The term t is the diffuse atmospheric transmittance from the surface to the satellite sensor. The separation of the ocean and atmosphere terms assumes that L_w is sufficiently small where the two components can be decoupled.

For clear waters the sensor signal in the near-infrared (NIR, i.e., in the 765 and 865 nm SeaWiFS bands) can reasonably be assumed to come from the atmosphere alone. In other wavelengths, L_A is derived from L_A(NIR). L_w in all wavelengths is then obtained. For turbid coastal waters, the assumption of L_w(NIR) ≈ 0 often does not hold true. In these cases, there are several alternatives to processing the SeaWiFS data (Arnone et al. 1998; Hu et al. 2000; Ruddick et al. 2000; Siegel et al. 2000; Stumpf et al. 2003). In the SeaWiFS version 3 data (used in this book) the approach of Siegel et al. (2000) was used. This models L_w(NIR) from a certain prescribed CHL, and the proper atmospheric properties and L_w are derived iteratively to later compute a final CHL. This algorithm results in uncertainties in the CHL estimated in coastal and marine waters around the globe where the optical properties are not dominated by phytoplankton (Sathyendranath 2000).

Finally, from the spectral L_w, CHL is estimated using a bio-optical algorithm. There are several options to conduct this step, which include algorithms based on purely empirical regression analyses (e.g., O'Reilly et al. 2000; Kahru and Mitchell 2001) as well as semi-analytical approaches (e.g., Carder et al. 1999; Maritorena et al. 2002). The algorithm used to derive the SeaWiFS version 3 data is an empirical algorithm, based on thousands of field measurements that relate the ratio of L_w between two bands to CHL. The two bands used in the OC4v4 algorithm (O'Reilly et al. 2000) are 555 nm as the denominator, and one of the three bands (443, 490, and 510 nm) that has the maximum L_w value as the numerator.

These steps are carried out for each satellite pass and for each pixel with the SeaWiFS Data Analysis System (SeaDAS), developed by NASA's Goddard Space Flight Center. The calibrated, atmospherically corrected water-leaving radiances, CHL, and other products of the process outlined above are usually considered to be the "Level-2" products. Here, information from each geographical location is stored without re-projection.

The SeaWiFS data obtained from the NASA DAAC and which are used and distributed with this book are the Standard Mapped Image (SMI) data products (Level-3). These Level-3 data were created by spatially and temporally "binning" the Level-2 data (Campbell et al. 1995). A fixed global grid of equal area bins of approximate $9 \times 9\,\text{km}^2$ is first created. For each pixel of the grid, all valid Level-2 data within a predefined period (1 day, 1 month, etc.) are used to derive the statistics (mean, standard deviation, median, etc.). Because CHL is largely log-normally distributed, the logarithm of CHL data was used to derive statistics, and then transformed back for reporting in units of milligrams per cubic meter [$\text{mg}\,\text{m}^{-3}$] (Campbell et al. 1995). The data used in the book are the global monthly mean CHL data.

Post-launch validation studies (McClain et al. 2000, 2004) have shown that for CHL ranging between 0.06 and $4.7\,\text{mg}\,\text{m}^{-3}$ the uncertainty is generally within ± 35 to $\pm 50\%$ (standard deviation) without significant bias (about 6%) for phytoplankton-dominated waters (typically found in the open ocean or coastal upwelling region). More recent effort for global validation (Gregg and Casey 2004) also showed that the RMS difference of CHL (after logarithmic transformation) is about 0.2 for most open ocean waters, corresponding to 50–60% uncertainty in CHL. Note that some (or most) of these uncertainty may not indicate an error in the satellite estimates, but rather result from the inherent difference in the two measurements: while the satellite measures about $1\,\text{km}^2$ the in situ sample is only a small point. For a particular region, the results need to be interpreted with caution (e.g., Marrari et al. 2006). For coastal waters where other constituents such as colored dissolved organic matter (CDOM) or suspended sediments dominate the optical signal, the uncertainty can be much larger (Sathyendranath 2000). Also, the SeaWiFS CHL indicates an integrated effect for the surface layer (typically down to one optical depth, which corresponds to approximately 40–60 m in the clear open ocean and shallower in coastal waters) where the weight of each depth decreases exponentially from the surface.

