A comparison of phytoplankton spring bloom fitting methods using MODIS satellite-derived chlorophyll-a concentration for the Maritimes region

Chantelle Layton, Emmanuel Devred, and Brendan DeTracey

Fisheries and Oceans Canada
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, Nova Scotia
Canada B2Y 4A2

2022

Canadian Technical Report of
Hydrography and Ocean Sciences 340

Canadian Technical Report of Hydrography and Ocean Sciences

Technical reports contain scientific and technical information of a type that represents a contribution to existing knowledge but which is not normally found in the primary literature. The subject matter is generally related to programs and interests of the Oceans and Science sectors of Fisheries and Oceans Canada.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts.*

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Regional and headquarters establishments of Ocean Science and Surveys ceased publication of their various report series as of December 1981. A complete listing of these publications and the last number issued under each title are published in the *Canadian Journal of Fisheries and Aquatic Sciences*, Volume 38: Index to Publications 1981. The current series began with Report Number 1 in January 1982.

Rapport technique canadien sur l'hydrographie et les sciences océaniques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles mais que l'on ne trouve pas normalement dans les revues scientifiques. Le sujet est généralement rattaché aux programmes et intérêts des secteurs des Océans et des Sciences de Pêches et Océans Canada.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques*.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page de titre.

Les établissements de l’ancien secteur des Sciences et Levés océaniques dans les régions et à l'administration centrale ont cessé de publier leurs diverses séries de rapports en décembre 1981. Vous trouverez dans l'index des publications du volume 38 du *Journal canadien des sciences halieutiques et aquatiques*, la liste de ces publications ainsi que le dernier numéro paru dans chaque catégorie. La nouvelle série a commencé avec la publication du rapport numéro 1 en janvier 1982.

Canadian Technical Report of
Hydrography and Ocean Sciences 340

2022

A COMPARISON OF PHYTOPLANKTON SPRING BLOOM FITTING METHODS USING MODIS SATELLITE-DERIVED CHLOROPHYLL-A CONCENTRATION FOR THE MARITIMES REGION

by

Chantelle Layton, Emmanuel Devred, and
Brendan DeTracey

Ocean and Ecosystem Sciences Division,
Fisheries and Oceans Canada
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, Nova Scotia
Canada, B2Y 4A2

Cat. No. Fs 97-18/340E-PDF ISBN 978-0-660-42988-5 ISSN 1488-5417

Correct Citation for this publication:

Layton, C., Devred, E., DeTracey B.. 2022. A comparison of phytoplankton spring bloom fitting methods using MODIS satellite-derived chlorophyll-a concentration for the Maritimes region. Can. Tech. Rep. Hydrogr. Ocean Sci. 340: vii + 22 p.

CONTENTS

List of Tables

List of Figures

ABSTRACT

Layton, C., Devred, E., DeTracey B.. 2022. A comparison of phytoplankton spring bloom fitting methods using MODIS satellite-derived chlorophyll-a concentration for the Maritimes region. Can. Tech. Rep. Hydrogr. Ocean Sci. 340: vii + 22 p.

Satellite-derived chlorophyll-a concentration time series have proven to be a valuable source of information to monitor marine ecosystems. In the Atlantic regions of Canada, these data are used to quantify the phytoplankton spring bloom timing, duration, and intensity. In this report, three different models were tested to characterize the spring bloom: the shifted Gaussian, currently implemented in the Atlantic Zone Monitoring Program reporting, the threshold method, and rate of change method. The objectives of the study were to test the robustness of each method to deal with missing data and the consistency of the bloom metrics when changing the number of parameters, or their value, in the optimization procedure. The three methods showed consistent results, but the choice of initial parameters, can have an impact on the spring bloom indices. The set of parameters that are optimized in the model also influenced the success in retrieving bloom parameters. Finally, we found that the shifted Gaussian is the best of the three methods for the retrieval of the spring bloom metrics as it can deal with time series with missing data, and thus should be used in the Atlantic Zone Monitoring Program and Atlantic Zone offshelf Monitoring Program reporting.

RÉSUMÉ

Layton, C., Devred, E., DeTracey B.. 2022. A comparison of phytoplankton spring bloom fitting methods using MODIS satellite-derived chlorophyll-a concentration for the Maritimes region. Can. Tech. Rep. Hydrogr. Ocean Sci. 340: vii + 22 p.

Les séries temporelles de concentration en chlorophylle-a dérivées par satellites ont prouvé être une source d'informations précieuses pour monitorer l'écosystème marin. Dans les régions de l'Atlantique du Canada, ces données sont utilisées pour quantifier la floraison printaniére de phytoplancton de l'initiation, de la durée et de l'intensité. Dans ce rapport, nous avons testé trois modèles différents pour caractériser la floraison printanière, la Gaussienne décalée, couramment implémentée dans les rapports du Programme de Monitorage de la Zone Atlantique, ainsi que les méthodes du seuil et du taux de changement. Les objectifs de cette étude étaient de tester la robustesse de chaque méthode face aux données manquantes et de tester la cohérence des métriques de floraison quand le nombre de paramètres, ou leur valeur, était changé lors de la procédure d'optimisation. Les trois méthodes ont montré des résultats cohérents, mais le choix des paramétres initiaux peut avoir un impact sur les indices de floraison printanière. Les paramètres initiaux qui sont optimisés influencent le succès à retrouver les paramètres de la floraison. Finalement, nous avons trouvé que la Gaussienne décalée est adaptée au calcul des métriques de la floraison printanière dû à sa capacité de traiter les séries temporelles avec des données manquantes, et devrait être utilisée dans les rapports des Programmes de Monitorage de la Zone Atlantique et de la zone Atlantique au large du plateau continental.

1 Introduction

The Atlantic Zone Monitoring program (AZMP) is a DFO led initiative that began in 1999 to monitor the physical, chemical and biological state of the marine ecosystem in the Northwest Atlantic and report the findings to management. This program is a joint effort from the four Atlantic regions, namely, Maritimes, Québec, Gulf, and Newfoundland and Labrador. Several sea-going expeditions are carried out every year to take measurements and collect water samples that help define the state of the Northwest Atlantic Ocean. An important biological property of the ocean is the amount (or concentration) of primary producers present in the upper layer of the water column at any time of the year, and in particular during the spring when phytoplankton growth is exponential, sustaining an important part of the marine ecosystem: this phenomenon is referred to as the phytoplankton spring bloom. Reporting on the spring bloom in the AZMP has evolved throughout the years for each of the three Maritimes, Newfoundland, and Québec regions. These regions started to report on any bloom parameters using satellite derived data in 2007, but only using figures of the bi-weekly spatially averaged values for defined boxes. Starting for the reporting year 2008 in the Maritimes region, standardized anomalies were provided for duration, magnitude, and timing of the spring bloom at fixed stations using water-column integrated chlorophyll-a concentration data measured from water samples (Harrison et al., 2009). The bloom parameters were estimated using a threshold of $40\,\mathrm{mg\,m^{-2}}$ (counting the number of days when chlorophyll-a concentration remains over the threshold) and standardized anomalies were based on the reference period of 1999-2006 (Harrison et al., 2009). For the reporting years 2009 and 2010, the Newfoundland region adopted the method used in the Maritimes region, but with a threshold of $80\,\mathrm{mg\,m^{-2}}$, and the standardized anomalies of bloom parameters were based on a reference period of 1999-2010 (Pépin et al., 2011).

For the reporting years of 2011 and 2012 in the Newfoundland region, for the first time, the shifted Gaussian method was applied to satellite-derived measurements of chlorophyll-a concentration in a systematic manner for 10 reference boxes (Pépin et al., 2013). Satellite-derived data used to retrieve the phytoplankton bloom metrics included Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) data from September 1997 to December 2009 and Moderate Resolution Imaging Spectroradiometer (MODIS) data from January 2004 to December 2011 (Figure 1). The standardized anomalies were based on a reference period ranging from 1998 to 2012. For the reporting year of 2014, the Québec region also started to apply the shifted Gaussian method to satellite-derived chlorophyll-a concentration including SeaWiFS from September 1997 to December 2007 and MODIS from January 2008 to December 2014. The standardized anomalies were based on a reference period spanning from 1999 to 2010 (Devine et al., 2015). For the reporting year of 2015, to be consistent with the two other regions, the Maritimes region adopted the shifted Gaussian method to determine spring bloom metrics (Johnson et al., 2017). Satellite-derived data used to perform the fit of the spring bloom included SeaWiFS from January 1998 to December 2007, MODIS from January 2008 to December 2014, and Visible Infrared Imaging Radiometer Suite (VIIRS) from January 2015 to December 2015, and the standardized anomalies were based on the reference period of 1999 to 2010. The period for which a given sensor was used to derive the bloom indices varied between reporting years and regions (Figure 1). For example, in the reporting year 2015, the Newfoundland region only used data from SeaWiFS, September 1997 to December 2009, and MODIS, July 2002 to December 2013, and no VIIRS, to report on the spring bloom using the shifted Gaussian method, unlike the Maritimes region (Pépin et al., 2017). Note that the chlorophyll-a concentration was computed as the average of both sensors for the overlapping period, specifically, July 2002 to December 2009. By 2016, all three regions adopted the same data set and reference period to derive bloom metrics and anomalies, with the exception of the Québec region, which used an extra four months corresponding to the start of the SeaWiFS time series (i.e, September to December 1997) to derive the reference period.

