Gudmundsson, Kristinn and Heath, Mike R. and Clarke, Elizabeth D., Marine Laboratory, Aberdeen, Scotland (Funder), Marine Research Institute, Reykjavik, Iceland (Funder) (2009) Average seasonal changes in chlorophyll a in Icelandic waters. ICES Journal of Marine Science, 66 (10). pp. 2133-2140. ISSN 1054-3139 , http://dx.doi.org/10.1093/icesjms/fsp208

This version is available at https://strathprints.strath.ac.uk/18622/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk

The Strathprints institutional repository (https://strathprints.strath.ac.uk) is a digital archive of University of Strathclyde research outputs. It has been developed to disseminate open access research outputs, expose data about those outputs, and enable the management and persistent access to Strathclyde's intellectual output.
Average seasonal changes in chlorophyll $a$ in Icelandic waters

Kristinn Guðmundsson, Mike R. Heath, and Elizabeth D. Clarke


The standard algorithms used to derive sea surface chlorophyll $a$ concentration from remotely sensed ocean colour data are based almost entirely on the measurements of surface water samples collected in open sea (case 1) waters which cover $\sim60\%$ of the worlds oceans, where strong correlations between reflectance and chlorophyll concentration have been found. However, satellite chlorophyll data for waters outside the defined case 1 areas, but derived using standard calibrations, are frequently used without reference to local in situ measurements and despite well-known factors likely to lead to inaccuracy. In Icelandic waters, multiannual averages of 8-d composites of SeaWiFS chlorophyll concentration accounted for just 20% of the variance in a multiannual dataset of in situ chlorophyll $a$ measurements. Nevertheless, applying penalized regression spline methodology to model the spatial and temporal patterns of in situ measurements, using satellite chlorophyll as one of the predictor variables, improved the correlation considerably. Day number, representing seasonal variation, accounted for substantial deviation between SeaWiFS and in situ estimates of surface chlorophyll. The final model, using bottom depth and bearing to the sampling location as well as the two variables mentioned above, explained 49% of the variance in the fitting dataset.

Keywords: chlorophyll $a$, modelling, North Atlantic, remote sensing, seasonal, Subarctic.

Introduction

Chlorophyll concentrations derived from the sea surface reflectance data gathered by satellite borne sensors have become widely accepted as measures of phytoplankton abundance in oceanic waters. Indeed for many ecological applications, no ground truth data sources are now available to supplement these data. Nevertheless, there are a number of well known unsolved problems with satellite collected data (Siegel et al., 2005a; Werdell and Bailey, 2005; Alvain et al., 2006; Lee and Hu, 2006) that sometimes are dismissed and may lead to considerable uncertainty, e.g. for the results of primary production estimates (Siegel et al., 2005b). First, the observations apply to just a thin layer at the ocean surface, and subsurface chlorophyll concentrations remain undetected. Second, the standard algorithms for deriving chlorophyll concentration from spectral data are almost entirely based on concurrent reflectance and in situ measurements at open ocean sites, referred to as case 1 waters (O’Reilly et al., 1998; Hooker and McClain, 2000). Considerable regional and seasonal deviations in satellite surface chlorophyll $a$ have been noted when compared with measurements from near surface water samples (Dierssen and Smith, 2000; Burenkov et al., 2001; Sathyendranath et al., 2001; Gregg and Casey, 2004), especially from high latitudes and areas characterized as case 2 waters. The variable accuracy of calculated satellite chlorophyll estimates (Gregg and Casey, 2004), especially for case 2 areas, has been attributed to material of different colour in the water that may be misinterpreted as chlorophyll (Mobley et al., 2004). The definition applies especially to shallow and turbid waters over continental shelves, or otherwise influenced from land (Kirk, 1994), but also to boreal and Subarctic areas of the North Atlantic (Lee and Hu, 2006). Others have cited biological variation as a possible explanation because different phytoplankton communities vary in their absorption characteristics (Cota et al., 2003; Sathyendranath et al., 2004; Alvain et al., 2006).

