AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2014JC010323

Key Points:
- Phenology metrics variability is greater than that between treatments
- Precision is a greater concern than accuracy
- Preprocessing data improves precision of phenology metrics

Supporting Information:
- Readme
- Supplementary material

Correspondence to:
M. R. Payne,
mpay@aqua.dtu.dk

Citation:
Ferreira, A. S., A. W. Visser, B. R. MacKenzie, and M. R. Payne (2014), Accuracy and precision in the calculation of phenology metrics, *J. Geophys. Res. Oceans*, *119*, 8438–8453, doi:10.1002/2014JC010323.

Received 21 JUL 2014
Accepted 12 NOV 2014
Accepted article online 14 NOV 2014
Published online 12 DEC 2014

Accuracy and precision in the calculation of phenology metrics

A. Sofia Ferreira[1], Andre W. Visser[1], Brian R. MacKenzie[1,2], and Mark R. Payne[1,3]

[1]Centre for Ocean Life, National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund, Copenhagen, Denmark, [2]Center for Macroecology, Evolution and Climate, Technical University of Denmark, Charlottenlund, Denmark, [3]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zurich, Switzerland

Abstract Phytoplankton phenology (the timing of seasonal events) is a commonly used indicator for evaluating responses of marine ecosystems to climate change. However, phenological metrics are vulnerable to observation-(bloom amplitude, missing data, and observational noise) and analysis-related (temporal resolution, preprocessing technique, and phenology metric) processes. Here we consider the impact of these processes on the robustness of four phenology metrics (timing of maximum, 5% above median, maximum growth rate, and 15% of cumulative distribution). We apply a simulation-testing approach, where a phenology metric is first determined from a noise- and gap-free time series, and again once it has been modified. We show that precision is a greater concern than accuracy for many of these metrics, an important point that has been hereto overlooked in the literature. The variability in precision between phenology metrics is substantial, but it can be improved by the use of preprocessing techniques (e.g., gap-filling or smoothing). Furthermore, there are important differences in the inherent variability of the metrics that may be crucial in the interpretation of studies based upon them. Of the considered metrics, the 15% of cumulative distribution metric best satisfies the precision criteria. However, the 5% above median metric is comparable in terms of precision and exhibits more inherent variability. We emphasize that the choice of phenology metric should be determined by the specific nature of the question being asked. We believe these findings to be useful to the current discussion on phenology metrics of phytoplankton dynamics.

1. Introduction

The biological components of ecosystems are often characterized by annually recurring cycles. While these cycles show a regular pattern, variations in the timing of key features occur from year to year. The study of the timing of such features within annual cycles is referred to as phenology.

Phenology can provide great insight into the dynamics of marine ecosystems. Phenology is an emergent property, resulting from the integrated interplay of physical and biological processes in nature, such as water column stabilization, nutrient availability, and predation. Variations in any one of these processes can potentially propagate through to the bloom dynamics, i.e., initiation, peak, duration, and amplitude [*Platt and Sathyendranath*, 2008]. Consequently, this expected sensitivity to external processes leads to phenology often being proposed as an indicator of both climate variability and climate change [*Edwards and Richardson*, 2004; *Thackeray et al*., 2010].

The phenology of phytoplankton blooms is also linked to the survival of many fish species during their larval and juvenile stages by the so-called "match-mismatch hypothesis" [*Cushing*, 1990]. The hypothesis states that a key factor in determining the survival of larval fish is the match (or mismatch) between the timing of the period when larvae feed on plankton and the timing of the local spring phytoplankton bloom. In years when the bloom is delayed relative to the timing of larval feeding, there is, therefore, insufficient food available during the larval stage [*Cushing*, 1990], and larval mortality is greatly increased. Conversely, a precondition for strong year classes is that they can only occur in years when there is a good match and sufficient food. Despite limited empirical data, there have been some studies supporting this hypothesis [e.g., *Platt et al*., 2003; *Koeller et al*., 2009]. The timing of the spring phytoplankton bloom can therefore be considered as a key concept in understanding the response of marine ecosystems to environmental change.

The study of phenology typically involves extracting representative metrics from time series of observations. Here we distinguish between phenology indicators (the idealized characteristics of the phytoplankton bloom) and the phenology metrics (real-world estimators of the indicators). *Platt and Sathyendranath* [2008] have suggested a suite of indicators suitable for use in the study of phytoplankton production cycles. The two most common indicators are the timing of the maximum chlorophyll concentration during the spring bloom [e.g., *Yoder and Kennelly*, 2003; *Rolinski et al*., 2007], and the timing of initiation of the spring bloom [e.g., *Cole et al*., 2012; *Brody et al*., 2013]. Each of these indicators represents a different aspect of the bloom dynamics. For example, the timing of the chlorophyll maximum results from a combination of the reassertion of predation control and nutrient limitation, whereas the timing of bloom initiation is typically governed by the balance between (light-controlled) growth and respiration losses [*Sverdrup*, 1953]. *Brody et al*. [2013] proposed that the definition from *Siegel et al*. [2002] (a 5% increase in chlorophyll concentration above the annual median) would be more appropriate for studies on the match-mismatch hypothesis, whereas the definition described by *Roerink et al*. [2000] (a rate of change method based on harmonic analysis of the time series) would be useful for studies on seasonal physical and biological mechanisms. The choice of indicator (and its corresponding metric), therefore, needs to be made in the context of the question being asked.

Furthermore, each individual phenology metric (e.g., timing of initiation) can be expected to have its own unique degree of variability. Some may vary widely, while others may be nearly constant, depending on what they are indicating and the variability of the processes that contribute to them. An ideal phenology metric needs to exhibit some form of variability that can be linked to environmental variation rather than simply being noise.

The majority of commonly used phenological metrics apply algorithmic approaches rather than statistical estimators [*Ji et al*., 2010]. Algorithmic approaches [e.g., *Siegel et al*., 2002; *Racault et al*., 2012] assume that all data points correspond to the bloom signal. Statistically based estimators [e.g., *Zhai et al*., 2012, 2011], on the other hand, explicitly include the observational process in their model structure and therefore allow the model to distinguish between signal and noise. Algorithmic phenology metrics are, thus, vulnerable to observational noise (and potentially to nonuniform sampling-missing data), resulting in possible biases [*Cole et al*., 2012; *Behrenfeld et al*., 2013], and poor precision [*Maritorena and Siegel*, 2005; *Verbesselt et al*., 2010; *Ballabrera-Poy et al*., 2003].

Multiple processes can impact the resulting reliability and robustness of these estimators. Some of these processes are inherent to the observation process and are therefore beyond the direct control of the experimenter, such as missing data, the presence of observational noise and the amplitude of the underlying signal relative to the noise. In satellite-based phenological studies, missing data mainly result from the pattern of the satellite track over the surface of the earth, the presence of clouds and winter darkness, especially in high-latitude regions [*Cole et al*., 2012; *Sasaoka et al*., 2011]; changes in the satellites, including outages, can also be important. Missing data may reduce the precision of the estimated metric, and lead to systematic biases in the calculation method [*Cole et al*., 2012]. In addition, appreciable observational noise is present in satellite-derived estimations of chlorophyll concentrations [*Maritorena et al*., 2002]. Comparisons with *in situ* observations suggest that individual chlorophyll retrievals typically have a coefficient of variation (or noise) of 0.3–0.4 in each individual retrieval [*Maritorena and Siegel*, 2005; *Maritorena et al*., 2002]. Such a high coefficient of variation can readily disguise an underlying chlorophyll signal, especially in algorithmic approaches. Furthermore, chlorophyll bloom amplitude (and shape) is not constant in space and the importance of observational noise is expected to be less in regions with large bloom amplitudes [*Behrenfeld et al*., 2013], but become more significant in less variable regions (with low signal-to-noise ratios, i.e., low seasonal amplitude versus noise level) [*Verbesselt et al*., 2010; *Ballabrera-Poy et al*., 2003]. Therefore, the impacts of missing data, observational noise, and bloom amplitude need to be considered when assessing the accuracy and precision of bloom phenology metrics.

The manner in which the analysis is performed and the phenology algorithms are applied can also impact their robustness. Some authors, but not all [e.g., *Cole et al*., 2012], use gap-filling (as discussed by *Brody et al*. [2013]) and/or smoothing/filtering techniques [*Racault et al*., 2012] to preprocess the data prior to algorithm application. Furthermore, the choice of the temporal resolution, i.e., the time window over which observations in a given pixel are aggregated, can also be important. Although satellite ocean color instruments typically provide products at daily resolutions [*Polovina et al*., 2001; *Werdell and Bailey*, 2005], albeit with high

gap proportions, most recent studies have used 8 day binned data [*Cole et al*., 2012; *Brody et al*., 2013]. However, since phytoplankton biomass can exhibit significant increases at a much higher rate (daily or even faster), particularly in the runup to bloom conditions, some authors [*Rolinski et al*., 2007] suggested that weekly data may not capture the underlying dynamics sufficiently well. Notwithstanding, there are also phenology studies using monthly in situ data [*Edwards and Richardson*, 2004; *Beaugrand and Reid*, 2012]. Ideally, the choice of preprocessing technique and time resolution of the product should be appropriate to capture the processes involved [*Ji et al*., 2010] and is therefore closely linked to the underlying question being studied.