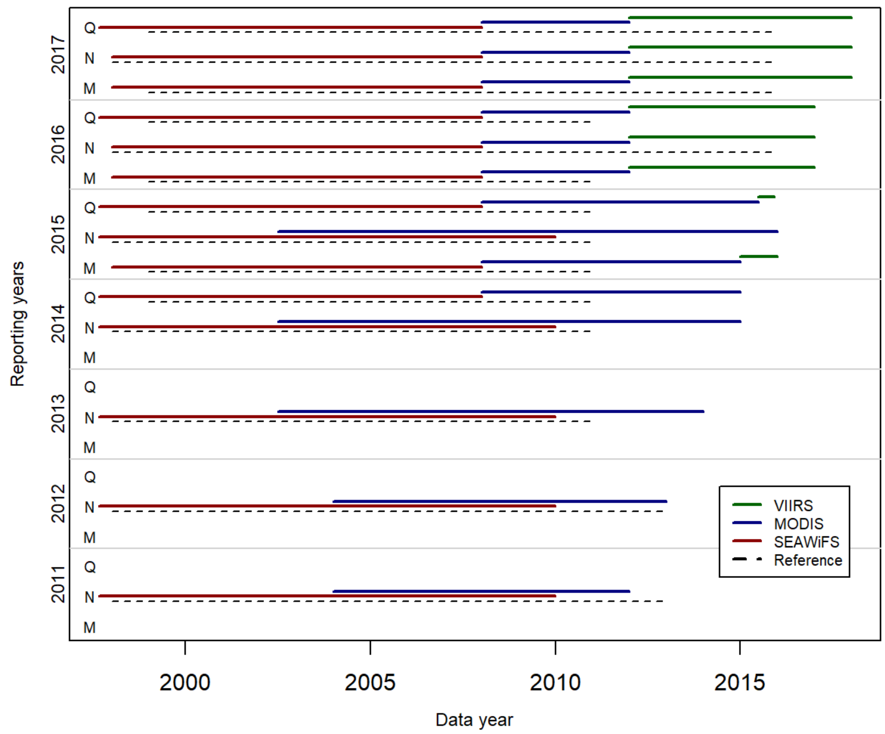

Figure 1: Summary of reference period used to calculate anomalies and satellite data used by each region, Maritimes (M), Newfoundland (N), and Québec (Q), since the implementation of reporting on spring bloom parameters using the shifted Gaussian method. Note that data usage date ranges may appear to go beyond the reporting year due to two years being reported on in a single report.

While the Shifted Gaussian method is currently used by the three regions, there are several other methods to estimate phytoplankton bloom parameters using satellite-derived chlorophyll-a data including i) the threshold method (Siegel et al., 2002), which was implemented for the fixed sampling stations in the Maritimes and Newfoundland region (Harrison et al., 2009; Pépin et al., 2011), and ii) the rate of change method (Brody et al., 2013). While the shifted Gaussian has proven reliable at the lower latitude boxes, particularly those for the Maritimes and Québec region, its application to high-latitude boxes of the Newfoundland region (Pépin et al., 2017) has been challenging mainly due to the lack of data as a result of continuous cloud cover. Here we propose to evaluate the three different methods cited above to infer phytoplankton spring bloom indices, using six pre-defined boxes of the Maritimes region as an illustration.

2 Data

2.1 Description

Although the bloom metrics were historically computed using satellite-derived chlorophyll-a concentration computed at the Bedford Institute of Oceanography from raw data (top-of-atmosphere) with a spatial resolution of 1.5 km, the satellite-based data set used in this study relies on data generated by the National Aeronautics and Space Administration (NASA). The MODIS-Aqua daily Level3 4-km resolution data for the global ocean were downloaded from the NASA ocean colour website (http://oceancolor.gsfc.nasa.gov/) for the period from year 2003 to 2017 (NASA Goddard Space Flight Center, 2014). The MODIS OC3 algorithm, which relates chlorophyll-a concentration to remote sensing reflectance through a four-degree polynomial, was selected as it provides the highest number of valid pixels compared to other algorithms (Clay et al., 2019). The global data were subset to a pan-Canadian grid within the bounds limited by 39°N to 85°N and 42°W to 146°W. Valid data points for the grid were defined to be those in a water depth below sea level, as interpolated using ETOPO1 bathymetry data, with a positive distance (i.e., greater than 0 km) from land, and a minimum of one neighboring valid data point. It is noteworthy that the decrease in spatial resolution of the satellite data from 1.5 to 4 km had a minimal impact on the results, given that the data are further averaged in boxes spanning over one or more degree in latitude and longitude (*e.g.*, Johnson et al. (2017)).

Table 1: Geographical boundary for the AZMP Maritime boxes on the Scotian Shelf.

Box name	Latitude min (° N)	Latitude max (° N)	Longitude min (° W)	Longitude max (° W)
Georges Bank (GB)	41	42	68	66.5
Central Scotian Shelf (CSS)	43.33	44.33	64	62
Eastern Scotian Shelf (ESS)	44.2	45.67	60	58
Western Scotian Shelf (WSS)	42.5	43.33	65.5	64.5
Cabot Strait (CS)	46.9	48	60.4	59
Lurcher Shoal (LS)	43	44	66.7	66

The performance of the three bloom fitting methods was assessed following the AZMP standard reporting protocol using the annual spring bloom statistics on the Scotian Shelf for the standard six boxes for the Maritimes region for 15 years of observations from years 2003 to 2017 (Table 1 and Figure 2). Two years, namely 2007 and 2016, were selected to illustrate the application of the methods and support the results and discussion sections. The year 2007 provides an ideal case with good data coverage and a marked spring bloom using the 8-day chlorophyll-a concentration composite, while the year 2016 is more challenging in terms of fitting the data with a small spring bloom peak and a lot of variability in the 8-day composite.

2.2 Processing and Quality Control

Prior to applying the bloom fitting methods, decisions on the minimum data coverage for each box need to be considered as satellite-derived ocean colour data availability is spatially and temporally variable due to cloud cover. The AZMP boxes were designed to represent a homogeneous region (*e.g.* Central Scotian Shelf) with a size that was optimized to ensure good coverage. However, the patchy nature of phytoplankton blooms and cloud cover made the selection of an area of interest

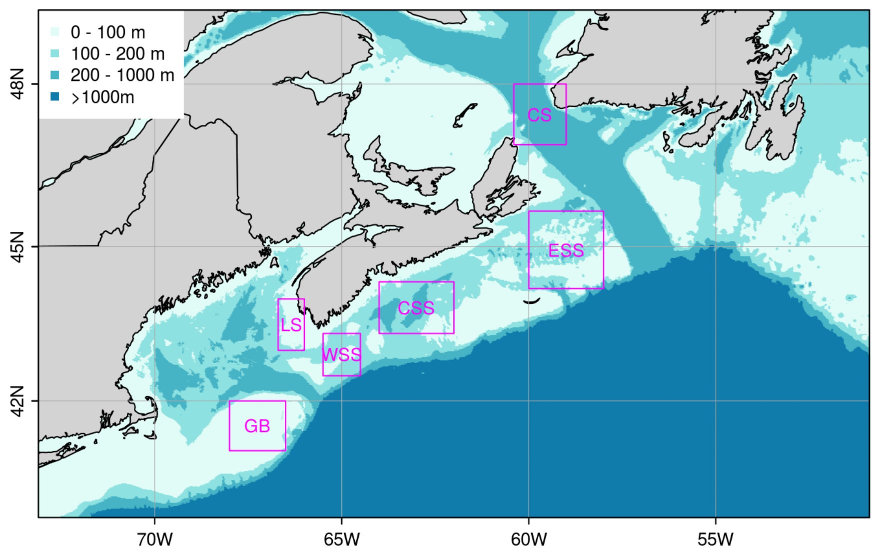

Figure 2: Location of the six boxes for which satellite-derived bloom metrics are computed for AZMP reporting. Table 1 provides latitudes and longitudes of the boundaries.

difficult as on any given day, phytoplankton biomass can vary by several orders of magnitude within a short distance, on the order of tens of kilometers. To ensure that sufficient data are used when fitting the bloom metrics, at least 10% of the pixels in any given box on any given day must be valid.

For all boxes, the daily geometric mean was computed using data lying within two standard deviations of the geometric mean. The geometric mean required log transforming the data to ensure that they followed a normal distribution. Removing data outside of two standard deviations removed any bias due to abnormally high or low values.

3 Methods

Here, we introduce the mathematical formulation and narrative of the three approaches, as well as various parameterizations to improve the fitting of the satellite data to derive the spring bloom metrics. Figure 3 provides a schematic of the definition of all the spring bloom metrics derived from applying the three methods to daily chlorophyll-a concentration.

3.1 Shifted Gaussian

The generalized biomass profile function, which is assumed to follow a shifted Gaussian as a function of depth as introduced by Platt and Sathyendranath (1988), was adapted for fitting surface chlorophyll-a concentration as a function of time with the addition of a linear term to account for background chlorophyll-a variation (Zhai et al., 2011) according to,

$$B(t) = B_0 + \beta t + \frac{h}{\sqrt{2\pi}\sigma} exp\left(-\frac{(t-t_m)^2}{2\sigma^2}\right) \tag{1}$$

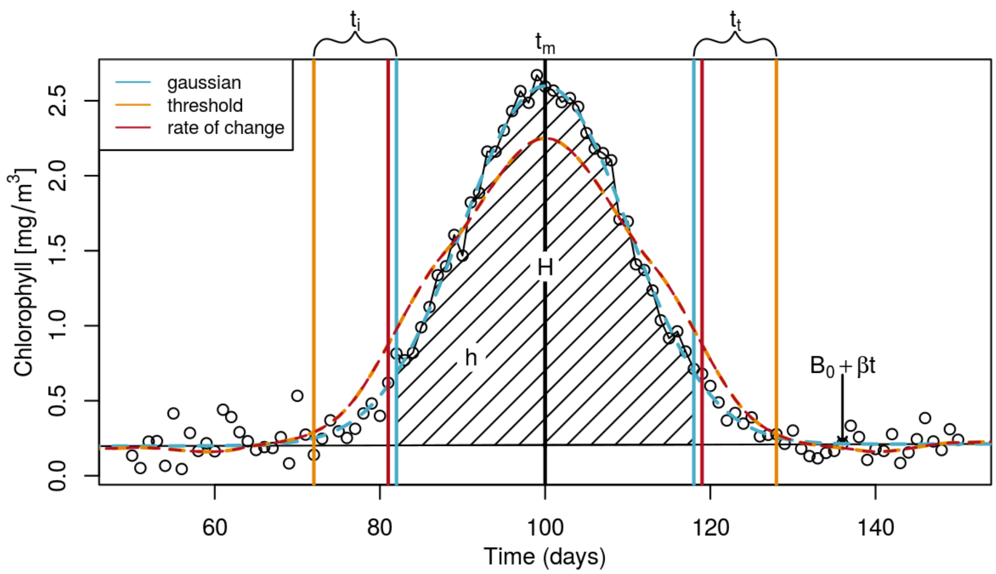

Figure 3: Resulting fit and parameters of interest for the shifted Gaussian, threshold, and rate of change methods using idealized Gaussian shaped data. Open circles represent idealized data points. The dashed coloured lines indicate the resulting fit of the data for each method and vertical coloured lines indicate the inferred initiation and termination time for each method. The vertical black line is the timing of the maximum chlorophyll concentration which is the same for all methods. Relevant parameters in Equation 1 are also labeled for context.

