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

Marie-Fanny Racault[1*], Shubha Sathyendranath[1], Trevor Platt[1]

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

*Corresponding author:

mfrt@pml.ac.uk, Phone +44 175 263 34 34, Fax +44 175 263 31 01

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%), while 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.

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

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; Platt and Sathyendranath, 2008; Cardoso et al., 2010; Ferreira et al., 2011; 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 and McDonald, 2004). The state of the first trophic level of the marine ecosystem can be characterized by the annual cycle of phytoplankton. In-situ or remote-sensing observations of chlorophyll concentration, a proxy for phytoplankton biomass, have been used to depict changes in the annual cycle of phytoplankton (Platt and Sathyendranath, 1996; Platt and 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 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 (Hollman 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 et al. (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 et al., 2010; Mackas et al., 2011; Rykaczewski and Dunne, 2011; McQuatters-Gollop et al., 2011; Raitsos et al., 2013; Wernand et al., 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, Thomalla et al., 2011, Racault et al., 2012; González Taboada and Ricardo Anadón, 2014).

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° x 1° boxes (Fig. 1).

2.2 Estimation of ecological indicators

The annual cycle of phytoplankton was characterized by estimating six well-established ecological indicators from remote-sensing observations of chlorophyll concentrations (Platt and Sathyendranath, 1996; Platt and 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% (Siegel et al., 2002; Racault et al., 2012). 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 1998-2007.

Error in the estimation of ecological indicators was evaluated using two measures: the root-mean-square-error (RMSE) and the bias. The procedure to evaluate the error is presented in the flow diagram (Fig. 1) and described in the following steps: 1) a SeaWiFS nine-year chlorophyll time-series was selected as the reference from the 10 years of available data during 1998-2007 (by drawing out, without duplication); 2) the SeaWiFS nine-year time-series was sub-sampled to simulate the distribution of missing data in the nine-year CZCS time-series, generating a SeaWiFS time-series with CZCS-like gaps; 3) nine-year climatologies were computed for the SeaWiFS reference time-series and the SeaWiFS time series with CZCS-like missing data; 4) the six ecological indicators were estimated from each climatology; and 5) the difference δ defined as:

Equation (1): $$\delta = ind_{gaps} - ind_{ref}$$

was computed, with ind_{ref} representing the ecological indicator estimated from the SeaWiFS reference climatology and ind_{gaps} representing the same indicator estimated from the SeaWiFS climatology with CZCS-like gaps. The entire procedure was repeated for each one-degree pixel of the global oceans. In addition, a relative difference $\delta_r$was estimated for the indicators of maximum amplitude, annual mean, median and duration:

Equation (2): $$\delta_r = \frac{\delta}{ind_{ref}} .$$

To account for the sensitivity of difference estimates to the choice of the particular years in the time series, we generated a total of 25 unique SeaWiFS reference time-series by drawing out, without duplication, nine years from the 10 years of SeaWiFS data (1998-2007). Then, at each given pixel of the oceans where indicator estimates were available, the magnitude of the error was measured using the root-mean-square-error RMSE as follows:

Equation (3): $$RMSE(\delta) = \sqrt{\frac{\sum_{i=1}^{25} \delta^2}{25}}$$; and

Equation (4): $$RMSE(\delta_r) = \sqrt{\frac{\sum_{i=1}^{25} {\delta_r}^2}{25}}$$;

Moreover, at each pixel, the bias was computed as:

Equation (5): $$Bias(\delta) = \frac{\sum_{i=1}^{25} \delta}{25}$$; and

Equation (6): $$Bias(\delta_r) = \frac{\sum_{i=1}^{25} \delta_r}{25}$$.

Next, to reduce the number of missing data in the SeaWiFS reference and in the SeaWiFS with CZCS-like gaps time-series, a spatial and temporal linear interpolation was performed (see gap filling “option” in Fig. 1) and the error estimation procedure described in equations (1) to (6) was re-applied. The interpolation scheme was applied sequentially in the order: longitude, latitude, and time. Specifically, the gaps were filled with the average value of the surrounding grid points along the indicated axis. The averaging window had a width of three points and the surrounding points were weighted equally. Along the indicated axis, if one of the points bordering the gap was invalid, it was omitted from the calculation. If the two surrounding points were invalid, then the gap was not filled (the interpolation scheme is illustrated in Fig. 1).

The outcome of the sensitivity analysis is an evaluation of the RMSE (providing information on the precision of the error) and the bias (corresponding to a measure of accuracy of the error) in the estimation of the six ecological indicators due to the missing data in the CZCS time-series (with and without interpolation procedure). It is noteworthy that the CZCS time-series is used here only to identify the spatio-temporal distribution of the missing data. The indicators are actually estimated from the SeaWiFS observations (i.e. with and without CZCS-like sub-sampling, and with and without interpolation). Therefore, any difference in the estimated ecological indicators arises from differences in the gaps between the two datasets analysed.