Relatively little is known about the impact of the above-described processes on the reliability of commonly used phenology metrics. Previous studies have examined the effect of missing data on the accuracy of phenology metrics by using 8 day data to reveal systematic and spatially dependent biases [*Cole et al*., 2012]. However, many questions remain unresolved; in particular, the relationship between precision and accuracy has not been quantified, nor have the effects of observational noise. Furthermore, the relatively limited number of processes previously considered [*Cole et al*., 2012] makes it hard to draw general conclusions about the most appropriate choice of phenology metric.

In this study, we aim to identify the most robust phenology indicators, and their associated metrics. We consider the inherent variability of the phenology indicator, and how its perception in a real-world-like scenario is affected by missing data, observational noise, and bloom amplitude. We further aim to assess the effect of preprocessing technique and temporal resolution when generating phenology metrics. Our focus is on assessing the sensitivity, precision and accuracy of these metrics when applied to (i) a representative set of time series and (ii) over a large geographical extent. We follow the study design of *Cole et al*. [2012] and apply a simulation-testing approach. However, we greatly expand the range of processes and phenology metrics considered. The insights gained here should therefore greatly improve our understanding of the behavior, strengths, and weaknesses of commonly employed phenology indicators and metrics.

2. Material and Methods

2.1. Data Sets

The NASA Ocean Biogeochemical Model (NOBM, http://www.nasa.gov/) is a three-dimensional global ocean-biogeochemical model [*Nerger and Gregg*, 2008]. The biogeochemical component comprises four nutrient cycles, four phytoplankton functional types, a single zooplankton group, and three detrital reservoirs. The model is assimilative in nature and incorporates chlorophyll observations from SeaWiFS on a daily basis. The assimilative nature of the model means that it is well suited to the characterization of phytoplankton population dynamics and has previously been used in this type of phenology simulation-testing study [e.g., *Cole et al*., 2012]. We used daily composites of chlorophyll concentration from 1998 to 2007, with increments of 1.25° longitude and of $2/3^\circ$ latitude grid from the North Atlantic, from 0 to 70°N (northern boundary of NOBM data), and from -80 to 20°E.

We used observational products derived from the European Node for Global Ocean Colour (GlobColour Project, http://www.globcolour.info/). The GlobColour Project blends observational data from the SeaWiFS, MODIS-AQUA, and MERIS instruments by using the Garver-Siegel-Maritorena (GSM) algorithm [*Maritorena et al*., 2002] to generate a merged, global ocean color product. Combining the three sensors increases the data coverage in both time and space, thus providing significantly elevated spatiotemporal coverage [*Maritorena et al*., 2010], making it a common and appropriate choice for phenology studies [*Cole et al*., 2012]. GlobColour observational data were gathered from 1998 to 2012 for the three available temporal resolutions: daily, weekly, and monthly, with a $1/4^\circ$ grid. Averages of the values of all grid cells within each $2/3^\circ$ were calculated to match the NOBM grid ($2/3^\circ$).

2.2. Simulation-Testing Approach

We use a simulation-testing approach (Figure 1) to investigate the importance of various processes on the robustness of phenology metrics. This involves the identification of a data set to serve as the underlying truth in the first instance, and then its subsequent modification by adding observational noise and gaps to generate a set of synthetic observations. We apply the phenology algorithms both on the simulated "truth" and "observations" to generate two separate realizations of the phenology metric under study. The

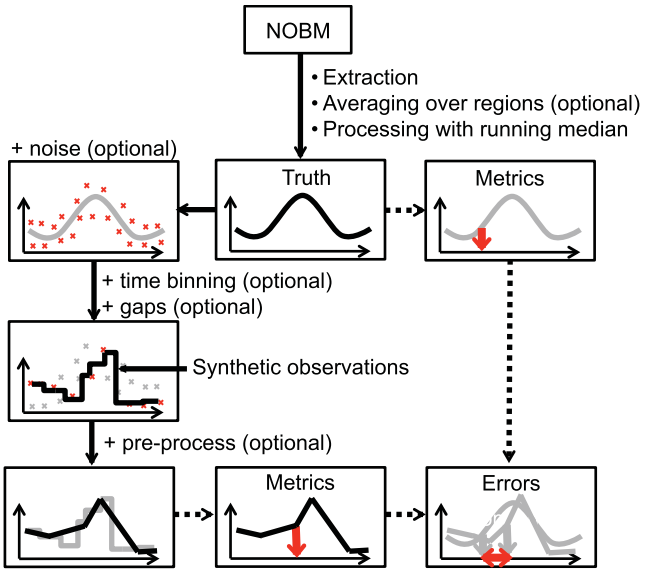

Figure 1. Schematic figure of methodology used. Optional indicates that the step was not applied in all data sets.

discrepancy between these measures (the phenology error) is the key quantity of interest that we use to quantify the robustness of a phenology algorithm. We used the NOBM output to act as the truth (following *Cole et al*. [2012]) and the characteristics (underlying missing data and observational noise structure) of the GlobColour products as the basis for generating simulated data.

Our study involves the exploration of six dimensions: observational noise, missing data, bloom amplitude, temporal resolution, preprocessing technique, and phenology metric. The first three of these (the observation-related processes) are inherent properties of satellite ocean color observations and are therefore outside the control of the experimenter. However, the remaining three (the analysis-related processes) are part of the process of generating a phenology metric and can therefore be controlled. It is important to note that some of these processes covary; for instance, bloom amplitude covaries in space with both missing data and observational noise. Furthermore, as we are interested in both the accuracy and precision of the phenology metrics, these two outputs need to be considered for each treatment. The high-dimensionality of the space that we wish to explore makes the problem extremely challenging.

To reduce the complexity of the problem, and allow ready visualization of the results, we split the problem into stages. We first focus on the analysis-related dimensions, i.e., temporal resolution, preprocessing technique, and phenology metric, and examine their impact on the precision and accuracy of phenology metrics in a set of representative case studies (the time series study). The results of this first stage are then used to refine the analysis and reduce the dimensionality of the problem in the second stage. The second study focuses on the observation-related dimensions, i.e., bloom amplitude, missing data, and observational noise. As each of these processes are spatially variable, we therefore choose to examine their impacts in a spatial context (the spatial study). Together, these two studies should give us fresh insight into the robustness of the various metrics.

2.3. Data Processing

The general work flow employed in each study (time series and spatial) was as follows (Figure 1). Details of each process are described in subsequent sections.

2.3.1. Generation of Synthetic Truth

1. Retrieve biogeochemical model daily data.
2. Calculate daily averages over the three spatial regions of interest (for use in the time series study).
3. Apply a 5 day running median per region (in the time series study) or per pixel (in the spatial study): the resulting data sets are the "truth" data.

2.3.2. Generation of Synthetic Observations

1. Starting with the appropriate "truth" time series, add appropriate level of observational noise, which is either fixed (time series study, supporting information Table ts01) or with a spatial structure (spatial study, Figure 2, bottom).
2. Aggregate the daily resolved data to weekly and monthly synthetic observations by binning and averaging (where appropriate).

3. Remove some of the observations, either by sampling following a uniform random distribution at the appropriate level of missing data (time series study, supporting information Table ts01), or by applying masks (spatially resolved structure of missing data) from the satellite data sets from the three temporal resolutions (spatial study, see an average in Figure 2, middle) to each cycle.

In the time series study, the above-mentioned steps were iterated 100 times for each "truth" to generate an ensemble of simulated observations for each set of treatment (combination of levels). In the spatial study, following the approach of *Cole et al*. [2012], the ensemble was generated by crossing the truths in each year estimated from NOBM (from 1998 to 2007, 9 years in total) with the missing data masks from the GlobColour product (from 1998 to 2012, 15 in total) to give a maximum of 135 simulated sets of observations for each treatment at each pixel.

2.3.3. Estimation of Phenology Metrics and Errors

1. Identify the timing of the climatological maximum for each region (time series study) or pixel (spatial study). A seasonal cycle was then defined in relation to this maximum covering a window from 6 months prior to the maximum to 6 months after, as previously performed by *Cole et al*. [2012]. Seasonal cycles where the data set was incomplete (i.e., at the ends of the time series) were not considered in the analysis.
2. Apply filtering and smoother preprocessing techniques to the synthetic observations, as appropriate.
3. Estimate phenology metrics for the truth data sets.
4. Estimate metrics for each set of synthetic observations (100 in the time series study, 135 in the spatial study).
5. Estimate the phenological errors by subtracting the phenological metric derived from the truth data from that obtained from the synthetic observations. Negative error values indicate earlier bloom indices, whereas positive values indicate delayed blooms.
6. Summarize result distributions for each analysis unit.

2.4. Observational Noise

Observational noise was applied to each truth data set by generating a normally distributed random noise data set with a mean of 0 and a standard deviation corresponding to the coefficient of variation (observational noise) level (see supporting information Table ts01 for the time series study, and see Figure 2 for the spatial study). This new data set was then added to the natural log-transformed data sets of chlorophyll concentration to generate synthetic observations with log-normally distributed noise (upon back-transformation).

2.5. Temporal Resolution

Daily data were binned into weekly and monthly resolutions by calculating geometric means of chlorophyll concentration (mean of natural log-transformed data) over the corresponding time bin.

