Density estimation of plankton size spectra: a reanalysis of IronEx II data

MARKUS SCHARTAU[1]*, MICHAEL R. LANDRY[2] AND ROBERT A. ARMSTRONG[3]

[1]GKSS FORSCHUNGSZENTRUM GEESTHACHT, INSTITUT FÜR KÜSTENFORSCHUNG, GEESTHACHT, GERMANY, [2]SCRIPPS INSTITUTION OF OCEANOGRAPHY, UNIVERSITY OF CALIFORNIA SAN DIEGO, CA, USA AND [3]SCHOOL OF MARINE AND ATMOSPHERIC SCIENCES, STONY BROOK UNIVERSITY, NY, USA

*CORRESPONDING AUTHOR: markus.schartau@gkss.de

Received August 4, 2009; accepted in principle May 24, 2010; accepted for publication May 26, 2010

Corresponding editor: Roger Harris

Many critical processes of ecosystem function, including trophic relationships between predators and prey and maximum rates of photosynthesis and growth, are size-dependent. Size spectral data are therefore precious to modellers because they can constrain model predictions of size-dependent processes. Here we illustrate a multi-step statistical approach to create size spectra based on a reanalysis of plankton size data from the IronEx II experiment, where iron was added to a marked patch of water and changes in productivity and community structure were followed. First, bootstrapping was applied to resample original size measurements and cell counts. Kernel density estimation was then used to provide nonparametric descriptions of density versus size. Finally, parametric distributions were used to obtain parameter estimates that can more easily be applied in models. A major advantage of this approach is that it provides confidence envelopes for the density distributions. Our analyses suggest three basic distributional patterns of cell concentration versus logarithm of equivalent spherical diameter for individual taxa. Composite size-densities of heterotrophs and photoautotrophs reveal important aspects of the coupling between protist grazing and the phytoplankton community.

KEYWORDS: plankton size; nonparametric kernel density estimation; parametric size density; plankton community structure; iron fertilization

INTRODUCTION

Most oceanic primary production and consumption occurs in the microbial base of the food web among bacteria and protists (Calbet and Landry, 2004). Trophic interactions within this microbial base are potentially complex (Azam and Malfatti, 2007), but the general structure suggests that size-based analyses and interpretation are needed (Armstrong, 2003a, b). The ability to construct, analyse and predict changes in size distributions may therefore allow a more precise description of interactions within plankton communities, where larger consumers typically eat smaller prey (Hansen *et al.*, 1994).

Interest in the size structure of aquatic systems has been stimulated by the introduction of technologies (Coulter Counter, flow cytometer, image analysis microscopy) for efficient enumeration of particles and organisms by size, and has sparked the development of additional novel approaches and instruments (Optical Plankton Counter, Underwater Video Profiler, FlowCAM, ZooScan) to fill measurement gaps.

Plankton size distributions, often pieced together from measurements optimized for different size ranges or types of organisms, are typically represented as size spectra, where the main attribute is the slope on a biomass-normalized logarithmic scale (Sheldon *et al.*, 1972). Numerous studies have established the usefulness of plankton size spectra for summarizing the inherent structural properties of a system's biota, and for comparing and contrasting system types and temporal/spatial

doi:10.1093/plankt/fbq072, available online at www.plankt.oxfordjournals.org.

variability within systems (e.g. Rodriquez and Mullin, 1986; Gaedke, 1993; Cavender-Bares *et al.*, 2001; Rinaldo *et al.*, 2002; San Martin *et al.*, 2006a; Reul *et al.*, 2006). Understanding the size structure of lower trophic levels has similarly provided mechanistic insights into the bioenergetic constraints and habitat partitioning of foraging predators (e.g. Kerr, 1979; Rykaczewski and Checkley, 2008). Parallel developments in spectral analyses of plankton communities and allometric theory have further established a basis for inferring biomass and energy/mass flux, trophic efficiencies and vital rates (growth and mortality) from biomass distributions in aquatic food webs (e.g. Kerr, 1974; Platt and Denman, 1977; Moloney and Field, 1989; Zhou and Huntley, 1997; San Martin *et al.*, 2006b; Zhou, 2006).