3. Results

3.1 Spatio-temporal distribution of ocean-colour observations

Large differences are apparent in the spatial coverage of the SeaWiFS and CZCS missions (Fig. 2a and 2b). In the SeaWiFS data collection, the number of scenes (i.e. 8-day composites) decreases markedly poleward of 30°N and 30°S, following the latitude-dependent increase in the solar zenith angle during the winter season (Fig. 2a). The tropics and subtropics are not affected by high-sun zenith angle, and the reduction in the number of scenes is caused mainly by atmospheric aerosols, sun-glint and persistent clouds (e.g. during the monsoon season). The tropical regions with lowest coverage include the coasts of Western Africa and South-Western America, the Arabian Sea and the Bay of Bengal (Fig. 2a). During the CZCS mission, in addition to the reduction of scenes due to all the same reasons as in the case of SeaWiFS, the collection of data was further limited by low duty cycle. The spatial coverage of the CZCS is better in coastal regions and in the Northern Hemisphere, with the highest density of 8-day composites observed in upwelling regions, the Arabian and Mediterranean Seas, and along the coasts of Europe, North-Eastern Africa, Northern America and Eastern as well as Western Australia (Fig. 2b). When a linear interpolation procedure (i.e. interpolating spatially- and temporally-adjacent values) is applied to 8-day composites of ocean-colour data from SeaWiFS and from CZCS, the density of data increases by 2% for SeaWiFS over the period 1997-2011 and by 91% for CZCS over the period 1978-1986 in the global oceans (Fig. 2c and 2d). In other words, the interpolation procedure nearly doubled the spatio-temporal coverage of CZCS data.

Since we are evaluating the gaps in CZCS data compared with SeaWiFS, the coverage of the CZCS is estimated as a percentage of the SeaWiFS climatological coverage (Fig. 3a). On average, during the period 1978-1986, the global ocean coverage of CZCS reaches 19% of the SeaWiFS climatological coverage, with 12% of the observations located in the Northern Hemisphere and 6.5% in the Southern Hemisphere. Moreover, a major reduction in sampling occurred in the global oceans in the Spring of 1982 after the volcanic eruption of "El Chichon" released large quantities of ash into the atmosphere (Michalsky et al., 1990; Antoine et al., 2005). In the following years, the sampling density remained low, particularly during the summers 1984 and 1985, when nearly no observations were recorded. When the linear interpolation procedure is applied, the mean (1978-1986) global ocean coverage of CZCS reaches 40% (Fig. 3b).

3.2 Error associated with missing data on the estimation of ecological indicators

The distribution of RMSE in the global oceans is shown in Figure 4 and as a function of the percentage of missing data of the CZCS sensor in Figure 5. The bias (a measure of accuracy) in the estimated indicators in the global ocean is shown in Figure 6 and as a function of the percentage of missing data of the CZCS sensor in Figure 7. Global averages of RMSE and bias are provided in Tables 1 and 2.

3.2.1 Phytoplankton biomass indicators

The distribution of RMSE values for peak chlorophyll shows large variations throughout the global oceans (Fig. 4a). Lower RMSE values tend to be observed where the percentage of missing data was lower (i.e. fewer missing data). This tendency is more clearly apparent after a linear interpolation has been applied to fill in some of the missing data. Peak chlorophyll RMSE values of 30% or less are generally observed in coastal regions, across the North Atlantic ocean, the eastern North Pacific ocean and the western coast of Australia. The percentage of ocean coverage with RMSE below 30% reaches 56% after applying linear interpolation (Table 1). Interestingly, the shape of the distribution of the RMSE remains similar before and after applying linear interpolation (Fig. 5c and 5d), indicating that the effect of interpolation is uniform across all the regions.

The missing data in the CZCS sampling induce a bias of +16% on average on the estimation of peak chlorophyll concentration (Table 2). The bias appears positive throughout most of the global oceans (Fig. 6a) indicating that peak chlorophyll concentrations tend to be over-estimated in the multi-year SeaWiFS climatology with missing data, compared with the multi-year SeaWiFS reference climatology. In regions of the oceans where the percentage of missing data (in the SeaWiFS time-series with CZCS-like gaps) is less than 65%, the bias in peak chlorophyll concentration is positive with values ranging between +5 and +25% (Fig. 7c). However, when the percentage of missing data is particularly high (i.e. greater than 95%), the bias in peak chlorophyll concentration appears negative (i.e. peak chlorophyll values estimated from the climatology with gaps tend to be lower, Fig. 7c). The positive biases in peak chlorophyll may be counter intuitive, in the sense that in any given year, one anticipates that missing data would lead to estimated peak values equal to, or less than the reference dataset. However, when dealing with multi-year climatologies, one of the consequences of missing data is that a high peak value in any single year does not get averaged with lower values from other years if data are missing from those other years. But as gaps in data increase, the probability of missing all peak values increases, leading to negative bias. Such negative bias values are observed over large regions of the Southern Ocean where the sampling coverage of the CZCS was particularly limited (Fig. 2b and 6a). After applying the interpolation procedure, the results are spatially more homogeneous (i.e. positive bias values throughout the global oceans), and the bias values are noticeably reduced in large regions of the global oceans (Fig. 6a). Interestingly, the interpolation procedure had limited influence on the shape of the bias distribution as a function of the percentage of missing data in the SeaWiFS time-series with CZCS-like gaps (Fig. 7c, d).

The RMSE values for climatological mean and median indicators are particularly low, with average values of 13 and 14% for the global oceans (before applying a linear interpolation; Fig. 4b, c and Table 1). RMSE values for these two indicators are below 30% in more than 90% of the oceans (Fig. 4b, c) and the interpolation procedure had a very limited influence on the spatial distribution of the

RMSE (Fig. 4b, c). The RMSE for the climatological mean and median indicators increases exponentially with increasing missing data (Fig. 5e-h). Moreover, the shape of the curve describing RMSE as a function of missing data remained similar before and after applying the interpolation procedure. The biases in the estimated climatological mean and median chlorophyll concentrations are -0.5% and -2% respectively on average in the global oceans (Fig. 6b, c; Table 2). These bias values change to +1% and -0.5% respectively after applying the interpolation procedure (Fig. 6b, c and Table 2). It is noteworthy that for these two indicators, the bias estimates alternate between positive and negative values throughout the global oceans. This pattern is also clearly apparent when the bias is plotted as a function of the percentage of missing data (Fig. 7e-h). For these two indicators, when the percentage of missing data is <50%, the bias is constrained within ±5%, which is particularly low compared with the bias associated with peak chlorophyll concentration.

