Predicting marine phytoplankton community size structure from empirical relationships with remotely-sensed variables

Carolyn Barnes, Xabier Irigoen, José a A de Oliveira, Simon Jennings

OXFORD UNIVERSITY PRESS Journal of Plankton Research

Predicting marine phytoplankton community size structure from empirical relationships with remotely-sensed variables

Journal:	*Journal of Plankton Research*
Manuscript ID:	JPR-2010-044.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	18-Jun-2010
Complete List of Authors:	Barnes, Carolyn; Centre for Environment, Fisheries and Aquaculture Science, Environment and Ecosystems Irigoen, Xabier; AZTI - Tecnalia, Marine Research Division De Oliveira, José; Centre for Environment, Fisheries and Aquaculture Science, Fisheries Jennings, Simon; Centre for Environment, Fisheries and Aquaculture Science, Environment and Ecosytems
Keywords:	Primary production, phytoplankton, trophic level, transfer efficiency, size composition, community, picoplankton, nanoplankton, microplankton

Predicting marine phytoplankton community size structure from empirical relationships with remotely-sensed variables

Carolyn Barnes[1], Xabier Irigoien[2], José A. A. De Oliveira[1] and Simon Jennings[1,3].

[1] Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, Suffolk, NR33 0HT, UK.

[2] AZTI - Tecnalia / Marine Research Division, Herrera kaia portualdea z/g 20110 Pasaia (Gipuzkoa), Spain

[3] School of Environmental Sciences, University of East Anglia, Norwich, NR7 4TJ, UK.

Abstract

The size composition of primary producers has a potential influence on the length of marine food chains and carbon sinking rates, thus on the proportion of primary production that is removed from the upper layers and available to higher trophic levels. While total rates of primary production are widely reported, it is also necessary to account for the size composition of primary producers when developing food web models that predict consumer biomass and production. Empirical measurement of size composition over large space and time scales is not feasible, so one approach is to predict size composition from environmental variables that are measured and reported on relevant scales. Here, we describe relationships between the environment and the size composition of phytoplankton communities, using a collation of empirical measurements of size composition from sites that include polar, tropical and upwelling environments. The size composition of the phytoplankton communities can be predicted using two remotely-sensed variables, Chl a concentration and sea surface temperature. Applying such relationships in combination allows prediction of the slope and location of phytoplankton size spectra and estimation of the percentage of different sized phytoplankton groups in communities.

Keywords: Primary production, community, phytoplankton, trophic level, transfer efficiency, size composition, picoplankton, nanoplankton, microplankton.

Introduction

Global primary production by marine phytoplankton is around 5×10^{10} tonnes C y^{-1} (Carr et al. 2006) but there are large regional variations in the proportion of this production being removed from the upper layer through sinking and therefore available to higher trophic levels. This is due to variations in absolute productivity among regions, with 50% of production estimated to come from 27% of ocean area (Longhurst et al. 1995), and to regional differences in phytoplankton community structure. Regionally, the factors that affect the availability of phytoplankton to a given size class of consumers are their (i) spatial and temporal distribution, (ii) palatability, (iii) abundance, and (iv) size composition (due to morphological constraints of consumers and because cell individual sinking rates are related to size (Smayda 1971)).

Marine pelagic food chains are strongly size-based, with larger predators eating smaller prey. This size-based predation is predominantly responsible for the transfer of energy from phytoplankton to

progressively larger animals and total production falls with body mass as trophic level rises (Sheldon et al. 1972). Mean ratios between predator and prey body mass (predator-prey mass ratio PPMR) are relatively constant in marine ecosystems (Barnes et al. 2010). Consequently, when primary producers are smaller, there are, on average, more steps in a food chain to a predator of given size. Further, since mean annual trophic transfer efficiency at each step is also relatively constant (Barnes et al. 2010), production by consumers of a given body size will be a smaller proportion of primary production in regions where the primary producers are smaller. Such an effect has been shown in lakes (Sprules and Munawar 1986), although it has not yet been reported in large-scale observations in marine systems (San Martin et al. 2006b).

With knowledge of PPMR and the factors that influence trophic transfer efficiency, estimates of primary production can be used to predict production at higher trophic levels (Dickie 1976). For global scale analyses of the transfer of energy from phytoplankton to higher trophic levels it is often convenient to use primary production estimates based on satellite measurements of surface chlorophyll concentration provided by the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) global time series (McClain et al. 2004). Some progress has been made with estimating phytoplankton cell sizes by linking phytoplankton absorption to phytoplankton size classes using a single variable, the optical absorption by phytoplankton at 443 nm, which can be derived from the inversion of ocean colour data (Hirata et al. 2008) but a complementary approach is to identify general relationships between remotely-sensed environmental variables and the size composition of phytoplankton communities.

Here, we seek to identify relationships between the observed size composition of phytoplankton communities and remotely-sensed environmental variables. Our aim is purely to use readily available remotely-sensed environmental variables such as surface chlorophyll and sea surface temperature to enable estimations of phytoplankton size parameters for input to models. To provide a comprehensive description of the size and relative abundance of cells we used 4 descriptors of community structure: (1) mean mass ($\log_{10}$), (2) the variance of mass ($\log_{10}$), (3) the slope of size spectra (relationship between the logarithm of total abundance by cell mass class and the logarithm of cell mass, with individuals binned to cell mass classes irrespective of species identity (Sheldon and Parsons 1967)) and (4) the range of cell masses that encompass a given proportion of total biomass or production. This information is necessary to predict phytoplankton production by size class in inputs to size-based models of production at higher trophic levels (Jennings et al. 2008; Blanchard et al. 2009).

Method

Abundance and species composition were determined for phytoplankton in 361 water samples collected at twelve sites: five transects from 48° N to 50° S in the Atlantic Ocean (hereafter AMT1-5, n = 125, taken at 7 m (second samples at each site, taken at the deep Chl a maximum were excluded from this analysis), the Benguela upwelling (n = 54), mesocosms in the Bergen fjord (n = 46), the Irminger Sea (n = 59), Long Island Sound (n = 7), the North Sea (n = 44), the Norwegian Sea (n =19) and the Oregon upwelling (n = 7). In all locations samples were taken in subsurface but see Irigoien et al. (Irigoien et al. 2005) for details. Sub-samples (100 ml) were settled (Utermőhl technique (Lund et al. 1958)) and individuals counted at the species level with an inverted microscope. Heterotrophic species were excluded from the analysis. Picoplankton was measured using flow cytometry (see Irigoien et al (Irigoien et al. 2004) for details). Biomass was calculated as the product of numerical abundance and cell mass. More details of sample positions, collection, processing and composition are provided by Irigoien et al. (Irigoien et al. 2004; Irigoien et al. 2005).

To assess the proportion of phytoplankton biomass (B) or production (P) that was attributable to cells in specified mass (M) ranges, we investigated expressing cumulative B or P as a function of M by attempting to fit various forms. All samples were successfully fitted by:

$$B = 100\,(1-\exp(-pM^{q})) \qquad \text{Equation 1}$$

The fitted relationships were used to predict the M ranges that contributed to 10%, 25%, 50%, 75%, 80% and 90% of total B or P (Fig. 1). For example,

$$M_{80\%} = (-\ln(0.2)/p)^{1/q} \qquad \text{Equation 2}$$

Since P was not measured directly, we calculated the net production rate (R) of individuals from M, accounting for photosynthetic active radiation (PAR) following López-Urrutia et al. (López-Urrutia et al. 2006):

$$\ln R = \ln(Nc) + \alpha \times \ln(M) - E \times (1/kT) + \ln\left(\frac{PAR}{PAR + K_m}\right) \qquad \text{Equation 3}$$

where Nc is the normalization constant (ln(Nc)= -11.28); α is the allometric exponent (α=1.05); E is the activation energy for photosynthetic reactions (E = 0.29); and PAR/(PAR + K_m) is the Michelis-Menten photosynthetic light response where PAR is the photosynthetically available radiation at the sample site and K_m is the half-saturation constant (K_m = 1.51) that represents the amount of quanta at which half the maximum photosynthetic activity is reached. R is expressed as metabolic rate in mmol of O_2 d^{-1} and M is expressed in pg C, *T* is the absolute temperature and PAR is irradiance in mol photons (Einsteins) m^{-2} d^{-1} (2003 PAR values for each sample site were taken from http://oceandata.sci.gsfc.nasa.gov/SeaWiFS/Mapped/Annual/par/). Rate in mmol of O_2 d^{-1} was converted to rate in pg of C $year^{-1}$ (molar mass of C = 12.01):

$$Rate\ in\ pg\ C\ year^{-1} = R / 12.01 \times 365 \times 10^{12} \qquad \text{Equation 4}$$

Biomass of each species of phytoplankton in each sample was calculated by multiplying the abundance by species mean M. Attempts to fit Pareto distributions to estimate the slopes of size-spectra (Vidondo et al. 1997) were unsuccessful since the data consisted of abundance by mean species' cell mass rather than individual mass measurements. Instead, traditional normalized biomass size spectra were defined by assigning each species to a $\log_2$ integer scale size class based on the mean cell size of the species. Total abundance of individuals at mass was then calculated by summing the numbers of individuals of all species in each size class. Biomass in each size class was normalized by dividing the biomass in that class by the width of the class. The size spectrum slope and intercept were calculated for each sample at each site after removing very small and very large classes so that all contiguous classes contained non-zero abundance. To ensure that slope and intercept were independent, mid-point heights of the spectra were zeroised (Daan et al. 2005). All data manipulation and analyses were performed using R (R Development Core Team 2007).

