Remote Sensing of Environment 152 (2014) 15–28

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Impact of missing data on the estimation of ecological indicators from satellite ocean-colour time-series

Marie-Fanny Racault *, Shubha Sathyendranath, Trevor Platt

Plymouth Marine Laboratory, Prospect Place, The Hoe, PL1 3DH, Plymouth, United Kingdom

ARTICLE INFO

Article history:
Received 14 January 2014
Received in revised form 21 May 2014
Accepted 26 May 2014
Available online xxxx

Keywords:
CZCS
SeaWiFS
Ecological indicators
Chlorophyll-a
Phenology
Missing data
Uncertainty

ABSTRACT

Ocean-colour remote sensing provides high-resolution and global-coverage of chlorophyll concentration, which can be used to estimate ecological indicators and to study inter-annual and long-term trends in the state of the marine ecosystem. To date, the record of ocean-colour observations is a rich one, including data from a number of sensors spanning more than three decades. The ESA Ocean-Colour Climate Change Initiative has advanced seamless merging of ocean-colour observations from missions during the period 1990s to 2010s. However, comparison of these more recent observations with records from 1970s to 1980s remains a complex undertaking, particularly for absolute values of chlorophyll concentration, primarily due to differences in the sensors. A further impediment to the analysis of the past records is the non-uniform distribution of gaps in the observations, in both time and space dimensions, when data from two or more sensors are compared. Here, we use the CZCS gap distribution from the Coastal Zone Color Scanner (CZCS, 1978–1986) as a mask to evaluate the impact that missing data may have on the estimation of six ecological indicators, when using the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data set. Specifically, we evaluate the precision and accuracy of indicators by computing the root-mean-square-error (RMSE) and the bias arising purely from missing data. We develop an original resampling method allowing comparison of indicator estimates between SeaWiFS reference time-series and SeaWiFS time-series with CZCS-like gaps. We reduce some of the sampling gaps by applying a linear interpolation procedure, and compute multi-year averages of the indicators for every one-by-one degree pixel where sufficient data are available. Indicators from SeaWiFS reference and SeaWiFS with CZCS-like gaps are compared. Lowest uncertainty arising from missing data is observed in the indicators of annual mean and median chlorophyll concentration (global mean RMSE of 8% and |bias| ≤1%), whilst higher uncertainty is recorded for the peak chlorophyll values and the duration of the phytoplankton growing period (global mean RMSE of 33 and 47% respectively and |bias| ≤20%). Timing of initiation of the increasing phase of chlorophyll concentration in the seasonal cycle and timing of peak chlorophyll are subject to a global mean RMSE of nearly two months and a bias of two weeks or less. The present quantitative evaluation of uncertainty due to missing data demonstrates that, when pooled to create a nine-year climatology at 8-day temporal resolution, the coverage of CZCS is adequate for many climate-related studies on the marine ecosystem. Phytoplankton annual mean biomass can be estimated with low error in approximately 95% of the global oceans (i.e. regions where the indicators can be estimated with RMSE values of less than 30% and bias within ±10%), and the phenological patterns can be estimated with low error in approximately 25% of the global oceans.

1. Introduction

In the marine environment, ecological indicators have been developed to provide specific information relevant to the evaluation of the state of the marine ecosystem (Borja et al., 2008; Cardoso et al., 2010; Ferreira et al., 2011; Platt & Sathyendranath, 2008; Tett et al., 2013). The function of an indicator may be to depict the condition of the environment, to provide early-warning signals or to register long-term trends (Niemi & McDonald, 2004). The state of the first trophic level of the marine ecosystem can be characterised by the annual cycle of phytoplankton. In-situ or remote-sensing observations of chlorophyll concentration, a proxy for phytoplankton biomass, have been used todepict changes in the annual cycle of phytoplankton (Platt & Sathyendranath, 1996; Platt & Sathyendranath, 2008). Some indicators, for instance, the mean, median and maximum concentrations or biomasses of phytoplankton in a given year, are generally expressed in units of mass of chlorophyll or carbon per unit volume of water. Other indicators correspond to the patterns of the annual cycle of phytoplankton, and are referred to as phenology (i.e. timing of periodic events). These phenological metrics describe phases in the annual cycle, and carry units of time (e.g. days, weeks, month…). Such indicators include