3.2.2 Phytoplankton phenology indicators

In this analysis, only the ocean pixels for which all phenological metrics (i.e. timings of initiation, peak, termination, and duration) could be estimated are shown on the maps of the distribution of the RMSE and bias (Fig. 4d-f and Fig. 6d-f). Before applying the interpolation procedure, the phenology indicators could be estimated over 25% of the global oceans. This figure increases to 70% after applying the interpolation (which was applied prior to the calculation of the climatology from which the indicators are estimated, as described in the method section). The identification of timings of specific events, such as those of initiation and termination, are particularly sensitive to the presence of missing data in the time-series. As a result, the calculation of the duration (which is estimated as the difference between the timings of initiation and termination) can fail. The increase in spatial coverage of the indicators achieved, once the interpolation is implemented, highlights the critical importance of estimating phenology indicators from the most temporally-complete time-series.

The missing data in the CZCS sampling induce on average, over the global oceans, a RMSE and a bias of 57% and -43% respectively (before applying the interpolation); and 47% and -20% respectively (after applying the interpolation) on the estimated duration of phytoplankton growing period (Fig. 4d, Fig. 6d, and Tables 1 and 2). Negative bias values are observed throughout most of the global oceans, indicating that when missing data are present in the SeaWiFS time-series with CZCS-like gaps, the duration tends to be under-estimated compared with the SeaWiFS reference time-series with more data. The RMSE values decreased in those ocean regions where the percentage of missing data was lower. Before applying the interpolation, 8% of all of the pixels in the oceans presented an RMSE of 30% or less, whereas after applying the interpolation, 26% of all of the ocean pixels showed an RMSE of 30% or less (Table 1). As with the indicators of climatological mean, median and peak chlorophyll, the plot of RMSE and bias in the estimated duration as a function of the percentage of missing data in CZCS, showed similar patterns with and without the linear interpolation procedure (Fig. 5i-j and Fig. 7i-j), except for percent missing data <30%. It is probable that the increase in RMSE observed in Fig. 5i is due to the low number of observations in those class intervals (i.e. only eight pixels in the global oceans presented missing data between 20-25%). The bias in the duration estimates drops below 10% when the percentage of missing data (in the SeaWiFS time-series with CZCS-like gaps) is less than 60% (Fig. 7i).

On average, the RMSE in the estimation of the timings of initiation and peak are 76 and 75 days respectively. After applying the interpolation procedure, the RMSE was reduced to 61 and 62 days for the timings of initiation and peak respectively (Fig. 4e, f; Table 1). Similar to the estimated duration, the RMSE decreased in ocean regions where the percentage of missing data was lower. In the case of the timings of initiation and peak, without applying interpolation, 11% of all of the ocean pixels presented an RMSE of 30 days or less, whereas, with interpolation, the percentage increased to 27 and 24% (Table 1).

The bias values were equal to -10 and -24 days on average in the global oceans for the timings of bloom initiation and peak chlorophyll respectively (Fig. 6e, f, Table 2). Negative bias values indicate that the estimated timings tend to be earlier in the climatological seasonal cycle with missing data compared with the SeaWiFS reference climatology data set. After applying the interpolation, the number of ocean pixels for which the timings of initiation and peak could be estimated increased markedly and their average bias values decreased to -1 and -13 days respectively. Ocean pixels with less than 65% missing data show a bias of less than 30 days (~1 month) for timings of both initiation and peak (Fig. 7k-n).

4. Discussion and Conclusions

4.1 Sensitivity of ecological indicators to the distribution of missing data in the CZCS time-series

The present sensitivity analysis provides an original assessment of the impact that the distribution of the missing data in the CZCS time-series is having on the estimation of six ecological indicators. The impact estimated here is based on multi-year composite of an annual cycle in phytoplankton dynamics, and not on a year-to-year basis, in which case the gaps, and hence the uncertainties, would be greater. The selected indicators are key to characterize and monitor the composition, structure and functioning of the marine ecosystem on seasonal, interannual, decadal and longer time-scales. Thus, an evaluation of the confidence range in the estimates is essential, especially for detection of trends influenced by large-scale environmental and climate drivers (Vantrepotte and Mélin, 2009; Martinez et al., 2011; Thomalla et al., 2011; Racault et al., 2012; Zhai et al., 2013; González-Taboada and Ricardo-Anadón, 2014).

