Deep-Sea Research II 51 (2004) 1053–1062

DEEP-SEA RESEARCH
PART II

www.elsevier.com/locate/dsr2

Deriving four parameters from patchy observations of ocean color for testing a plankton ecosystem model

Cheng-Chien Liu[a,*,1], John Woods[b]

[a]*Department of Earth Sciences, National Cheng Kung University, No. 1, Ta-Hsueh Road, Tainan, Taiwan 701, ROC*
[b]*Department of Earth Science and Engineering, Imperial College, London SW7 2AZ, UK*

Received 15 April 2002; accepted 8 October 2003
Available online 23 August 2004

Abstract

The success of ocean-color observations from space has raised interests in the application of this technology to ecosystem models. We investigated the time series of SeaWiFS data near the Azores during the first half of 1998, with the intention of testing a plankton ecosystem model. Results show that the data are very patchy, due to mesoscale variability and cloud mask gaps. The general approach of taking averages of these patchy observations introduces a bias and provides only limited information. Based on our knowledge and experience with spring blooms for characteristic subdivisions of the world's oceans, we propose a new approach: apply a four-parameter Gaussian curve fit to the gappy time series at each grid, and then extract for four parameters. These are the background chlorophyll, the increase in chlorophyll at the height of the bloom, the timing of the maximum, and a width parameter related to the duration of the bloom. Histograms of these parameters can then be compared to similar measures derived from an ecosystem model. This provides an effective way to test the model and compare its bloom dynamics with that of the satellite observations.

1. Introduction

Satellite ocean-color observations have been comprehensively employed in estimating global primary production (Longhurst, 1995; Platt et al., 1995). Recent international programs, such as the International Geosphere Biosphere Program and the Joint Global Ocean Flux Study, all emphasize the importance of deriving near-surface pigment fields from ocean-color data (Bricaud et al., 1999), in particular: (1) for initializing and validating numerical models of ecosystems and biogeochemical processes, (2) for calculating primary production from regional to global scales, and (3) when used in synergy with other remote-sensing

*Corresponding author. Tel.: +886-6-2757575x65422; fax: +886-6-2740285.

E-mail addresses: ccliu88@mail.ncku.edu.tw (C.-C. Liu), j.woods@ic.ac.uk (J. Woods).

[1]Also at Satellite Geoinformatics Research Centre, National Cheng Kung University, No. 1, Ta-Hsueh Road, Tainan, Taiwan 701, ROC.

0967-0645/$ - see front matter
doi:10.1016/j.dsr2.2003.10.007

techniques for forcing circulation fields in upper-ocean numerical models of biogeochemical processes. This provides a motivation for investigating ways of applying satellite ocean-color observations to test the plankton ecosystem model.

The general approach of using satellite ocean-color observations to test plankton ecosystem models is based on spatially and temporally composite maps. For example, Sarmiento et al. (1993) took the average chlorophyll concentration of the upper few layers simulated from their model at 3-month intervals, and then compared these data to CZCS level-3 seasonally composite maps to validate their plankton ecosystem model (see Plate 2 in Sarmiento et al., 1993). This approach, however, only compares general patterns and features. The dynamics of the plankton ecosystem, such as the timing and scale of blooming, cannot be quantitatively evaluated by comparing these seasonal composite maps.

This paper attempts to develop an effective way for testing the plankton ecosystem model through the use of satellite ocean-color data. The target model for testing is the WB plankton ecosystem model (Woods and Barkmann, 1994) based on the Lagrangian Ensemble method, which was successfully employed in simulating the annual variations of the plankton ecosystem around the Azores (41 °N, 27 °W). We investigated the time series of SeaWiFS data near the Azores during the first half of 1998, and found that the data are very patchy due to mesoscale variability and cloud mask gaps. The general approach of taking averages of these patchy observations introduces a bias and provides only limited information. Based on our knowledge and experience with spring blooms for characteristic subdivisions of the world oceans, we propose a new approach: apply a four-parameter Gaussian curve fit to the gappy time series at each grid, and then extract for four parameters. These are the background chlorophyll, the increase in chlorophyll at the height of the bloom, the timing of the maximum, and a width parameter related to the duration of the bloom. Histograms of these parameters can then be compared to similar measures derived from an ecosystem model. This provides an effective way to test the model and compare its bloom dynamics with that of the satellite observations.