To provide a long-term description of the environment at the sampling sites we estimated mean annual sea surface temperature (SST), surface chlorophyll a concentration (Chl a) and primary production (PP) at each site. SST data were derived from the Moderate-resolution Imaging Spectroradiometer (MODIS) aboard the NASA satellites. Monthly SST averages for 2003 were extracted through the Jet Propulsion Laboratory physical oceanography DAAC web portal (http://poet.jpl.nasa.gov/) and averaged to give a mean annual value at a scale of 36 km^2. Estimates of Chl a were taken from Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data (http://orca.science.oregonstate.edu/1080.by.2160.monthly.hdf.chl.seawifs.php) by calculating an annual mean value for each sampling location from data extracted for each month in 2003. Where no value was available for a given month at a particular location then the value from the nearest

available location for that month was used. PP was computed from a wavelength- and depth-resolved model (Mélin 2003), building on the approach of Longhurst et al. (Longhurst et al. 1995). The main biological input to the models was the surface concentration of chlorophyll a pigment provided by SeaWiFS (see above). All changes from the implementation of Longhurst et al. (Longhurst et al. 1995) are detailed in Mélin (Mélin 2003). Outputs were calculated on a 36km grid. Annual primary production was obtained by averaging positive values (mg C $m^{-2}d^{-1}$) over the number of available months, except at high latitudes where it was normalized to 12 (as ocean colour has no good coverage of wintertime high latitudes, owing to the presence of cloud cover and sea ice). The mean annual values were assigned to each station location. When station locations had to be matched to point locations rather than onto a grid, we attempted the match in the following order; (1) number of degrees to one decimal place for both latitude and longitude, (2) number of degrees to one decimal place for latitude, rounded whole number of degrees for longitude, (3) rounded whole number of degrees for latitude and number of degrees with one decimal place for longitude, (4) rounded whole number of degrees for both latitude and longitude. All were successfully matched (split between the 4 priorities 0%, 8%, 5% and 87% respectively).

For some comparisons, cell size had to be converted between equivalent spherical diameter (ESD) and M. For example, a phytoplankton cell with ESD 2µm has carbon content of approximately 0.8pg (-0.08 on the $\log_{10}$ pg scale used in our Figures) using the conversion

$$pg\ C\ year^{-1} = 0.216 \times volume^{0.939}\ \mu m^{-3} \qquad \text{Equation 5}$$

reported for taxonomically diverse protist plankton (Menden-Deuer and Lessard 2000).

Relationships were explored between the remotely-sensed environmental variables primary production, sea surface temperature and chlorophyll a and (a) cell mass that accounted for 50% of total phytoplankton biomass (i.e. 50% on cumulative biomass curve) (hereafter M_{B50}), (b) cell mass range that included 80% of total biomass (range of cell masses from 10% to 90% of biomass on cumulative biomass curve) (M_{B90-10}), (c) cell mass that accounted for 50% of total phytoplankton production (i.e. 50% on cumulative production curve) (M_{P50}), (d) cell mass range that included 80% of total production (M_{P90-10}), (e) size spectra slope and (f) size spectra mid-point height. Prior to performing ANOVA, where necessary the data were log-transformed to achieve normality and equality of variance.

To identify the remotely-sensed environmental variables that best predicted the properties of the phytoplankton communities we randomly divided the data into equal-sized 'training' and 'predicting' data sets (Tian et al. 2007). We fitted linear models to the cases in the training set and then evaluated how well these models predicted the properties of the phytoplankton communities in the predicting set. The training set models were of the form:

$$Y(t) = f(X(t);\theta) + \text{error} \qquad \text{Equation 6}$$

where the *Y (t)* variable was related to a function of a subset of the environmental variables *X(t)* and the parameters θ were estimated by least squares. The models were assumed to be normally distributed with mean zero and constant variance. The effect of each variable as a predictor was tested individually; then each was tested as the second variable along with the variable previously identified as the best predictor.

The performance of the training set models was evaluated using the prediction data set and we described performance with the following summary statistic (Tian et al. 2007):

$$D = |Y(p) - Z(p);\hat{\theta}| \qquad \text{Equation 7}$$

Where *Y(p)* are the Y values in the prediction data set and *Z(p)* contains the predicted values of *Y(p)* based on the training data set model with its estimated parameter vector $\hat{\theta}$. This statistic can be interpreted as prediction error.

We report the mean values of D from 1000 calculations based on different random choices of the training and prediction data sets. This approach ensures that our results are not affected by the selection of training and prediction data sets. We refer to the mean *D* as $\overline{D}$ and use it to evaluate our models (rather than a summary of the fit of the model) because we want to use our model for prediction (models with the best fit often include more variables than those that give the best prediction). We then investigate the explanatory power of the best prediction model by fitting the model to the whole dataset.

Applying these relationships in combination allows prediction of the slope, intercept and location of phytoplankton size spectra with respect to M. The location of the cell mass range is calculated so

that that the integrated biomass to either side of the mid-point is equal. Thus the value for mass at 10% (M_{B10}) is calculated as:

$$M_{B10} = M_{B50} \{(10^{(b+1)\ \log_{10}(M_{B90-10})} + 1) / 2\}^{-(1/(b+1))} \quad \text{Equation 8}$$

where M_{B50} is mid-point mass, b is slope and $\log_{10}(M_{B90-10})$ is $(\log_{10}(M_{B90}) - \log_{10}(M_{B10}))$, the values predicted by the models. The derivation is shown in the Supplementary Material. M_{B90} is then calculated as:

$$M_{B90} = 10\text{^}(\log_{10}(M_{B10}) + \log_{10}(M_{B90\text{-}10})] \quad \text{Equation 9}$$

A statistical model that adequately fitted the very variable tails of the cumulative distribution was not found, and so to ensure the integrated B was equal to 100% of cumulative B the values of M_{B0} and M_{B100} were estimated by plotting M_{B10}, M_{B50} and M_{B90} and fitting a second order polynomial relationship.

The percentage contribution of a particular size class to total community biomass or production can then be calculated for a given sea surface temperature and chlorophyll a concentration. The derivations of the equations are shown in the Supplementary Material.

Results

Mean cell mass ranged from 243pg of C (AMT3) to 4740pg of C (Bergen fjord) and variance was lowest on AMT1 and highest in the Bergen fjord (Table 1). M_{B50} ranged from 1.6pg C (AMT3) to 1209pg C (Long Island Sound) and $M_{B90\text{-}10}$ ranged from 6.2pg C (AMT1) to 4476pg C (AMT4) (Table 1). M_{P50} ranged from 3.0pg C (AMT3) to 1340pg C (Long Island Sound) and $M_{P90\text{-}10}$ ranged from 8.9pg C (AMT1) to 4070pg C (Bergen fjord) (Table 1).

Across all sites the mean size spectra slope was –1.19, ranging from –1.66 in the Norwegian Sea to –0.74 in Long Island Sound. Mean mid-point height ($\log_2$ C pg) was 22.11 ranging from 20.54 on AMT5 to 23.41 in the Benguela. The fits of all size spectra slopes were significant at the 0.1 level except Long Island Sound (p=0.14) and mean r^2 was 0.63 (Table 2).

There was a positive linear relationship between M_{B50} ($\log_{10}$) and both of the variables primary production ($\log_{10}$) and chlorophyll a ($\log_{10}$) and an inverse linear relationship between M_{B50} ($\log_{10}$)

and sea surface temperature. There was an inverse linear relationship between M_{B90-10} (log_{10}) and both of the variables primary production (log_{10}) and chlorophyll a (log_{10}) (Fig. 2 and Supplementary Table 1). There was a positive linear relationship between M_{B90-10} (log_{10}) and sea surface temperature but the p-value for the intercept was not significant (Fig. 2 and Supplementary Table 1). These biomass relationships were consistent with those for production (Fig. 3 and Supplementary Table 1).

There were significant positive linear relationships between size spectra slope and both the variables primary production (log_{10}) and chlorophyll a (log_{10}) and an inverse relationship between slope and sea surface temperature (Fig. 4 and Supplementary Table 1). The linear relationships between size spectra mid-point height and primary production (log_{10}) and chlorophyll a (log_{10}) were also significant and positive and again there was an inverse relationship with sea surface temperature (Fig. 4 and Supplementary Table 1).

Prediction results

Prediction models are only useful for application on large space and time scales when the explanatory variables are readily available so we only investigated prediction using primary production, sea surface temperature and chlorophyll estimates that can be easily derived from remote sensing.

Chl a and SST together successfully predicted M_{B50}, M_{B90-10}, M_{P50}, M_{P90-10} and the slope of the size spectrum, with all prediction models being significant at $p<0.001$. No environmental variables predicted the intercept of the size spectrum any better than by just taking the mean value. However, to enable an estimate to be made, we report a single variable model using Chl a (this gave as good a result as using SST or PP or any combination of the three variables) and this single variable prediction model was also significant at $p<0.001$. The prediction equations are given in Table 3 and the statistical results are given in Supplementary Table 2.

Applying these relationships in combination allows prediction of the slope, height and location of the phytoplankton size spectra over relevant size ranges (Fig. 5) and the make-up of communities by size class (see Supplementary Material). For example, where sea surface temperature is 15ºC and chlorophyll a concentration is 1.0 mg m^{-3} there is total biomass of 47 log_{10} pg C, of which 24% is picoplankton (ESD<2µm), 76% is nanoplankton (ESD>2 but <20µm) and there is no microplankton (ESD >20µm).

Discussion

The analyses reveal that relationships between the size composition of phytoplankton communities and the environment can be used to predict mean cell mass, the slopes of size spectra and the range of cell masses that encompass a given proportion of total biomass or production. Since predictions are based on temperature and chlorophyll estimates that are available at high resolution and over large spatial scales, for example from remote sensing, they can be used to predict community size structure at the same scales. Such predictions provide necessary inputs to models that seek to link primary production to production at higher trophic levels.

Chlorophyll concentration varies due to a variety of physical and chemical factors. Irradiance over the mixed layer depth, surface nitrate, sea-surface temperature, and latitude and longitude together can predict 83% of the variation in log chlorophyll in the North Atlantic (Irwin and Finkel 2008). We are not here investigating the mechanisms leading to the correlation between phytoplankton size composition and chlorophyll; instead we accept the practical value of the relationship pending research that identifies mechanistic relationships. Tight correlations between temperature, chlorophyll a and nitrate concentrations are known suggesting that the environmental factors are highly intertwined and strongly regulate the phytoplankton average cell size (Chen and Liu 2010).