2.6. Missing Data

Data gaps were generated in the time series study by a uniform random sampling of data, with percentages of missing data appropriate for each region (supporting information Table ts01): the sampled daily indices were removed from the corresponding data set. In the spatial study, masks of missing data were obtained from satellite observations (Figure 2). Missing observations in the satellite data set were removed from the corresponding pixel on the corresponding day/week/month in the synthetically generated data sets.

2.7. Preprocessing Techniques

On daily, weekly, and monthly data, we applied a linear interpolation between neighboring observations to fill in the gaps. In addition, a linear interpolation followed by a 3 week (21 days) centered-running mean was also used for daily and weekly data to fill in the gaps and smooth noise levels. For all temporal resolutions, the effect of not preprocessing the observational data was also considered.

2.8. Phenology Metrics

We chose to compare four commonly used metrics (Figure 3). The first is the timing when chlorophyll rises 5% above the median value of the particular seasonal cycle (TMED) [*Cole et al*., 2012; *Henson et al*., 2009; *Racault et al*., 2012; *Siegel et al*., 2002; *Brody et al*., 2013]. However, we do not require chlorophyll levels to

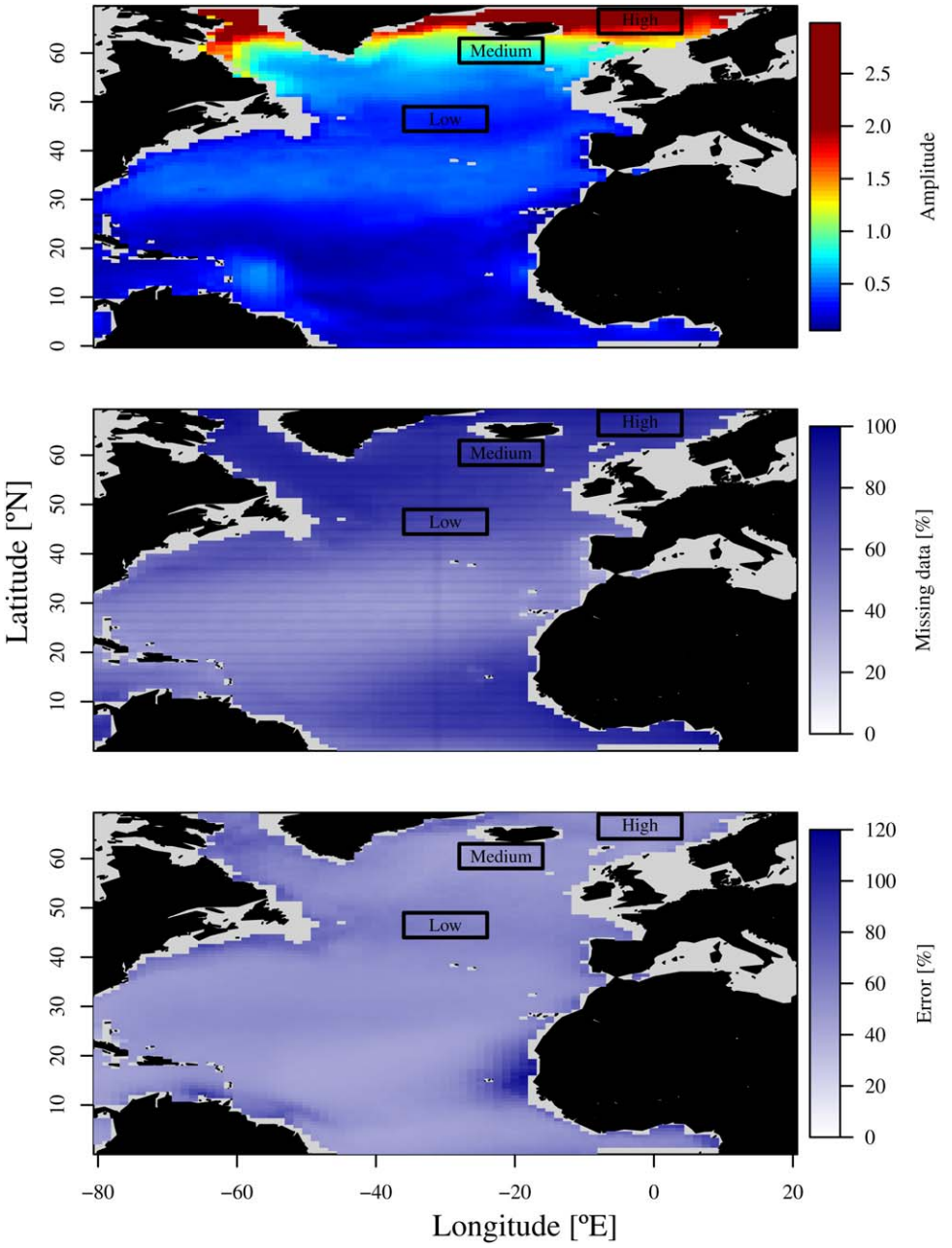

Figure 2. (top) Chlorophyll "amplitude" seasonality from NOBM data shown for each bloom amplitude (low, medium, and high amplitudes) in the time series study (black boxes). Chlorophyll seasonality is defined here as the standard deviation of the seasonal natural log-chlorophyll climatology and is therefore equivalent to a coefficient of variation. (middle) Percentage of missing daily data from the GlobColour Project used in the spatial study. (bottom) Average percentage of error (as an indicator of observational noise) from the GlobColour Project used in the spatial study.

stay above the threshold for two consecutive weeks as *Cole et al*. [2012] and *Brody et al*. [2013] did. The second metric is the timing when chlorophyll reaches 15% of its cumulative distribution (TCUD) [*Greve et al*., 2005; *Brody et al*., 2013]. The third metric is the timing of maximum growth rate (TMGR) using the approach by *Rolinski et al*. [2007] which notably lacks the harmonic smoother used elsewhere [*Brody et al*., 2013; *White et al*., 2009; *Sharples et al*., 2006; *Wiltshire et al*., 2008; *Brody and Lozier*, 2014]. However, a smoother is still required since there is high-frequency variability in both the truth and synthetic data sets, which the TMGR metric would otherwise pick out. We therefore apply a 7 day running mean to smooth the data set before estimating the highest daily increase. The fourth metric is the only metric assessing the chlorophyll peak

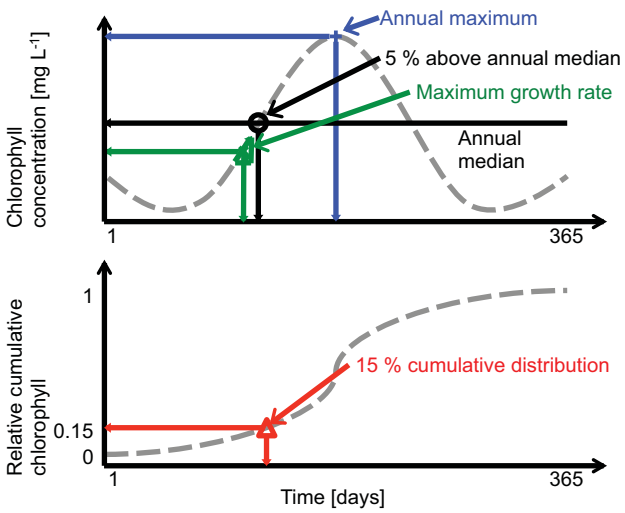

Figure 3. Schematic figure of metrics used. Metrics are defined as timing of: maximum (TMAX), 5% above median (TMED), maximum growth rate (TMGR), and 15% of cumulative distribution (TCUD).

and is straightforward: the timing of maximum chlorophyll concentration (TMAX) [*Yoder and Kennelly*, 2003; *Cole et al*., 2012; *Rolinski et al*., 2007]. These four metrics cover the majority of the approaches employed in the published literature [*Ji et al*., 2010].

2.9. Error Statistics Distribution

Phenological error distributions (in days) were summarized to facilitate interpretation. Means were used as a measure of accuracy, while standard deviations of the distribution were used as a measure of precision.

3. Results

Our analysis is based on two different approaches: a time series study, where the effect of analysis-related processes is examined; and a spatial study, where realistic and spatially varying observation-related processes are considered. In each case, we aim to quantify the accuracy and precision of the error metrics, and use this as a basis to assess their robustness. Here we consider each study in turn.

3.1. Time Series Study

The time series study attempts to quantify the impact of the analysis-related dimensions. To do so, it is first necessary to eliminate the impact of the observation-related processes, such as bloom amplitude, missing data, and observational noise (Figure 2, top, middle, and bottom, respectively), allowing the impact of the analysis-related processes to be compared independent of the observation-related processes.

The outputs of the NOBM and GlobColour products were used as the basis for selecting regions that are representative of the North Atlantic basin. Phytoplankton blooms from the NOBM data exhibit the well known increase in relative amplitude with increasing latitude in this region. The frequency of missing data from GlobColour, on the other hand, shows a dipolar pattern, with regions of high proportions of missing data in high latitudes (due to winter darkness) and in low latitudes (due to increased cloudiness associated with the tropics). The precision of GlobColour chlorophyll retrievals is typically lowest (highest error) in regions near land but is relatively constant. We therefore chose three representative regions based on these results (plotted in Figure 2), approximately corresponding to the intergyre region (or the transition as in *Henson et al*. [2009]), the subpolar gyre, and the Norwegian Sea (see supporting information Table ts01 and Figure 3).