2. Patchy observations

This research focuses on a site (41 °N, 27 °W) near the Azores. It is a place where the yearly heat budget is just balanced, and the slow ocean current near the island provides a water column similar to that used in the WB plankton ecosystem model. A systematic study of the physical and biogeochemical processes, as well as the time series of observations, near the Azores was conducted recently (Parrilla et al., 2002a, b). Considering the capability we now have for providing a two-dimensional synoptic view with high spatial resolution and making low-frequency observations over long periods of time, however, the satellite ocean-color observation still probably provides the most promising data for testing the plankton ecosystem model.

Fig. 1 shows the time series of SeaWiFS level-2 satellite-derived chlorophyll-*a* concentration, collected within 1 ° of latitude and longitude centered at (41 °N, 27 °W) during the first half of 1998. Note that the original SeaWiFS data were reprocessed by use of the version 2 of OC2 algorithm at the time of writing (McClain et al., 1998). The data for each pixel were examined using 16 algorithms, including land or cloud mask, tilt state of sensor, aerosol or chlorophyll algorithm error, stray light, and shallow water in order to reject suspicious values. A diamond symbol indicates that the quality test failed, but processing of the data was continued.

The first remarkable feature of this figure is that most of the observations were either masked by cloud or failed the quality test. Generally speaking, only one-fifth of the data collected are valid. This ratio is even lower during the bloom season. The operational orbit and swath of SeaWiFS are designed to cover the entire surface of the earth every 2 days. However, the weather conditions around the Azores are not stable during springtime, and being frequently masked by clouds. As a result, the time-sampling rate decreases, resulting in patchy observation data.

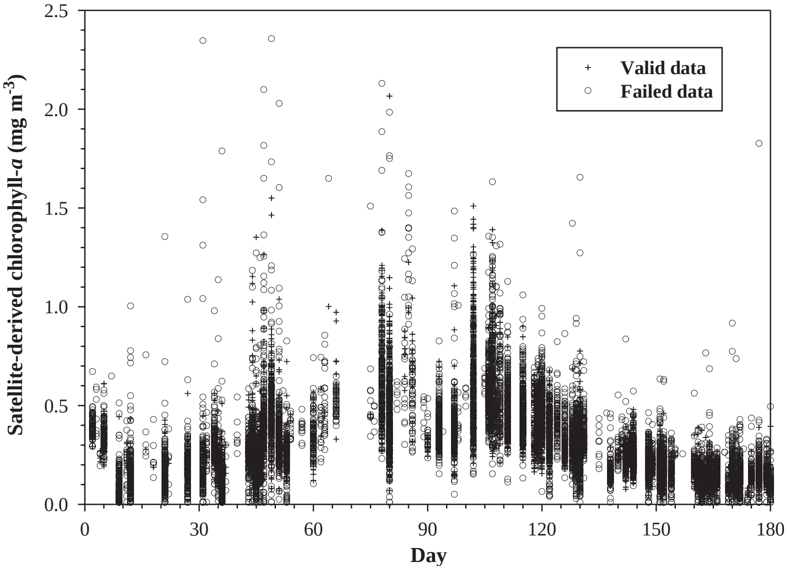

Fig. 1. Time series of SeaWiFS level-2 satellite-derived chlorophyll-a, collected within 1° of latitude and longitude centered at the Azores (41°N 27°W) during the first half of 1998.

The second remarkable feature of Fig. 1 is that high variability can be found on the same day within a small area. Generally speaking, the closer the time to spring bloom, the higher the variability. For example, on day 80, the satellite-derived chlorophyll-*a* ranged from 0.07 to 2.07 ($\mathrm{mg\,m^{-3}}$), while on day 102, it ranged from 0.21 to 1.51 ($\mathrm{mg\,m^{-3}}$). The nature of the marine ecological system is highly dynamic, both temporally and spatially. It is therefore not unreasonable to expect observations with such high variability. In spite of the aforementioned features, however, the patchy observations made by SeaWiFS still reveal the phenomenon of a spring bloom at the Azores during the first half of 1998. The question is how to extract information from such patchy observations, for the purpose of testing a plankton ecosystem model.