Picophytoplankton made up 50% of the biomass at chlorophyll a concentrations less than 0.25 mg m^{-3} and at temperatures over 20.4°C. At this temperature, 50% of production was from cells up to 2.5µm. The relationship to chlorophyll is consistent with the findings of Agawin et al. (Agawin et al. 2000) who reported that picophytoplankton dominated ($\geq$ 50%) where chlorophyll a concentration was lower than 0.3 mg m^{-3} but they reported picophytoplankton dominating in waters over 26°C compared to our estimate of 20.4°C. Using our prediction models, which use both temperature and chlorophyll a concentration, we estimate that at 26°C and 0.3 mg m^{-3} chlorophyll concentration, phytoplankton cells up to 1.3µm diameter (0.24 pg C) make up 50% of the biomass (see Fig. 5c) and up to 1.2µm diameter (0.20 pg C) make up 50% of the production.

Slopes of phytoplankton size spectra are predicted to be steeper when total biomass is low and at higher temperatures. If the size-spectrum slope is –1, then the trend is for all size classes to have equal biomass; if the slope is greater than –1 then biomass tends to increase with increasing cell mass; if the abundance–mass slope is less than –1 then biomass decreases with increasing cell mass. Modellers have often assumed slopes of either –0.75 or –1.0 when describing the size structure of

the phytoplankton community but the model: *"Slope = -0.007 SST (°C) + 0.114 log_{10} (Chl a (mg m^{-3})) – 1.049"* suggests a mean slope of –1.05 at 3°C but –1.20 at 25°C (for an intermediate level of Chl a of 1.5 mg m^{-3}). There are relatively few empirical estimates of size-spectrum slope for phytoplankton communities with which to compare our estimates, as most estimates are for communities that include zooplankton. However, those available provide qualitative support for our results. For example, Marañón et al. (Marañón et al. 2007) reported that unproductive ecosystems were characterized by more negative slopes (-1.3 to –1.1) than productive ones (–0.8 and –0.6), but the slopes were rather shallower than reported in our study in which 10 of the 12 sites had a slope steeper (more negative) than –1 and 3 sites steeper than -1.3. San Martin et al. (San Martin et al. 2006a) also found that the slopes of biomass size spectra for the picoplankton and microplankton size ranges were positively related to biomass (but the pattern disappeared with the addition of mesozooplankton). They did not report relationships with temperature.

Size spectrum slopes as steep as –1.39 have been reported for pure phytoplankton (Huete-Ortega et al. 2010). These values, like ours, are much steeper than the value of -0.75 that is theoretically predicted to result from the allocation of energy among competing individuals (Belgrano et al. 2002). Our values are also steeper than the slope of –0.78 (Li 2002) and slopes ranging from –0.74 to –1.06 (Cermeño and Figueiras 2008) reported for phytoplankton. Differences in slope could be attributed to sampling artefacts or ecological processes. Smaller plankton are not effectively sampled by some gears, are not so well preserved after sampling and can be under-sampled during microscopic identification so we expect that any bias in the data will lead to the under representation of small cells and lower rather than higher estimates of slope (Harris et al. 2000). Of necessity, deriving a size spectrum by pooling the abundance of cells with a mean mass that fall in that size class will introduce bias. Biases introduced by working with mean mass in studies of relationships between abundance and body mass are greatest when working with species that have indeterminate growth and when the range of body sizes considered is narrow (Jennings et al. 2007). These biases will be minimal for phytoplankton that have limited and determinate growth and a range in mean mass that spans over 5 orders of magnitude. Another factor that may have biased slope estimates is the exclusion of sizes at the extremes of the distribution. However, we excluded small and large species using the same rules. Since sampling and analytical artefacts are unlikely to explain why slopes are relatively negative, there may be an influence of ecological processes. For example, consumer density increases with temperature, leading to increased grazing pressure (O'Connor et al. 2009) thus the slopes may increase as a result of greater predation rates on larger phytoplankton or uniform predation rates that have a greater impact on larger phytoplankton owing to their slower turnover times.

Our analyses do not determine cause and effect and are primarily intended to predict the size composition of phytoplankton communities from readily available large scale and high-resolution remote sensing data to support parameterisation of food web models. However, direct linkages between temperature and the size composition of phytoplankton communities have been proposed (López-Urrutia 2008). He suggests that large cells may dominate in colder systems owing to the different temperature dependence of heterotrophic and autotrophic rates. In colder areas heterotrophs may not grow fast enough to control autotrophs while in warmer areas large phytoplankton cells are likely to have relatively longer division times than smaller ones and may not be able to attain high levels of abundance given the high abundance of heterotrophic predators. We also expect community structure to be influenced by nutrient supply and the colder seas (as represented in the database) often correspond to areas with more nutrients (e.g. upwellings and coastal areas) than the warmer ones (e.g. oligotrophic subtropical gyres).

Other methods are being developed to predict the size composition of phytoplankton communities at high spatial resolution over large spatial scales. For example, Uitz et al. (Uitz et al. 2006; Uitz et al. 2008) have determined size-spectra of phytoplankton communities from near-surface chlorophyll a concentration using accessory pigments as markers for pico, nano and micro-plankton to infer the column-integrated phytoplankton biomass, its vertical distribution, and ultimately the community composition by quantifying on a global scale the phytoplankton biomass associated with each of the three algal assemblages. However pigment analysis based on high-performance liquid chromatography although a widely used method for studying phytoplankton community composition does not provide any information on size structure in each group (Chen and Liu 2010).

Our approach can be used to allocate phytoplankton biomass and primary production to cell mass classes and thus improve the description of the primary producer community in size based models that have been used to link primary production and production at higher trophic levels. This is useful because estimates of phytoplankton production from NPZD (Nutrient Phytoplankton Zooplankton Detritus) models and remote sensing rarely provide information on the size composition of the phytoplankton community alone. At a sea surface temperature of 5°C, our predictions suggest that the phytoplankton cell size at 50% of biomass ranges from 31 to 47 pg C, at 15°C it ranges from 5.4 to 8.1 pg C and at 25°C it ranges from 0.9 to 1.4 pg C for intermediate chlorophyll a concentrations from 1 to 1.5 mg m^{-3}. At a sea surface temperature of 5°C the phytoplankton cell size at 50% of production ranges from 3.3 to 4.8 pg C, at 15°C from 0.9 to 1.3 pg C and at 25°C from 0.2 to 0.3 pg C over the same range of chlorophyll a concentrations. The

range of cell sizes that make up the majority of the community (the mid 80% of biomass/production) is smaller at more productive sites but wider at warmer temperatures. One consequence of the dominance of smaller primary producers in less productive waters is that food chains will be longer. The ratio between the sizes of consumers and their prey (reported as the predator-prey mass ratio, PPMR) does not depend on production or temperature (Barnes et al. 2010), so the mean trophic level of a given size class of consumers would be expected to be higher in low productivity areas such as ocean gyres. Since smaller primary producers are linked to lower primary production and PPMR and transfer efficiency may be unrelated to the environment, production at higher trophic levels is disproportionately low when primary productivity is low.

In addition to predicting fluxes of energy to higher trophic levels and the biomass of consumer communities based on measurements of primary production and temperature, our predictive relationships may also be valuable for predicting how the composition of phytoplankton communities may change in relation to environmental change. For example, Morán et al. (Morán et al. 2009) have reported consistent relationships among temperature, cell size and picophytoplankton abundance and speculate that the size of cells in phytoplankton assemblages will gradually decrease as temperatures rise. Li et al. (Li et al. 2009) concur that a reduction in community average cell size because of an increase in the abundance of individuals belonging to small-sized species may be a common response to increasing sea temperatures. There is evidence that reduced body size is the third universal ecological response to global warming besides the shift of species ranges toward higher altitudes and latitudes and the seasonal shifts in life-cycle events (Daufresne et al. 2009). Such influences need to be considered in models that seek to predict how future changes in primary production and temperature will affect production at higher trophic levels.

Conclusions

The size composition of primary producers is an essential input to enable estimations of food chain length, consumer biomass and production in any given location. We have described relationships between the environment and the size composition of phytoplankton communities using environmental variables that are easily estimated from ocean colour satellite measurements. Estimates of consumer biomass, production and trophic level depend on the length of food chains that support this biomass, a consequence of predator-prey size relationships and the size composition of primary producers. As mean predator-prey size ratios in marine ecosystems do not depend on temperature or primary production, the size composition of the phytoplankton

community has an overriding influence on food chain length which, in turn can be used to further explore fish production, fisheries catch potential and the bioaccumulation of contaminants.

Acknowledgements

We would like to thank Roger Harris, Kevin Flynn, Ángel López-Urrutia, Doug Glazier and Rodney Forster for their thought-provoking comments and suggestions on how to improve the manuscript. We also thank Dan Reuman, David Maxwell and Jon Barry for assistance with the statistical analyses.

Funding

This work was supported by Department for Environment, Food and Rural Affairs (M1001). The collection of the AMT data was supported by the UK Natural Environment Research Council through the Atlantic Meridional Transect consortium (NER/O/S/2001/00680). This is contribution number 192 of the AMT programme.

References

Agawin, N. S. R., Duarte, C. M. and Agusti, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnol. Oceanogr.* **45**, 591-600.

Barnes, C., Maxwell, D., Reuman, D. C. and Jennings, S. (2010). Global patterns in predator-prey size relationships reveal size-dependency of trophic transfer efficiency. *Ecology* **91**, 222-232.

Belgrano, A., Allen, A. P., Enquist, B. J. and Gillooly, J. F. (2002). Allometric scaling of maximum population density: A common rule for marine phytoplankton and terrestrial plants. *Ecol. Lett.* **5**, 611-613.

Blanchard, J. L., Jennings, S., Law, R., Castle, M. D., McCloghrie, P., Rochet, M.-J. and Benoit, E. (2009). How does abundance scale with body size in coupled size structured food webs? *J. Anim. Ecol.* **78**, 270-280.

Carr, M. E., Friedrichs, M. A. M., Schmeltz, M., Noguchi Aita, M., Antoine, D., Arrigo, K. R., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E. T., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Le Quéré, C., Lohrenz, S., Marra, J., Mélin, F., Moore, K., Morel, A., Reddy, T. E., Ryan, J., Scardi, M., Smyth, T., Turpie, K., Tilstone, G., Waters, K. and Yamanaka, Y. (2006). A comparison of global estimates of marine primary production from ocean color. *Deep-Sea Research Part II: Topical Studies in Oceanography* **53**, 741-770.