A.2.3 MODIS CHL

The MODIS instruments (Esaias et al. 1998) were onboard the US NASA satellites, Terra and Aqua, launched in 1999 and 2002, respectively. Of the 36 spectral bands, 9 were customized for the ocean, with higher sensitivity and more digitization bits than SeaWiFS. The wavelengths of the nine spectral bands, including their center wavelengths and bandwidth, are similar to SeaWiFS, except an additional narrow band at 678-nm to detect chlorophyll fluorescence. The center wavelengths of the spectral bands are 412, 443, 488, 531, 551, 667, 678, 748, and 869 nm, respectively.

The orbital characteristics and spatial resolution of the MODIS instruments are similar to those of SeaWiFS, with cross-track swath width of about 2300 km. Equatorial crossing time of Terra (descending node, i.e., from the north to the south) and Aqua (ascending node) are 10:30 am and 1:30 pm local time, respectively. Unlike SeaWiFS, MODIS instruments are not tilted, which lead to significant sun glint patterns in low latitude oceans. Although moderate sun glint can be corrected, most glint-contaminated pixels are masked to prevent retrieval of CHL due to large uncertainties in the algorithms.

The MODIS data-processing algorithms, after taking account of the sensor specifications, are nearly identical to those used for SeaWiFS. After atmospheric correction, band ratios between 443, 488, and 551 nm (OC3, O'Reilly et al. 2000) are used to derive CHL empirically.

Because of the sophisticated design of MODIS, it took the research community several years to improve the data quality for scientific use. Earlier validation results showed large uncertainties in MODIS CHL (Blondeau-Patissier et al. 2004; Darecki and Stramski 2004), but more recent comparisons between MODIS and SeaWiFS CHL showed consistent products between the two instruments (Zhang et al. 2006). Indeed, there is ongoing effort to merge the CHL products from various sensors, including SeaWiFS and MODIS (Antoine 2004; Maritorena and Siegel 2005).

The MODIS global CHL data products, in the same format as those for SeaWiFS (SMI Level-3), are also available at NASA GSFC (http://oceancolor.gsfc.nasa.gov) at both 9-km and 4-km resolutions. To date, however, only data products from MODIS/Aqua are deemed as science quality, while MODIS/Terra products are still provisional.

A.2.4 IDL Analysis Tool

An IDL program has been developed to facilitate data analysis based on the SeaWiFS Level-3 CHL Standard Mapped Image data (SMI in HDF format). The program is used primarily to extract time-series data at one or at multiple oceanic locations. The tool, however, can also be used to display the data in graphical or imagery format.

Data at the specified locations are extracted from each CHL file, and a time-series is thereby derived. The CHL data occasionally show image "speckling" (e.g., Hu et al. 2001), in which the CHL at individual pixels varies significantly from the neighboring pixels. Therefore, for each location, the median value of the surrounding 3×3 or 5×5 pixels can be used to obtain more robust statistics. If multiple locations are specified, statistics of these locations (mean, standard deviation, etc.) for each CHL file are also derived. For multi-year time-series, a climatology may also be derived.

The results are stored in computer files in both ASCII and graphical formats. For each CHL file, an image (based on a pre-defined color lookup table) in rectangular projection is also generated and stored. Some sample results are provided in the electronic supplement (See Appendix C.6).

A.2.5 GIOVANNI Online Analysis Tool

Funded by the U.S. NASA to facilitate data sharing and distribution, the GES DISC DAAC recently developed an ocean color time-series online visualization and analysis system, based on the GES-DISC Interactive Online Visualization and ANalysis Infrastructure (GIOVANNI, http://reason.gsfc.nasa.gov/Giovanni/). The web-based interface allows a user to specify the time span and region of interest to generate time-series results in both ASCII and graphical formats. The system currently supports 9-km SeaWiFS data products only, but in the future will support MODIS and other higher-resolution products. Additionally, the NASA Ocean Biology Processing Group (OBPG) has generated monthly mean, climatology, and anomaly products that are ready to download for regional and global studies (http://oceancolor.gsfc.nasa.gov).