where $B(t)$ is chlorophyll-a concentration (mg m^{-3}) as a function of time, t, in day of the year, B_0 is the background value of chlorophyll-a concentration (mg m^{-3}), β is the linear rate of change of background chlorophyll-a concentration ($\mathrm{mg\,m^{-3}\,d^{-1}}$), h is the integral of chlorophyll-a concentration under the Gaussian peak (mg m^{-3}) and was calculated by integrating between the time of initiation (t_i in day of the year) and termination (t_t in day of the year), and between the background concentration and the daily spatially averaged chlorophyll-a concentration (i.e., the continuous curve in Figure 3). The parameter σ corresponds to the width of the Gaussian curve at mid-height, and t_m is the time at which chlorophyll-a concentration reaches its maximum. This maximum is referred to as the amplitude, H, in $\mathrm{mg\,m^{-3}}$ and is defined by:

$$H = \frac{h}{\sqrt{2\pi}\sigma} - B_0 \tag{2}$$

From Equation (1), a few additional parameters of significance can be derived, such as the time of bloom initiation, t_i, which occurs when chlorophyll-a concentration reaches 20% of the bloom amplitude, or when $B(t) - (B_0 + \beta t) = 0.2H$, and can be expressed as:

$$t_i = t_m - (-2\log(0.2))^{1/2}\sigma, \tag{3}$$

also, the bloom duration, t_d, assuming the same threshold as the bloom initiation, is :

$$t_d = 2(t_m - t_i), \tag{4}$$

and the bloom termination, t_t, is defined by,

$$t_t = t_i + t_d. \quad (5)$$

For a given year and box, Equation (1) was fitted to the daily chlorophyll-a concentrations using the nonlinear least squares function, *nls*, in the R programming language (R Core Team, 2017) (Figure 3). The "port" algorithm was selected in order to constrain the range of variation of the optimized parameters and ensure convergence to realistic values (*e.g.*, positive B_0). The range of variation, or bounds, for each optimized parameter was chosen based on literature values, and a number of initial values for the algorithm were supplied (Table 2). In some cases, the *nls* function was unable to fit the given model to the satellite time series due to inappropriate starting values for the optimized parameters. When this happened, a new set of starting values was supplied (Table 2) and this step was repeated until a successful fit was obtained. An iteration through four sets of starting values was enough to ensure convergence and successful fit of the satellite data by a shifted Gaussian in most cases.

Table 2: Bounds and start parameters supplied to the *nls* function using the port algorithm.

	bounds		start values			
	lower	upper				
B_0	-1	5	0	0	0	0
h	0	350	50	50	10	10
σ	0	100	10	2	2	1
β ($\times 10^{-3}$)	-20	10	-2	-2	-1	-1
t_m	30	180	100	100	100	100

Several combinations of the number of optimized parameters in Equation 1 were tested to assess the robustness of the method, in particular, for the t_m and β parameters. Since t_m can easily be inferred from the time series, even in an automated manner, this parameter was supplied in a version of the model. The linear term, βt, representing the background concentration, provides useful information on chlorophyll-a concentration before and after the bloom, however it is not a critical term for the purpose of obtaining the bloom metrics. Therefore, four possible optimization schemes were tested (Table 3), Model G1 is the reference model as implemented in the current approach to derive bloom metrics for AZMP reporting. In the second model, G2, the term t_m is provided in the equation, while the model G3 has the additional term βt and a total of five parameters are optimized. Finally, the model G4 is similar to the model G2 with the addition of the β parameter.

Table 3: Description of the three optimisation schemes used to infer the bloom metrics using the shifted Gaussian method.

Model	Equation	optimized parameters
G1	$B(t) = B_0 + \frac{h}{\sqrt{2\pi}\sigma} exp\left(-\frac{(t-t_m)^2}{2\sigma^2}\right)$	B_0, h, σ and t_m
G2	$B(t) = B_0 + \frac{h}{\sqrt{2\pi}\sigma} exp\left(-\frac{(t-t_m)^2}{2\sigma^2}\right)$	B_0, h and σ
G3	$B(t) = B_0 + \beta t + \frac{h}{\sqrt{2\pi}\sigma} exp\left(-\frac{(t-t_m)^2}{2\sigma^2}\right)$	B_0, β, h, σ and t_m
G4	$B(t) = B_0 + \beta t + \frac{h}{\sqrt{2\pi}\sigma} exp\left(-\frac{(t-t_m)^2}{2\sigma^2}\right)$	B_0, β, h, and σ

For consistency between all approaches compared in this analysis, the magnitude, h, (although optimized in the model) and amplitude, H, derived from the shifted Gaussian approach were calculated in a similar manner to the threshold and rate-of-change methods, which is described in detail in Sections 3.2 and 3.3. Here, the background concentration is defined by the B_0 or the $B_0 + \beta t$ term, depending on the version of the model being used.

Satellite time series of ocean colour data are imperfect, meaning they are subject to natural variations both spatially and temporally in phytoplankton biomass and cloud cover that results in data gaps and thus a possible bias in the time series. To circumvent these issues, the performance of the methods to derive phytoplankton bloom metrics was also assessed against a theoretical time series that was generated using a Gaussian curve (Equation 1) described by the following parameters: $B_0 = 0.2$, $\beta = 5 \times 10^{-5}$, $h = 60$, $\sigma = 10$, and $t_m = 100$, with the addition of a random noise of zero mean and a standard deviation of 0.08 (Figure 3).

3.2 Threshold

Among popular methods used to derive bloom metrics, the threshold method was introduced by Siegel et al. (2002) to find the initiation of a bloom, which is defined when chlorophyll-a concentration reaches a given percentage of the annual median of chlorophyll-a concentration. Since then, modifications to the method have been made to avoid false detection of the spring bloom (Brody et al., 2013) due to short lived spikes in chlorophyll-a concentration occurring in the winter, and to detect both the primary bloom, which will be referred to as the spring bloom, and the secondary bloom, here defined as the fall bloom (Racault et al., 2015). Here, the former method was used, and was adapted to infer additional bloom parameters such as duration and magnitude. First the time series was subset to avoid contamination by the fall bloom when the spring bloom metrics, for the Maritimes region it was found suitable to discard data past day of year 200. The day when chlorophyll-a concentration reached its maximum, t_m, was found within this time period of day of year 1 to 200. The initiation time was defined following two criteria: 1) time when chlorophyll-a concentration is above a given threshold, defined as a percentage of the annual median, and 2) the chlorophyll-a concentration had to remain above that threshold for at least 14 consecutive days (Figure 3). This approach identifies the timing of the maximum chlorophyll-a concentration and the initiation of the bloom. Assuming that the bloom shape is symmetric, the duration of the bloom and the termination day are calculated using Equations (4) and (5) respectively. If it is assumed that the bloom shape is asymmetric, the steps outlined above are inverted: the time series is subset to after the bloom maximum, and the termination day is defined as the day closest to the day of the maximum of chlorophyll-a concentration following which the chlorophyll-a concentration remains below the threshold values for at least 14 consecutive days.

The magnitude, h, of the bloom was calculated by integrating the area bounded by t_i and t_t on the x-axis and between B_0 and H on the y-axis. The amplitude of the bloom, H, was calculated by subtracting the the background concentration from the maximum chlorophyll-a concentration, $B(t_m)$.

The threshold value of 5% of the annual median set in Siegel et al. (2002) has also been used in other studies (Henson and Thomas, 2007; Henson, 2007; Kim et al., 2009; Racault et al., 2012; Sapiano et al., 2012; Groetsch et al., 2016; Lemos et al., 2018). Here, two threshold values of 5 and 20% were tested to evaluate the impact of the threshold on the bloom parameters, these models are hereafter referred to as T5 and T20. The threshold of 20% was selected to remain consistent with the shifted Gaussian approach. In addition, the method was applied to daily spatially averaged data, and to a local regression model (LOESS) applied to the daily data, these models are referred to as TL5 and TL20 for a 5 and 20% threshold respectively. For the LOESS fit, the *loess* function in the R programming language was used (R Core Team, 2017). The default parameters for the function were used, and a suitable span was determined by doing an anova test of five models with span values of 0.083, 0.167, 0.333, 0.5, and 1 respectively. The span corresponds to a degree of smoothing, the larger the span, the higher the degree of smoothing. A span value of 0.333 was selected as it resolved relatively low bandpass variations in chlorophyll-a concentration such as the spring bloom but did not resolve high frequency variations in chlorophyll-a concentration that result from data processing which could arise from the differing number of data points available. Both thresholds, with or without LOESS smoothing, were applied to the 8-day time series of chlorophyll-a concentration in 2007 and 2016, for the central Scotian shelf (see section 4.3.2).

The background chlorophyll-a concentration was calculated using a quantile regression function, *rq* from the *quantreg* package in the R programming language (Koenker, 2018). The function has a parameter, *tau*, that can be adjusted to define the quantile(s). Here we used a *tau* of 0.25 which corresponded to the 25^{th} quantile (Figure 3).