To allow the development of new size-based formulations of trophic interaction and energy flow, detailed information on size densities of individual plankton groups are needed. One attractive possibility is to characterize producer and consumer species in terms of parametric size-density functions, for example, using a normal probability density function (pdf) with two parameters (mean and variance) for each species or group. Such parametric pdfs could be used in describing plankton interaction by convolving parametric predator- and prey-density functions (Armstrong, 2003a). To derive parametric size-density functions from observation is not straightforward. Historically, microscopical data have been plotted as discrete histograms to combine species information with size measurements. However, this approach produces results of uneven quality, particularly when cell abundance is low and when sampled water volumes are small.

In this study, we describe a sequential procedure in which we first derive nonparametric kernel density estimates and then fit parametric functions to the nonparametric distributions. We illustrate this analysis with size-based data for plankton taxa observed during the IronEx II mesoscale iron-enrichment experiment in the eastern equatorial Pacific (Landry *et al.*, 2000a). Our data reconstruction reveals taxon-specific details of community size responses to iron fertilization and grazing impacts. More generally, the approach supports size-structure inter-comparison of disparate data sets and the development of size-based plankton models.

METHOD

Our primary data are plankton counts, cell lengths and widths, distinguished by taxonomic group. We analysed the data using kernel density estimators (Parzen, 1962; Wegman, 1972; Silverman, 1986), which provide smooth nonparametric plankton size-densities, even from sparse discrete measurements. In addition, we assessed confidence limits of these nonparametric density estimates. The confidence limits were obtained by applying a bootstrapping approach (Efron and Tibshirani, 1993), where original plankton counts and size measurements were resampled with the objective of separating size-structure information from observational noise. Plankton counts and corresponding size measurements were assumed to follow specified error distributions: Poisson distribution for counts, and Uniform, and Gaussian distributions for size measurements. For each set of original and resampled data, we computed a nonparametric representation of density versus size. Computations are extensive, but the resample series (ensemble) of nonparametric estimates allowed us to calculate mean size densities for each plankton group and to specify their confidence limits. We then fit unimodal and bimodal normal distribution functions, on a logarithmic scale, to mean density estimates of dominant plankton groups. In the sections below, we describe the assumptions inherent in our bootstrapping approach, provide details for the optimum bandwidth selection needed for the kernel density estimation and define a misfit function for fitting parametric functions to plankton size-density estimates. Figure 1 depicts the steps in these analyses.

Plankton counts and sizes

During the IronEx II project in the eastern equatorial Pacific (Coale *et al.*, 1996), data on species identity, abundance and size were collected at a series of stations, both inside and outside a mesoscale iron-fertilized patch. At all stations, samples were collected from the mixed layer (15 m) and analysed as described in Landry *et al.* (Landry *et al.*, 2000a). Briefly, small volume (3 mL) samples were preserved with 0.9% paraformaldehyde and analysed by dual-beam flow cytometry for populations of *Prochlorococcus, Synechococcus* and eukarytoic phytoplankton. Samples of 20–50 mL were preserved with 1% gluteraldehyde, stained with DAPI and prepared as slides (0.2-μm filters) for analysis by epifluorescence microscopy. At least 2000 cells per slide were enumerated and sized in specific recognizable categories (species, genera or classes) according to the method of Chavez *et al.* (Chavez *et al.*, 1991), with the vast majority of cells viewed at ×1000 magnification, but with large (>40 μm) and rare types counted on a larger portion of the slide at ×400. For the present analysis, picophytoplankton cell categories were grouped for data processing, with direct size measurements of *Synechococcus* and eukaryotes (abundance estimates from microscopy and flow cytometry were in good agreement). For