Simultaneous measurements of in situ chlorophyll $a$ and ocean colour records from satellites (Gregg and Casey, 2004; Werdell and Bailey, 2005) are in some regions too few for detailed comparison (e.g. in Subarctic Atlantic waters). Therefore, blending methodologies have been applied to produce climatological models of the seasonal distribution of surface chlorophyll. For example, Clarke et al. (2006) used a penalized regression spline analysis to model in situ measurements of chlorophyll as a three dimensional function of day of the year, seabed depth, and multiannual average (1997–2002) of 8 d composites of SeaWiFS chlorophyll (CHL$_{sat}$). The model was applied to predict surface chlorophyll for any day of the year, and at any location in the domain, in this case the Northeastern Atlantic including the northern North Sea, 56°–72°N and 30°–80°E, and later extended (Speirs et al., 2006) to the northern North Atlantic, 30°–80°N and 80°–85°E. The variable seabed depth was a proxy for the reflectance water type, i.e. case 1 water in open oceans or case 2 water over continental shelves.
The productive waters around Iceland and bordering the Arctic basin present a particular challenge for estimating the distribution of chlorophyll concentration from reflectance data. The continental shelf there is narrow with an anticyclonic coastal current driven by freshwater run off, which is occasionally loaded with glacial clay. Ocean currents of contrasting temperature and salinity flow along sections of the continental slope (Figure 1a): the warm saline North Atlantic Current (NAC) and the Irminger Current (IC) in the southwest, and the cold, less saline East Icelandic Current (EIC) in the northeast (Valdimarsson and Malmberg, 1999). As a result, the northern and southern Icelandic shelves have characteristically different water column stability and seasonal cycles of phytoplankton abundance (Gudmundsson, 1998). In general, the area may be characterized as case 2 waters, according to the definition of Lee and Hu (2006).

Clarke et al.’s (2006) Northeast Atlantic model (NEA model) covers the water around Iceland, but although the assembled water sample analyses used for the study included some 13 000 stations visited by multinational survey vessels between 1986 and 1999, just 203 stations were in Icelandic waters. We located additional chlorophyll data at the Marine Research Institute (MRI) in Iceland that had not previously been collated for spatial and temporal syntheses. The new dataset was treated as an independent test of the chlorophyll distributions predicted by the NEA model, then used to produce an alternative model (IS model) based on the penalized regression spline methodology. An analysis of the results and an interpretation of the findings are presented.

Material and methods

The study area, 62° 69′N and 30° 6.5′W, covers the shelf and slope waters around Iceland. Chlorophyll a data were collated from spectrophotometric and fluorimetric analysis of pigment extracts from MRI’s water samples in this region from 1986 to 2005. In all, 1470 stations (Figure 1b) were collected (i) during annual hydrographic monitoring surveys in May/June, (ii) from the pumped seawater supply to flow through instruments aboard MRI vessels, and (iii) at fixed stations on the shelf sampled at varying intervals throughout the year. At each station, the interpolation scheme described by Clarke et al. (2006) was applied to estimate the average concentration of chlorophyll in the upper 5 m (CHLsurf) from the various discrete depth water samples.

Few stations were sampled during winter, and the SeaWiFS sensor is unable to provide data from high latitudes then because of the low angle of the sun. Therefore, the analysis was restricted to a 9 month period from mid February to mid November. The final number of CHLsurf observations used for the analysis was 1614, of which 179 were common with the dataset of Clarke et al. (2006).

In common with Clarke et al. (2006), we used the calibrated output from the 2002 NASA OC4v4 reprocessing (O’Reilly et al., 1998) of the SeaWiFS data archive, compiled into multiannual (1997-2002) averages over 5′ latitude x 5′ longitude pixels, composites for successive 8 d intervals throughout the year (CHLsurf). Full details of the procedures for in filling missing pixels and processing of these data are provided by Clarke et al. (2006).

Aiming for an optimal distribution of the data for fitting a new model, the whole data assemblage was allocated to intervals of 8 d throughout the year, then one observation was selected per interval, at random, from each 1/4′ latitude x 1/2′ longitude cell in the model region. This created a subset of 910 observations (stations) which were, as far as possible, evenly distributed in space and time. The remaining data were used as a secondary dataset for subsequent testing. Obviously, because of the over weighting of data from the annual monitoring in late May, the dataset was biased with regard to temporal distribution, especially the secondary dataset.

Seabed depths for each cell with SeaWiFS data were determined from the ETOPO2 (2′ latitude resolution) global relief dataset (National Geophysical Data Centre; http://www.ngdc.noaa.gov/mgg/global/global.html).

The dominant circulation regime around Iceland is a clockwise flow along the shelf and shelf edge. Major rivers discharge glacier meltwater at various points, and ocean water masses are entrained into the circulation over the shelf (Figure 1a), causing changes in temperature, salinity, and nutrient concentrations. To caricature the possible effect of river discharge and entrainment on the patterns of chlorophyll distribution, the angular bearing of each sampling station from a central position in Iceland (65°N 19°W) was used as an additional covariate in the IS model.