3.3 Rate of Change

The first two methods relied directly on absolute values of chlorophyll-a concentration, but the third method, referred to as rate of change, RoC, differs in the way that it is based on temporal changes in chlorophyll-a concentration. In particular, this method assumes that the initiation of a bloom is indicated by exponential growth, or when there is a sudden and sustained increase in chlorophyll-a concentration. The rate of change approach was introduced by Brody et al. (2013), who determined the initiation date using a discrete fast-Fourier transform reconstruction on a 10-year time series. Here, the method was applied to a fitted curve using a single year and the method was repeated for all years (Figure 3). The maximum chlorophyll-a concentration was identified from the fitted curve, and the time series was subset prior to the day of the maximum. The initiation was defined as the maximum of the derivative of chlorophyll-a concentration over time. As introduced by Brody et al. (2013), this method only provides a way to find the initiation date of the bloom. Since additional parameters, such as termination time, duration, and magnitude, are needed for AZMP reporting, the algorithm was extended to calculate these additional parameters. Similarly to the threshold method, assuming that the bloom is symmetric, the duration and termination of the bloom were calculated using Equations 4 and 5 respectively. For asymmetric blooms, the steps outlined to find the initiation date are inverted such that the time series is subset after the peak of the bloom identified from a fitted curve, and the termination date is defined as the negative minimum $\partial B(t)/\partial t$.

For the rate of change method, similar to the threshold method, the fitted curve was generated by smoothing the data using a LOESS function with a span of either 0.333 or an optimized span that minimized the root-mean-square difference between the original data and the LOESS smoothing. These models are referred to as RoCFS for the fixed span and RoCOS for an optimized span. The background chlorophyll-a concentration was determined using the 25^{th} quantile, same as for the threshold method.

4 Results

4.1 Cloud cover impact on data availability

Across the Scotian Shelf, temporal data coverage for chlorophyll-a concentration from the MODIS satellite from 2003 to 2017 does not exceed 30%, meaning that on average, for a pixel within a given box, one observes that pixel only at most 30% of the time. One also observes a southwest-northeast gradient on the Scotian Shelf, with highest temporal coverage in the southwest roughly between 20 and 25% for GB, LS, WSS, and CSS, minimum temporal coverage between 15 and 20% for ESS, and between 10 to 15% for CS (Figure 4). For each box, the temporal distribution of the spatial percent coverage, which provides information on how many times in a given year there will be more than 10% of spatial coverage of a given box, follows a similar bi-modal distribution for the six boxes (Figure 5). The maximum daily percent coverage (i.e. greater than 50% except for CS, which is greater than 40%) occurs around day of year 100 and 250, and minimum daily percent coverage occurs during the beginning and end of the year, which corresponds to winter which is known for persistent cloud cover (Figure 5).

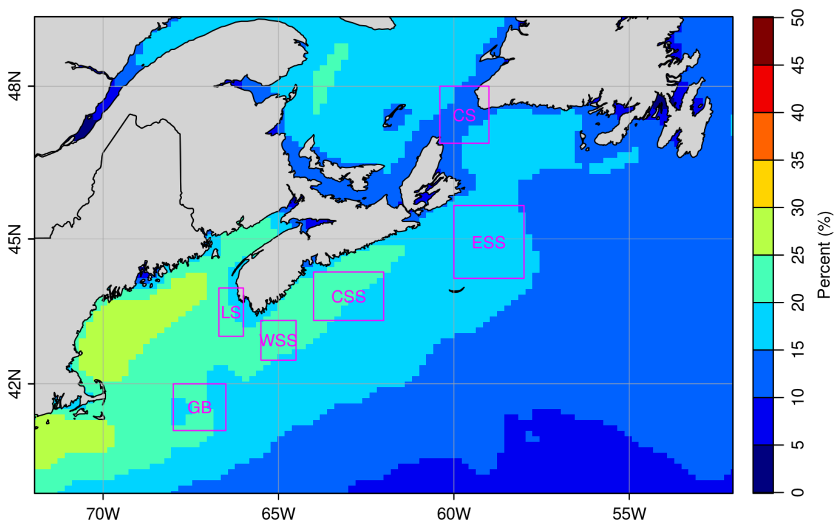

Figure 4: Total percent data availability from 2003 to 2017 including boxes across the Scotian Shelf for the Maritimes region, Cabot Strait (CS), central Scotian Shelf (CSS), eastern Scotian Shelf (ESS), Georges Bank (GB), Lurcher Shoal (LS), western Scotian Shelf (WSS).

4.2 Phytoplankton biomass spatial variation

Phytoplankton blooms are, by nature, very patchy (Denman and Platt, 1976) such that recording data in a large area results in a wide range of chlorophyll-a concentrations, which can span over several orders of magnitude. For instance, concentrations in year 2007 ranged from 0.2 to 20 mg m^{-3} on day 100 in the CSS box (Figure 6a). Averaging within a region leads to a first smoothing of the signal (Figure 6b, solid dark grey circles) that reduces the range of variations within the box to a single value for a given day. The annual range of variation is further decreased when the 8-day binning is computed (Figure 6b, solid dark circles). For instance in the image from year day 100 in 2007, the initial range of variation of chlorophyll-a concentration from 0.2 to 20 mg m^{-3},

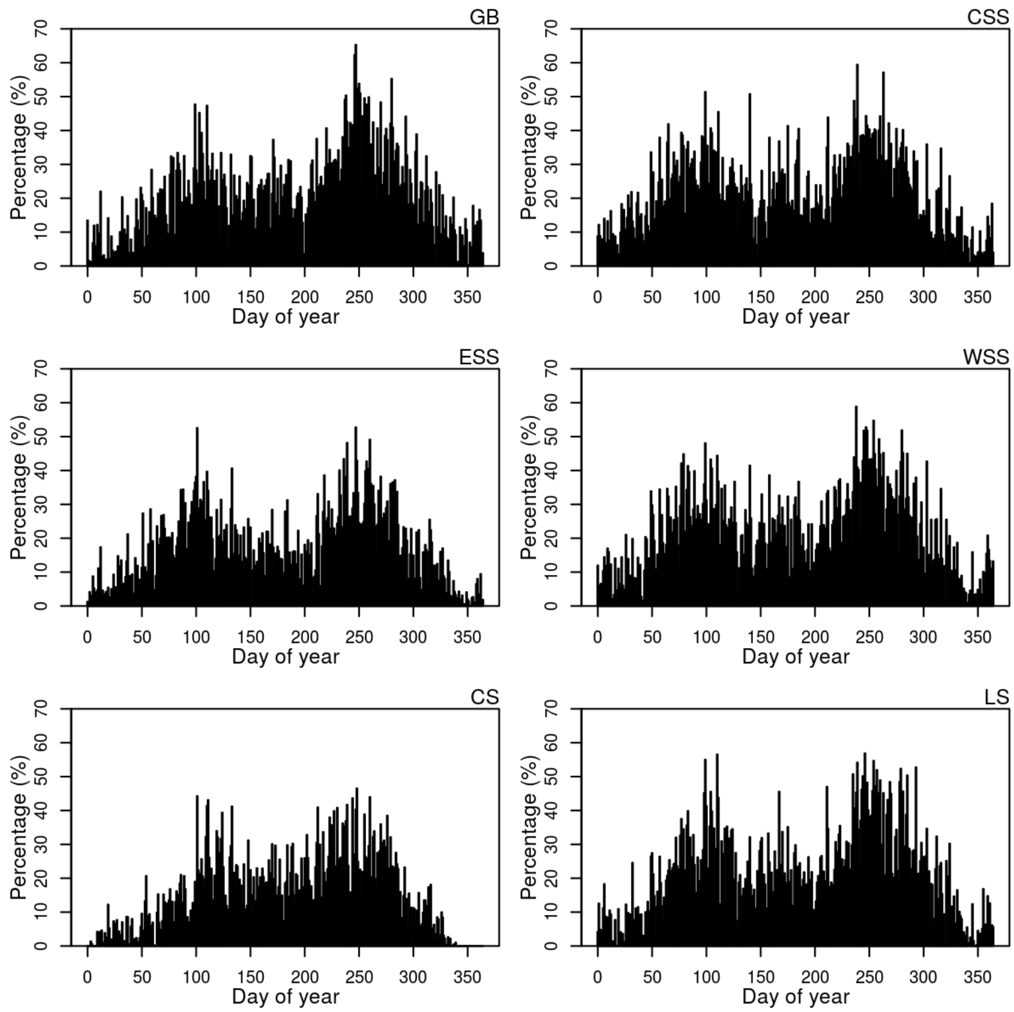

Figure 5: Total daily percent data availability from 2003 to 2017 for all boxes in Maritimes regions.

after averaging to create an 8-day composite, the maximum chlorophyll value is reduced to a single value of about 3 mg m^{-3}, and data vary from 0.3 to 3 mg m^{-3} over the course of the year 2007.

4.3 Model parameters impact on bloom fit metrics retrieval

4.3.1 Shifted Gaussian

Model G1, currently used in AZMP reporting, provides spring bloom parameters for both 2007 and 2016, despite not fitting the background concentration of chlorophyll-a (Figure 7a-b). Excluding dubious high concentration values prior to day of year 50 removes the effect of high chlorophyll-a concentration early in the winter as seen in 2016 (Figure 7b). Model G2 (Figure 7c-d), with the lowest number of parameters to fit given that t_m is provided, was not able to converge for the 2016 study case (Figure 7d). Model G3 provided bloom parameters for both case studies with the additional information of changing background with time (Figure 7e-f). Model G1 and G3 provided similar results regarding the timing and duration of the bloom, while model G2 provided an earlier initiation and peak of the bloom with a longer duration. As with model G2, model G4 was not able to fit the data in 2016 (Figure 7g-h), perhaps due to the constraint of a fixed t_m. In

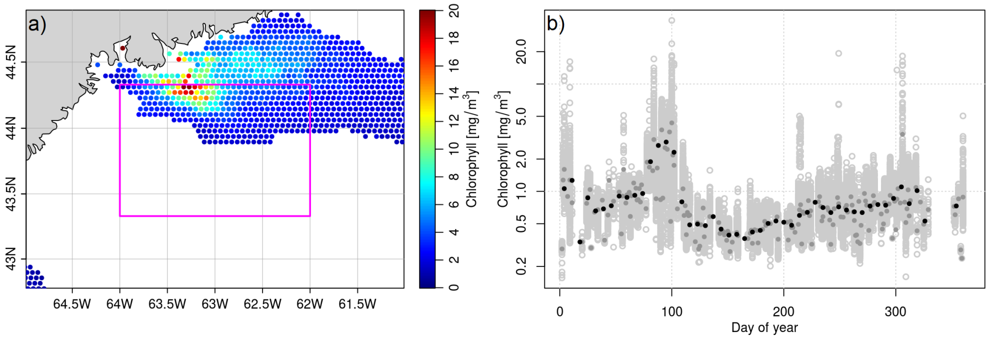

Figure 6: (a) MODIS chlorophyll-a concentration for the Central Scotian Shelf (CSS) box for day of year 100, or April 10, 2007. (b) Time series of averaged MODIS chl-a for the same box, light grey open circles represent all daily data available in the CSS box, dark grey solid circles represent the daily average within the same box and solid black circles represent the 8-day composites.

the remainder of the study, results from model G2 are not analyzed, as this model has the lowest number of parameters and does not always converge when fitting the chlorophyll-a time series.