* Corresponding author. Tel.: +44 175 263 34 34; fax: +44 175 263 31 01.
E-mail address: mfrt@pml.ac.uk (M.-F. Racault).

http://dx.doi.org/10.1016/j.rse.2014.05.016

the timings of initiation, peak, termination and the duration of phytoplankton growing period (blooming period) in a given season.

The most cost-efficient datasets available to implement ecological indicators are provided by ocean-colour remote sensing observations (Platt et al., 2009). These data sets have the additional advantage of having high spatial resolution, high sampling frequency and global coverage. The first satellite sensor developed specifically to study ocean-colour properties was the Coastal Zone Color Scanner (CZCS). It was launched by NASA in October 1978 and remained operational for seven and a half years, until June 1986. A decade later, the Ocean Colour and Temperature Scanner (OCTS) was launched by the Japanese Space Agency (NASDA) in November 1996 and it collected ocean colour data until June 1997. The next major satellite instrument for ocean colour was the Sea-viewing Wide Field-of-View Sensor (SeaWiFS), which functioned for more than 13 years from September 1997 until December 2010. The spacecraft and SeaWiFS were owned and operated by Orbital Sciences and subsequent commercial entities. NASA purchased the data, and was then responsible for processing, quality control, and data distribution to approved researchers. In 2002, two additional sensors began acquiring ocean-colour data: the Moderate Resolution Imaging Spectroradiometer (MODIS) launched by NASA, and the MEdium Resolution Imaging Spectrometer (MERIS) launched by the European Space Agency (ESA). MERIS ceased operations in early 2012, but MODIS is still operating, though well past its design lifespan. Further information about historical, current and scheduled ocean-colour sensors can be found on the International Ocean Colour Coordinating Group (IOCCG) website at http://www.ioccg.org/sensors_ioccg.html.

The use of data from the CZCS period could possibly allow us to extend the ocean-colour-based record of ecological indicators backwards in time to the period 1978–1986, when CZCS was operational. However, the CZCS mission was exploratory: it had limited spatial coverage and spectral bands, and it did not overlap with other ocean-colour sensors (making it difficult to correct for any potential inter-sensor bias). Because of the absence of overlapping periods, the merging of ocean-colour data such as implemented by the ESA Ocean Colour-Climate Change Initiative using SeaWiFS, MODIS and MERIS (Hollmann et al. 2013), is not possible with the CZCS. Nevertheless, a number of efforts have been made to improve the precision and accuracy of the CZCS archive and effectively compare it with ocean-colour data from follow-on missions. Gregg and Conkright (2002) re-analysed the archive by blending the CZCS ocean-colour data with in-situ chlorophyll measurements to minimise possible bias in the satellite-derived fields. In the re-analysis effort of Antoine, Morel, Gordon, Banzon, and Evans (2005), the authors revised the CZCS data processing algorithms to generate an improved, revised CZCS chlorophyll data set. Then, to allow an inter-comparison between the CZCS and SeaWiFS sensors, they applied the same revised algorithms to SeaWiFS data over the period 1998–2002. However, the regional increases and decreases in absolute values of chlorophyll shown in these two publications are not straightforward to reconcile. More generally, taking into account also the findings based on in-situ observations, the debate on multi-decadal trends in phytoplankton biomass is still open (Boyce, Lewis, & Worm, 2010; Mackas, 2011; McQuatters-Gollop et al., 2011; Raitsos et al., 2013; Rykaczewski & Dunne, 2011; Wernand, Van der Woerd, & Gieskes, 2013).