The distribution of the phenology errors was determined for each region and temporal resolutions considered (supporting information Figure fs01). It is generally broad and includes zero error in nearly all cases. There are, however, large differences occurring in the spread of the error distributions between metrics and regions. The broad distribution of phenology errors is clearly problematic for the interpretation of phenology metrics and in most practical cases can be expected to be more important than any potential bias. We therefore focus on the precision of the phenology metrics.

Calculating the standard deviation of the phenology errors (as a measure of precision) for each combination of treatments allows the impact of each treatment to be simplified for each metric and visualized (Figure 4). The cases where the phenology metrics are the least precise is in the low bloom amplitude region: TMED is particularly sensitive to low bloom amplitudes. TMGR is consistently the least precise (highest standard deviation). TMAX and TCUD are consistently the most precise metrics across all treatments. Increasing the degree of preprocessing improves the precision of the error estimations in all cases. We also see a weak trend of decreasing precision (increasing standard deviation) with decreasing temporal resolution (as we

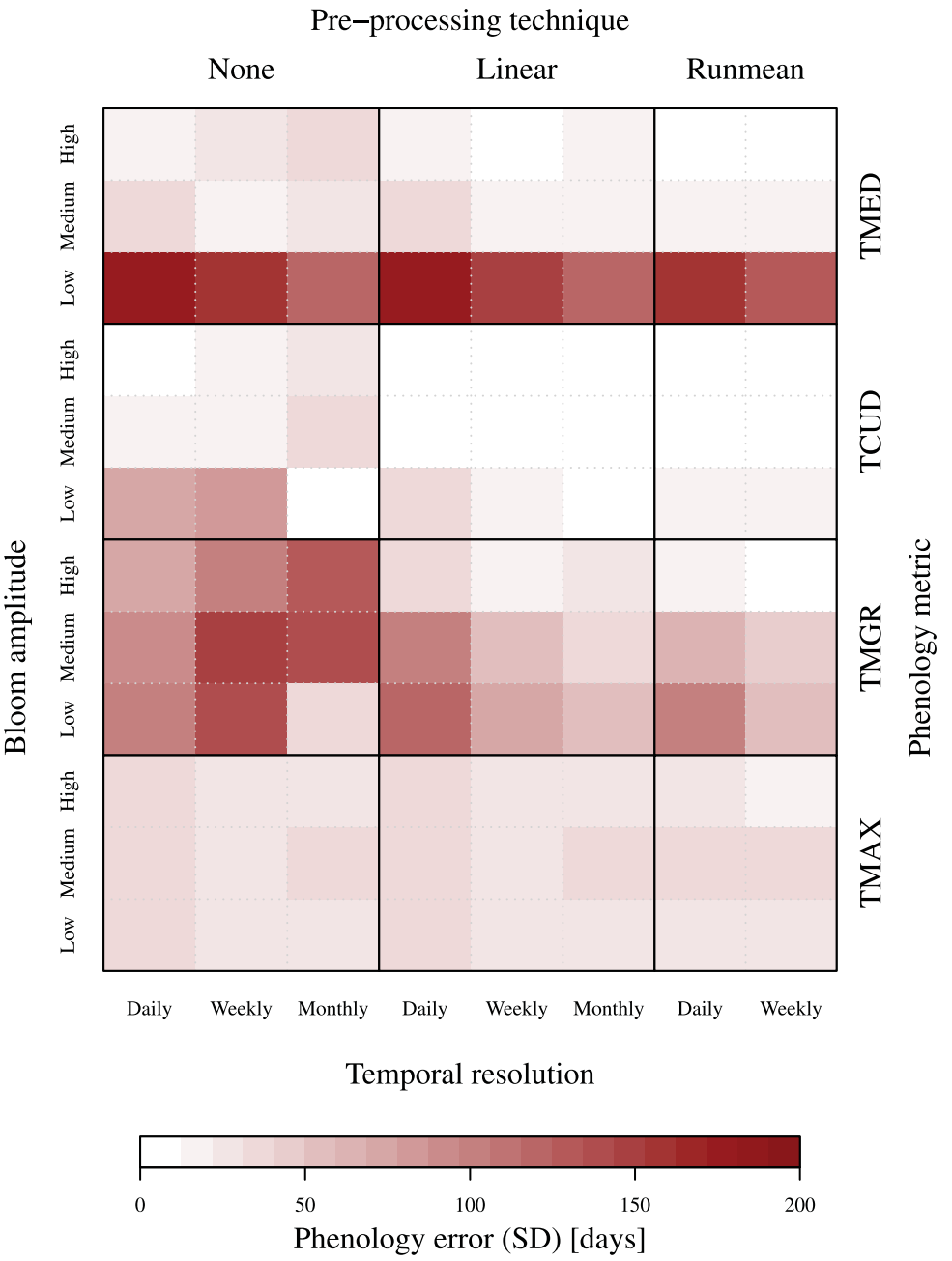

Figure 4. Spread of phenology errors (standard deviation as a measure of precision) from the time series study for each phenology metric, bloom amplitude, temporal resolution, and preprocessing technique.

move from daily to monthly). These general trends are also reflected in accuracy, although less pronounced (supporting information Figure fs02).

3.2. Spatial Study

The time series study suggests several useful conclusions that can be used to refine the spatial study. In particular, the poor precision in many of these metrics appears to be a greater concern than any systematic biases. The addition of preprocessing also appears to generally improve the precision of the metrics, while temporal resolution does not have a strong effect. The spatial study therefore focuses on a single temporal resolution (daily) and preprocessing technique (preprocessing with both linear interpolation and 3 week

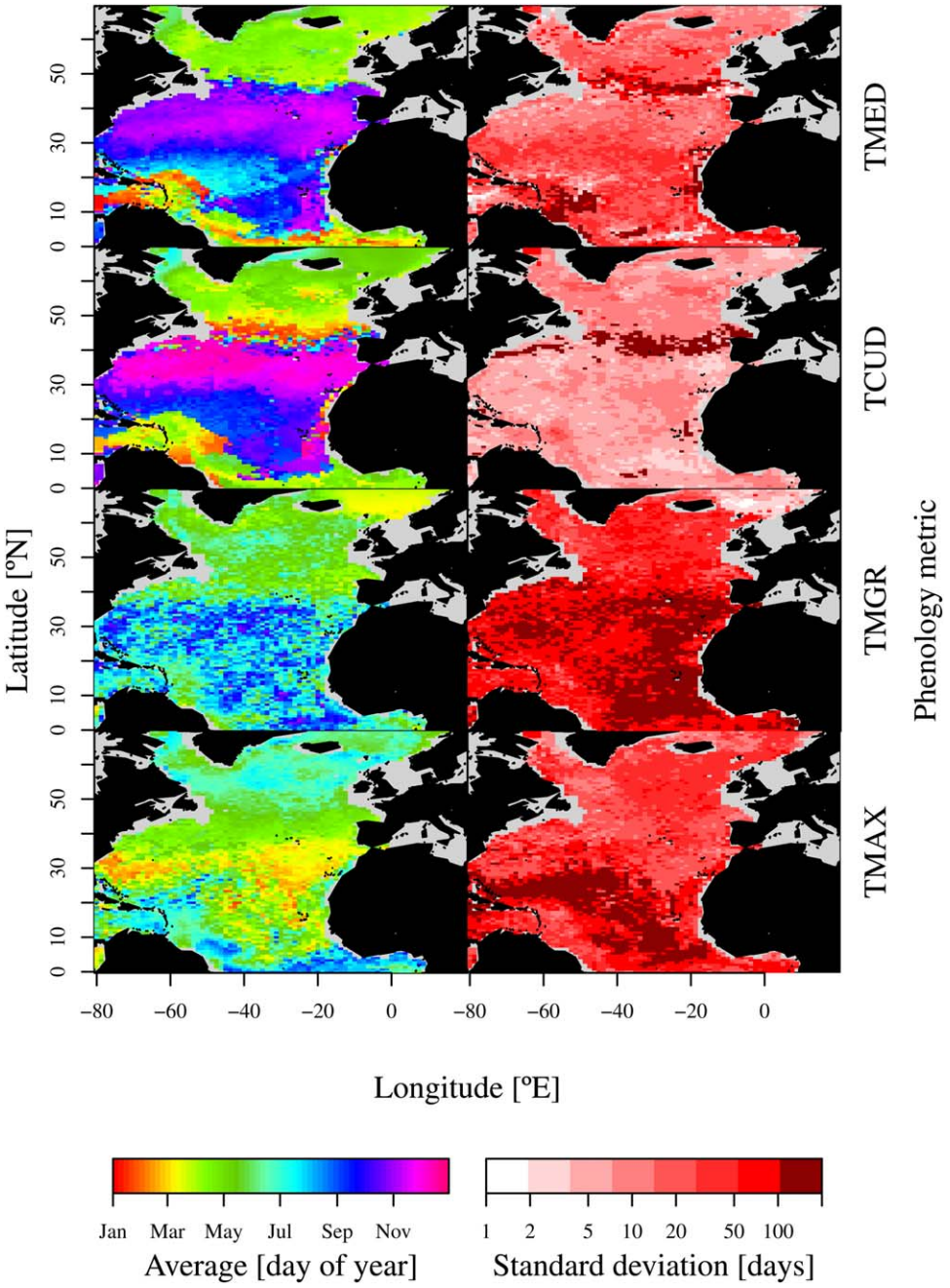

Figure 5. Average (or bias, as a measure of accuracy, left, linear scale) and standard deviation (or spread, as a measure of precision, right, logarithmic scale) of the true metric estimations (i.e., prior to the addition of noise and gaps) for each phenology metric from the spatial study. Dark red indicates the errors higher than 100 days.

running mean). The distribution of errors for other temporal resolutions and preprocessing techniques has been calculated, and is generally similar, but the results are not presented here.