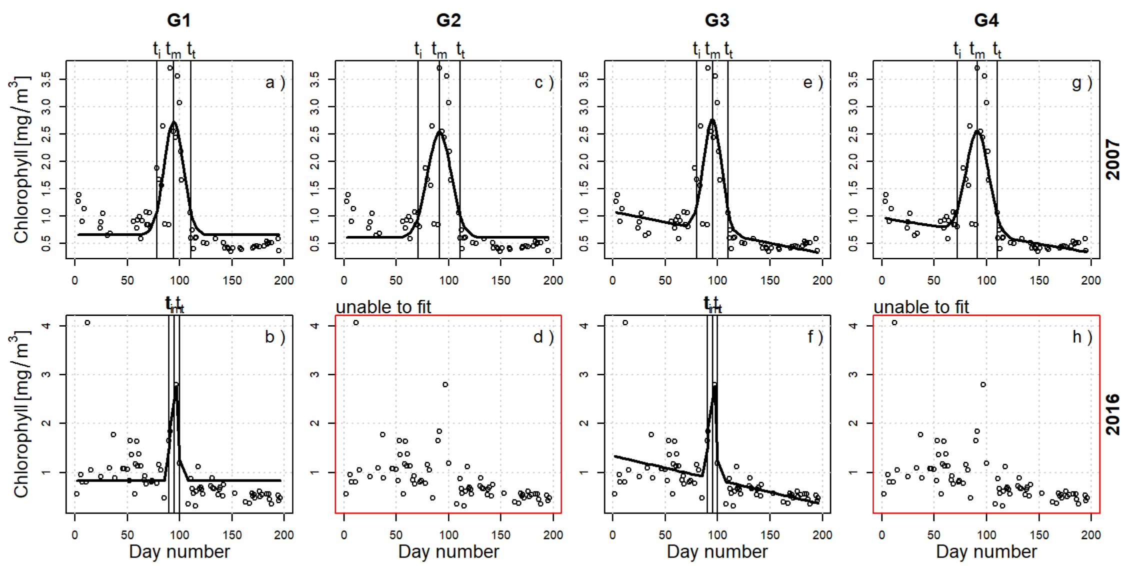

Figure 7: Shifted Gaussian model G1 (a & b), G2 (c & d), G3 (e & f), G4 (g & h) fitted to using annual data from the Central Scotian shelf box for 2007 (a, c, e, & g) and 2016 (b, d, f, & h). Open circles correspond to the daily geometric mean of chlorophyll-a within the box, the thick vertical black lines correspond to the terms t_i, t_m and t_t in Equation 1. Red lines framing (d) and (h) indicates failure of the fit.

The three optimization schemes for the Gaussian fitting (i.e., G1, G3 and G4) were tested against each other to investigate if the method was sensitive to the inclusion of parameters β_t and t_m. Model G1 was taken as the reference and variations in t_i, t_d, h and H were studied for the three models and for the six boxes on the Scotian Shelf (Figure 8). Note that the parameters B_0, β_t and t_m are not included in the comparison given that they are not optimized in all models and they are not reported in the AZMP research document. Models G1 and G3 show the best agreement for all four optimized parameters with most of the points for all six boxes located on, or close to, the 1:1 line (Figure 8 a,c,e,g). Note that here, we did not perform linear regression for a given parameter between two optimization schemes to obtain statistical confidence, as this would break the 'dependent-independent' assumption for the parameters and the results would be meaningless. The main difference between models G1 and G3 occurs for the Lurcher Shoal box and in particular for the initiation and duration of the bloom, with a difference of up to 20-25 days (Figure 8a-c). The parameter H shows the best agreement as the values are similar for all the boxes and years (Figure 8g). Results from model G4 show some discrepancies with the ones from model G1 (Figure 8b, d, f, h), in particular for the initiation and duration of the bloom in the Lurcher Shoal and Georges Bank boxes. In some instances, optimization of model G4 returns null values for the initiation and duration of the blooms in the Lurcher Shoal box or there is a late initiation of the bloom that occurs more than 50 days later than the initiation derived from model G1 (Figure 8b).

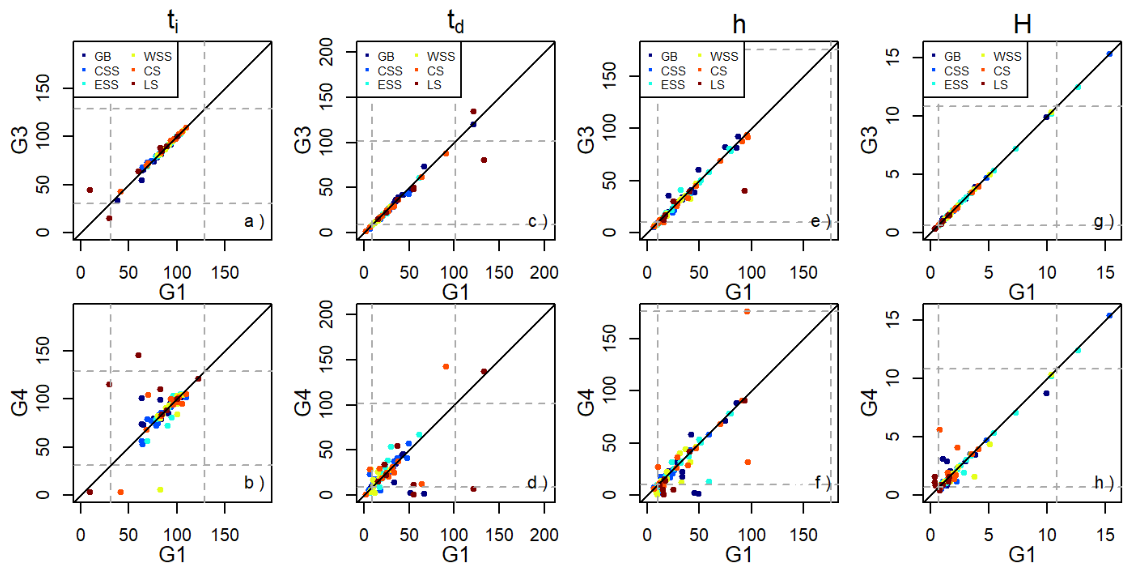

Figure 8: Fitted parameters t_i, t_d, h, and H for model G3 against the same parameters derived from model G1 (a, c, e, and g respectively) and fitted parameters t_i, t_d, h, and H for model G4 against the same parameters derived from model G1 (b, d, f, and h respectively). Dashed grey lines indicate the range of values for each parameter from Johnson et al. (2017).

The box and years when the *nls* was unable to fit were investigated, particularly when t_m was inferred from the data (Table 4). The primary reason for a failed fit was the occurrence of an early maximum chlorophyll-a concentration, mostly before day of year 50, as seen in one of the examples fits in Figure 7. Omitting these early maxima and optimizing the peak timing, t_m decreased the total number of failed fits from 7 to 4 (Table 4).

4.3.2 Threshold

Bloom metrics were obtained for two threshold values of 5% (Model T5) and 20% (Model T20) as well as when smoothing the original data using a LOESS approach (Models TL5 and TL20)

Table 4: Number of years (out of 15) when models G1 to G4 failed to converge for each region. Under each model variation, the left column indicates a fit using all data, and the right indicates excluding maximum values prior to day of year 50.

	G1		G2		G3		G4	
GB	2	2	2	2	2	2	2	2
CSS	2	2	1	1	1	0	1	0
ESS	0	0	0	0	0	0	0	0
WSS	0	0	0	0	1	0	2	1
CS	2	2	3	3	0	0	0	0
LS	7	6	7	7	3	2	4	4
Total	13	12	13	13	7	4	9	7

(Figure 9). Both threshold values provided similar bloom characteristics in 2007 (Figure 9a, e and g). However, when winter concentration values are rather high, as observed in 2016, the threshold method provided a very early initiation of the bloom (around day of year 10), which is unrealistic (Figure 9b, d, f, h)). When smoothing the data with a LOESS function (Figure 9c,d), the threshold method did not return any initiation date for either 2007 or 2016 with a 5% threshold.