Cermeño, P. and Figueiras, F. G. (2008). Species richness and cell-size distribution: Size structure of phytoplankton communities. *Mar. Ecol. Prog. Ser.* **357**, 79-85.

Chen, B. and Liu, H. 2010. Relationships between phtoplankton growth and cell size in surface oceans: Interactive effects of temperature, nutrients and grazing. *Limnology & Oceanography* **55**, 965-972.

Daan, N., Gislason, H., Pope, J. G. and Rice, J. C. (2005). Changes in the North Sea fish community: Evidence of indirect effects of fishing? *ICES J. Mar. Sci.* **62**, 177-188.

Daufresne, M., Lengfellner, K. and Sommer, U. (2009). Global warming benefits the small in aquatic ecosystems. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12788-12793.

Dickie, L. M. (1976). Predation, yield, and ecological efficiency in aquatic food chains. *Journal of Fisheries Research Board of Canada* **33**, 313-316.

Harris, R., Wiebe, P., Lenz, J., Skjoldal, H. R. and Huntley, M. (2000). *ICES Zooplankton Methodology Manual*, Academic Press. 684pp.

Hirata, T., Aiken, J., Hardman-Mountford, N. J., Smyth, T. J. and Barlow, R. G. (2008). An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sens. Environ.* **112**, 3153-3159.

Huete-Ortega, M., Marañón, E., Varela, M. and Bode, A. (2010). General patterns in the size scaling of phytoplankton abundance in coastal waters during a 10-year time series. *J Plankton Res* **32**, 1-14.

Irigoien, X., Flynn.K.J. and Harris, R. P. (2005). Phytoplankton blooms: a 'loophole' in microzooplankton grazing impact? *J Plankton Res* **27**, 313-321.

Irigoien, X., Huisman, J. and Harris, R. P. (2004). Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature* **429**, 863-867.

Irwin, A. J. and Finkel, Z. V. (2008). Mining a sea of data: Deducing the environmental controls of ocean chlorophyll. *PLoS ONE* **3**.

Jennings, S., De Oliveira, J. A. A. and Warr, K. J. (2007). Measurement of body size and abundance in tests of macroecological and food web theory. *J. Anim. Ecol.* **76**, 72-82.

Jennings, S., Mélin, F., Blanchard, J. L., Forster, R. M., Dulvy, N. K. and Wilson, R. W. (2008). Global-scale predictions of community and ecosystem properties from simple ecological theory. *Proceedings of the Royal Society B* **275**, 1375-1383.

Li, W. K. W. (2002). Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. *Nature* **419**, 154-157.

Li, W. K. W., McLaughlin, F. A., Lovejoy, C. and Carmack, E., C. (2009). Smallest algae thrive as the Arctic Ocean freshens. *Science* **326**, 539.

Longhurst, A., Sathyendranath, S., Platt, T. and Caverhill, C. (1995). An estimate of global primary production in the ocean from satellite radiometer data. *J Plankton Res* **17**, 1245-1271.

López-Urrutia, Á. (2008). The metabolic theory of ecology and algal bloom formation. *Limnology & Oceanography* **53**, 2046-2047.

López-Urrutia, Á., San Martin, E., Harris, R. P. and Irigoien, X. (2006). Scaling the metabolic balance of the oceans. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8739-8744.

Lund, J. W. G., Kipling, C. and Le Cren, E. D. (1958). The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia* **11**, 143-170.

Marañón, E., Cermeño, P., Rodríguez, J., Zubkov, M. V. and Harris, R. P. (2007). Scaling of phytoplankton photosynthesis and cell size in the ocean. *Limnology & Oceanography* **52**, 2190-2198.

McClain, C. R., Feldman, G. C. and Hooker, S., B. (2004). An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. *Deep Sea Res. (II Top. Stud. Oceanogr.)* **51**, 5-42.

Mélin, F. (2003). Potential of remote sensing for the analysis of the optical properties of the ocean-atmosphere system and application to estimates of phytoplankton photosynthesis. PhD Thesis. Université Paul Sabatier, Toulouse, France**:** pp 514.

Menden-Deuer, S. and Lessard, E. J. (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnol. Oceanogr.* **45**, 569-579.

Morán, X. A. G., López-Urrutia, Á., Calvo-Díaz, a. and Li, W. K. W. (2009). Increasing importance of small phytoplankton in a warmer ocean. *Global Change Biol.* **In press**.

O'Connor, M. I., Piehler, M. F., Leech, D. M., Anton, A. and Bruno, J. F. (2009). Warming and resource availability shift food web structure and metabolism. *PLoS Biol.* **7**, e1000178.

R-Development-Core-Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

San Martin, E., Harris, R. P. and Irigoien, X. (2006a). Latitudinal variation in plankton size spectra in the Atlantic Ocean. *Deep-Sea Research Part II: Topical Studies in Oceanography* **53**, 1560-1572.

San Martin, E., Irigoien, X., Harris, R. P., López-Urrutia, Á., Zubkov, M. V. and Heywood, J. L. (2006b). Variation in the transfer of energy in marine plankton along a productivity gradient in the Atlantic Ocean. *Limnology & Oceanography* **51**, 2084-2091.

Sheldon, R. W. and Parsons, T. R. (1967). A continuous size spectrum for particulate matter in the sea. *J Fish Res Board Can* **24**, 909-915.

Sheldon, R. W., Prakash, A. and Sutcliffe, W. H. (1972). The size distribution of particles in the Ocean. *Limnol. Oceanogr.* **17**, 327-340.

Smayda, T. J. (1971). Normal and accelerated sinking of phytoplankton in the sea. *Mar. Geol.* **11**, 105-122.

Sprules, W. G. and Munawar, M. (1986). Plankton size spectra in relation to ecosystem productivity, size, and perturbation. *Can. J. Fish. Aquat. Sci.* **43**, 1789-1794.

Tian, L. U., Cai, T., Goetghebeur, E. and Wei, L. J. (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika* **94**, 297-311.

Uitz, J., Claustre, H., Morel, A. and Hooker, S., B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research* **111**.

Uitz, J., Huot, Y., Bruyant, F., Babin, M. and Claustre, H. (2008). Relating phytoplankton photophysiological properties to community structure on large scales. *Limnology & Oceanography* **53**, 614-630.

Vidondo, B., Prairie, Y. T., Blanco, J. M. and Duarte, C. M. (1997). Some aspects of the analysis of size spectra in aquatic ecology. *Limnol. Oceanogr.* **42**, 184-192.

Figure 1: A graphical summary of the definitions of cell mass ranges that account for various proportions of biomass and production. M_{B50} (M_{P50}) is the cell mass at which 50% of biomass (production) is reached. $M_{B90\text{-}10}$ ($M_{P90\text{-}10}$) is the range of cell masses that make up the mid 80% (i.e. $\log_{10}$(cell mass at 90%) - $\log_{10}$(cell mass at 10%) of biomass (production).

Figure 2: Relationships between (a) M_{B50}, the mass of phytoplankton cells that account for 50% of total biomass and (b) $M_{B90\text{-}10}$, the range of phytoplankton cell masses that account for the mid 80% of total biomass and remotely-sensed estimates of primary production, sea surface temperature and chlorophyll a at 12 sites.

Figure 3: Relationships between (a) $M_{P50,}$ the mass of phytoplankton cells that account for 50% of total production and (b) $M_{P90\text{-}10,}$ the range of phytoplankton cell mass that account for the mid 80% of total production and remotely-sensed estimates of primary production, sea surface temperature and chlorophyll a at 12 sites

Figure 4: Relationships between (a) size spectra slopes and (b) size spectra mid-point heights and remotely-sensed estimates of primary production, sea surface temperature and chlorophyll a at 12 sites.

Figure 5: Size spectra predictions for (a) a range of sea surface temperatures at an intermediate chlorophyll a level of 1mg m^{-3}; (b) a range of chlorophyll a levels at 10°C and (c) a range of sea surface temperatures and a range of chlorophyll a levels. • show position of M_{B50}, the mass of phytoplankton cells that account for 50% of total biomass. See text for explanation for mid-point prediction comparison with Agawin et al. (2000).

Table 1: Phytoplankton cell mass (pg C) statistics from 12 sites that include polar, tropical and upwelling environments.

Table 2: Properties and statistical analysis of normalised biomass size spectra from 12 sites that include polar, tropical and upwelling environments.

Table 3: Equations for predicting phytoplankton cell sizes from remotely-sensed variables.