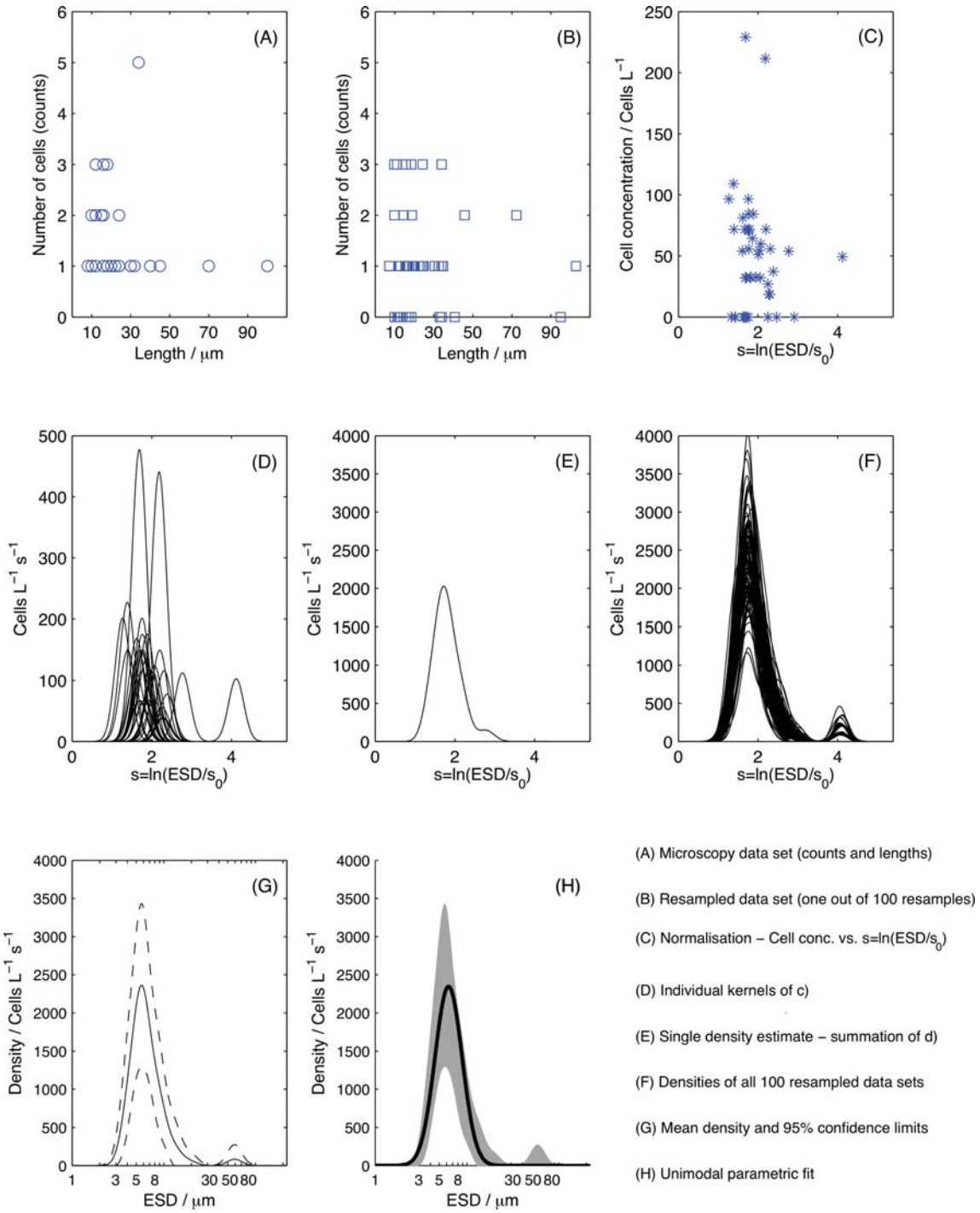

Fig. 1. (**A**–**H**) The sequential analysis procedure. The example includes resampling of counts and size measurements of centric diatoms observed at stations within the iron-fertilizsed patch. Original and resample data are similarly treated throughout the analysis. Details are explained in the Method section.

Prochlorococcus, we assumed a mean size of 0.735 μm, similar to that of Blanchot *et al*. (Blanchot *et al*., 2001). Figure 1A shows an example of original cell counts and the corresponding length measurements of centric diatoms observed at stations within the iron-fertilized patch. To compare plankton responses, we combined measurements from stations inside the fertilized patch into one data set, and those from stations outside the

Table I: Resolutions of microscopic size measurements during the IronEx II experiment

Resolution of microscopic measurement, μm	Size range for length, μm	Size range for width, μm
1	2–6	2–6
2	6–40	6–40
5	40–80	40–50
10	80–180	50–100
20	180–300	>100
40	300–500	–
100	>500	–

Resampled data sets were drawn from uniform error distributions based on the microscopic resolution within each length and width size range.

fertilized patch into a second ("control") set. Accordingly, we did not resolve variations due to temporal and spatial variability inside and outside the patch.