First, we examine the inherent variability within, and between, these metrics in a spatial setting. Each of the four metrics (TMED, TCUD, TMGR, and TMAX) are applied to the "truth" data in the absence of both observational noise and missing data across the entire North Atlantic basin. All metrics exhibit a systematic trend of increasingly later blooms from low latitudes to high latitudes and a systematic difference between the subtropical and subpolar gyres (Figure 5). The Gulf Stream region is notable for the high degree of interannual variability, particularly in the bloom initiation indicators. With regards to the individual metrics, the TMGR metric shows an extremely high degree of interannual variability, and is consistently the most variable.

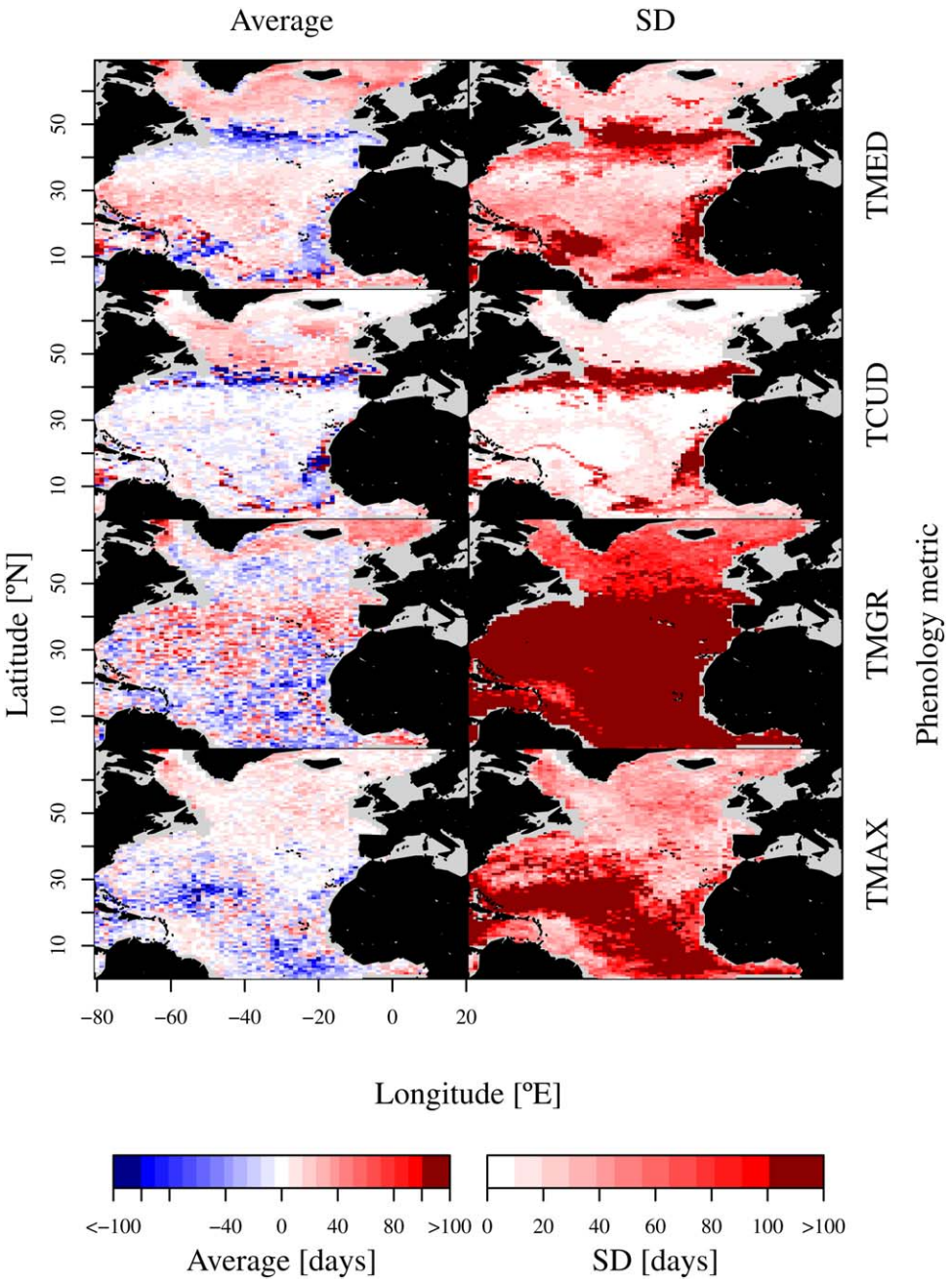

Figure 6. (left) Average and (right) standard deviation of daily errors from the spatial study estimated for each metric (TMED, TCUD, TMGR, and TMAX). Only results on data preprocessed with both linear interpolation and 3 week (21 days) running mean are shown. Negative values indicate earlier bloom indices, whereas positive values indicate delayed indices.

TMAX is highly variable in low latitude waters, consistent with the extremely low amplitude seasonal dynamics (and therefore poorly defined "blooms") [*Cole et al.*, 2012] in these regions. These patterns agree well with those reported elsewhere in the literature [*Henson et al.*, 2009; *Racault et al.*, 2012; *Brody et al.*, 2013]. The TCUD metric exhibits an interannual variability that was appreciably lower than any of the other metrics except in the midlatitudes.

The accuracy and precision of each metric can be visualized readily as a function of space (Figure 6). The spatial patterns associated with each phenology metric were generally quite different. TMED and TCUD show similar patterns, with generally positive average errors but with a region of high spread (low precision) and negative mean (negative bias, thus low accuracy) errors around the Gulf Stream. In contrast with the

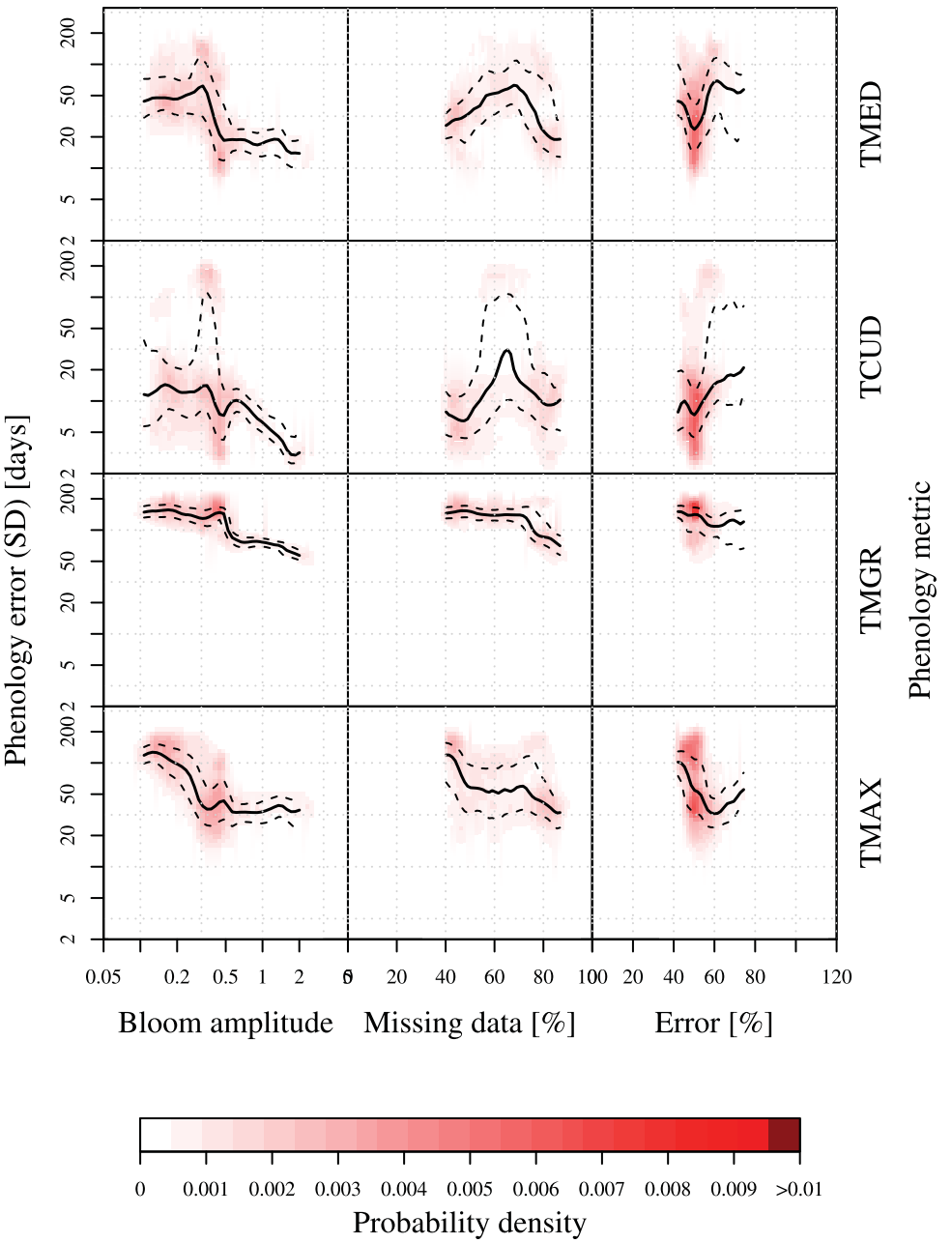

Figure 7. Spread of phenology errors (standard deviation as a measure of precision) from the spatial study plotted as a function of bloom amplitude, missing data, and observational noise (error) for each metric. Estimation errors were binned for plotting purposes and expressed as a probability density distribution. Results are based on daily data preprocessed with both linear interpolation and a 3 week (21 days) running mean smoother. Thick, black line shows the median values and the dotted lines show the first and third quartiles.