Annual coverage of chlorophyll data in the CZCS record is low compared with that of more recent sensors such as SeaWiFS and the distribution of the CZCS missing data is non-uniform both spatially and temporally (Fig. 2 and 3). These gaps make it difficult to estimate phenology indicators on annual time-scale, which is why the present analysis is limited to multi-year climatologies. To further limit the negative impact of missing data, data were averaged spatially (i.e. re-gridding 4 km x 4 km to 1° x 1° grid-box), though we have maintained the 8-day temporal resolution as a requirement for studying phenology (Fig. 1). Other approaches to reducing missing data include implementation of interpolation procedures: gaps can be filled by interpolating spatially and temporally-adjacent values (e. g., Beaugrand et al., 2008; Pottier et al., 2008; Racault et al., 2012) or by using the climatology of the annual cycle as a basis for interpolating across gaps for particular years in a time-series (Land et al., 2014). The use of climatology allows us to constrain potential errors in phenology estimates, which are associated with missing data in annual time-series (Cole et al., 2012; Land et al., 2014). In spite of the limitations imposed by missing data, the 8-day climatology of CZCS provides the most comprehensive dataset available to compute many ecological indicators during the 1970s and 1980s and to study their long-term changes in relation to climate drivers by comparison with later satellite sensors such as SeaWiFS, MERIS or MODIS.

The indicators of annual mean and median chlorophyll concentrations showed the lowest RMSE and bias associated with the presence of missing data (Fig. 4b, c, Fig. 5e-h and Tables 1 and 2). Low RMSE and bias values indicate that the mean and median chlorophyll concentrations can be estimated with relatively high confidence from a climatology which includes the gap distribution of the CZCS time-series. This feature is consistent throughout the global oceans. The linear interpolation procedure (i.e. spatial and temporal filling of missing data with adjacent values performed before calculating the climatology) reduced the global average of RMSE values in the median and the mean from 12 and 13% respectively to 8% (for both). The magnitude of the RMSE increased with increasing missing data (Fig. 5e-h). In addition, for these two indicators, the mean bias for the global oceans was particularly low (within ±2%) regardless of application of the interpolation procedure. But of course the global averages do not tell the whole story, and what is really important is the regional distribution of uncertainties. In fact, regionally, the bias could be greater, reaching ±8%. Even though the uncertainties in these indicators are relatively low, they are based on absolute values of chlorophyll concentrations, and hence would be vulnerable to any inter-sensor biases in estimated chlorophyll values arising from differences in sensor design or in algorithms. Such potential errors would also have to be quantified before these indicators derived from CZCS and SeaWiFS can be compared.

The indicators of peak chlorophyll concentration and duration of the bloom have higher RMSE (33 and 47% on average respectively for the global oceans) and bias (+18 and -20% on average respectively for the global oceans) associated with the presence of missing data even after interpolation (Fig. 4a, 4d, Fig. 6a, 6d, and Tables 1 and 2). The RMSE and bias values are lower in oceanic regions where the density of data collected during the CZCS time-series is higher, demonstrating the sensitivity of these indicators to missing data. As a result, the reduction of gaps in data using linear interpolation significantly decreases the RMSE and bias for both the peak chlorophyll concentration and the duration estimates (Fig. 4a, 4d, Fig. 6a, 6d). The regions with higher confidence (i.e. RMSE values < 30% and bias < 10%) on the estimations of peak chlorophyll concentration and duration of the growing period include the North Atlantic Ocean between 10°N-50°N, the Pacific Ocean between 10°N-40°N, the western coast of North America, the eastern coast of Africa, and the eastern and western coast of Australia and New Zealand. Outside of these regions, the RMSE and bias tend to increase markedly, because of reduction in the density of observations, rendering difficult the detection of long-term trends in these indicators.

The timing of bloom initiation and timing of peak chlorophyll estimated from SeaWiFS with CZCS-like gaps climatology had RMSE values of 62 and 61 days respectively on average for the global oceans (Table 1). The high RMSE values reported here underline the sensitivity of indicators of timing of events to the missing data in the CZCS sampling. The mean biases for the global oceans in the timings of initiation and peak were -1 and -13 days respectively, after applying linear interpolation (Table 2). The linear interpolation used here to fill gaps in data nearly doubles (Table 1) the number of pixels in the global oceans where these phenology indicators can be estimated with an RMSE of less than one month (~30 days).

Moreover, the linear interpolation allows the phenological estimates to gain coherence in most of the coastal regions, across the North Atlantic Ocean, the eastern North Pacific Ocean and the western coast of Australia (Fig. 4e, f and Fig. 6e, f). Increased confidence in the phenology estimates, even over limited regions of the oceans, is extremely useful for the detection of long-term trends or differences.

The error estimates (RMS uncertainty and bias) presented here are specifically designed to evaluate the impact of the distribution of missing data in the CZCS sampling, compared with the SeaWiFS coverage. The computed biases provide a basis for correcting for systematic differences in estimates of these ecological indicators for every one degree grid for which the computations have been carried out. The RMSE, once corrected for the bias, yields the standard deviation in the results, which can then be used to constrain interpretation of differences in indicators estimated from SeaWiFS with CZCS-like gaps and SeaWiFS reference data sets: the observed differences cannot be significant if they are less than the standard deviation in the results.

Cole et al. (2012) estimated the differences between phenology metrics from the GlobColor time-series and those from the NASA Ocean Biogeochemical Model (treated as the gap-free time series). In sub-polar regions, where the percentage of missing data is high, the authors showed typical differences of 30 days for the timing of initiation and 15 days for the timing of peak. The differences were lower (typically below 20 days for the timing of initiation and less than 10 days for the timing of peak) in the tropics and the subtropics where the percentage of missing data is low. Though their measures of errors are different from ours, their results are coherent with ours, in the sense that the RMSE and bias values shown here decrease when the percentage of missing data decreased.