Resampling of lengths, widths and counts

Obtaining simultaneous data on cell abundances, size and taxonomic identity requires labour-intensive microscopy; and it is therefore seldom possible, within time and effort constraints, to replicate microscopic analyses. On the other hand, we desire confidence limits for nonparametric size-density estimates, while considering uncertainties in counts and size measurements. We calculated these confidence limits from an ensemble of kernel density estimates. The ensemble of densities was determined using resample data sets, which were obtained by randomly resampling the original data set with replacement (bootstrapping). In addition to the original data set, the bootstrapping produces sets of pseudo-data, having the same statistical properties as the data. In total, we produced 100 resample data sets from each primary data set.

The absolute accuracies of microscope measurements of cell length and width depend on the magnification chosen. By examining the data from IronEx II, we found that the precision of length and width measurements varied with different magnifications used. Table I shows the size dependency in precision, rounded to the nearest 1, 2, 5, 10, 20, 40 and 100 μm. The precisions in length and width measurements increase with increasing size. We imposed uniform error distributions whose interval was equal to the resolution of the size range. For sizes below 2 μm, we assumed a normal error distribution with a standard deviation of 0.2 μm. While length and width were resampled, we approximated cell thickness based on different shape factors, which varied among plankton groups. For approximation of cell thickness, we adopted the conversion factors used in Landry *et al.* (Landry *et al.*, 2000a), who related width to thickness according to Chavez *et al.* (Chavez *et al.*, 1991). Cell volumes were calculated using the formula for a scalene ellipsoid, combining the two semi-axes (measured or resampled) length and width with the calculated thickness. Cell volumes were then rescaled to equivalent spherical diameter (ESD) of a sphere.

A single plankton count can be interpreted as a discrete event during a certain time period of measurement, where each cell is found within a microscopic sub-sample that has been taken from a larger water volume. If no event is found, the sub-sample may have been too small to capture cells that were highly dispersed at this station. On the other hand, if a cell is counted, it may not be representative of the larger water volume and an extrapolation to cell concentration (cells per litre) may be biased. Thus, a single plankton count within a sub-sample should be regarded as a rare event, and a repetition of this event should remain rare during resampling. The Poisson error distribution is an appropriate description of uncertainty when accounting for such rare events. As individual counts were resolved in the data set, we therefore imposed the Poisson expected value of "one". We also obtained zero counts while drawing from the Poisson distribution (Fig. 1B). The resampled data were then extrapolated to cell concentrations in units of cells per litre (Fig. 1B and C) as in Landry *et al.* (Landry *et al.*, 2000a), accounting for different water volumes filtered and varying portions of the slide areas analysed. Individual cell counts correspond to concentrations on the order of 10^2 cells per litre (Fig. 1C).

Nonparametric density estimation

Histograms are often used to depict plankton size densities. Defining the bins of histograms requires defining points of origin and interval widths, subjective choices that can yield different distributional patterns. Such subjectivity can be justified; but it does hinder intercomparison among studies. In the present study, we chose a different statistical approach, based on a density estimation method first proposed by Parzen (Parzen, 1962) and now commonly termed "kernel density estimation". In comparison to other density estimation techniques, the kernel density estimator generally provides greater detail of distribution structure (Wegman, 1972; Silverman, 1986). A nonparametric density obtained from a kernel density estimator is less susceptible to the intuitive specifications needed for histograms. Also, if the density estimate is smooth rather than discontinuous, the intermediate step of kernel

density estimation provides an ideal basis for devising parametric functions.

A kernel density estimate is derived from the sum of symmetric probability density functions (kernels) defined around each data point. For each data point, the kernel can be regarded as a weighting function for the observation. For example, Fig. 1D shows all kernels of the resampled set depicted in Fig. 1C. The summation over all kernels leads to a smoothed distribution (Fig. 1E), which can then be normalized to fulfil criteria of a probability density function (pdf). In plankton ecology, the utility of kernel density estimation has not been widely appreciated. Our analysis joins recent studies by Sanvicente-Añorve *et al.* (Sanvicente-Añorve *et al.*, 2003) and Quintana *et al.* (Quintana *et al.*, 2008) in applying kernel density estimation to plankton size spectra. We complement the method with a bootstrapping approach that relates uncertainties in plankton measurements to confidence intervals of the density estimates (Fig. 1F and G).