other metrics, TMGR shows a noisy spatial pattern in the mean error, and a broad spread of values, particularly south of 40°N. TMAX shows relatively accurate and precise estimations, although with reduced precision and potential bias at low latitudes. Similar results are observed for weekly data (supporting information Figure fs03), but with less extreme values (thus more accurate and precise errors).

The processes driving the phenology errors can be understood by individually investigating the impact of the bloom amplitude, missing data, and observational noise (Figure 7). The bloom amplitude is clearly the most important of the three observation-related processes, where the spread of errors increases with increasing bloom amplitude. Important to note is the increased precision of TMAX with increasing missing data. However, the general effect of both missing data and observational noise (error) is less clear and lacks

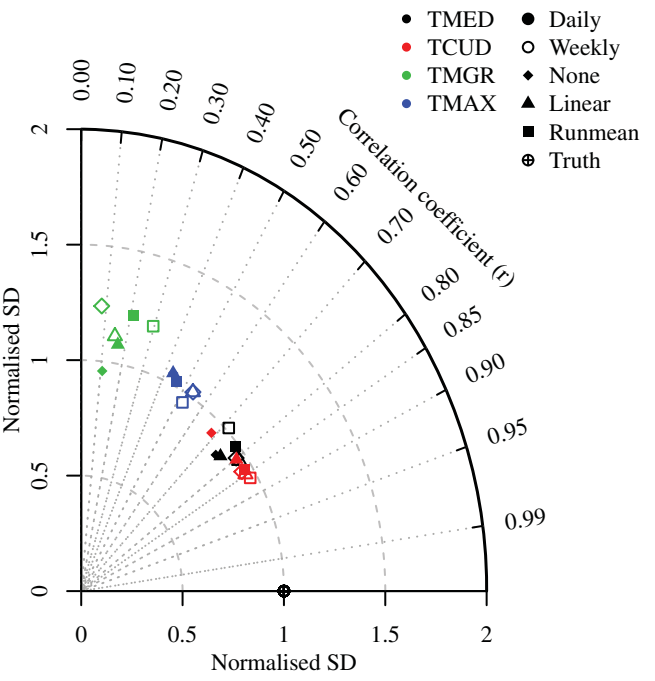

Figure 8. Taylor diagram of estimations of the four phenology metrics (TMED, TCUD, TMGR, and TMAX colors) for the 135 simulations (9 × 15 seasonal cycles) for each combination of temporal resolution (filled/nonfilled symbols), and preprocessing technique (shapes of symbols) from the spatial study integrated for each pixel. Standard deviations were normalized to allow comparison between indicators. Correlations represent the relationship between each estimation and the truth for each combination of phenology metric, temporal resolution, and fitting technique.

systematic trends across metrics. Similar results are observed for the accuracy of the same data (supporting information Figure fs04), although less extreme.

Finally, insight into the performance of each metric across the entire basin can be gathered from a Taylor diagram (Figure 8). In contrast to previous ways of visualizing these results, the Taylor diagram [*Taylor*, 2001] incorporates the correlation between the phenology metrics obtained from the synthetic observations and those obtained from the original time series. This method is particularly useful, as it indicates the ability of a metric to identify a basin-level signal (underlying variability in phenology) in the presence of noise. This diagram shows that the TCUD and TMED metrics are closest to the underlying truth, with TCUD being marginally superior. All metrics reproduce the amplitude of the variability well (i.e., are close to a normalized standard deviation of 1). However, the correlation between the truth and the estimated phenology metric is highly variable: TMGR shows the correlation coefficients between 0.1 and 0.3, TMAX between 0.4 and 0.6, TMED between 0.7 and 0.85, and TCUD between 0.7 and 0.9. It is also interesting to note that both temporal resolution and preprocessing techniques have relatively minor effects on this result compared to the inherent differences between the metrics.

4. Discussion

One of the first and most important results in this work is the relative importance of accuracy and precision. In many situations, the precision of the phenology metrics (i.e., spread or standard deviation of the phenology errors) was found to be relatively poor and was typically much greater than any systematic bias (i.e., the accuracy or mean of the phenological errors). Other similar studies [e.g., *Cole et al*., 2012] examining the properties of phenology metrics have focused solely on the accuracy of the metric, and presented mean biases. However, our results clearly indicate that an important aspect of these metrics has been overlooked. In the following sections, we examine each of the processes that can potentially impact the precision of these metrics.

4.1. Analysis-Related Processes

The preprocessing techniques applied to the data before estimating phenology metrics proved to be the most important factor that can potentially be controlled by the experimenter for a given metric. Our analysis suggested that preprocessed data are likely to provide estimates of bloom initiation date with improved precision (Figure 7 and supporting information Figure fs04). When looking at the same analysis using data with no preprocessing technique applied (supporting information Figures fs05 and fs06), increasing the intensity of the preprocessing methods considered here to include both gap-filling and smoothing showed improvement in the precision of the metrics in all cases. In contrast, *Brody et al*. [2013] found large effects on bloom initiation dates when time series were filled with the annual minimum, and advised against this

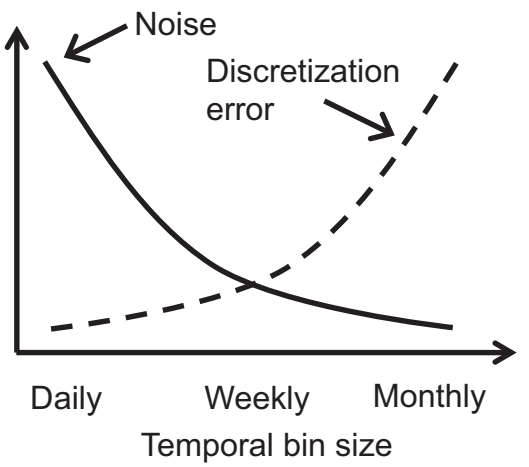

Figure 9. Observational noise (full line) decreases with temporal resolution, while discretization error (dashes line) increases. Therefore, there is a trade-off between these two processes when choosing between low and high temporal resolutions. Weekly data seem to have an intermediate impact of noise and discretization error, due to an averaged noise level and an uncertainty error that is still not too high.

technique. Similarly, it is possible to envisage a situation where the smoothing is too severe, and metric performance begins to deteriorate as a result. The choice of preprocessing therefore needs to be made with care, and preferably tested, prior to use.

The temporal resolution of the data was, surprisingly, one of the least important variables. Part of the reason for this may lie in the overwhelming impact of the other processes. Once these issues have been resolved, temporal resolution can still be expected to be important. On one hand, reducing the temporal resolution can be considered as a form of preprocessing: the binning procedure fills gaps, reduces noise and smooths the data (by averaging over the bin) (Figure 9). However, as *Ji et al*. [2010] noted, short-term variability also needs to be taken into account in order to avoid biased conclusions drawn from phenology studies, and, at some point, the discretization error (due to the low temporal resolution) can be expected to become important. In our study, we observed a general reduction in precision with decreasing the temporal resolution from daily to monthly data. However, noting that temporal resolution is the analysis-related variable that appeared to have the least impact on the phenology metric, it is unclear whether our survey has sufficient power to resolve such shifts. A more focused study is most likely required to identify the optimal temporal resolution.

4.2. Observation-Related Processes

The precision of phenology metrics is influenced by numerous processes beyond the control of the experimenter, including bloom amplitude, observational error, and missing data. The bloom amplitude was found to be the most important of these processes, with an increase in amplitude leading to improved precision for all metrics. An important concept in understanding this result is the signal-to-noise ratio of the data. As suggested by *Verbesselt et al*. [2010], phenological changes may not be effectively detected when noise levels are larger than the signal amplitude. Therefore, the relative impact of observational noise is expected to be higher in lower latitudes, where the bloom amplitude is small, as the signal may be masked by large relative uncertainties in the data. Similarly, high amplitude regions have higher precision.