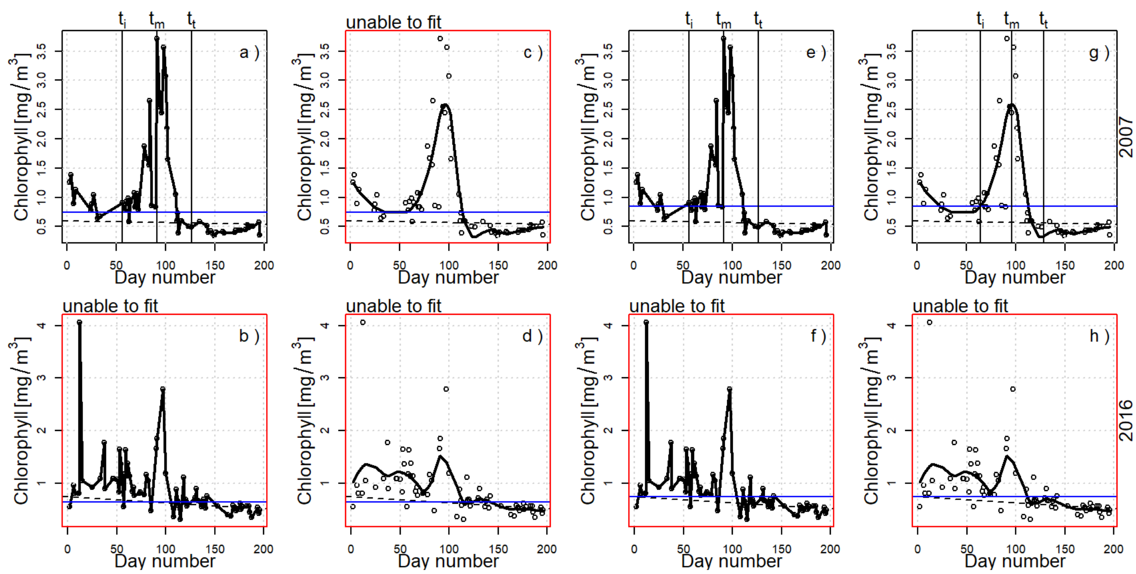

Figure 9: Case study of the threshold method using annual data from the Central Scotian shelf box for 2007 and 2016: 5% threshold with no fit (a-b), 5% threshold with LOESS fit (c and d), 20% threshold with no fit (e-f) and 20% threshold with a LOESS fit (g-h). The black vertical lines indicate a fit, red frames indicate failure of the method. The blue line indicates the annual median of chlorophyll-a concentration times the threshold value, and the dashed black line indicates the background using the 25th quantile regression.

In general, using a 20% threshold rather than a 5% threshold resulted in similar or later initiation of the spring bloom and a similar or earlier termination of the bloom, resulting in a similar or shorter duration of the spring bloom of up to 100 days or more (Figure 10a and i). The magnitude, h, and amplitude, H, of the spring bloom remained relatively similar when using either the 5 or 20% threshold (Figure 10e and g). Note that the magnitude computed using the 20%

threshold can be slightly smaller than the one computed using the 5% threshold. The TL5 model had an important impact on the retrieval of all four spring bloom indices (Figure 9b, d, f and h). The initiation of the bloom using TL5 and TL20 predicted a later initiation than T5 and T20, and thus a shorter duration (Figure 10d and j). The magnitude retrieved from the three models was consistent. TL20 model provided the smallest number of failed attempts across all regions (Table 5) with 16 failures out of 90 cases with or without including the first 50 days of the year.

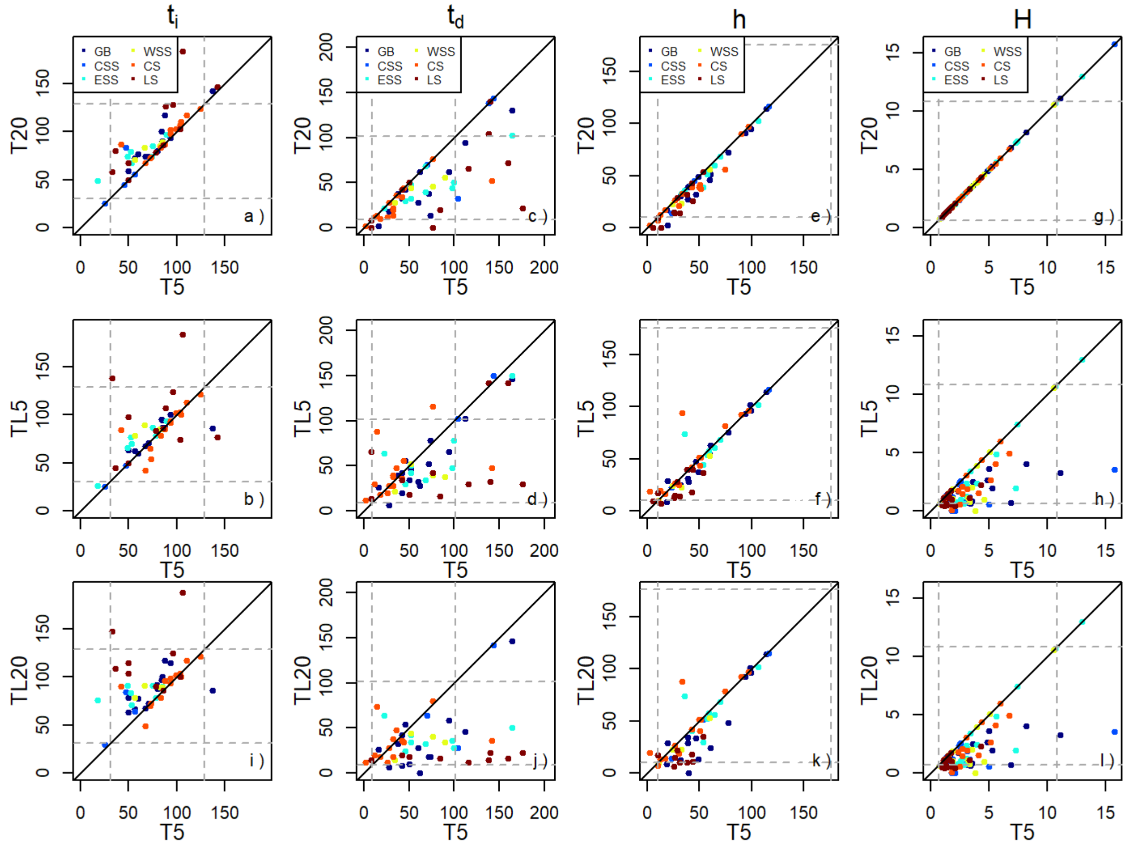

Figure 10: Intercomparison of the threshold method by looking at the residuals about the 1:1 line for four parameters, t_i (a, b, i), t_d (c, d, j), h (e, f, k), and H (g, h, i), reported on an annual basis. The top row (a to d) compares the two thresholds, 5 and 20%. The middle row (b to h) compares the results of the 5% threshold with or without the LOESS smoothing. And the bottom row (i to l) compares the results of the 5% threshold without LOESS smoothing and the 20% with LOESS smoothing. Dashed grey lines indicate the range of values for each parameter from Johnson et al. (2017).

The boxes and years when the threshold method was unable to infer the initiation time were investigated. As seen for the shifted Gaussian method, the threshold method also failed when the apparent maximum chlorophyll-a concentration occurred prior to day of year 50. However, results when omitting this maximum value prior to day of year 50 had negligible impact on the reliability, unlike the shifted Gaussian method (Table 5). In addition to removing dubious high values prior to the anticipated spring bloom, subtracting the background chlorophyll-a concentration decreased the convergence of the method (Table 5).

Table 5: Summary of the number of times out of 15 years the threshold method was unable to fit for each region using various threshold values and fitting methods. Under each model variation, the left column indicates results using all data, the middle indicates results that exclude maximum values prior to day of year 50, and the right indicates results where the background values were subtracted.

	TL5			TL20			T5			T20		
GB	0	0	1	0	0	1	0	0	3	0	0	1
CSS	9	9	8	7	7	5	11	11	9	8	8	9
ESS	2	2	6	2	2	5	6	6	6	1	1	6
WSS	8	8	7	4	4	7	11	10	9	7	6	9
CS	1	1	3	0	0	2	1	1	6	1	1	6
LS	3	3	4	3	3	4	4	4	5	1	1	4
Total	23	23	29	16	16	24	33	32	38	18	17	35

4.3.3 Rate of Change

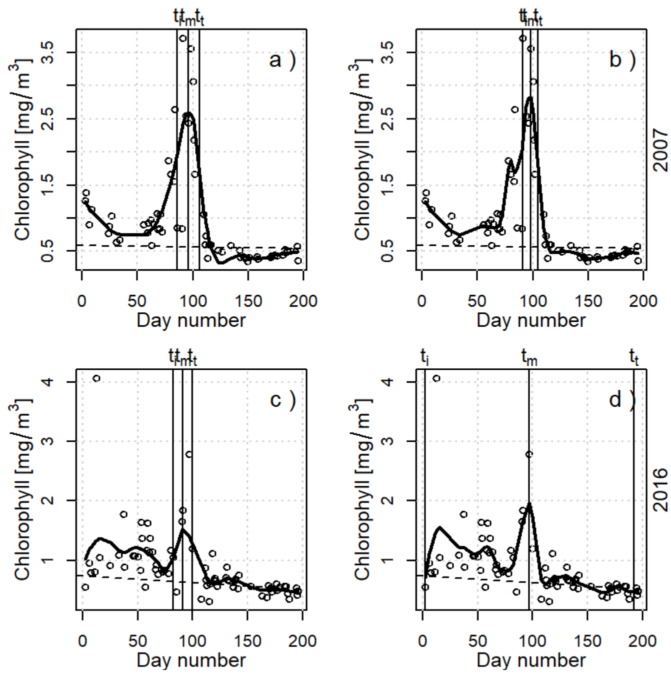

Figure 11: Case study for the rate of change method using annual data from the Central Scotian shelf box for 2007 (a and b) and 2016 (c and d) and for a fixed span of 1/3 (a and c) and an optimized span (b and d) for the LOESS fit. Black lines indicate LOESS fits. Black dashed lines correspond to the background chlorophyll-a concentration inferred using the 25$^{\text{th}}$ quantile regression.

The rate of change method requires a continuous annual time series since it detects changes in chlorophyll-a concentration; such criteria is met when a LOESS smoothing is applied. Similar to the shifted Gaussian and threshold methods, the rate of change method was applied to data for the years 2007 and 2016 (Figure 11). Two spans were tested when smoothing the time series: a fixed span set arbitrarily to a value of 0.33 as for the threshold method and an optimized span that minimized the root-mean-square difference between the original data and the loess smoothing. In

2007, the optimized span provided a more detailed smoothing of the chlorophyll-a concentration time series with a secondary peak that can be observed around day of year 85 (Figure 11a and b). In 2007, the initiation and timing of the maximum chlorophyll-a concentration were slightly earlier when the fixed span was used, however the date of termination remained similar perhaps due to the fact that the decrease in phytoplankton biomass is more monotonic than the increase. In 2016, the start and termination of the bloom are very different between the fixed and optimized span (Figure 11c and d), certainly due to the high variability in chlorophyll-a concentration. In fact, when the optimized span is used, the rate of change method provides, in many cases, the first day of the year as the start of the bloom due to a sudden decrease in chlorophyll-a concentration at a higher rate than the spring bloom.