3. Geometric average

A study was conducted in 1992–93 by NASA to address statistical questions related to binning algorithms for level-3 data in terms of averaging SeaWiFS data spatially and temporally. Three estimators were considered: the arithmetic mean,

$$\bar{X}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{1}$$

the geometric mean,

$$\bar{X}_{\mathrm{geom}} = e^{m_x}, \tag{2}$$

and the maximum likelihood,

$$\bar{X}_{\mathrm{mle}} = e^{(m_x + (1/2)s_x^2)}, \tag{3}$$

were compared using CZCS data and from the Shelf Edge Exchange Program II moored fluorometer data (Campbell et al., 1995). Here m_x is the sample mean of log-transformed data,

$$m_x = \frac{1}{n}\sum_{i=1}^{n} \ln(X_i), \tag{4}$$

and s_x^2 is the standard variance given by

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} [\ln(X_i) - m_x]^2. \tag{5}$$

It was concluded that $\bar{X}_{\mathrm{mle}}$ is a reasonably accurate estimator for the mean of satellite-derived variables within sampling domains. This statistic approach also was recommended as the level-3 binning algorithm for averaging SeaWiFS data spatially and temporally. Note that only cloud-free scenes were used in their study, and moored fluorometer data provided a continuous sample in the time domain. Neither of the cases has proven to be valid in current processing of SeaWiFS observations. Nonetheless, we took this approach to process the SeaWiFS observations, with the intention of extracting information for testing a plankton ecosystem model.

The time series of the recommended $\bar{X}_{\mathrm{mle}}$ is plotted against all valid pixels in Fig. 2, which indeed shows the trend of in and out of bloom. However, suspicious values are frequently found during the blooming period. For example, among a total of 694 pixels collected on day 64, only one comparatively high value of 1.001 ($\mathrm{mg\,m^{-3}}$) is valid. It introduces a non-negligible bias. Likewise, another peak value that occurred on day 85 is actually calculated from merely three valid pixels. Because only one-fifth of SeaWiFS data points are valid, the sample mean derived from these small samples might be a poor estimator of the true population mean. In other words, statistical averages of patchy SeaWiFS observations may not represent the true population mean. For the purpose of testing a plankton ecosystem model, it is difficult to derive reliable information from these under-sampled observations.

4. Generalized pattern of spring bloom

Platt et al. (1988) proposed a generalized pigment profile for assessing the errors in the estimation of primary production by remote sensing due to non-uniformity in the biomass profiles. Basically, their profile is a four-parameter Gaussian function (see Eq. (14) in their paper). They claimed this profile is sufficiently versatile to mimic a large variety of profiles from coastal, upwelling, open ocean and Arctic waters, as long as the profiles contain a single peak (Platt et al., 1988). The site we focused on, the Azores, is located at mid-latitudes in the eastern North Atlantic region. According to the definition of characteristic subdivisions of the world ocean (Platt and Sathyendranath, 1988; Longhurst, 1995), it can be categorized as falling in the *westerly winds domain*, where the pattern of chlorophyll accumulation follows the classical pattern with a spring peak and a subsidiary peak in the fall. Analogous to the vertical pigment profile, the time series of satellite-derived chlorophyll-*a* during the spring bloom also follows a similar distribution with a single peak. Therefore, it leads us to extend the generalized pigment profile as the following expression:

$$\mathrm{Chl}a(t) = \mathrm{Chl}a_0 + \frac{h}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(t-d_{\max})^2}{2\sigma^2}\right] \quad (6)$$

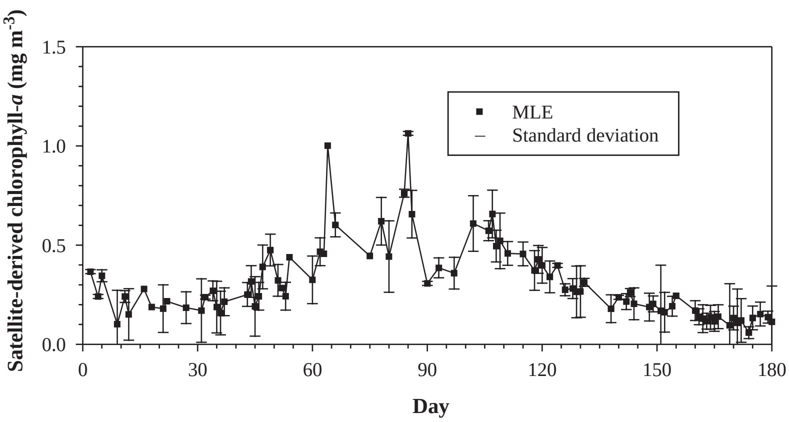

Fig. 2. The maximum likelihood estimator of the time series of the SeaWiFS level-2 satellite-derived chlorophyll-a, collected within 1° of latitude and longitude centered at the Azores (41°N 27°W) during the first half of 1998.

to describe the time series of satellite-derived chlorophyll-a during the springtime.