A further cautionary note is that the present study identifies and quantifies only one source of uncertainty: gaps in data. Other factors will have an influence on the uncertainty associated with the estimation of phenological indicators. Although it is beyond the scope of the present study, it would be extremely interesting to provide a comprehensive analysis of the propagation of uncertainties associated with: (1) the presence of missing data due to persistent cloud cover, high-sun zenith angle, and sensor sampling; (2) the variability of the annual chlorophyll cycle; and (3) the uncertainties in the calibration of satellite sensors and in the chlorophyll-retrieval algorithm (Moore et al., 2009).

All the results presented here are based on analyses carried out using multiple sets of years. This was done to increase the generality of results and to avoid the impact of any particular year or a particular combination of years on the results. However, when actual comparisons are made between phytoplankton indicators from particular sets of CZCS years and SeaWiFS years, it would be more useful to repeat the analyses presented here, but for those particular sets of years, to evaluate the uncertainties for that special case.

In summary: 1) lowest uncertainty due to missing data is observed in the indicators of annual mean and median chlorophyll concentration (global mean RMSE < 10% and |bias| ≤ 1%) while higher uncertainty is observed for peak chlorophyll and duration (global mean RMSE < 50% and |bias| ≤ 20%) and for timing metrics (global mean RMSE < 2 months and |bias| ≤ 2weeks); 2) gap filling (by linear interpolation) increases precision by 4-10% and ~2 weeks (global mean RMSE) and increases accuracy by 0.5-13% and ~10 days (global mean |bias|); 3) regional differences are apparent, and lowest uncertainty is recorded where CZCS coverage is greater than 40%.

4.2 Implications for estimation of long-term trends in ecological indicators

The low error values for annual mean and median chlorophyll concentrations indicate a low sensitivity of these two indicators to the distribution of missing data in the CZCS time-series, lending confidence that the assessments of decadal changes reported in the re-analysis efforts of Gregg and Conkright (2002) and Antoine et al. (2005) were not affected much by the missing data. They had applied the CZCS data distribution to SeaWiFS to minimize, if not eliminate, the impact of differing data distributions in comparing average chlorophyll levels. Therefore, the discrepancies in the decadal changes reported in the two publications are probably a consequence of the differences in the approaches followed by the two authors. For example, Gregg and Conkright (2002) blended in situ results with remotely-sensed data, whereas Antoine et al. (2005) avoided using in-situ data, relying instead on an improved algorithm. Other factors influencing the estimation of long-term trends include the direction or sign of the dominant climate drivers (such as El Niño Southern Oscillation, or ENSO) occurring during the periods under assessment (Martinez et al., 2009). In fact, Gregg and Conkright (2002) compared the CZCS 1979-1986 archive with SeaWiFS 1997-2000 data, whereas Antoine et al. (2005) compared the CZCS (1978-1986) and SeaWiFS (1998-2002) records. Both the CZCS and SeaWiFS periods were marked by major El Niño (1997) and La Niña (1998) events, which profoundly influence phytoplankton production, composition and phenology in the global oceans (Dandonneau et al., 1986; Harris 1987; Comiso et al., 1993; Chavez et al., 1999; Behrenfeld et al., 2001; Yoder and Kennelly, 2003; Hirawake et al., 2005; Behrenfeld et al., 2006; Chavez et al., 2011; D'Ortenzio et al., 2012). This also raises the possibility that the characteristics of the errors associated with missing data may also be specific to the pairs of CZCS and SeaWiFS years considered in any particular analysis. It would therefore be prudent to repeat the analysis presented here, but for the particular years relevant for any analyses, to lend further confidence that the missing data do not introduce any significant errors into the results.

Given the rapid response of phytoplankton chlorophyll concentration to these variations in climate and environmental conditions, as well as the sensitivity of absolute chlorophyll values to sensor-specific differences in chlorophyll retrieval, indicators of phytoplankton phenology (which are not sensitive to errors in the absolute values of chlorophyll) may be robust for studying long-term climate change impacts on the state of the first trophic level of the marine ecosystem. However, phenological studies do require data well distributed in time, to enable resolution of timings of seasonal events with sufficient precision. The sensitivity analysis presented here provides the first comprehensive and quantitative evaluation of errors in ecological (including phenological) indicators associated with gaps in the CZCS data, when pooled to create a nine-year climatology at 8-day temporal resolution. The results demonstrate that 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 from the nine-year climatology in approximately 95% of the global oceans and the phenological patterns can be estimated with low error in approximately 25% of the global oceans (i.e. regions where the indicators can be estimated with RMSE values of less than 30% and bias within ±10%). In particular, oceanic regions where estimates of ecological indicators can be used reliably to extend the remote-sensing record back three decades and thus assess long-term trends in the state of the marine ecosystem, include the North Atlantic Ocean between 10°N-

50°N, the Pacific Ocean between 10°N-40°N, the western coast of North America, the eastern coast of Africa, and the eastern and western coast of Australia and New Zealand.

It is noteworthy that the surest way to avoid errors of the type discussed here is to limit the analysis to areas where the CZCS observations are matched in time with SeaWiFS, and where the temporal resolution is sufficient to extract the indicators with sufficient confidence. But, as one can see from Figure 3, for any given 8-day composite, areas of the world where we have both SeaWiFS and CZCS data are limited to approximately 30-40%, and even in these areas the uncertainties due to missing data can be high for some of the indicators (Tables 1 and 2). The analysis carried out here suggests ways in which the areal coverage can be extended by linear interpolation. Furthermore, having an idea of the potential bias (Table 2), this type of errors can be corrected for, and knowing the RMSE allows us to place confidence intervals on the results. Finally, these results demonstrate some of the issues associated with comparing or blending phytoplankton datasets with different spatial and temporal coverage. The method developed here helps to assess uncertainties in comparison of two phytoplankton datasets (CZCS and SeaWiFS) arising from this source, and thus, to improve confidence in inferred long-term trends (Mackas et al., 2011).

