GLOBAL BIOGEOCHEMICAL CYCLES, VOL. 16, NO. 3

Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance

Janet Campbell,[1] David Antoine,[2] Robert Armstrong,[3] Kevin Arrigo,[4] William Balch,[5] Richard Barber,[6] Michael Behrenfeld,[7] Robert Bidigare,[8] James Bishop,[9] Mary-Elena Carr,[10] Wayne Esaias,[7] Paul Falkowski,[11] Nicolas Hoepffner,[12] Richard Iverson,[13] Dale Kiefer,[14] Steven Lohrenz,[15] John Marra,[16] Andre Morel,[2] John Ryan,[17] Vladimir Vedernikov,[18] Kirk Waters,[19] Charles Yentsch,[5] and James Yoder[20]

Received 21 May 2001; revised 8 March 2002; accepted 8 March 2002; published 17 July 2002.

[1] Results of a single-blind round-robin comparison of satellite primary productivity algorithms are presented. The goal of the round-robin exercise was to determine the accuracy of the algorithms in predicting depth-integrated primary production from information amenable to remote sensing. Twelve algorithms, developed by 10 teams, were evaluated by comparing their ability to estimate depth-integrated daily production (IP, mg C m^{-2}) at 89 stations in geographically diverse provinces. Algorithms were furnished information about the surface chlorophyll concentration, temperature, photosynthetic available radiation, latitude, longitude, and day of the year. Algorithm results were then compared with IP estimates derived from ^{14}C uptake measurements at the same stations. Estimates from the best-performing algorithms were generally within a factor of 2 of the ^{14}C-derived estimates. Many algorithms had systematic biases that can possibly be eliminated by reparameterizing underlying relationships. The performance of the algorithms and degree of correlation with each other were independent of the algorithms' complexity. *INDEX TERMS:* 4894 Oceanography: Biological and Chemical: Instruments and techniques; 4275 Oceanography: General: Remote sensing and electromagnetic processes (0689); 4805 Oceanography: Biological and Chemical: Biogeochemical cycles (1615); 4806 Oceanography: Biological and Chemical: Carbon cycling; 4853 Oceanography: Biological and Chemical: Photosynthesis; *KEYWORDS:* primary productivity, algorithms, ocean color, remote sensing, satellite, chlorophyll

1. Introduction

[2] Global maps of the upper-ocean chlorophyll concentration are now being generated routinely by satellite ocean color sensors. These multispectral sensors are able to map the chlorophyll concentration, a measure of phytoplankton biomass, by detecting spectral shifts in upwelling radiance. As the chlorophyll concentration increases, blue light is increasingly absorbed, and thus less is scattered back into space. Although global coverage can nominally be achieved every 1–2 days, the actual temporal resolution is reduced to ~5–10 days because of cloud cover. Nevertheless, the coverage afforded by satellite remote sensing is vastly greater than that obtainable by any other means.

[3] A principal use of the global ocean chlorophyll data is to estimate oceanic primary production [*Behrenfeld et al.*, 2001]. The mathematical models or procedures for estimating primary production from satellite data are known as primary productivity algorithms. In the early days of the Coastal Zone Color Scanner (CZCS), simple statistical relationships were proposed for calculating primary production from the surface chlorophyll concentration [e.g., *Smith and Baker*, 1978; *Eppley et al.*, 1985]. Such empirically derived algorithms are still considered useful when applied to annually averaged data [*Iverson et al.*, 2000], but they are not sufficiently accurate to estimate production at seasonal timescales. The surface chlorophyll concentration explains only ~30% of the variance in primary production at the scale of a single station [*Balch et al.*, 1992; *Campbell and O'Reilly*, 1988].

[4] Over the past 2 decades, scientists have sought to improve algorithms by combining the satellite-derived chlorophyll data with other remotely sensed fields, such as sea surface temperature (SST) and photosynthetic available radiation (PAR). These algorithms incorporate models

Table 1. Algorithm Testing Subcommittee of NASA's Ocean Primary Productivity Working Group[a]

Participant	Affiliation
Robert Armstrong	Stony Brook University
Richard T. Barber	Duke University
James Bishop	Lawrence Berkeley National Laboratory
Janet W. Campbell	University of New Hampshire
Mary-Elena Carr	Jet Propulsion Laboratory
Wayne E. Esaias	NASA Goddard Space Flight Center
Richard Iverson	Florida State University
Charles S. Yentsch	Bigelow Laboratory for Ocean Sciences

[a] These individuals were responsible for conducting the primary productivity algorithm round-robin experiment. They agreed not to participate by testing algorithms of their own.

of the photosynthetic response of phytoplankton to light, temperature, and other environmental variables, and some also incorporate models of the vertical distribution of these properties within the euphotic zone [*Balch et al.*, 1989; *Morel*, 1991; *Platt and Sathyendranath*, 1993; *Howard*, 1995; *Antoine and Morel*, 1996a; *Behrenfeld and Falkowski*, 1997a; *Ondrusek et al.*, 2001]. Algorithms have been used to estimate global oceanic primary production from CZCS data [*Antoine and Morel*, 1996b; *Longhurst et al.*, 1995; *Behrenfeld and Falkowski*, 1997a; *Howard and Yoder*, 1997], and more recently from Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data [*Behrenfeld et al.*, 2001]. Global maps of the average daily primary production for varying periods (weeks, months, and years) are now being produced from Moderate Resolution Imaging Spectroradiometer (MODIS) data.