Although the formulation of (Eq. (6)) is the same as the generalized pigment profile (Platt et al., 1988), each parameter is given a new meaning. $\text{Chl}a_0$ ($\text{mg}\,\text{m}^{-3}$) is the background value of satellite-derived chlorophyll-a that is never less than 0. σ (day) defines the width of the peak. h ($\text{mg}\,\text{m}^{-3}$ day) is the integral of total satellite-derived chlorophyll-a above the background. $\text{Chl}a_0 + h/(\sigma\sqrt{2\pi})$ gives the height of the peak, and $d_{\max}$ (day) is the day when the peak occurs. Implementation of Eq. (6) can be imagined as opening a Eulerian window at the site we considered, then utilizing all available observations within this window to fit the four-parameter Gaussian distribution. The set of parameters is determined by iteration to achieve the least-square error between the predictions and observations. The size of window can be selected to be as small as the size of a single pixel.

At least two advantages can be obtained through this approach of fitting the satellite ocean-color observations with a four-parameter Gaussian curve. First, the scattered observations, both spatially and temporally, can be refined into four meaningful parameters for testing the plankton ecosystem models. Second, all valid data collected are used to test the generalized pattern of spring bloom and to derive the required parameters for the pattern, rather than to derive the pattern itself. To have a theoretical pattern based on our previous knowledge and experience of spring blooms, and then to validate this pattern with all observed data, such an approach can compensate for a lack of observations.

It is better to fit the pattern of spring bloom through the use of the time series of observations made from the smallest Eulerian window, i.e. the single pixel at the same location. However, due to the progression of satellite orbits, the SeaWiFS position is slightly different for each day's overpass. Nadir pixels observed today might fall on the edge of the swath during the next observation. Consequently, the size and position of the observing window cannot be considered constant over time. As a result, the valid level-2 data as shown in Fig. 1 cannot be used directly to fit the pattern. One way of addressing the above problem is to introduce some kind of mapping scheme. The level-3 data (another SeaWiFS data product) are simply derived by mapping level-2 data onto a fixed global grid whose resolution element is approximately $9\times 9\,\text{km}^2$. The observation windows can be selected at each fixed grid such that the time series of level-3 data can be used to fit the generalized pattern of the spring bloom.

5. Results and discussion

In total there are 10×10 grids within 1° of latitude and longitude centered at the Azores (41°N 27°W). Because the pattern is aimed at describing the spring bloom signal, the time series of data processing are selected to be from days 30 to 181. For each grid, the generalized pattern of spring bloom Eq. (6) is applied to fit the time series for all valid data. The quality of each curve fitting can be assessed by calculating the Pearson correlation coefficient R and the variance explanation VE (R^2). The standard deviation SD of chlorophyll-a is also shown in Fig. 3. Among these 100 cases, most of the results can be illustrated by Fig. 3(a) and (b), which give very high variance explanations of 90% and 87%, respectively. In a few cases, the four-parameter Gaussian curve seems to be too simple to describe the observations. For example, in those distributions with two peaks, such as Fig. 3(c), the current curve fit tends to miss the actual peaks and generates higher deviations. Another example can be seen in Fig. 3(d), where the blooming occurs within a shorter period of time. However, a few cases suggest that the variance explanation could be improved by using a more complex formulation, such as a double-peak distribution or a non-Gaussian distribution with additional parameters for expressing kurtosis and/or skewness. Results from this research show that the four-parameter Gaussian curve fit of spring bloom offers a fairly good fit to first order. In the aforementioned examples, 68% and 75% of variance is explained respectively for Fig. 3 (c) and (d). The worst result of curve fitting is found at the grid (41.35 °N, 27.45 °W), where the value of variance explanation