Acknowledgments

The authors acknowledge the NASA Ocean Color Processing Group for providing SeaWiFS and CZCS Chlorophyll data. This work is a contribution to the Ocean Colour Climate Change Initiative of the European Space Agency and GreenSeas, a project of the European Commission Seventh Framework Programme (265294[FP7-ENV-2010]).

References

Antoine, D., Morel, A., Gordon, H. R., Banzon, V. F., & Evans, R. H. (2005). Bridging ocean color observations of the 1980s and 2000s in search of long-term trends. Journal of Geophysical Research, 110(C6), C06009. doi:10.1029/2004JC002620

Beaugrand, G., Edwards, M., Brander, K., Luczak, C., & Ibanez, F. (2008). Causes and projections of abrupt climate-driven ecosystem shifts in the North Atlantic. Ecology letters, 11(11), 1157–68. doi:10.1111/j.1461-0248.2008.01218.x

Behrenfeld, M. J., Randerson, J. T., McClain, C. R., Feldman, G. C., Los, S. O., Tucker, C. J., Falkowski, P. G., Field, C. B., Frouin, R., Esaias, W. E., Kolber, D. D., & Pollack, N. H. (2001). Biospheric primary production during an ENSO transition. Science, 291(5513), 2594–7. doi:10.1126/science.1055071

Behrenfeld, M. J., Michael J., O’Malley, R. T., Siegel, D., McClain, C. R., Sarmiento, J. L., Feldman, G. C., Milligan, A. J., Falkowski, P. G., Letelier, R. M., & Boss, E. S. (2006). Climate-driven trends in contemporary ocean productivity. Nature, 444(7120), 752–5. doi:10.1038/nature05317

Borja, A., Bricker, S. B., Dauer, D. M., Demetriades, N. T., Ferreira, J. G., Forbes, A. T., Hutchings, P., Jia, X., Kenchington, R., Marques, J. C., & Zhu, C. (2008). Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. Marine Pollution Bulletin 56, 1519–1537

Boyce, D., Lewis, M., & Worm, B. (2010). Global phytoplankton decline over the past century. Nature, 466(7306), 591–596. doi:10.1038/nature09268

Cardoso, A. C., Cochrane, S., Doerner, H., Ferreira, J. G., Galgani, C., Hagebro, C., Hanke, G., Hoepffner, N., Keizer, P. D., Law, R., Olenin, S., Piet, G. J., Rice, J., Rogers, S. I., Swartenbroux, F., Tasker, M. L., & Van de Bund, W. (2010). Scientific support to the European commission on the marine strategy (pp. 65). doi:10.2788/86430

Chavez, F. P., Messié, M., & Pennington, J. T. (2011). Marine Primary Production in Relation to Climate Variability and Change. Annual Review of Marine Science, 3(1), 227–260. doi:10.1146/annurev.marine.010908.163917

Chavez, F., Strutton, P., Friederich, G., Feely, R., Feldman, G., Foley, D., & McPhaden, M. (1999). Biological and chemical response of the equatorial pacific ocean to the 1997-98 El Niño. Science, 286(5447), 2126–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10591638

Chiba, S., Aita, M. N., Tadokoro, K., Saino, T., Sugisaki, H., & Nakata, K. (2008). From climate regime shifts to lower-trophic level phenology: Synthesis of recent progress in retrospective studies of the western North Pacific. Progress in Oceanography, 77(2-3), 112–126. doi:10.1016/j.pocean.2008.03.004

Cole, H., Henson, S., Martin, A., & Yool, A. (2012). Mind the gap: The impact of missing data on the calculation of phytoplankton phenology metrics. Journal of Geophysical Research, 117(C8), C08030. doi:10.1029/2012JC008249

Comiso, J. C., Mcclain, C. R., Sullivan, C. W., Ryan, J. P., & Leonard, C. L. (1993). Coastal Zone Color Scanner Pigment Concentrations in the Southern Ocean and Relationships to Geophysical Surface Features, Journal of Geophysical Research, 98(92), 2419–2451.

D'Ortenzio, F., Antoine, D., Martinez, E., & Ribera d'Alcalà, M. (2012). Phenological changes of oceanic phytoplankton in the 1980s and 2000s as revealed by remotely sensed ocean-color observations. Global Biogeochemical Cycles, 26(4), doi:10.1029/2011GB004269

Dandonneau, Y. (1986). Monitoring the sea surface chlorophyll concentration in the tropical pacific: consequences of the 1982-83 El Niño. Fishery Bulletin, 84(3), 687–695.

Ferreira, J. G., Andersen, J. H., Borja, A., Bricker, S. B., Camp, J., Cardoso da Silva, M., Garcés, E., Heiskanen, A.-S., Humborg, C., Ignatiades, L., Lancelot, C., Menesguen, A., Tett, P., Hoepffner, N., & Claussen, U. (2011). Overview of eutrophication indicators to assess environmental status within the European Marine Strategy Framework Directive. Estuarine, Coastal and Shelf Science, 93(2), 117–131. doi:10.1016/j.ecss.2011.03.014

González Taboada, F., & Anadón, R. (2014). Seasonality of North Atlantic phytoplankton from space: impact of environmental forcing on a changing phenology (1998-2012). Global Change Biology, doi:10.1111/gcb.12352

Gregg, W. W., & Conkright, M. E. (2002). Decadal changes in global ocean chlorophyll. Geophysical Research Letters, 29(15), 20–1–20–4. doi:10.1029/2002GL014689

Harris, G., Nilsson, C., Clementson, L. & Thomas, D. (1987). The water masses of the east coast of Tasmania: seasonal and interannual variability and the influence on phytoplankton biomass and productivity. Australian Journal of Marine and Freshwater Research, 38, 569-590.