Table 1

	%	AMT1	AMT2	AMT3	AMT4	AMT5	Benguela	Bergen fjord	Irminger	Long Island Sound	North Sea	Norwegian Sea	Oregon upwelling
Mean cell mass		253	279	243	316	246	912	4740	381	742	788	267	702
1 standard deviation		670	1150	817	941	954	4844	99418	1083	2133	3379	1052	1728
1 standard error		5.5	9.9	6.6	7.3	7.5	28	635	6.4	33	25	14	24
Smallest cell mass		0.1	0.1	0.01	0.01	0.01	0.1	1.0	0.2	1.0	1.0	1.0	0.2
Largest cell mass		8206	27123	10686	15221	27123	49930	2799208	14709	27123	143195	7000	14709
Cell mass that accounts for % of total biomass	10	19	0.31	0.06	0.03	0.22	5.3	4.9	3.4	175	20	2.2	4.6
	25	25	1.8	0.30	0.19	0.6	20	15	6.4	466	41	5.3	15
	50	35	10	1.6	1.8	3.8	165	117	19	1209	123	12	81
	75	54	47	11	17	35	1324	961	81	2722	496	24	505
	80	64	69	18	30	60	2181	1558	122	3265	708	28	789
	90	118	184	71	132	229	8051	5174	361	5109	1769	41	2419
Cell mass range that accounts for the mid 80% of total biomass		6.2	590	1098	4476	1049	1509	1050	107	29	90	19	529
Cell mass that accounts for % of total production	10	19	0.81	0.15	0.35	0.58	7.2	6.0	3.8	225	24	2.6	5.6
	25	25	3.1	0.59	0.61	1.3	35	26	7.4	564	52	6.0	22
	50	36	14	3.0	3.2	7.4	265	208	24	1340	161	13	118
	75	63	61	19	33	53	2056	2590	124	2766	627	25	641
	80	78	89	30	58	85	3470	4895	190	3252	881	29	951
	90	165	247	101	258	269	25325	24322	590	4848	2093	41	2526
Cell mass range that accounts for the mid 80% of total production		8.9	305	657	745	463	3506	4070	155	22	88	16	453

Table 2

	AMT1	AMT2	AMT3	AMT4	AMT5	Benguela	Bergen fjord	Irminger	Long Island Sound	North Sea	Norwegian Sea	Oregon upwelling
No. samples	25	25	25	26	24	54	46	59	7	44	19	7
Mean slope	-1.18	-1.32	-1.36	-1.28	-1.22	-1.16	-0.94	-1.08	-0.74	-1.10	-1.66	-1.21
Mean mid-point zeroised height (log_2 pg C)	21.38	21.91	21.78	21.37	20.54	23.41	22.56	22.48	23.24	21.11	22.35	23.22
Mean r squared	0.64	0.68	0.75	0.71	0.67	0.62	0.49	0.65	0.38	0.50	0.78	0.71
Mean p-value	0.018	0.014	0.005	0.009	0.016	0.017	0.058	0.015	0.143	0.067	0.004	0.021
Slope coefficient of variation	0.19	0.17	0.13	0.18	0.19	0.26	0.27	0.15	0.49	0.45	0.13	0.28
Mid-point zeroised height coefficient of variation	0.06	0.05	0.07	0.06	0.07	0.09	0.03	0.06	0.08	0.07	0.07	0.07
R squared coefficient of variation	0.28	0.18	0.11	0.17	0.24	0.24	0.39	0.22	0.63	0.38	0.09	0.34

Table 3

	Predicted Value	SST (℃) multiplier	Log_{10} (Chl a (mg m^{-3})) multiplier	Constant
M_{B50}	Log_{10} (cell size (pg C) for 50% of biomass)	–0.076	0.999	1.873
M_{B90-10}	Log_{10} (cell size range (pg C) for mid 80% of biomass)	0.184	– 1.082	0.447
M_{P50}	Log_{10} (cell size (pg C) for 50% of production)	-0.058	0.909	1.813
M_{P90-10}	Log_{10} (cell size range (pg C) for mid 80% of production)	0.142	-0.837	0.906
Slope	Slope of the size spectrum	-0.007	0.114	– 1.049
Intercept	Log_2 (intercept of the size spectrum)	0	0.816	29.802

Figure 1

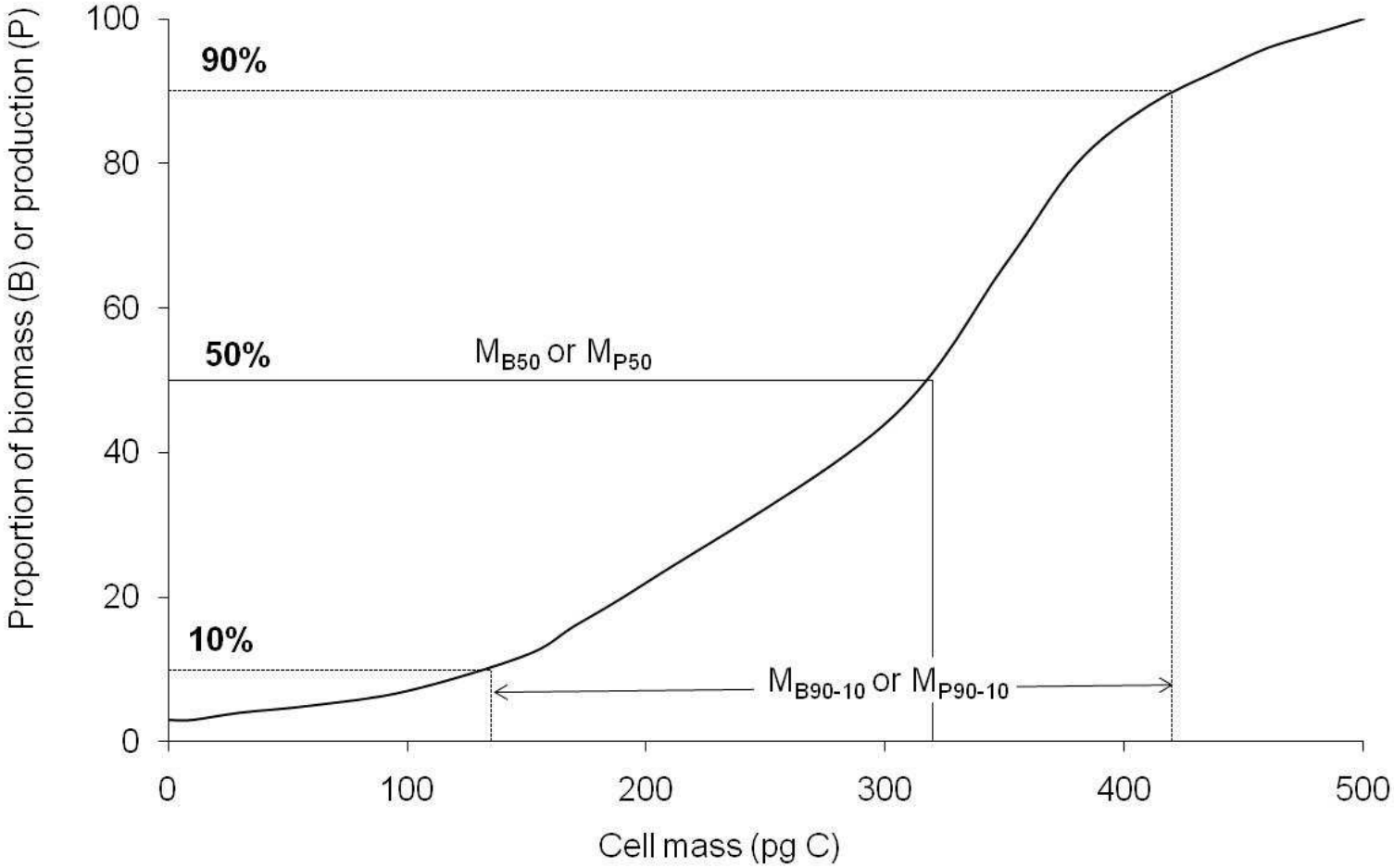

Proportion of biomass (B) or production (P)
100
80
60
40
20
0
90%
50%
10%
M_{B50} or M_{P50}
M_{B90-10} or M_{P90-10}
0
100
200
300
400
500
Cell mass (pg C)

Figure 2

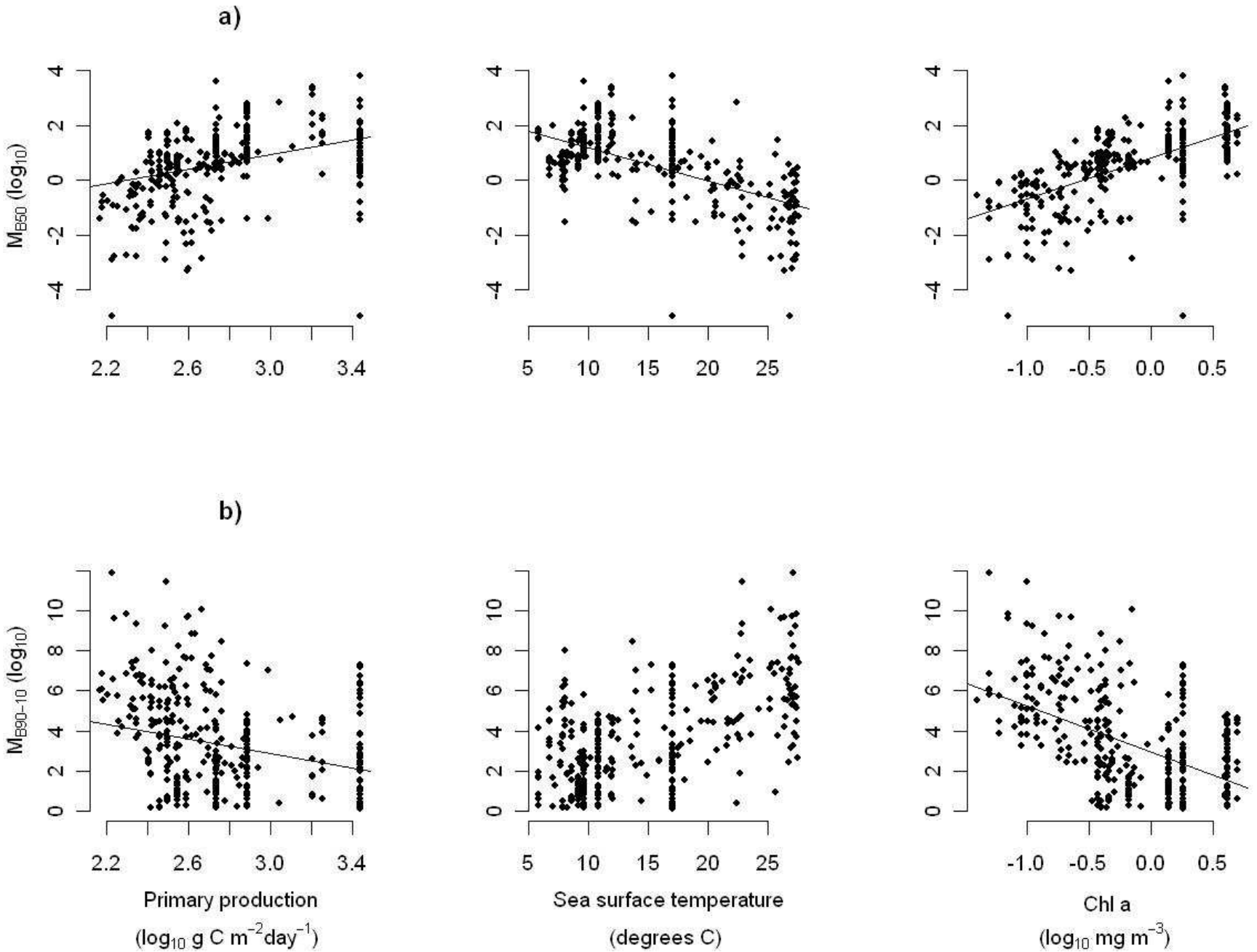

a)
b)
M_{B50} (log_{10})
M_{B90-10} (log_{10})
Primary production
(log_{10} g C $m^{-2}day^{-1}$)
Sea surface temperature
(degrees C)
Chl a
(log_{10} mg m^{-3})