Both span values for the LOESS smoothing resulted in a small number of failures to retrieve the bloom metrics for all regions (Table 6). The number of failures was further reduced when data before day of year 50 were removed from the analysis, and only the ESS, WSS and LS regions failed to provide bloom parameters in two years out of the six boxes out of the 15-year time series used in this study.

Table 6: The number of times out of 15 years the rate of change method was unable to fit for each region using a fixed span of 1/3 (RoCFS) and an optimized span (RoCOS) for the *loess* fit. Under each model variation, the left column indicates a fit using all data, and the right indicates excluding maximum values prior to day of year 50.

	RoCFS		RoCOS	
GB	0	0	0	0
CSS	0	0	0	0
ESS	1	1	1	1
WSS	1	0	2	1
CS	0	0	0	0
LS	1	1	1	1
Total	3	2	4	3

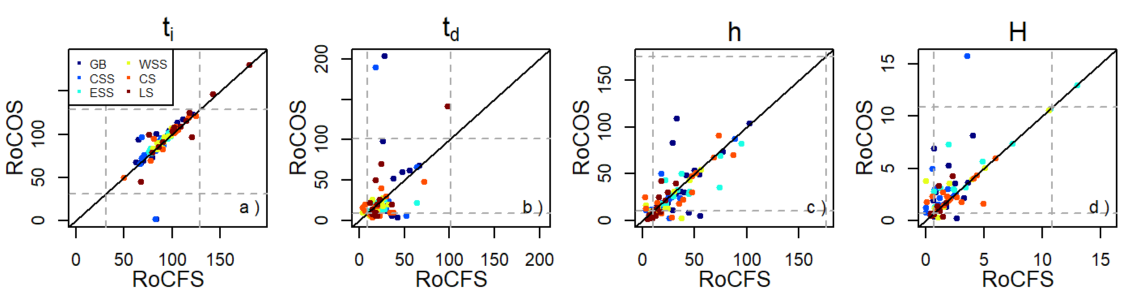

Figure 12: Intercomparison of the rate of change method by looking at the residuals about the 1:1 line for four parameters, t_i, t_d, h, and H, reported on an annual basis. Here, the various values of *span* used in the *loess* function are compared. Dashed grey lines indicate the range of values for each parameter from Johnson et al. (2017).

When applied to all boxes on the Scotian Shelf for all years, the initiation of the bloom remains similar between both RoCFS and RoCOS models with the exception of one year on Georges Bank (Figure 12a). Retrieval of the duration, magnitude and amplitude showed more discrepancies between the two models (Figure 12b-d). In general, the RoCOS model estimated a longer duration of the bloom than the RoCFS model, in particular on GB and the CSS with a duration of up to 200 days (Figure 12b). As a result of the long duration, the amplitude was often higher for the optimized-span model than for the fixed-span model. This is not surprising given that the amplitude is an integration of chlorophyll-a concentration over the length of the duration. The optimized span tends to better resolve the 8-day variations in chlorophyll-a concentration during a given year, therefore, it has the ability to account for high chlorophyll-a concentration as evidenced by the amplitude derived from the RoCOS model, which is often larger than for the RoCFS model (Figure 12d).

4.4 Method Comparison

For the three methods and their variations, those that converged the most were chosen to be compared with the other methods. These models are G4, TL20 and RoCFS. In addition, the G4 model was used for the reference when compared against TL20 and RoCFS as it has a term to account for the linear variation of the background chlorophyll-a concentration with time, which impacts the computation of the magnitude and amplitude of the spring bloom. For the rate of change method, both models provided consistent bloom metrics, after removing dubious high chlorophyll-a values prior to day of year 50, and the fixed span model was selected over the optimized span to remain consistent with the threshold model (i.e., TL20), which also used a LOESS fitting of the original data.

When applied to the theoretical time series, Figure 3, the threshold method inferred an earlier initiation time than the shifted Gaussian, and thus estimates a later termination due to the assumption of symmetry of the bloom, and resulting in a larger estimate of bloom magnitude. The rate of change method inferred an initiation around the same time as the shifted Gaussian, and consequently, other parameters were similar. Assuming the bloom shape follows a theoretical shifted Gaussian curve, the relationship of the parameters for the threshold and rate of change is similar to that of the threshold and shifted Gaussian, given that the shifted Gaussian and rate of change approaches are consistent, however note that this comparison should be taken with caution as the chosen threshold (*e.g.*, 5, 10 or 20%) will inherently impact all the bloom metrics.

Given that the number of successful fits increases when data points before day of year 50 are discarded, this criteria was applied when comparing the methods to each other. For the initiation of the bloom, comparing the Gaussian with both the threshold and rate of change methods graphically indicates an even spread about the 1:1 line (Figure 13a, e and i). Most of the differences between the three types of models occur for the derivation of the duration of the bloom with the shifted Gaussian approach tending to provide shorter duration of blooms than the two other approaches (Figure 13b, f and j). This results in a magnitude that is often smaller for the shifted Gaussian method than for the rate of change or threshold methods, though the difference is less pronounced between the Gaussian and RoC. Both *RoC* and threshold methods provide the same amplitude given that they both rely on a LOESS fit before applying the model (Figure 13h). However, the amplitude remains similar to the one derived using the shifted Gaussian (Figure 13d and l).

5 Discussion

Satellite ocean colour data are only collected in clear sky conditions, which impacts the continuity of the chlorophyll-a concentration time series, even when averaged over large areas such as the AZMP boxes of the Scotian Shelf. In addition, averaging the data within a given area drastically reduces the range of variation of daily and 8-day chlorophyll-a concentration, which can dampen phytoplankton blooms (Figure 6b, around day 300), or results in abnormally high-chlorophyll-a concentrations that can be treated as outliers (Figure 7b, around day 10). This emphasizes the fact that data processing and quality control are important steps when applying models to retrieve phytoplankton metrics to avoid unrealistic values (*e.g.*, bloom that lasts 200 days).

Overall, all three methods adequately estimated spring bloom parameters that can be annually reported in the AZMP Maritimes region. However, each method presents advantages and limitations. The ideal method should be able to deal with gaps in the time series and short term variations, on the order of a few weeks, and have the ability to identify weak blooms as recently observed on the Scotian Shelf (Johnson et al., 2018). The shifted Gaussian appears to be a suitable choice as the method assumes a functional form for the shape of the bloom, therefore, in general,

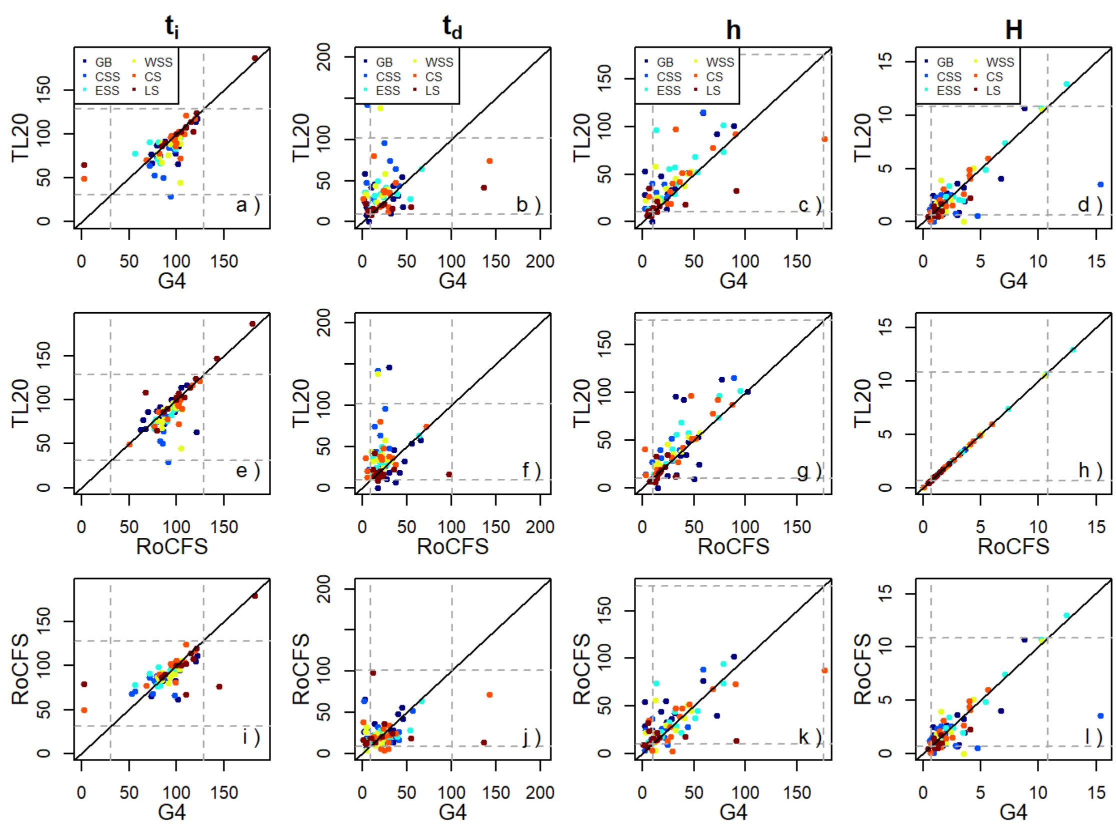

Figure 13: Comparison of bloom fit methods by looking at the residuals about the 1:1 line for four parameters, t_i, t_d, h, and H, reported on an annual basis. Here, the *loess* function uses a fixed span of 1/3. Dashed grey lines indicate the range of values for each parameter from Johnson et al. (2017).

it can withstand the variable nature of remotely-sensed data. The method proved to be the least sensitive of the three methods when it came to altering the function form. However, one of the main downsides of the method lies in the assumption of a symmetrical shape for the bloom, while the bloom decay is often less abrupt than the growth phase.