Models were fitted to this new Icelandic dataset using the same approach as Clarke et al. (2006) with thin plate splines and tensor product splines (Wood, 2006). The models were fitted using the package mgcv, version 1.3-1, in R 2.1.1 (R Development Core

![Figure 1. Maps of (a) seabed depth and the main system of currents around Iceland, the North Atlantic Current (NAC), the Irminger Current (IC), the East Icelandic Current (EIC), and the East Greenland Current (EGC), and sampling stations (dots) for time-series of chlorophyll. (b) Geographical distribution of sampling stations for the collated CHLsurf used in this study.](image-url)
Several variants of the model were tested to determine the most appropriate smoother. The degrees of freedom assigned to smoothing each explanatory variable were chosen by recursively fitting the model and systematically varying the degrees of freedom associated with each variable in turn and comparing GCV scores.

**Results**

Correlation between the full dataset of Icelandic sampling stations, CHLsurf ($n = 1614$), and the raw SeaWiFS composite values CHLsat ($r^2 = 0.23$) was not markedly different from that with NEA model predictions (CHLNEA) at the corresponding locations and days of the year ($r^2 = 0.27$). From this, we conclude that although the NEA model provides significant improvement in predictions of average CHLsurf over the whole Northeast Atlantic compared with composite SeaWiFS data, the same does not apply for Icelandic waters because of the sparseness of the available data in this region.

A GAM with one dimensional cubic spline smoothing and a normal error distribution was used for an initial exploration of the dependence of log(CHLsurf) on each of the explanatory variables, log(CHLsat), square root transformed seabed depth, angular bearing around Iceland, and day of the year. The two last variables were treated as cyclic. The predictive influence and distribution of each variable over the range of values measured is shown in Figure 2.

The variance explained by the model, as covariates were added sequentially, was 31% for CHLsat and increasing to 41% with the addition of day of the year. Addition of either one or both of the remaining variables (seabed depth and angular bearing) only increased the variance explained by a further 2%. Nonetheless, the best model was a four dimensional tensor product smoother of all covariates, optimized for the degrees of freedom assigned to each explanatory variable. This IS model explained 49% of the variance in log(CHLsurf), using the fitting dataset. Seabed depth and bearing contributed substantially to the final model, because the variance explained was 44% for a two dimensional tensor product smooth using CHLsat and day of the year as predictors. The CHLsat was, as expected, the primary predictive variable, but the seasonal variation was obviously important too.

The CHLsurf in the fitting dataset and the corresponding composite SeaWiFS CHLsat values, along with the predicted values from the NEA model (CHLNEA) and the new IS model (CHLIS), were plotted (Figure 3) for examination of the variation in the scatter. This was then repeated for CHLsurf in the secondary dataset, and the corresponding squared correlation coefficients ($r^2$) were calculated. Although the $r^2$ values for the secondary dataset were low, the scores were highest for the IS model for both datasets (Table 1). To assess whether the three predictors (CHLsat, CHLNEA, and CHLIS) were biased, we calculated the average difference between each predictor and the observed CHLsurf value it was being used to predict (Table 1). This value would be negative when the predictor consistently underestimated the values observed and positive when the predictor consistently overestimated those values. The IS model performed better than both the NEA model and the satellite values, apparently being almost unbiased for both the fitting and testing datasets. However, this apparent lack of bias was really the result of the model predictions being negatively biased at high observed chlorophyll values and positively biased at low values (Figure 3). This is

![Figure 2](image-url)
a well known problem with satellite data, which the model has partially overcome for these data. The NEA model performs the worst, being the most (negatively) biased, again a well known phenomenon in that using predictions from inappropriate models is likely to cause bias. The fact that the results are similar for both the fitting and the test datasets supports our contention that the high $r^2$ in the test dataset is attributable to high variability in the data rather than poor predictions.

Extreme values of CHL$_{surf}$ were poorly predicted by all predictors (Figure 3). Broken down by month (Figure 4), the correlations for the IS model during the months May–July were clearly weaker than for the rest of the year. For closer examination, we performed separate model analyses, on the one hand restricted to data from waters north and east of Iceland (cold Arctic waters) and on the other hand to waters to the south and west (warm Atlantic waters), to determine whether systematic differences in seasonality between these regions might account for the weaker fit in May/June. However, the analyses did not result in any substantial change in the overall explained variance or suggest any obvious and plausible hypotheses.