Figure 3

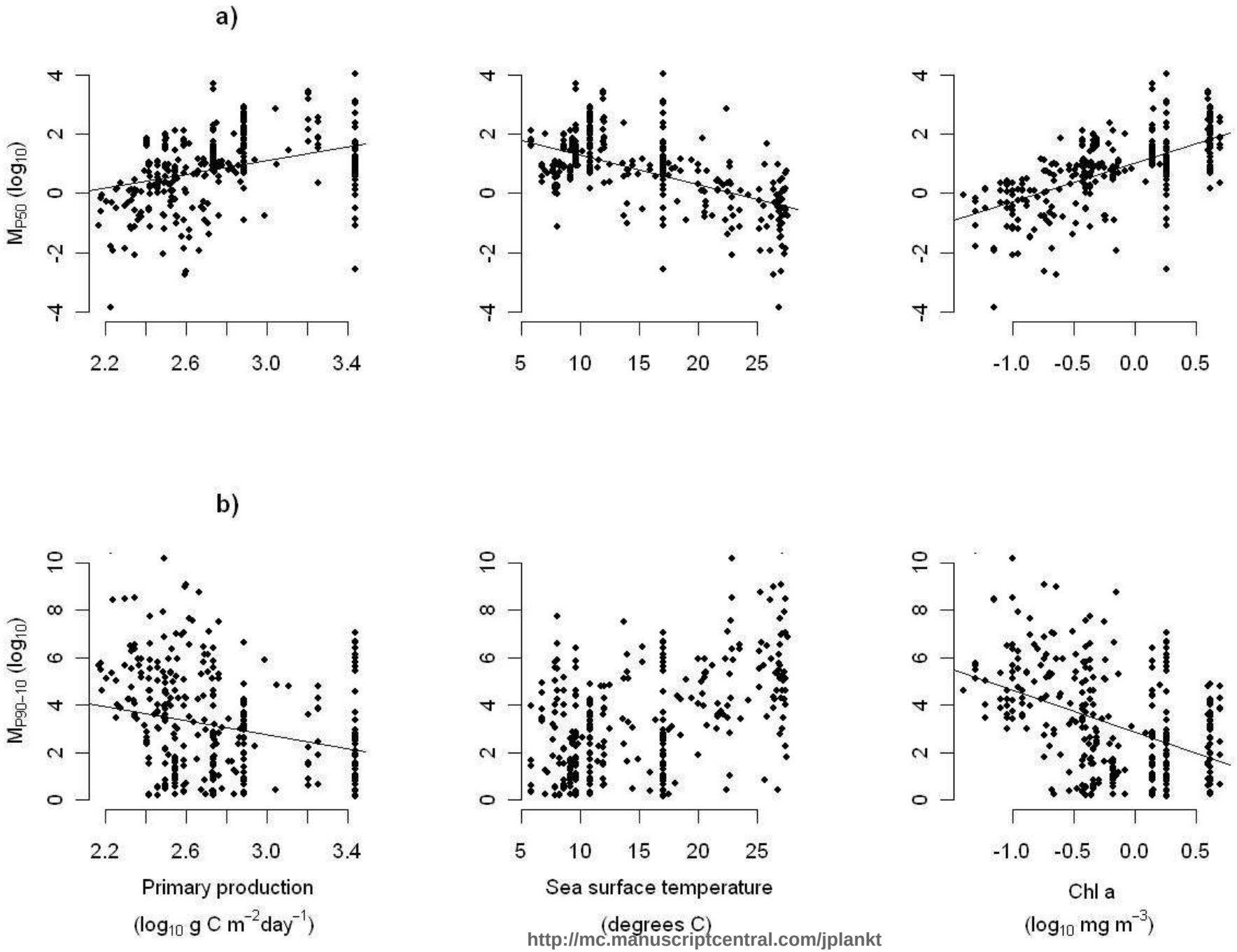

a)
M_{P50} (log_{10})
-4
-2
0
2
4
2.2
2.6
3.0
3.4
5
10
15
20
25
-1.0
-0.5
0.0
0.5
b)
M_{P90-10} (log_{10})
0
2
4
6
8
10
Primary production
(log_{10} g C $m^{-2}day^{-1}$)
Sea surface temperature
(degrees C)
Chl a
(log_{10} mg m^{-3})

Figure 4

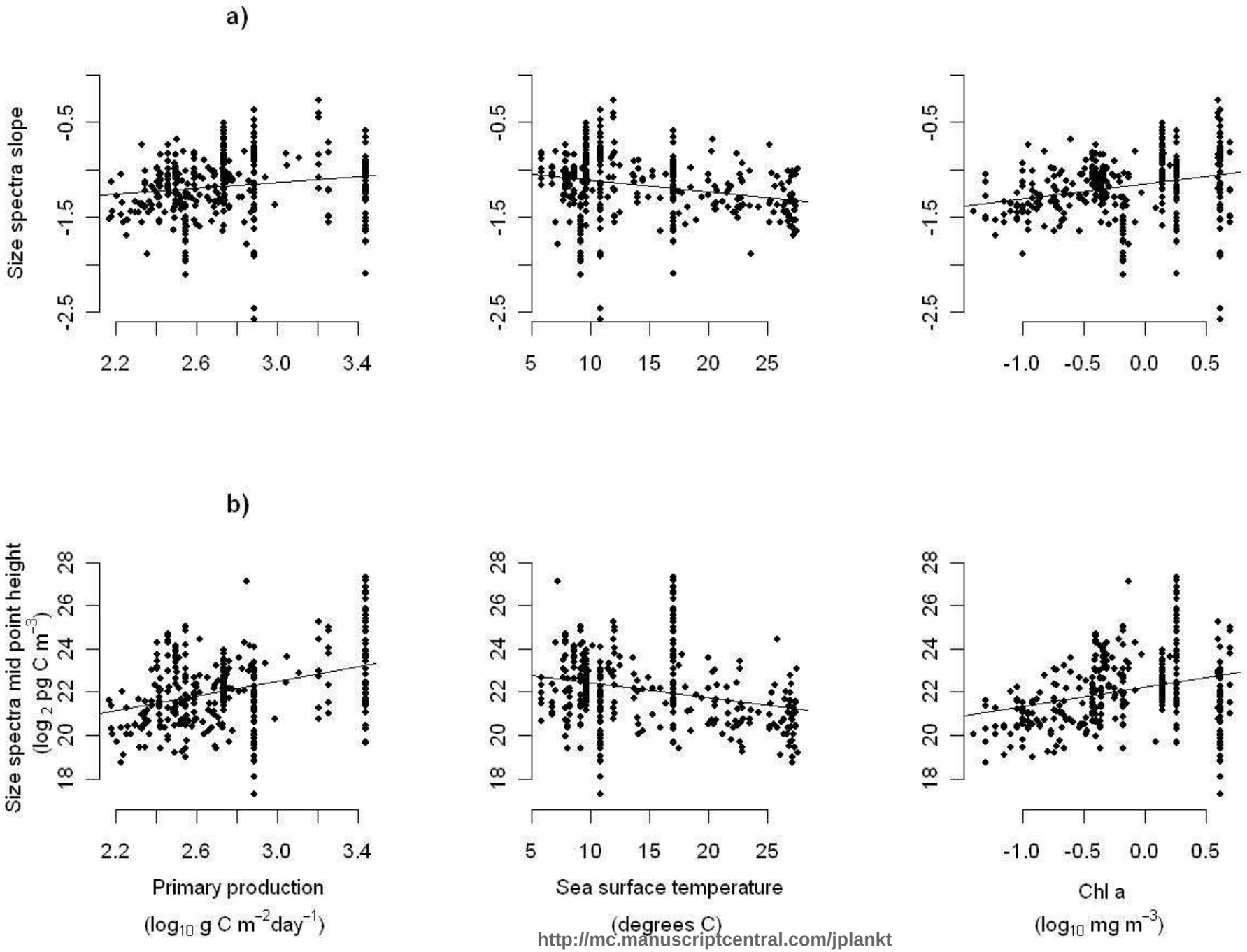
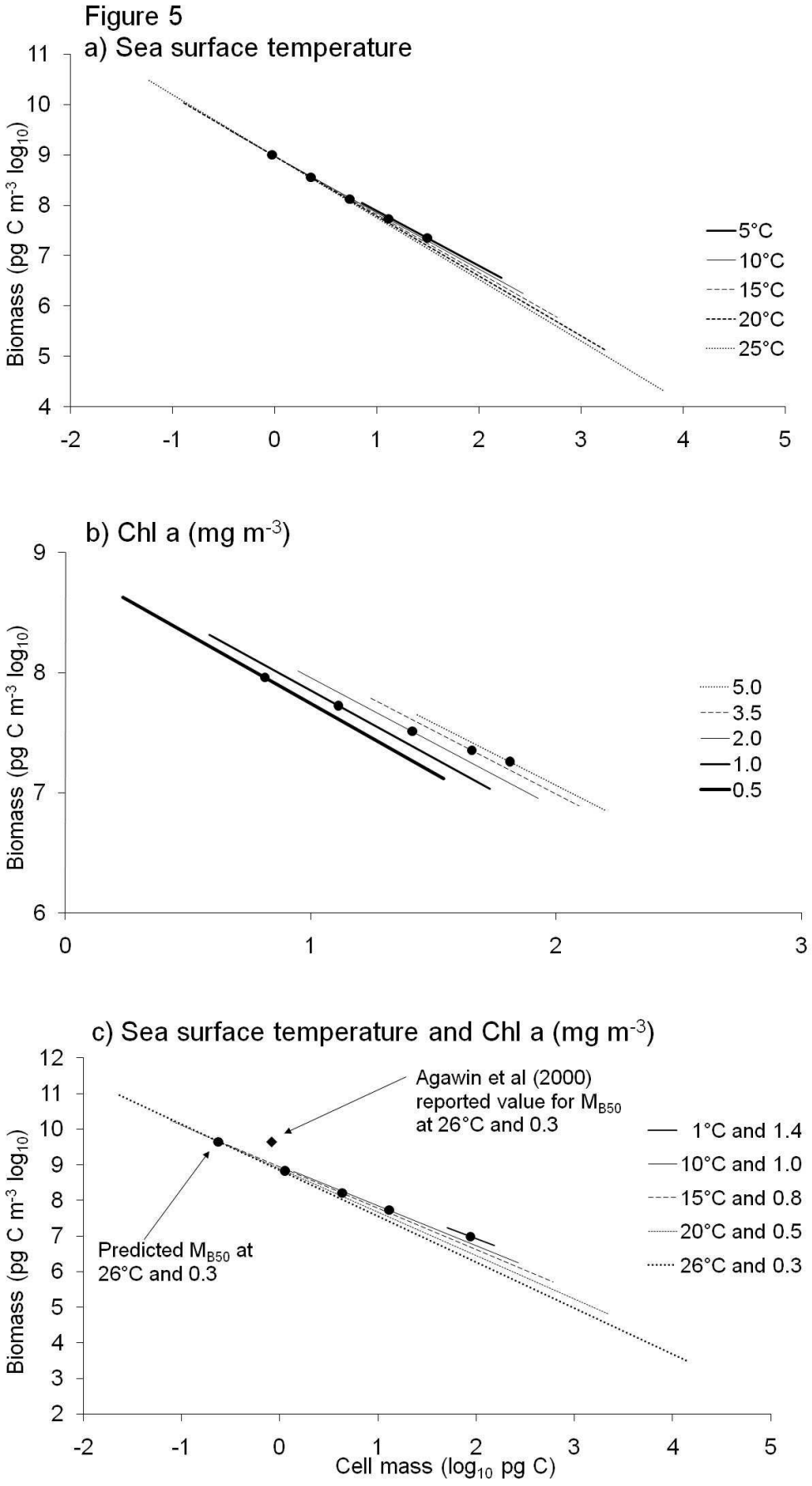