The threshold method proved to be the least robust approach as it is sensitive to the choice of threshold value, though it is the most widely used method in the literature, perhaps due to its ease of application to satellite data. The lack of reliability may be due to the lack of data prior to the bloom and the constraint that chlorophyll-a concentrations must remain below the threshold value for at least 14 consecutive days. The convergence increases when the data are smoothed, for instance using a LOESS smoother as in the current study. However, this step requires an additional parameter, a span window, which is set arbitrarily. The method also fails when chlorophyll-a concentrations during the bloom are abnormally low. This can be addressed by using a higher threshold, for example 30% rather than 20%, to distinguish the bloom from the background chlorophyll and successfully obtain the spring bloom metrics. An advantage of this method compared to the shifted Gaussian method, is that it can resolve the asymmetry of the phytoplankton bloom.

When comparing the results for the two threshold values (5 and 20%), a relationship is expected between t_i, t_d, and h. Application of a low threshold will result in estimating an earlier initiation of the bloom than for a high threshold. As a result, the estimates for the duration will be longer and the magnitude will be higher for a low threshold than for a high threshold. Application of the LOESS smoothing mainly affects estimates of the duration and amplitude of the bloom (Figure 10d, j, f, k), and tends to reduce the differences due to the choice of the threshold (result not shown here). Finally, the computation of the amplitude is independent to the threshold value as the relationship for both the 5% and 20% is nearly 1:1.

The rate of change method proved to be the most reliable method as it failed only in one case out of 90 possible cases. Like the threshold method, it requires that the data be smoothed to obtain a continuous time series and the choice of the degree of smoothing influences the computation of the bloom metrics. This method has the ability to resolve asymmetry of the phytoplankton bloom. Graphically, the method showed the best agreement with the shifted Gaussian method.

Based on these findings, we suggest, when using daily MODIS level 3 binned data, to:

1. Discard days when less than 10% of the pixels are valid within the region of interest,

2. Remove data points outside two standard deviations when computing the geometric mean (note that taking the $\log_{10}$ of chlorophyll-a concentration ensures the normal distribution of the data within a box),

3. Remove dubiously high chlorophyll-a concentrations prior to year day 50.

4. Provide the time (in days) of occurrence of the maximum chlorophyll-a concentration.

Keeping in mind these initial steps, we recommend to continue using the shifted Gaussian method. While the shifted Gaussian method can supply all parameters, only the initiation of the bloom and thus the termination should be used. The magnitude, h, and amplitude, H, should be calculated using the daily chlorophyll-a concentration, rather than the fitted curve. We showed that the inclusion of a linear term to resolve annual variation in the background chlorophyll-a concentration had a minimal impact on the bloom metrics, however, this value could be reported in the annual AZMP research document.

6 Future Work

The main objective of this study was to investigate three different methods to retrieve phytoplankton bloom metrics. In addition, we have documented in detail the method that is currently used in AZMP reporting. The findings presented here have only focused on the Maritimes region, however, the AZMP also extends to the Québec and Newfoundland regions, both of which deal with different bloom dynamics. The Québec region focuses on areas that are constrained by land and sea ice, which can impact the amount of reliable satellite data. In addition, the Gulf of St-Lawrence is subject to important runoff from the St-Lawrence river resulting in high concentrations of dissolved organic matter that affects the performance of generic chlorophyll-a algorithms (Laliberté et al., 2018). The Newfoundland region is similar to the Maritimes region in the sense that their boxes are located in open water, however, the latitude range is much larger and extends further north into the Labrador Sea from about 55 to 65°N, and thus the blooms are generally later and shorter. Additional attention should be given to the range of days of the year used to calculate bloom metrics, the initial values for the parameters in the optimization, and the lower and upper limits of the parameters that are supplied to the shifted Gaussian method for each region, as the ones provided for the Maritimes region might not work for the Newfoundland and Labrador region.

The data used in this report were restricted to the MODIS-Aqua ocean colour sensor, while two other satellites are used in the AZMP reporting (Figure 1). The methods should therefore be tested on the VIIRS and SeaWiFS data sets. In addition, attention should be given to the merging of the time series from the three different sensors to ensure a unique time series that is free from sensor bias.

Currently, only bloom parameter scorecards are presented in the annual reports. We recommend that additional details on the satellite data should be included, perhaps in an annex. In particular, the annual time series of chlorophyll-a concentration should be supplied, along with additional details on how the data were processed to provide more context to the reader. This would only add one to three pages of additional figures, depending on the region, and an extra paragraph or two of text.

An evaluation of the box boundaries should be completed. For instance, some of the boxes are currently in close proximity or overlapping land, which introduces bias in the data. Some of the boxes do not adequately summarize the region in question, especially when a box encompasses a region of high production and a region of lower production (*e.g.*, Georges bank). A novel approach would be to compute bloom metrics on a pixel-per-pixel basis, such that bloom parameters would be presented over the entire Northwest Atlantic rather than discrete areas and annual information could be presented in anomaly map plots in order to see if there is larger regional variability in bloom dynamics, as done for the sea-surface temperature.

REFERENCES

Brody, S., Lozier, M., and Dunne, J. (2013). A comparison of methods to determine phytoplankton bloom initiation. *Journal of Geophysical Research: Oceans*, 118(5):2345–2357.

Clay, S., Peña, A., DeTracey, B., and E.Devred (2019). Evaluation of satellite-based algorithms to retrieve chlorophyll-a concentration in the canadian atlantic and pacific oceans. *Remote Sensing*, 11(22):2609.

Denman, K. and Platt, T. (1976). The variance spectrum of phytoplankton in a turbulent ocean. *Journal of Marine Research*, 34(2):593–601.

Devine, L., Plourde, S., Starr, M., St-Pierre, J.-F., St-Amand, L., Joly, P., and Galbraith, P. S. (2015). *Chemical and Biological Oceanographic Conditions in the Estuary and Gulf of St. Lawrence during 2014*. Canadian Sceince Advisory Secretariat.

Groetsch, P., Simis, S., Eleveld, M., and Peters, S. (2016). Spring blooms in the baltic sea have weakened but lengthened from 2000 to 2014. *Biogeosciences*, 13(17):4959–4973.

Harrison, G., Johnson, C., Head, E., Spry, J., Pauley, K., Maass, H., Kennedy, M., Porter, C., and Soukhovtsev, V. (2009). *Optical, chemical and biological oceanographic conditions in the Maritime region in 2008*. Canadian Science Advisory Secretariat.

Henson, S. (2007). Water column stability and spring bloom dynamics in the gulf of alaska. *Journal of Marine Research*, 65(6):715–736.

Henson, S. and Thomas, A. (2007). Interannual variability in timing of bloom initiation in the california current system. *Journal of Geophysical Research: Oceans*, 112(C8).

Johnson, C., Devred, E., Casault, B., Head, E., and Spry, J. (2017). *Optical, chemical and biological oceanographic conditions on the Scotian Shelf and in the eastern Gulf of Maine in 2015*. Canadian Science Advisory Secretariat.

Johnson, C., Devred, E., Casault, B., Head, E., and Spry, J. (2018). *Optical, chemical and biological oceanographic conditions on the Scotian Shelf and in the eastern Gulf of Maine in 2016*. Canadian Science Advisory Secretariat.

Kim, H.-J., Miller, A., McGowan, J., and Carter, M. (2009). Coastal phytoplankton blooms in the southern california bight. *Progress in Oceanography*, 82(2):137–147.

Koenker, R. (2018). *quantreg: Quantile Regression*. R package version 5.38.

Laliberté, J., Larouche, P., Devred, E., and Craig, S. (2018). Chlorophyll-a concentration retrieval in the optically complex waters of the st. lawrence estuary and gulf using principal component analysis. *Remote Sensing*, 10(2).

Lemos, A., Ghisolfi, R., and Mazzini, P. (2018). Annual phytoplankton blooming using satellite-derived chlorophyll-a data around the vitória-trindade chain, southeastern brazil. *Deep Sea Research Part I: Oceanographic Research Papers*, 136:62–71.

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, O. B. P. G. (2014). MODIS-Aqua Ocean Color Data. Accessed on 2018/07/16.

Pépin, P., Maillet, G. L., Fraser, S., Doyle, G., Robar, A., Shears, T., and Redmond, G. (2017). *Optical, chemical, and biological oceanographic conditions on the Newfoundland and Labrador Shelf during 2014- 2015.* Canadian Science Advisory Secretariat.

Pépin, P., Maillet, G. L., Fraser, S., Shears, T., and Redmond, G. (2011). *Optical, chemical, and biological oceanographic conditions on the Newfoundland and Labrador Shelf during 2009 and 2010.* Canadian Science Advisory Secretariat.

Pépin, P., Maillet, G. L., Fraser, S., Shears, T., and Redmond, G. (2013). *Optical, chemical, and biological oceanographic conditions on the Newfoundland and Labrador Shelf during 2011-12.* Canadian Science Advisory Secretariat.

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

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Racault, M.-F., Le Quéré, C., Buitenhuis, E., Sathyendranath, S., and Platt, T. (2012). Phytoplankton phenology in the global ocean. *Ecological Indicators*, 14(1):152–163.

Racault, M.-F., Raitsos, D., Berumen, M., Brewin, R., Platt, T., Sathyendranath, S., and Hoteit, I. (2015). Phytoplankton phenology indices in coral reef ecosystems: Application to ocean-color observations in the Red Sea. *Remote Sensing of Environment*, 160:222–234.

Sapiano, M., Brown, C., Schollaert Uz, S., and Vargas, M. (2012). Establishing a global climatology of marine phytoplankton phenological characteristics. *Journal of Geophysical Research: Oceans*, 117(C8).

Siegel, D. A., Doney, S. C., and Yoder, J. A. (2002). The north Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis. *Science*, 296(5568):730–733.

Zhai, L., Platt, T., Tang, C., Sathyendranath, S., and Hernández Walls, R. (2011). Phytoplankton phenology on the scotian shelf. *ICES Journal of Marine Science*, 68(4):781–791.