For visual examination of the changes in horizontal distribution during the growth season, the fitted model was used to
predict average surface chlorophyll (CHLsurf) on the first day of each month between March and November over a 5° latitude × 5° longitude grid of seabed (Figure 5). Further, at four locations where detailed annual time series were available in the dataset (Figure 1a), the relationships between in situ measurements (CHLsurf), SeaWiFS CHLsurf, and predicted CHLsurf according to the models were examined (Table 2). The IS model predictions provided the closest correlation with the observed variable CHLsurf at three of the locations and similar or lower bias than the composite SeaWiFS data and the NEA model predictions.

**Discussion**

The highly variable correlations between CHLsurf and CHLsat found for case 2 waters are attracting increased attention (Hu et al., 2000; Cota et al., 2003; Gregg and Casey, 2004; Magnuson et al., 2004; Maritorena and Siegel, 2005; Brown et al., 2008; Komick et al., 2009). The variations are for the most part considered to be caused by material of variable colour dissolved or suspended in the seawater (Morel and Belanger, 2006) that is misinterpreted as chlorophyll when using standard algorithms for calculating CHLsurf from satellite records. Some suggestions for resolving the problems have been proposed (Hu et al., 2000; Magnuson et al., 2004; Siegel et al., 2005a; Wynne et al., 2006), but await further tests and evaluation. Given the shortage of available CHLsurf data that meet the criteria set for validating algorithms (Werdell and Bailey, 2005) for our study area, an alternative is to analyse climatological satellite data and multiannual observations.

The analysis described here produced a fitted statistical model (IS model), based on averaged 8 d composites of SeaWiFS chlorophyll data, day of the year, seabed depth, and the angular bearing to the location of water sampling around Iceland. The model, adjusted to the data available inside the study area, may be used to predict the surface chlorophyll for any day and location inside the region modelled. The predictions were tested against measurements of chlorophyll a in water samples, and the calculated values of $r^2$ were compared with that of predictions according to a model for the whole Northeast Atlantic (NEA model) and the average SeaWiFS composite values, testing the relative performance (Figure 3, Table 1). Further, the correlations between water sample measurements of surface chlorophyll a and either the values inverted from satellite ocean colour records or those predicted by the IS model at four locations used for seasonal studies (Table 2) support the

<table>
<thead>
<tr>
<th>Source</th>
<th>Fitting data (n = 910)</th>
<th>Testing data (n = 704)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeaWiFS</td>
<td>$r^2 = 0.26$ Bias 0.26</td>
<td>$r^2 = 0.17$ Bias 0.25</td>
</tr>
<tr>
<td>NEA model</td>
<td>$r^2 = 0.31$ Bias 0.42</td>
<td>$r^2 = 0.21$ Bias 0.50</td>
</tr>
<tr>
<td>IS model</td>
<td>$r^2 = 0.49$ Bias 0.00</td>
<td>$r^2 = 0.30$ Bias 0.04</td>
</tr>
</tbody>
</table>

**Figure 4.** The scatter of chlorophyll a, measured from water samples vs. that predicted according to the IS-model for the months March – November.
notion that the IS model is an improvement on the perception of climatological spatial and temporal patterns of surface chlorophyll around Iceland. The predicted CHL$_{IS}$ explains more of the variance in the available in situ measurements of surface chlorophyll measured from water sampled around Iceland than the raw SeaWiFS composite data and the NEA model predictions. The paramount reason for the high variability, found when analysing the relationship between spectral reflectance and chlorophyll, is connected to seasonal change (Figures 2 and 4). An obvious explanation is the variation in chlorophyll concentration observed in these waters, such as may be caused by storms (Thórdardóttir, 1986).

The angular bearing of locations around Iceland was used to represent possible differences between the Arctic waters masses overlying the shelf north and east of Iceland and the Atlantic waters overlying the southern and western shelf, as well as other variable environmental influence near land (e.g. silt in glacial rivers and wind borne dust). Seabed depth, as a single predictive variable, contributed least to explaining the variability in CHL$_{surf}$. Leaving seabed depth out of the analysis, however, resulted in a greater reduction in overall variance explained than the contribution of the variable alone implied, indicating an interaction with other covariates, probably the angular bearing, because the two together act as a spatial index.

Like the NEA model, the IS model was based on the SeaWiFS chlorophyll data averaged for the years 1998–2002, and predictions from the NEA- and IS-models calculated for available time-series at GR (Grimsey 66°15'N 18°34'W), EY (Eyjafjörður 66°28'N 18°04'W), SW1 (63°46'N 21°04'W), and SW2 (63°29'N 20°20'W).