a)
Size spectra slope
-0.5
-1.5
-2.5
2.2
2.6
3.0
3.4
5
10
15
20
25
-1.0
-0.5
0.0
0.5
b)
Size spectra mid point height
(log 2 pg C m-3)
18
20
22
24
26
28
Primary production
(log10 g C m-2day-1)
Sea surface temperature
(degrees C)
Chl a
(log10 mg m-3)

Figure 5


a) Sea surface temperature
Biomass (pg C m-3 log10)
5°C
10°C
15°C
20°C
25°C
b) Chl a (mg m-3)
Biomass (pg C m-3 log10)
5.0
3.5
2.0
1.0
0.5
c) Sea surface temperature and Chl a (mg m-3)
Biomass (pg C m-3 log10)
Agawin et al (2000) reported value for MB50 at 26°C and 0.3
Predicted MB50 at 26°C and 0.3
1°C and 1.4
10°C and 1.0
15°C and 0.8
20°C and 0.5
26°C and 0.3
Cell mass (log10 pg C)

Supplementary Table 1: Analysis of variance statistics for linear relationships between remotely-sensed estimates of environmental variables and the phytoplankton cell masses that account for 50% of total biomass/production, the range that accounts for 80% of total biomass/production and the slope and mid-point height of size spectra of phytoplankton.

	Predicted Value	Environmental variable	F-stat	DF	P-value	Slope	Slope P-value	Intercept	Intercept P-value	R^2
M_{B50}	Cell mass (log_{10} pg C) that accounts for 50% of total biomass (log_{10})	Primary production (log_{10} g C m^{-2} day^{-1})	55.71	1,359	<0.001	1.331	<0.001	-3.063	<0.001	0.13
		Sea surface temperature (°C)	217.1	1,359	<0.001	-0.121	<0.001	2.388	<0.001	0.38
		Chlorophyll a (log_{10} mg m^{-3})	235.1	1,359	<0.001	1.501	<0.001	0.834	<0.001	0.39
M_{B90-10}	Cell mass (log_{10} pg C) that accounts for the mid 80% of total biomass (log_{10}(90%) – log_{10}(10%))	Primary production (log_{10} g C m^{-2} day^{-1})	21.13	1,359	<0.001	-1.810	<0.001	8.304	<0.001	0.05
		Sea surface temperature (°C)	162.5	1,359	<0.001	0.232	<0.001	-0.109	0.709	0.31
		Chlorophyll a (log_{10} mg m^{-3})	93.95	1,359	<0.001	-2.299	<0.001	2.965	<0.001	0.21
M_{P50}	Cell mass (log_{10} pg C) that accounts for 50% of total production (log_{10})	Primary production (log_{10} g C m^{-2} day^{-1})	56.86	1,359	<0.001	1.166	<0.001	-2.392	<0.001	0.13
		Sea surface temperature (°C)	179.5	1,359	<0.001	-0.099	<0.001	2.280	<0.001	0.33
		Chlorophyll a (log_{10} mg m^{-3})	227.7	1,359	<0.001	1.292	<0.001	1.021	<0.001	0.39
M_{P90-10}	Cell mass (log_{10} pg C) that accounts for the mid 80% of total	Primary production (log_{10} g C m^{-2} day^{-1})	16.78	1,359	<0.001	-1.477	<0.001	7.192	<0.001	0.04
		Sea surface temperature (°C)	103.8	1,359	<0.001	0.179	<0.001	0.475	0.094	0.22

	production (log_{10}(90%) – log_{10}(10%))	Chlorophyll a (log_{10} mg m^{-3})	63.06	1,359	<0.001	-1.776	<0.001	2.850	<0.001	0.15
Slope	Size spectrum slope	Primary production (log_{10} g C m^{-2} day^{-1})	9.207	1,359	0.003	0.147	0.003	-1.576	<0.001	0.02
		Sea surface temperature (°C)	22.86	1,359	<0.001	-0.012	<0.001	-0.990	<0.001	0.06
		Chlorophyll a (log_{10} mg m^{-3})	27.03	1,359	<0.001	0.161	<0.001	-1.147	<0.001	0.07
Intercept	Size spectrum intercept (log_2 cells m^{-3})	Primary production (log_{10} g C m^{-2} day^{-1})	54.16	1,359	<0.001	1.714	<0.001	17.378	<0.001	0.13
		Sea surface temperature (°C)	29.21	1,359	<0.001	-0.070	<0.001	23.146	<0.001	0.07
		Chlorophyll a (log_{10} mg m^{-3})	32.53	1,359	<0.001	0.897	<0.001	22.244	<0.001	0.08

Supplementary Table 2: Prediction statistics for remotely-sensed environmental variables used to predict phytoplankton cell sizes.

M_{B50}		M_{P50}	
Explanatory Variables	$\overline{D}$	**Explanatory Variables**	$\overline{D}$
Mean	0.93	Mean	0.80
Primary production (log_{10})	0.88	Primary production (log_{10})	0.77
Sea surface temperature	0.74	Sea surface temperature	0.69
Chl a (log_{10})	0.72	Chl a (log_{10})	0.65
Chl a (log_{10}), primary production (log_{10})	0.71	Chl a (log_{10}), primary production (log_{10})	0.64
Chl a (log_{10}), sea surface temperature	0.66	Chl a (log_{10}), sea surface temperature	0.61
Prediction models			
Variables	SST + Chl a (log10)	Variables	SST + Chl a (log10)
F-statistic	178.8	F-statistic	157.6
Degrees of freedom	2,358	Degrees of freedom	2,358
Multipliers (SST, log_{10} Chl a)	-0.076, 0.999	Multipliers (SST, log_{10} Chl a)	-0.058, 0.909
Intercept	1.873	Intercept	1.813
P-value	<0.001	P-value	<0.001
Adjusted R-squared	0.50	Adjusted R-squared	0.47
M_{B90-10}		M_{P90-10}	
Explanatory Variables	$\overline{D}$	**Explanatory Variables**	$\overline{D}$
Mean	2.12	Mean	1.94
Primary production (log_{10})	2.01	Primary production (log_{10})	1.85
Sea surface temperature	1.69	Sea surface temperature	1.66
Chl a (log_{10})	1.84	Chl a (log_{10})	1.74
Sea surface temperature, primary production (log_{10})	1.67	Sea surface temperature, primary production (log_{10})	1.63
Sea surface temperature, Chl a (log_{10})	1.67	Sea surface temperature, Chl a (log_{10})	1.65
Prediction models			
Variables	SST + Chl a (log10)	Variables	SST + Chl a (log10)
F-statistic	93.82	F-statistic	58.88
Degrees of freedom	2,358	Degrees of freedom	2,358
Multipliers (SST, log_{10} Chl a)	0.184, -1.082	Multipliers (SST, log_{10} Chl a)	0.142, -0.837

Intercept	0.447	Intercept	0.906
P-value	<0.001	P-value	<0.001
Adjusted R-squared	0.34	Adjusted R-squared	0.24

Supplementary Table 3: Prediction statistics for remotely-sensed environmental variables used to predict slope and intercept of phytoplankton size spectra.