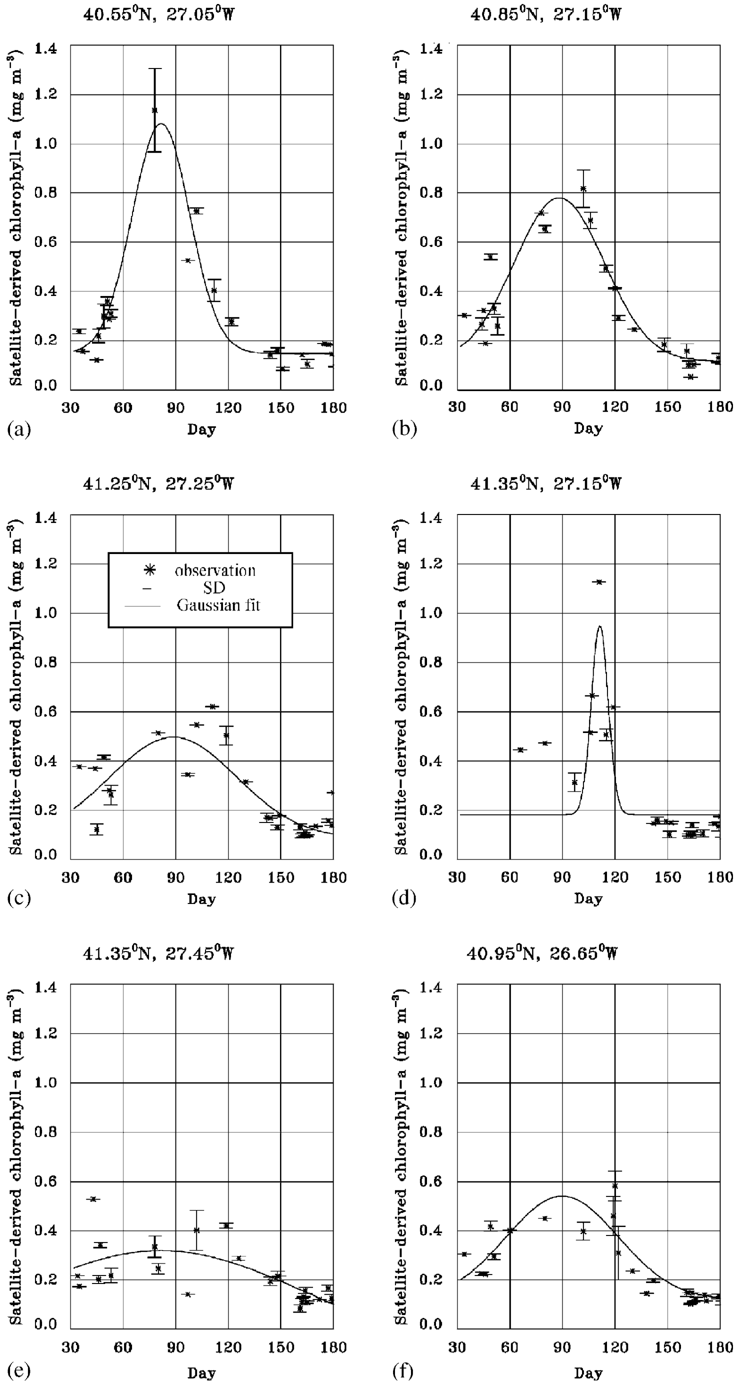

Fig. 3. Examples of fitting the SeaWiFS level-3 data with a spring bloom pattern of four-parameter Gaussian distribution.

is only 44%. This result can be explained by examining the data plotted in Fig. 3(e). Except for a higher value detected on day 43, all data collected in that grid have relatively low values. No clear signal of blooming is found and the fitted curve is fairly flat. This result suggests that either the observation misses the bloom signal, or the relatively low signal is a typical phenomenon at this particular position. Another possibility is that the fitted curve is an accurate description of this particular data set, but that the pattern it shows is not typical. This cannot be concluded based on only 1 year of observation. Another interesting result can be seen in Fig. 3(f). Overall, the time series of observations in that grid fits the pattern very well, except for some abrupt high values detected within a very short period of time. This anomaly may result from blue-absorbing Saharan dust being present (e.g. see Evans et al., 2000), or is an example of patchiness caused by mesoscale upwelling.

One way of expressing the uncertainty of these parameters is to plot their histograms (Fig. 4). Each parameter can be used as an independent index for testing the plankton ecosystem model. In addition, the ranges of uncertainty for these parameters can be described clearly. It should be noted that to gain a clearer idea of the peak of the spring bloom pattern, the last figure gives the maximum chlorophyll-*a* $\mathrm{Chl}a_{\max}$, which is calculated by $\mathrm{Chl}a_{\max} = \mathrm{Chl}a_0 + h/(\sigma\sqrt{2\pi})$ when $t = d_{\max}$. These histograms can be employed to evaluate a plankton ecosystem model. This research employed the WB plankton ecosystem model (Woods and Barkmann, 1994) as an example. The same four-parameter Gaussian function was first applied to fit the time series of surface chlorophyll concentration simulated from the WB model. The derived indices $\mathrm{Chl}a_0$, σ, h and $d_{\max}$ were then plotted onto the correspondent histogram in Fig. 4. Compared with SeaWiFS observations, the spring bloom predicted by the WB model was $1\frac{1}{2}$ months late, lasted one-third of the time and had a peak chlorophyll concentration 10 times higher (Table 1). The goal of quantitative comparison between model predictions and satellite observations is therefore achieved by comparing these indices. A detailed description of the comparison can be found in (Liu, 2000).