Table 2. Correlation coefficients ($r^2$) and bias estimate calculated for in situ measurements of chlorophyll a, the corresponding values from the 8-d composite SeaWiFS (OC4v4) averages for the years 1997–2002, and predictions from the NEA- and IS-models calculated for available time-series at GR (Grimsey 66°15'N 18°34'W), EY (Eyjafjörður 66°28'N 18°04'W), SW1 (63°46'N 21°04'W), and SW2 (63°29'N 20°20'W).

<table>
<thead>
<tr>
<th>Source</th>
<th>GR (n = 79)</th>
<th>EY (n = 12)</th>
<th>SW1 (n = 117)</th>
<th>SW2 (n = 115)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r^2$</td>
<td>Bias</td>
<td>$r^2$</td>
<td>Bias</td>
</tr>
<tr>
<td>SeaWiFS</td>
<td>0.12</td>
<td>0.36</td>
<td>0.38</td>
<td>0.40</td>
</tr>
<tr>
<td>NEA model</td>
<td>0.15</td>
<td>0.78</td>
<td>0.60</td>
<td>1.01</td>
</tr>
<tr>
<td>IS model</td>
<td>0.22</td>
<td>0.23</td>
<td>0.49</td>
<td>0.29</td>
</tr>
</tbody>
</table>

As most of the collated data were collected during the latter half of May, during the annual regional monitoring surveys, there is an unavoidable bias in terms of temporal distribution. Moreover, as the first selection for the fitting dataset aimed for a uniform spatial and temporal distribution, the remaining secondary dataset was obviously and inevitably biased towards the sampling in May. The monthly plots of CHL$_{IS}$ vs. CHL$_{surf}$ (Figure 4) illustrate the scatter in the months May–July, during the high growth season. In light of the uneven distribution towards the latter half of May at the time of the spring bloom in the region (Gudmundsson, 1998), one may expect low values of $r^2$, especially when testing the correlations for the secondary dataset (Table 1). Therefore, a
reason for the poor fit during the high growth season may be that CHL$_{sat}$ are averaged values, from several years of SeaWiFS records, whereas CHL$_{surf}$ are highly variable in situ measurements. Obviously, the two subsamples of data are not entirely comparable.

Visual examination of isolvels drawn according to the results of measurements of chlorophyll in water samples during annual cruises around Iceland, and comparison with corresponding 8 d composite images of chlorophyll distribution made available by NASA, had demonstrated some correlation. However, for a detailed study on the exact correlation between CHL$_{sat}$ and CHL$_{surf}$, one needs simultaneous high quality datapoints, which are not yet available. Calculating the average values of satellite chlorophyll for a number of years was a mean to obtain a complete dataset on satellite chlorophyll for the whole region, needed because the persistent cloud cover results in poor coverage of satellite data in the region (Clarke et al., 2006).

Our study has demonstrated the need for local corrections of CHL$_{sat}$ shown here for a multiannual average of 8 d composites from SeaWiFS ocean colour data. The results confirm the method of Clarke et al. (2006) as a valuable approach to adjust inverted chlorophyll from satellite ocean colour records to average CHL$_{surf}$ and show that a locally adapted model is needed to produce realistic predictions of CHL$_{surf}$ in a regional domain. The next rational step will be to initiate a sampling scheme for high quality sea truth measurements (Gregg and Casey, 2004; Yuan et al., 2005), intended to construct a regionally adapted algorithm to correct regional CHL$_{sat}$ values or to test some general algorithms that may be able to cope with both case 1 and 2 waters and seasonal and local variations. To date, the IS model predictions presented here are the best available information (interpolation) on spatial and temporal distribution of surface chlorophyll around Iceland.

**Acknowledgements**

Our study was facilitated by the Marine Laboratory, Aberdeen, Scotland, and the Marine Research Institute, Reykjavik, Iceland. We are grateful for access to the data on in situ chlorophyll and for the satellite records made accessible by the SeaWiFS project (Code 970.2) and the Distributed Active Archive Center (Code 902) at the Goddard Space Flight Center, Greenbelt, MD, USA, sponsored by the National Aeronautics and Space Administrations "Mission to Planet Earth" programme. The Remote Sensing Group at the Plymouth Marine Laboratory is acknowledged for compiling the SeaWiFS composite data as part of the MarProd Programme funded by the UK Natural Environment Research Council.

**References**


abundance, and seasonal dynamics of the copepod Calanus finmarchicus. Marine Ecology Progress Series, 131: 183–192.