Size spectra slope		**Size spectra intercept**	
Explanatory Variables	$\overline{D}$	**Explanatory Variables**	$\overline{D}$
Mean	0.26	Mean	1.70
Primary production (log_{10})	0.25	Primary production (log_{10})	1.70
Sea surface temperature	0.24	Sea surface temperature	1.70
Chl a (log_{10})	0.24	Chl a (log_{10})	1.71
Sea surface temperature, primary production (log_{10})	0.24	Chl a (log_{10}), primary production (log_{10})	1.71
Sea surface temperature, Chl a (log_{10})	0.23	Chl a (log_{10}), sea surface temperature	1.72

Prediction models				
Variables	SST + Chl a (log_{10})		Variable	Chl a (log_{10})
F-statistic	16.48		F-statistic	13.52
Degrees of freedom	2,358		Degrees of freedom	1,359
Multipliers (SST, log_{10} Chl a)	-0.007	0.114	Multiplier (log_{10} Chl a)	0.816
Intercept	-1.049		Intercept	29.802
P-value	<0.001		P-value	<0.001
Adjusted R-squared	0.08		Adjusted R-squared	0.03

Supplementary Material

Calculation of biomass in defined cell size classes

The empirical models derived in the main paper are used to predict the slope (b) and intercept of the size spectrum, location of size spectrum midpoint (M_{B50}) and endpoints (M_{B10} and M_{B90}). M_{B50} is defined as M at 50% cumulative biomass based on the relationship between cumulative B and M. M_{B10} and M_{B90} are similarly defined for 10% and 90% respectively. All points are derived by fitting the relationship $B = 100\ (1- \exp\ (-pM^{q}))$ to data then rearranging and substituting to estimate M as $M = (-\ln(1-x/100)/p)^{1/q}$ where x is the relevant percentage.

The prediction equations were:

$$Slope\,(b) = -0.007\,SST\,(^{\circ}C) + 0.114 \log_{10} Chl\,a\,(mg\,m^{-3}) - 1.049$$

$$Log_{2}\,(intercept) = 0.816 \log_{10} Chl\,a\,(mg\,m^{-3}) + 29.802$$

$$Log_{10}\,(M_{B90-10}) = 0.184\,SST\,(^{\circ}C) - 1.082 \log_{10} Chl\,a\,(mg\,m^{-3}) + 0.447$$

$$Log_{10}\,(M_{B50}) = -0.076\,SST\,(^{\circ}C) + 0.999 \log_{10} Chl\,a\,(mg\,m^{-3}) + 1.873$$

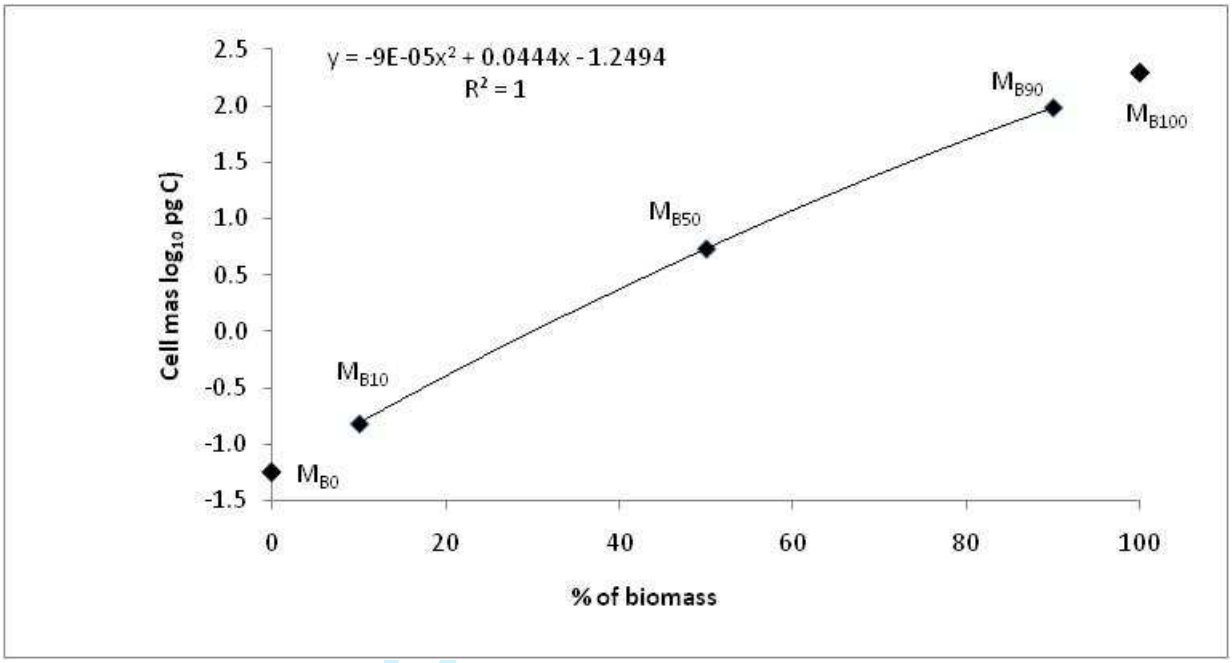
The location of the cell mass range is calculated so that that the integrated biomass to either side of the mid-point is equal. Thus the value for mass at 10% (M_{B10}) is calculated as:

$$M_{B10} = M_{B50}\,\{(10^{(b+1)\ \log_{10}(M_{B90-10})} + 1)/2\}^{-(1/(b+1))}$$

And 90% (M_{B90}):

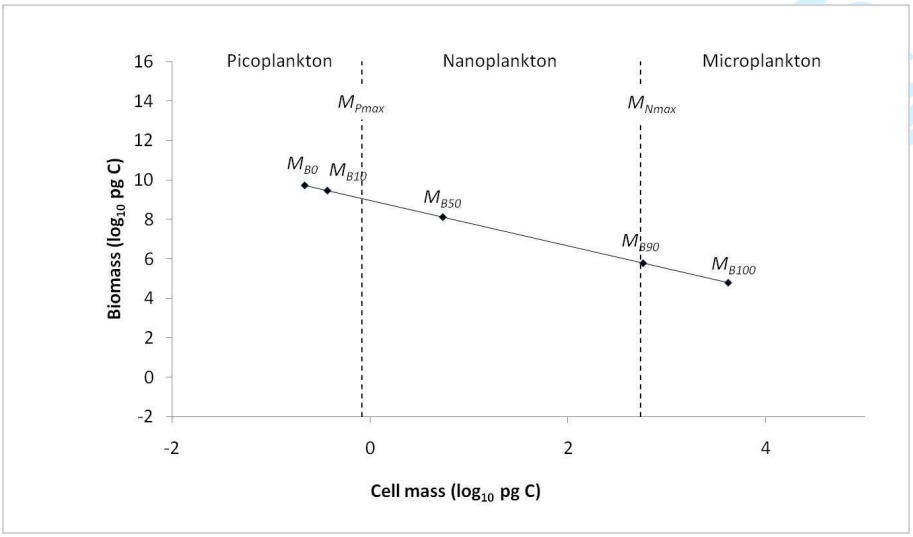
$$M_{B90} = 10\,\hat{}\,(\log_{10}(M_{B10}) + \log_{10}(M_{B90-10}))$$

A statistical model that adequately fitted the very variable tails of the cumulative distribution was not found, and so to ensure the integrated B was equal to 100% of cumulative B the values of M_{B0} and M_{B100} were estimated by plotting M_{B10}, M_{B50} and M_{B90} and fitting a second order polynomial relationship (Supplementary Fig. 1).



Supplementary Figure 1: Example plot of M_{B10}, M_{B50} and M_{B90} with a fitted second order polynomial relationship to enable estimation of M_{B0} and M_{B100}.

If the mass class boundaries between picoplankton and nanoplankton, and between nanoplankton and microplankton, are defined as M_{Pmax} (=M_{Nmin}) and M_{Nmax} (=M_{Mmin}) respectively then when (1) $M_{B0} < M_{\mathrm{Pmax}}$ and $M_{B100} > M_{\mathrm{Nmax}}$ then picoplankton, nanoplankton and microplankton are all present. When (2) $M_{B0} > M_{\mathrm{Pmax}}$ and $M_{B100} > M_{\mathrm{Nmax}}$ then nanoplankton and microplankton are present. When (3) $M_{B0} < M_{\mathrm{Pmax}}$ and $M_{B100} < M_{\mathrm{Nmax}}$ then picoplankton and nanoplankton are present and when (4) $M_{B0} > M_{\mathrm{Pmax}}$ and $M_{B100} < M_{\mathrm{Nmax}}$ then only nanoplankton are present (Supplementary Fig. 2).



Supplementary Figure 2: Example of the relationships between M_{B0}, M_{B10}, M_{B50}, M_{B90} and M_{B100} for a hypothetical size spectrum. The biomass of phytoplankton in picoplankton, nanoplankton and

microplankton size classes is defined as the integrated biomass between M_{B0} and M_{B100} that falls within class boundaries (broken lines denote mass class boundaries M_{Pmax} (=M_{Nmin}) and M_{Nmax} (=M_{Mmin})).

So, to calculate the biomass of each size class that is present as a percentage of total biomass, the following approaches are adopted in each of these cases: (1), calculate the biomass of picoplankton, nanoplankton and microplankton as a proportion of total biomass, (2) picoplankton are not present, so calculate the biomass of nanoplankton and microplankton as a proportion of total biomass, (3) microplankton are not present, so calculate the biomass of nanoplankton and picoplankton as a proportion of total biomass and (4) only nanoplankton are present, so they must constitute 100% of total biomass.

For biomass, calculate the area under the curve y = mx + c between x_1 and x_n:

$$\int_{x_1}^{x_n} (mx + c)\, dx \quad = \quad \left[\frac{1}{2} mx^2 + cx \right]_{x_1}^{x_n} \quad = \quad \frac{1}{2} mx_n^{\,2} + cx_n - \frac{1}{2} mx_1^{\,2} - cx_1$$

Total biomass is found between M_{B0} and M_{B100}; biomass in the picoplankton group is found between M_{B0} and M_{Pmax}, biomass in the nanoplankton group is found between M_{Pmax} and M_{Nmax} and biomass in the microplankton group is found between M_{Nmax} and M_{B100}.

For example, where sea surface temperature is 15ºC and chlorophyll a concentration is 1.0 mg m^{-3} there is total biomass of 47 $\log_{10}$ pg C, of which 24% is picoplankton, 76% is nanoplankton and there is no microplankton.