Inevitably, using level-3 data not only reduce the resolution, but also introduce a basic statistical processing in the first place; the level-3 products are derived by taking statistical means and mapping level-2 data onto a fixed global grid with resolution $9 \times 9\,\mathrm{km}^2$. The essence of this research is to have a theoretical pattern based on our knowledge and experience with the plankton ecosystem, and then utilize all available data to determine the parameters of the pattern, without any statistical analysis. However, as mentioned previously, level-2 data cannot provide continuous observations at a fixed location. One concern is whether using level-3 rather than level-2 data gives a different result for testing a plankton ecosystem model. Fig. 5 shows the result of fitting all valid level-2 data, as shown in Fig. 1, with one four-parameter Gaussian curve. This implies that all data collected within 1 $^\circ$ of latitude and longitude centered at the Azores (41°N 27°W) are equally weighted. Results shown in Table 1 indicate that all four parameters, derived from Fig. 5, have approximately the same value as the median values expressed in Fig. 4. Therefore, it can be concluded that using either level-2 or level-3 data provides similar results; however, level-3 data retain more information at the scale of global grid ($9 \times 9\,\mathrm{km}^2$) than level-2 data (one degree of latitude and longitude).

The satellite-derived chlorophyll-*a* collected during the spring bloom is a good choice for evaluating the plankton ecosystem. Because the spring bloom dominates the primary annual production of the North Atlantic Ocean, it is always the focus of a lot of research interest. It is a well-understood, regular, and important phenomenon, and so has received much attention. Most of all, it is a comparatively strong and clear signal that can be detected remotely. It should be noted that the four-parameter Gaussian distribution is not the only pattern for describing the temporal aspects of SeaWiFS-derived chlorophyll-*a* for spring blooms. Results from this research suggest that in some cases the variance explanation could be improved by using a more complex formulation, such as a double-peak distribution or a non-Gaussian distribution with additional