Given the unique availability of observations from the CZCS during the period 1978–1986, and the critical importance of determining long-term trends in the marine ecosystem, scrutiny is required to determine the impact of missing data in the CZCS record on the estimation of ecological indicators. The spatial and temporal coverage of remotely-sensed data is limited by sun-glint, clouds, atmospheric aerosol, sensor saturation over ice, sand or snow, and high solar zenith angle. During the exploratory mission of the CZCS sensor, the collection of observations was limited for all the reasons above, but in addition, also by power and data recorder limitations, which led to the priority being set on observations in the coastal regions and in the Northern Hemisphere. The distribution of missing data in the CZCS time-series has been evaluated at monthly resolution (Antoine et al., 2005). However, monthly resolution is not sufficient to assess inter-annual variability and trends in phytoplankton phenology, which are driven by natural or anthropogenic forcing (Chiba et al., 2008; González Taboada & Anadón, 2014; Racault, Le Quéré, Buitenhuis, Sathyendranath, & Platt, 2012; Thomalla, Fauchereau, Swart, & Monteiro, 2011).

The present study aims to: 1) evaluate the distribution of missing data in the CZCS 1978–1986 time-series at a resolution of 8-days in the global oceans; 2) perform a sensitivity analysis for assessing the error that the distribution of missing data in the CZCS time-series may have on the estimation of six ecological indicators; and 3) compare the error associated with missing data when estimating the indicators from time-series, with and without applying an interpolation scheme to fill some of the missing data.

2. Material and methods

2.1. Remotely-sensed ocean-colour data

Synoptic fields of chlorophyll concentration were retrieved for the periods 1978–1986 and 1997–2010 from NASA Ocean Color Web http://oceancolor.gsfc.nasa.gov. The R2010.0 reprocessing of Level 3 Mapped chlorophyll concentrations from both CZCS and SeaWiFS were both downloaded at 9-km spatial resolution and 8-day temporal resolution. To reduce gaps in the global oceans time-series, the data were re-gridded to $1^\circ \times 1^\circ$ boxes (Fig. 1).

2.2. Estimation of ecological indicators

The annual cycle of phytoplankton was characterised by estimating six well-established ecological indicators from remote-sensing observations of chlorophyll concentrations (Platt & Sathyendranath, 1996; Platt & Sathyendranath, 2008). The selected indicators are: 1) annual mean chlorophyll; 2) median chlorophyll; 3) annual maximum chlorophyll; 4) timing of initiation of the phytoplankton growing period; 5) timing of peak of the phytoplankton growing period; and 6) duration of the phytoplankton growing period. The first three indicators are based on absolute values of chlorophyll concentration, whereas the last three can be calculated using relative changes in the field of chlorophyll. Timing of the peak in the phytoplankton growing period corresponds to when chlorophyll concentration reaches maximum amplitude in the annual cycle. The timings of initiation and termination of phytoplankton growth are detected using changes relative to a threshold of the long-term median plus 5% (Racault et al., 2012; Siegel, Doney, & Yoder, 2002). The duration of the growing period is estimated as the time elapsed between initiation and termination. Phenology estimates are calculated using 8-day composites, which is the temporal resolution of the chlorophyll data used.

2.3. Sensitivity analysis of the impact of missing data

The question we wish to address is whether the additional gaps in CZCS data compared with SeaWiFS data could lead to differences in the estimation of ecological indicators. Therefore, in the sensitivity analysis presented here, we treat SeaWiFS as the reference data set, and we use the CZCS gap distribution as a mask to create a SeaWiFS data set with CZCS-like gaps. Thus, we can investigate the impact that missing data may have on determination of ecological indicators from two consistent ocean-colour data sets (i.e. SeaWiFS reference and SeaWiFS with CZCS-like gaps) in terms of calibration and algorithms. To avoid bias associated with the significant increase in missing data in chlorophyll observations after 2007 in the SeaWiFS sensor, the sensitivity analysis was performed using SeaWiFS data from 1998 to 2007.

Error in the estimation of ecological indicators was evaluated using two measures: the root-mean-square-error (RMSE) and the bias. The