Increasing observational noise does not have a clear impact on the precision of the metrics. However, the range of observational errors encountered was relatively small (the majority of points lie between 40 and 60%) and there might therefore not be sufficient contrast to reveal these patterns. This is not to say that noise is not important. However, we recognize that the effect of observational noise may be masked out by the bloom amplitude. In regions where there is a high bloom amplitude (high latitudes), the observational noise becomes less important and both precision and accuracy are not as high as in low bloom amplitude regions. This is also a potential explanation for the breaks (discontinuity around 50°N) in estimated phenology metrics both from the truth (Figure 5) and from simulated observations (Figure 6).

The frequency and pattern of missing data are expected to be of importance when calculating phenological metrics. However, missing data had a limited and inconsistent effect on the precision of the phenology metrics. Such a result is surprising when taken in light of other results [e.g., *Cole et al*., 2012]. However, the explanation arises from the fact that we have mainly focused on data that has been preprocessed (i.e., gap-filled and smoothed) and therefore suggests that these preprocessing steps are effective. Indeed, when the preprocessing is removed, gaps in the data do have stronger effects (note different patterns in supporting information Figures fs05 and fs06). Additionally, for some cases (e.g., TMAX in Figure 7), the increase of missing data showed an increased precision. This is possibly due to the covariance nature between missing data

and bloom amplitude at high latitudes, where high levels of missing data do not affect the estimation of the phenology metric. Gaps in the data therefore need to be filled prior to analysis, at least for the range of phenology metrics considered here.

4.3. Phenology Metrics

Our study revealed several key points about the inherent variability of phenology metrics. First, there was a substantial difference in the interannual variability exhibited by the individual metrics (Figure 5). The TCUD metric, for example, exhibited a degree of interannual variability that was substantially less than the TMED and TMAX metrics, while the TMGR metric was clearly the most variable. These differences are fundamental characteristics of the metrics themselves, occurring even in the absence of both gaps and noise, and are most likely related to the way they are calculated. For example, this is mainly due to TMGR being calculated as a derivative of the time series, and thus highly sensitive to noise and local fluctuations. TCUD, on the other hand, is based on an integrated biomass (cumulative distribution) and thus has an inherently smoothing nature. As in many applied instances from statistics to particle size distributions [e.g., *Jackson et al*., 1997] to life-history theory [e.g., *Leaman*, 1991], cumulative metrics present robust means by which rates of change can be estimated. The remaining two metrics, TMED and TMAX, which are based on direct analysis of the time series (i.e., are neither integrated nor differentiated) are intermediate between the two extremes.

It is worth emphasizing the "poor" performance and high variability of the TMGR metric. The timing of the maximum growth rate is an inherently attractive phenology indicator, as it represents the point in the bloom cycle where growth conditions are optimal, and where the system is changing the fastest. It is also one of the most visually apparent and intuitive features of the bloom. However, in spite of the attractiveness of this indicator, the metric that attempts to characterize it is highly variable and imprecise, as it is based, in the first instance, on derivatives of (noisy) time series. This observation also points toward future opportunities. The implementation of the TMGR metric [*Rolinski et al*., 2007] that we have chosen here is one of the simpler versions published in the literature, and has been applied with a limited range of preprocessing techniques. A wide range of more sophisticated preprocessing techniques exists, and could potentially improve the precision of the TMGR metric: other authors have preprocessed their data prior to phenology extraction using harmonic analysis [*Brody et al*., 2013], the Census X-11 time series method [*Vantrepotte and Mélin*, 2011], the Hilbert-Huang filter [*Palacz et al*., 2011], and generalized linear models [*Vargas et al*., 2009; *Sapiano et al*., 2012], amongst others. Unfortunately, it was not practical to take all of these possible methods into account within the scope of this work. The appropriate choice of such preprocessing techniques, and its impact on the precision of the TMGR metric, is therefore an avenue for future research.

We earlier distinguished between phenology indicators and phenology metrics. The phenology metrics considered in this study are each representative of a different bloom characteristic, and thus they should rather be considered as indicators. However, we are not aware of any indicators in the literature that can truly be considered as indicators of bloom initiation, i.e., the point at which biomass increases outweigh those due to losses. Both TMED (5% above the median) and TMGR (maximum growth) have been previously used; however, they describe the midpoints of the growth phase, rather than its "initiation" per se (Figure 3).

It is also important to note that the substantial range of interannual variability (Figure 5) between metrics represents similar aspects of the seasonal cycle. For example, even though TMED, TMGR, and TCUD are supposedly representative of the midpoint of the bloom, there is a strong difference in their interannual variability. Moreover, TCUD and TMED exhibit poor precision in the intergyre region (or transition zone). The lack of precision in this region has been previously reported by *Henson et al*. [2009] to be due to a high range of starting dates (high interannual variability in bloom timing). Such differences can be deeply problematic: investigations attempting to explain environmentally driven variability in bloom phenology, for instance, can be fatally undermined by the subtleties inherent to the chosen phenology metric. We therefore reiterate the conclusion of *Brody et al*. [2013] and emphasize that the choice of phenology indicator should be driven in the first instance by *a priori* reasoning and the research question at hand.

The best candidate metrics seem to be TCUD and TMED. However, even though the TCUD metric showed the best precision and accuracy, we believe it is not representative of bloom initiation, as it is merely an indicator of when the seasonal cycle reaches 15% of its cumulative distribution (integrated biomass). This metric seems to fail at capturing the spatial variability in the seasonal cycles potentially due to its integrating nature. For instance, depending on the starting date for the integration, information from fall blooms

may contaminate the cumulative distribution used in the estimation of TCUD. Additionally, TCUD relies on having the whole year's chlorophyll data available. This may be a drawback for some applied situations related to real-time detection of bloom onset and its consequences. For example, many fish species spawn in the first quarter of the year, but the effects of phenological changes in bloom timing on larval survival using TCUD cannot be assessed until the end of the year. Notwithstanding, TCUD is a reasonable metric if one wishes to analyze how far the bloom is in terms of its concentration, but not as an indicator of bloom initiation per se.

TMED, on the other hand, appeared to work well across large parts of the study domain. However, its performance was severely degraded in the intergyre (transition) region, and also in some coastal zones. Systematic biases were also present in some regions. These issues may be problematic for local studies in these regions.

We recognize that some of the dimensions considered in this study are correlated, particularly in space. For instance, bloom amplitude is spatially correlated with both missing data and observational noise, as observed in Figure 2. High latitude regions are characterized by high bloom amplitudes and degree of missing data. If the observational noise is significantly higher than the bloom amplitude, the phenology metric will not be properly estimated. The spatial patterns of both missing data and observational noise will thus be less extreme with increasing bloom amplitude. Therefore, isolating a single process when looking at satellite data is not an easy task, simply because the processes lack independence from one another. For this reason, we choose to use space as a proxy for these processes to be able to test their effect on the estimation of phenology metrics in a reasonable approach.

Finally, our study focused solely on algorithmic-based approaches (application of algorithms); however, estimation-based metrics (statistical models with associated estimators) offer an alternative, and, potentially, superior approach. Estimation-based approaches explicitly incorporate noise into their modeling framework, and can be expected to be robust to gaps. Furthermore, as *Ji et al.* [2010] noted, there is a need for the development of more statistical metrics for indexing trophic interactions. For instance, the extent of overlap between different trophic levels would be more appropriate for match-mismatch studies, rather than simply estimating the timing of changing properties (e.g., biomass). Future phenology studies can potentially benefit by focusing efforts in this direction.

5. Conclusions

We have shown that there are important differences in the inherent variability of phenology metrics and in their precision. The variability between phenology metrics is the most important effect in our study, and is greater than that between treatments. However, the precision of many of these metrics is poor, and is a greater concern than their lack of accuracy. The application of preprocessing techniques to the data can help to improve the precision of phenology metrics. Bloom amplitude shows to be the most important of the observation-related processes, with increasing amplitude improving the precision of phenology metrics.

Ideally, a phenology metric should be accurate, precise, and simultaneously sensitive to the underlying environmental processes (physical or biological). Our results suggest that, of the metrics considered, the TCUD metric (15% of cumulative distribution) best satisfies all of these criteria, and is closely followed by TMED (5% above the annual median). Nevertheless, the choice of phenology metric should be determined by the research question being addressed, and whether the metrics are equally robust to observational noise, missing data, and bloom amplitude. We believe these findings to be useful to the current discussion on phenology metrics of phytoplankton.

Acknowledgments

GlobColour data were provided by accessing the GlobColour's FTP server available at ftp://ftp.acri.fr/, supported by EU FP7 MyOcean and ESA GlobColour projects, using ESA ENVISAT MERIS, NASA MODIS, and SeaWiFS data. The NOBM output was obtained using the FTP server available at ftp://gmaoftp.gsfc.nasa.gov/pub/data/NOBM/, developed and maintained by the NASA GES DISC. This paper is a deliverable of the Nordic Centre for Research on Marine Ecosystems and Resources under Climate Change (NorMER), which is funded by the Norden Top-level Research Initiative subprogramme "Effect Studies and Adaptation to Climate Change." A.W.V. acknowledges the Danish Council for Strategic Research and its support for the NAACOS project. B.R.M. thanks the Danish National Research Foundation (*Dansk Grundforskningsfond*) for support to the Center for Macroecology, Evolution, and Climate. The author would also like to thank Harriet Cole and Stephanie Henson, National Oceanography Centre, Southhampton, UK, for inspiring conversations that lead to this work and two anonymous reviewers for their insightful comments and constructive criticisms. The research leading to these results has received funding from the European Union 7th Framework Programme (FP7 2007–2013) under grant agreement 308299 (NACLIM project).