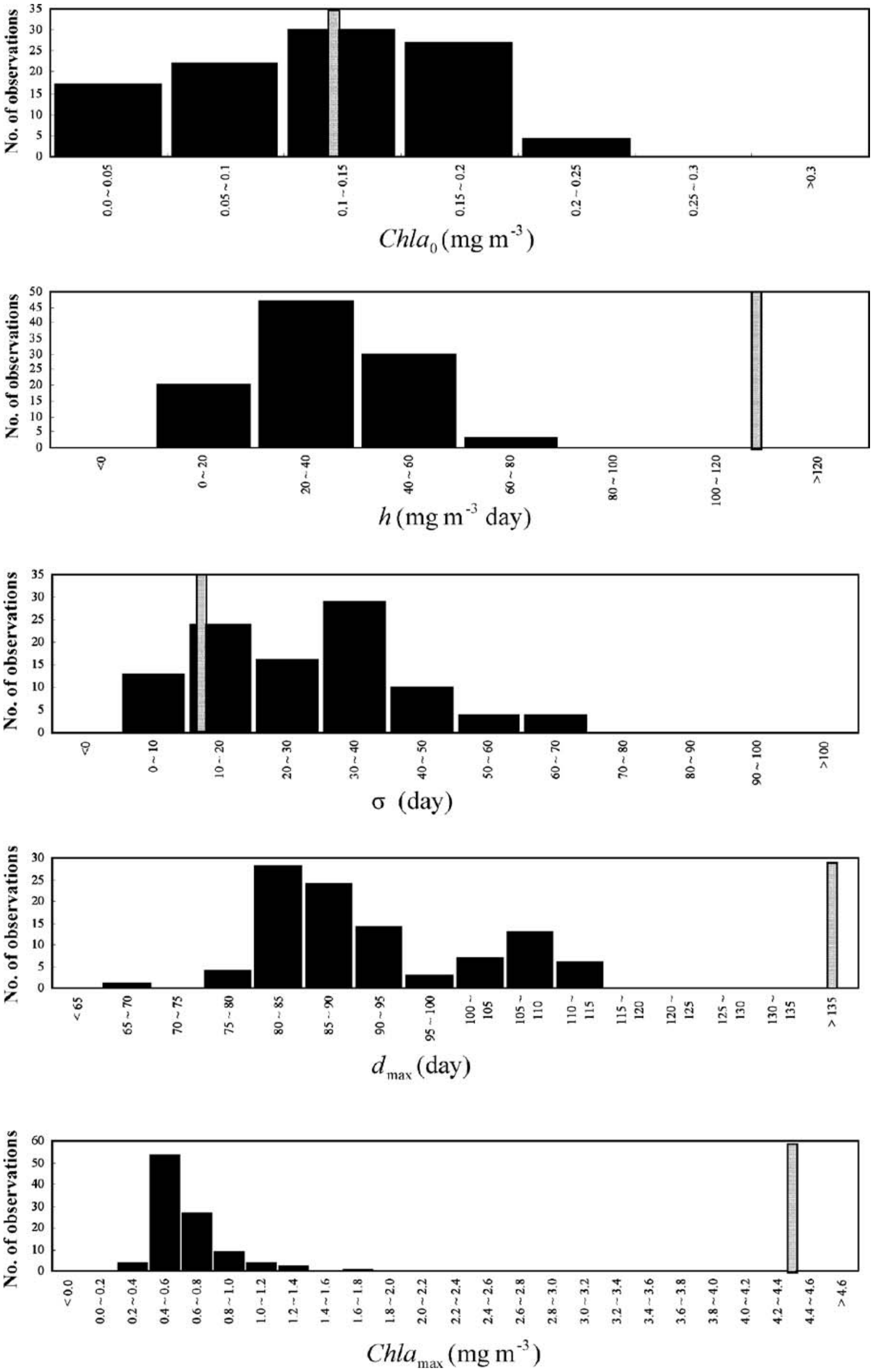

Fig. 4. Histograms of four parameters derived by fitting the time series of SeaWiFS level-3 data with a four-parameter Gaussian model. The similar measures derived from the WB model are plotted in diagonal-filled lines.

Table 1
Comparison of four parameters (Chla_0, h, σ and d_{max}) derived by fitting a four-parameter Gaussian curve to various time series of chlorophyll-a, including (a) WB model simulation, (b) SeaWiFS level-3 data, and (c) all SeaWiFS level-2 data collected within 1° of latitude and longitude centered at the Azores (41 °N, 27 °W)

	Chla_0	h	σ	d_{max}	R	VE
(a) WB model simulation	0.12	121.77	11.12	136.01	0.99	98.53
(b) SeaWiFS observation (level-3)	0.12	34.01	28.99	87.81	0.85	72.98
(c) SeaWiFS observation (level-2)	0.10	37.56	33.20	88.70	0.76	58.31

Note that (b) shows only the median values of the results shown in Fig. 4.

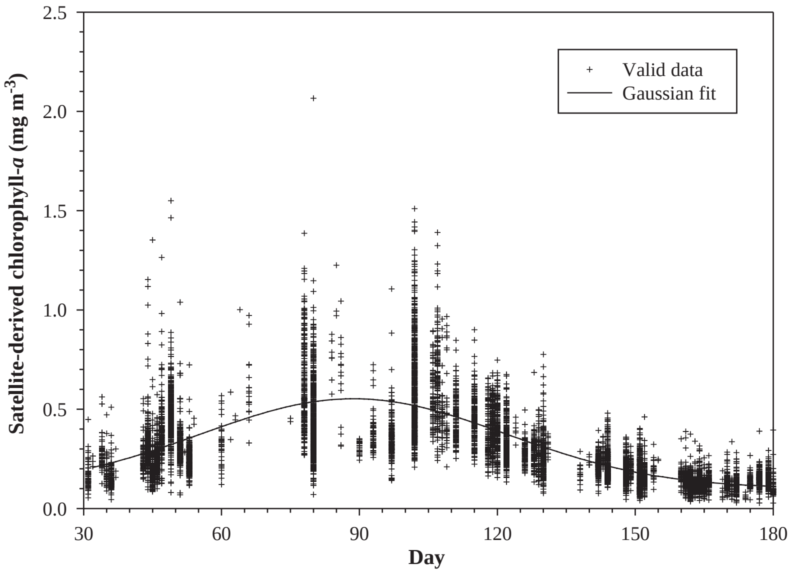

Fig. 5. Results of fitting all SeaWiFS level-2 data collected within 1° of latitude and longitude centered at the Azores (41 °N, 27 °W) with a four-parameter Gaussian curve.

parameters for expressing kurtosis and/or skewness. Nevertheless, this research also shows that current four-parameter Gaussian pattern offers a fairly good fit to first order. For the purpose of evaluating a plankton ecosystem model, this research provides a practical approach that can be treated as a good starting point. Other forms of spring bloom patterns might be of help in improving the model-data fit and providing more parameters about the characteristics of spring bloom, but uncertainties in the accuracy of retrievals after Saharan-dust events over the region suggest that the data may be too noisy for more specificity.