Hirawake, T., Odate, T., & Fukuchi, M. (2005). Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965–2002. Geophysical Research Letters, 32(5), L05606. doi:10.1029/2004GL021394

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, a., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., Van Roozendael, M., & Wagner, W. (2013). The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables. Bulletin of the American Meteorological Society, 94(10), 1541–1552. doi:10.1175/BAMS-D-11-00254.1

Land, P., Shutler, J., Platt, T., & Racault, M.-F. (2014). A novel method to retrieve oceanic phenology from satellite data in the presence of data gaps. Ecological Indicators, 37: 67–80.

Mackas, D. L. (2011). Does blending of chlorophyll data bias temporal trend? Nature, 472, E4-E5. doi:10.1038/nature09951

Martinez, E., Antoine, D., D'Ortenzio, F., & De Boyer Montégut, C. (2011). Phytoplankton spring and fall blooms in the North Atlantic in the 1980s and 2000s. Journal of Geophysical Research, 116(C11), C11029. doi:10.1029/2010JC006836

Martinez, E., Antoine, D., D'Ortenzio, F., & Gentili, B. (2009). Climate-driven basin-scale decadal oscillations of oceanic phytoplankton. Science, 326(5957), 1253–6. doi:10.1126/science.1177012

McQuatters-Gollop, A., Reid, P. C., Edwards, M., Burkill, P. H., Castellani, C., Batten, S., Gieskes, W., Beare, D., Bidigare, R. R., Head, E., Johnson, R., Kahru, M., Koslow, J. A., & Pena, A. (2011). Is there a decline in marine phytoplankton? Nature, 472, E5-E6. doi:10.1038/nature09950

Moore, T. S., Campbell, J. W., & Dowell, M. D. (2009). Remote Sensing of Environment: A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. Remote Sensing of Environment, 113(11), 2424–2430. doi:10.1016/j.rse.2009.07.016

Niemi, G. J., & McDonald, M. E. (2004). Application of Ecological Indicators. Annual Review of Ecology, Evolution, and Systematics, 35(1), 89–111. doi:10.1146/annurev.ecolsys.35.112202.130132

Platt, T., & Sathyendranath, S. (1996) Modelling primary production. Aquabiology, 18: 378-380.

Platt, T., & Sathyendranath, S. (2008). Ecological indicators for the pelagic zone of the ocean from remote sensing. Remote Sensing of Environment, 112(8), 3426–3436. doi:10.1016/j.rse.2007.10.016

Platt, T., Sathyendranath, S., White, G. N., Fuentes-Yaco, C., Zhai, L., Devred, E., & Tang, C. (2009). Diagnostic Properties of Phytoplankton Time Series from Remote Sensing. Estuaries and Coasts, 33(2), 428–439. doi:10.1007/s12237-009-9161-0

Pottier, C., Turiel, a, & Garcon, V. (2008). Inferring missing data in satellite chlorophyll maps using turbulent cascading. Remote Sensing of Environment, 112(12), 4242–4260. doi:10.1016/j.rse.2008.07.010

Racault, M.-F., Le Quéré, C., Buitenhuis, E., Sathyendranath, S., & Platt, T. (2012). Phytoplankton phenology in the global ocean. Ecological Indicators, 14(1), 152–163. doi:10.1016/j.ecolind.2011.07.010

Raitsos, D. E., Walne, A., Lavender, S. J., Licandro, P., Reid, P. C., & Edwards, M. (2012). A 60-year ocean colour data set from the continuous plankton recorder. Journal of Plankton Research, 35(1), 158–164. doi:10.1093/plankt/fbs079

Rykaczewski, R. R. & Dunne, J. P. (2011). A measured look at ocean chlorophyll trends. Nature, 472, E5-E6. doi:10.1038/nature09952

Siegel, D., Doney, S. C., & Yoder, J. (2002). The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis. Science, 296(5568), 730–3. doi:10.1126/science.1069174

Tett, P., Gowen, R., Painting, S., Elliott, M., Forster, R., Mills, D., Bresnan, E., Capuzzo, E., Fernandes, T. F., Foden, J., Geider, R. J., Gilpin, L. C., Huxham, M., McQuatters-Gollop, A. L., Malcolm, S. J., Saux-Picart, S., Platt, T., Racault, M.-F., Sathyendranath, S., van der Molen, J., & Wilkinson, M. (2013). A framework for understanding marine ecosystem health. Marine Ecology Progress Series. doi:10.3354/meps10539

Thomalla, S. J., Fauchereau, N., Swart, S., & Monteiro, P. M. S. (2011). Regional scale characteristics of the seasonal cycle of chlorophyll in the Southern Ocean. Biogeosciences, 8(10), 2849–2866. doi:10.5194/bg-8-2849-2011

Vantrepotte, V., & Mélin, F. (2009). Temporal variability of 10-year global SeaWiFS time-series of phytoplankton chlorophyll a concentration. ICES Journal of Marine Science, 66(7), 1547–1556, doi:10.1093/icesjms/fsp107