References

Ballabrera-Poy, J., R. Murtugudde, J. Christian, and A. Busalacchi (2003), Signal-to-noise ratios of observed monthly tropical ocean color, *Geophys. Res. Lett.*, *30*(12), 1645, doi:10.1029/2003GL016995.

Beaugrand, G., and P. C. Reid (2012), Relationships between North Atlantic salmon, plankton, and hydroclimatic change in the northeast Atlantic, *ICES J. Mar. Sci.*, *69*(9), 1549–1562.

Behrenfeld, M. J., S. C. Doney, I. Lima, E. S. Boss, and D. A. Siegel (2013), Annual cycles of ecological disturbance and recovery underlying the subarctic Atlantic spring plankton bloom, *Global Biogeochem. Cycles*, *27*, 526–540, doi:10.1002/gbc.20050.

Brody, S. R., and M. S. Lozier (2014), Changes in dominant mixing length scales as a driver of subpolar phytoplankton bloom initiation in the North Atlantic, *Geophys. Res. Lett.*, *41*, 3197–3203, doi:10.1002/2014GL059707.

Brody, S. R., M. S. Lozier, and J. P. Dunne (2013), A comparison of methods to determine phytoplankton bloom initiation, *J. Geophys. Res. Oceans*, *118*, 2345–2357, doi:10.1002/jgrc.20167.
Cole, H., S. Henson, A. Martin, and A. Yool (2012), Mind the gap: The impact of missing data on the calculation of phytoplankton phenology metrics, *J. Geophys. Res.*, *117*, C08030, doi:10.1029/2012JC008249.
Cushing, D. H. (1990), Plankton production and year-class strength in fish populations: An update of the match/mismatch hypothesis, *Adv. Mar. Biol.*, *26*, 249–293.
Edwards, M., and A. J. Richardson (2004), Impact of climate change on marine pelagic phenology and trophic mismatch, *Nature*, *430*(7002), 881–884.
Greve, W., S. Prinage, H. Zidowitz, J. Nast, and F. Reiners (2005), On the phenology of north sea ichthyoplankton, *ICES J. Mar. Sci.*, *62*(7), 1216–1223.
Henson, S. A., J. P. Dunne, and J. L. Sarmiento (2009), Decadal variability in North Atlantic phytoplankton blooms, *J. Geophys. Res.*, *114*, C04013, doi:10.1029/2008JC005139.
Jackson, G. A., R. Maffione, D. K. Costello, A. L. Alldredge, B. E. Logan, and H. G. Dam (1997), Particle size spectra between 1 m and 1 cm at Monterey Bay determined using multiple instruments, *Deep Sea Res., Part I*, *44*(11), 1739–1767.
Ji, R., M. Edwards, D. Mackas, J. Runge, and A. Thomas (2010), Marine plankton phenology and life history in a changing climate: Current research and future directions, *J. Plankton Res.*, *32*(10), 1355–1368.
Koeller, P., et al. (2009), Basin-scale coherence in phenology of shrimps and phytoplankton in the North Atlantic ocean, *Science*, *324*(5928), 791–793.
Leaman, B. M. (1991), Reproductive styles and life history variables relative to exploitation and management of Sebastes stocks, *Environ. Biol. Fishes*, *30*(1–2), 253–271.
Maritorena, S., and D. Siegel (2005), Consistent merging of satellite ocean color data sets using a bio-optical model, *Remote Sens. Environ.*, *94*(4), 429–440.
Maritorena, S., D. A. Siegel, and A. R. Peterson (2002), Optimization of a semianalytical ocean color model for global-scale applications, *Appl. Opt.*, *41*(15), 2705–2714.
Maritorena, S., O. H. F. d'Andon, A. Mangin, and D. A. Siegel (2010), Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues, *Remote Sens. Environ.*, *114*(8), 1791–1804.
Nerger, L., and W. Gregg (2008), Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter, *J. Mar. Syst.*, *73*(1), 87–102.
Palacz, A. P., H. Xue, C. Armbrecht, C. Zhang, and F. Chai (2011), Seasonal and inter-annual changes in the surface chlorophyll of the south china sea, *J. Geophys. Res.*, *116*, C09015, doi:10.1029/2011JC007064.
Platt, T., and S. Sathyendranath (2008), Ecological indicators for the pelagic zone of the ocean from remote sensing, *Remote Sens. Environ.*, *112*(8), 3426–3436.
Platt, T., C. Fuentes-Yaco, and K. T. Frank (2003), Marine ecology: Spring algal bloom and larval fish survival, *Nature*, *423*(6938), 398–399.
Polovina, J. J., E. Howell, D. R. Kobayashi, and M. P. Seki (2001), The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources, *Prog. Oceanogr.*, *49*(1–4), 469–483.
Racault, M.-F., C. Le Quéré, E. Buitenhuis, S. Sathyendranath, and T. Platt (2012), Phytoplankton phenology in the global ocean, *Ecol. Indicators*, *14*(1), 152–163.
Roerink, G., M. Menenti, and W. Verhoef (2000), Reconstructing cloudfree ADVI composites using Fourier analysis of time series, *Int. J. Remote Sens.*, *21*(9), 1911–1917.
Rolinski, S., H. Horn, T. Petzoldt, and L. Paul (2007), Identifying cardinal dates in phytoplankton time series to enable the analysis of long-term trends, *Oecologia*, *153*(4), 997–1008.
Sapiano, M. R. P., C. W. Brown, S. Schollaert Uz, and M. Vargas (2012), Establishing a global climatology of marine phytoplankton phenological characteristics, *J. Geophys. Res.*, *117*, C08026, doi:10.1029/2012JC007958.
Sasaoka, K., S. Chiba, and T. Saino (2011), Climatic forcing and phytoplankton phenology over the subarctic north pacific from 1998 to 2006, as observed from ocean color data, *Geophys. Res. Lett.*, *38*, L15609, doi:10.1029/2011GL048299.
Sharples, J., O. Ross, B. Scott, S. Greenstreet, and H. Fraser (2006), Inter-annual variability in the timing of stratification and the spring bloom in the north-western north sea, *Cont. Shelf Res.*, *26*(6), 733–751.
Siegel, D. A., S. C. Doney, and J. A. Yoder (2002), The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis, *Science*, *296*(5568), 730–733.
Sverdrup, H. U. (1953), On conditions for the vernal blooming of phytoplankton, *J. Cons. Int. Explor. Mer*, *18*(3), 287–295.
Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192.
Thackeray, S., T. Sparks, M. Frederiksen, S. Burthe, P. Bacon, J. Bell, M. Botham, T. Brereton, P. Bright, and L. Carvalho (2010), Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments, *Global Change Biol.*, *16*(12), 3304–3313.
Vantrepotte, V., and F. Mélin (2011), Inter-annual variations in the SeaWiFS global chlorophyll a concentration (1997–2007), *Deep Sea Res., Part I*, *58*(4), 429–441.
Vargas, M., C. Brown, and M. Sapiano (2009), Phenology of marine phytoplankton from satellite ocean color measurements, *Geophys. Res. Lett.*, *36*, L01608, doi:10.1029/2008GL036006.
Verbesselt, J., R. Hyndman, A. Zeileis, and D. Culvenor (2010), Phenological change detection while accounting for abrupt and gradual trends in satellite image time series, *Remote Sens. Environ.*, *114*(12), 2970–2980.
Werdell, P. J., and S. W. Bailey (2005), An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation, *Remote Sens. Environ.*, *98*(1), 122–140.
White, M. A., D. Beurs, M. Kirsten, K. Didan, D. W. Inoye, A. D. Richardson, O. P. Jensen, J. O'kefee, G. Zhang, and R. R. Nemani (2009), Intercomparison, interpretation, and assessment of spring phenology in north America estimated from remote sensing for 1982–2006, *Global Change Biol.*, *15*(10), 2335–2359.
Wiltshire, K., A. Malzahn, K. Wirtz, W. Greve, S. Janisch, P. Mangelsdorf, B. Manly, and M. Boersma (2008), Resilience of North Sea phytoplankton spring bloom dynamics: An analysis of long-term data at Helgoland Roads, *Limnol. Oceanogr. Methods*, 1294–1302.
Yoder, J. A., and M. A. Kennelly (2003), Seasonal and ENSO variability in global ocean phytoplankton chlorophyll derived from 4 years of SeaWiFS measurements, *Global Biogeochem. Cycles*, *17*(4), 1112, doi:10.1029/2002GB001942.
Zhai, L., T. Platt, C. Tang, S. Sathyendranath, and R. Hernández Walls (2011), Phytoplankton phenology on the Scotian Shelf, *ICES J. Mar. Sci.*, *68*(4), 781–791.
Zhai, L., et al. (2012), Phytoplankton phenology and production around Iceland and Faroes, *Cont. Shelf Res.*, *37*, 15–25.