[5] While many of the photosynthetic responses (to light, temperature, etc.) are commonly represented, model-based algorithms differ with respect to structure and computational complexity [*Behrenfeld and Falkowski*, 1997b]. Models may be similar in structure but require different parameters depending on whether they describe daily, hourly, or instantaneous production, and even where these aspects are similar, algorithms often yield different results because of differences in their parameterization. *Balch et al.* [1992] evaluated a variety of algorithms (both empirical and model based), using in situ productivity measurements from a large globally distributed data set, and found that they generally accounted for <50% of the variance in primary production.

[6] In January 1994 the National Aeronautics and Space Administration (NASA) convened an Ocean Primary Productivity Working Group with the goal of developing one or more "consensus" algorithms to be applied to satellite ocean color data. The working group initiated a series of round-robin experiments to evaluate and compare primary productivity algorithms. The approach was to use in situ data to test the ability of algorithms to predict depth-integrated daily production (IP, mg C m^{-2}) based on information amenable to remote sensing. It was decided to compare algorithm performances with one another and with estimates based on ^{14}C incubations.

[7] Our understanding of primary productivity in the ocean is largely based on the assimilation of inorganic carbon from ^{14}C techniques [*Longhurst et al.*, 1995], and thus it was considered appropriate to compare the algorithm estimates with ^{14}C-based estimates. However, it was recognized that the ^{14}C-based estimates are themselves subject to error [*Peterson*, 1980; *Fitzwater et al.*, 1982; *Richardson*, 1991]. The ^{14}C incubation technique measures photosynthetic carbon fixation within a confined volume of seawater, and there are no methods for absolute calibration of bottle incubations [*Balch*, 1997]. Furthermore, there is no universally accepted method for measuring and verifying vertically integrated production derived from discrete bottle measurements. Despite this fact, here we treat the ^{14}C-based estimates as "truth" and refer to the differences between algorithm-derived and ^{14}C-derived estimates as "errors." In all statistical analyses, however, the two are recognized as being subject to error.

[8] Participation in the round robin was solicited through a widely distributed "Dear Colleague" letter. A central ground rule was that the algorithms tested would be identified only by code numbers. The first round-robin experiment involved data from only 25 stations and was thus limited in scope. It was decided that a more comprehensive second round was needed. In this paper, we present results of the second round-robin experiment involving data from 89 stations with wide geographic coverage. Round two was open to all participants of round one, as well as to others who had responded positively to the initial invitation.

6. Conclusions

[45] Conclusions related to the four questions addresseed by this study are summarized as follows:

[46] 1. How do algorithm estimates of primary production derived strictly from surface information compare with estimates derived from ^{14}C incubation methods? The 12 algorithms tested varied widely in performance. The best-performing algorithms agreed with the ^{14}C-based estimates within a factor of 2. Two of these algorithms have been adapted by NASA for producing primary productivity maps with MODIS data. Most of the algorithms had significant biases causing them to differ systematically from the in situ data. A concerted effort should be made to understand the cause of the biases and to eliminate them if possible.

[47] 2. How does the error in satellite-derived chlorophyll concentration affect the accuracy of the primary productivity algorithms? The relative errors in primary productivity (Δ_{sat}) resulting from the simulated errors in surface chlorophyll concentration (Δ_B) were highly correlated with Δ_B. This fact reflects the deterministic relationship between production and chlorophyll in the underlying models. The slopes of the regressions (Δ_{sat} versus Δ_B) ranged between 0.3 and 0.8, indicating that errors in surface chlorophyll produce less-than-proportionate errors in IP.

[48] 3. Are there regional differences in the performance of algorithms? There were significant regional differences, as well as algorithm-region interactions, indicated by the ANOVA results. No one region was uniformly better or worse for all algorithms. The region with the most serious biases was the equatorial Pacific, where algorithms underestimated in situ measurements by a factor of 2.

[49] 4. How do algorithms compare with each other in terms of complexity vis-a-vis performance? Many of the algorithms were highly correlated with one another. This was not surprising, since several are based on the same models, but what was surprising was that the level of agreement had no apparent relationship to the mathematical structure or complexity of the algorithms. In some cases, complex algorithms based on depth-, time- and wavelength-resolved models were highly correlated with simpler algorithms that were time and/or depth integrated. There were distinct systematic differences between algorithms. A future effort to understand systematic differences is strongly recommended.