Wernand, M. R., Van der Woerd, H. J., & Gieskes, W. W. C. (2013). Trends in ocean colour and chlorophyll concentration from 1889 to 2000, worldwide. PloS one, 8(6), e63766. doi:10.1371/journal.pone.0063766

Yoder, J., & Kennelly, M. (2003). Seasonal and ENSO variability in global ocean phytoplankton chlorophyll derived from 4 years of SeaWiFS measurements. Global Biogeochemical Cycles, 17(4), 1112. doi:10.1029/2002GB001942

Zhai, L., Platt, T., Tang, C., Sathyendranath, S., & Walne, A. (2013). The response of phytoplankton to climate variability associated with the North Atlantic Oscillation. Deep Sea Research Part II: Topical Studies in Oceanography, 93, 159–168. doi:10.1016/j.dsr2.2013.04.009

Zhai, L., Platt, T., Tang, C., Sathyendranath, S., Fuentes-Yaco, C., Devred, E., & Wu, Y. (2010). Seasonal and geographic variations in phytoplankton losses from the mixed layer on the Northwest Atlantic Shelf. Journal of Marine Systems, 80(1-2), 36–46. doi:10.1016/j.jmarsys.2009.09.005

List of Figure Captions

Fig. 1: Schematic view of the data processing steps to estimate the six ecological indicators used to quantify the uncertainty due to the distribution of gaps in the CZCS time-series. The steps numerated 1) to 5) are further described in the method section. The gap-filling step is marked with a star (*) as it was only applied in the analyses labelled "after applying a linear interpolation" (shown in Figs. 2, 3, 4, 5, 6, 7).

Fig. 2: Spatial density of ocean-colour data from CZCS (1978-1986) and SeaWiFS (1997-2010) in the global oceans. (a) SeaWiFS coverage before applying linear interpolation to fill gaps; (b) CZCS coverage before applying linear interpolation to fill gaps; (c) SeaWiFS coverage after applying a linear interpolation to fill some of the missing data; (d) CZCS coverage after applying a linear interpolation to fill some of the missing data. The colour scale indicates the number of 8-day composites available during the sensors' periods of operation.

Fig. 3: Temporal density of CZCS 8-day composites expressed as percentage of SeaWiFS climatological coverage (i.e. the latter is treated as the reference against which the former is compared). (a) CZCS percentage coverage before applying linear interpolation; (b) CZCS coverage after applying linear interpolation to both CZCS and SeaWiFS time-series. In black, coverage for the global oceans and in blue, coverage for the Northern Hemisphere. The coverage for the Southern Hemisphere corresponds to the difference between global and Northern Hemisphere coverage. An assessment of the temporal density of CZCS data at monthly resolution is available from the NASA

ocean color webpage about CZCS mission at: http://oceancolor.gsfc.nasa.gov/CZCS/czcs_datacollect.html.

Fig. 4: Root-mean-square-error (RMSE) on the estimation of six ecological indicators arising solely from missing data. The RMSE is calculated as the difference between the estimates from the SeaWiFS time-series with CZCS-like gaps minus the estimates from the SeaWiFS reference time-series. (a-d) RMSE are expressed in percent and (e-f) RMSE are expressed in days. Left panel: RMSE before applying linear interpolation to fill missing data; Right panel: RMSE after applying linear interpolation to fill missing data (see Fig. 2 and 3 for changes in data coverage). Black colour indicates that the indicators could not be estimated (because there were too few data available).

Fig. 5: Root-mean-square-error (RMSE) of each indicator as a function of the gaps. The percentage of missing data is estimated at each pixel as the fraction of the total number of 8-day composites in the SeaWiFS nine-year climatology with the CZCS-like gaps to the total number of 8-day composites in the SeaWiFS reference nine-year climatology. Left panel: Before applying linear interpolation to fill missing data; Right panel: After applying linear interpolation to fill missing data. (a) and (b) Number of pixels in the global oceans for every increment of 5% in missing data. (c) to (n) Median RMSE values (plain black line) and upper and lower quartiles (dashed black lines) for each class interval of 5% missing data for the six ecological indicators discussed in this paper. Note that, for the left panel, no RMSE values are presented for percentage of missing data <20% because of lack of data. It is probable that the increase in RMSE at the low end of missing values for the phenology metrics (i, k and m) is associated with low number of observations in those class intervals.

Fig. 6: Bias on the estimation of six ecological indicators arising solely from missing data. The bias is calculated as the difference between the indicator estimates from the SeaWiFS time-series with CZCS-like gaps minus the estimates from the SeaWiFS reference time-series. (a-d) Bias values are expressed in percent and (e-f) Bias values are expressed in days. Left panel: Bias before applying linear interpolation to fill missing data; Right panel: Bias after applying linear interpolation to fill missing data (see Fig. 2 and 3 for changes in data coverage). Black colour indicates that the indicators could not be estimated (because of there were too few data available).

Fig. 7: Bias of each indicator as a function of the gaps. The percentage of missing data is estimated at each pixel as the fraction of the total number of 8-day composites in the SeaWiFS nine-year time-series with the CZCS-like gaps to the total number of 8-day composites in the SeaWiFS reference nine-year climatology. Left panel: Before applying linear interpolation to fill missing data; Right panel: After applying linear interpolation to fill missing data. (a) and (b) Number of pixels in the global oceans for every increment of 5% in missing data. (c) to (n) Median bias values (plain black line) and upper and lower quartiles (dashed black lines) for each class interval of 5% missing data for the six ecological indicators discussed in this paper.