Another issue when using different spring bloom patterns is related to the concept of 'biogeochemical provinces'. It is well known that marine equivalents of forest, tundra and grassland can be identified, based on the seasonal patterns of primary production and chlorophyll accumulation (Platt and Sathyendranath, 1988; Longhurst, 1995). This research suggests that different temporal functions of satellite-derived chlorophyll-a for different biogeochemical provinces can be specified as well, based on our knowledge of the pelagic ecosystem. For those sites where continuous satellite ocean-color data are available, observations can be used in examining and deriving a theoretical pattern. For those sites where satellite observations are limited, modeling is able to compensate for lack of observations, if the model is an accurate simulator of surface chlorophyll fields with time. In both cases, satellite ocean-color observations can be refined into some indices, which are very useful in quantitatively testing the plankton ecosystem model.

Acknowledgements

This research was supported by the UK Natural Environment Research Council, an Overseas Research Student award from the UK committee of vice-chancellors and principals, a research studentship from the Ministry of Education, the National Science Council of Taiwan (NSC 92-2611-M-006-002), and NCKU-NSYSU Research Center of Ocean Environment and Technology. Ocean-color data used in this study were produced by the SeaWiFS Project at Goddard Space Flight Center. The data were obtained from the Goddard Distributed Active Archive Center under the auspices of the National Aeronautics and Space Administration. Use of these data is in accord with the SeaWiFS Research Data Use Terms and Conditions Agreement.

References

Bricaud, A., Morel, A., Barale, V., 1999. MERIS potential for ocean colour studies in the open ocean. International Journal of Remote Sensing 20 (9), 1757–1769.

Campbell, J.W., Blaisdell, J.M., Darzi, M., 1995. Level-3 SeaWiFS Data Products: Spatial and Temporal Binning Algorithms. NASA Tech. Memo. 104566, NASA Goddard Space Flight Center, Greenbelt, MD, USA.

Evans, R., Kilpatrick, K., Kearns, E., 2000. Evaluation of the spectral sensitivity and performance of MODIS and pathfinder sea surface temperature algorithms in the presence of aerosols. In: Ackleson, S., Marra, J. (Eds.), Ocean Optics XV. Proc. SPIE (CDROM), Monaco.

Liu, C.-C., 2000. Using SeaWiFS ocean colour data to test a plankton ecosystem model. Ph.D. Thesis, Imperial College, University of London, London, 232 pp.

Longhurst, A., 1995. An estimate of global primary production in the ocean from satellite radiometer data. Journal of Plankton Research 17, 1245–1271.

McClain, C.R., Cleave, M.L., Feldman, G.C., Gregg, W.W., Hooker, S.B., Kuring, N., 1998. Science quality SeaWiFS data for global biosphere research. Sea Technology 39 (9), 10–16.

Parrilla, G., Neuer, S., Le Traon, P.-Y., Fernandez-Suarez, E., 2002a. Topical studies in oceanography: Canary Islands Azores Gibraltar observations (CANIGO). Vol. 2: studies of the Azores and Gibraltar regions. Deep-Sea Research II 49(19), 3951–3955.

Parrilla, G., Neuer, S., Le Traon, P.-Y., Fernandez-Suarez, E., 2002b. Topical studies in oceanography: Canary Islands Azores Gibraltar Observations (CANIGO): Vol. 1: Studies in the northern Canary Islands basin. Deep-Sea Research II 49(17), 3409–3413.

Platt, T., Sathyendranath, S., 1988. Oceanic primary production: estimation by remote sensing at local and regional scales. Science 241 (4873), 1613–1620.

Platt, T., Sathyendranath, S., Caverhill, C.M., Lewis, M.R., 1988. Ocean primary production and available light: further algorithms for remote sensing. Deep-Sea Research 35 (6), 855–879.

Platt, T., Sathyendranath, S., Longhurst, A., 1995. Remote sensing of primary production in the ocean: promise and fulfillment. Philosophical Transactions of the Royal Society B 348, 191–202.

Sarmiento, J.L., Slater, R.D., Fasham, M.J.R., Ducklow, H.W., Toggweiler, J.R., Evans, G.T., 1993. A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic euphotic zone. Global Biogeochemical Cycles 7 (2), 417–450.

Woods, J.D., Barkmann, W., 1994. Simulating plankton ecosystems by the Lagrangian Ensemble method. Philosophical Transactions of the Royal Society B (343), 27–31.