Kernel density estimation of plankton size

ESD data (esd_{ij}; μm, with index i for every observation of plankton group j) were calculated from length and width data that were originally rounded to discrete sizes (Table I), depending on the magnifications used for microscopy. After resampling of cell counts and sizes (Fig. 1B), we calculated cell concentrations (c_{ij}; cells L^{-1}), multiplying counts (n_{ij}) by an extrapolation factor (E) (Fig. 1C). The total cell concentration (C_j^{tot}) is the sum of M individual cell concentrations:

$$C_j^{\mathrm{tot}} = \sum_{i=1}^{M_j} c_{ij}(\mathrm{esd}_{ij}) = \sum_{i=1}^{M_j} E \cdot n_{ij}(\mathrm{esd}_{ij}) \qquad (1)$$

The size distribution ($\hat{f}_j$) of a single plankton group can be described as the product of total cell concentration and the pdf of ESD:

$$\hat{f}_j = C_j^{\mathrm{tot}} \cdot \mathrm{pdf}_j(\mathrm{ESD}). \qquad (2)$$

By definition, a pdf is required to integrate to one. For our size range of interest, 0−327 μm, the following equation fulfils that constraint:

$$\int_{0\,\mu\mathrm{m}}^{327\,\mu\mathrm{m}} \mathrm{pdf} \times \mathrm{dESD} = 1. \qquad (3)$$

Kernel density estimation theory allows one to construct a nonparametric plankton size distribution that is "smooth" (i.e. continuous with continuous derivatives of all orders). A smooth density estimate ($\hat{f}_j$) can be normalized simply by dividing by total cell abundance (cells L^{-1}); the result can then be interpreted as a data-based, nonparametric pdf. Since phytoplankton size measurements extend over a range of three orders of magnitude, we apply the kernel density estimator on a logarithmic scale. As used traditionally in aquatic ecology, log-scale analysis implicitly accounts for the doubling of cell volume during the cell cycle, allowing each species to have the same width on the log-scale, rather than giving larger species larger apparent variances. For kernel density estimation, we first transform ESD to a dimensionless size variable s:

$$s = \ln\left(\frac{\mathrm{ESD}}{s_0}\right) = \ln\left(\frac{\mathrm{ESD}}{1\,\mu\mathrm{m}}\right). \qquad (4)$$

A kernel density estimate for a pdf is constructed by adding together all individual kernel functions (here Gaussian distribution functions on a logarithmic scale, Fig. 1D). Each Gaussian kernel is centred around $\ln(esd_{ij}/s_0)$, the normalized and log-transformed observation of esd_{ij}, with index i for each observation within plankton group j. The Gaussian kernel itself is a smooth distribution that integrates to one. Each kernel is then multiplied by the corresponding cell concentration (c_{ij}; cells L^{-1}). Therefore, the sum of Gaussian kernels has to be normalized (dividing by the total cell concentration, C_j^{tot}) in order to yield again a probability density estimate for plankton group j:

$$pdf_{\,\mathrm{est}}(s)_j = \frac{1}{C_j^{\mathrm{tot}}} \sum_{i=1}^{M_j} \frac{c_{ij}}{h_j\sqrt{2\pi}} \exp\left[\frac{(s - \ln(esd_{ij}/s_0))^2}{2h_j^2}\right] \qquad (5)$$

The width of each kernel distribution is expressed by the smoothing parameter h_j, called *bandwidth*. When a Gaussian kernel is chosen for the transformed variable s [equation (5)], the bandwidth h_j, which remains constant for a single plankton species or group, resembles a standard deviation on a logarithmic scale. Overall, it is the summation over a series of kernels that yield a final distribution where each data point is represented by itself, which is the major difference from the application of a histogram.

Estimation of optimal bandwidths

The main task in kernel estimation theory is the selection of an optimum bandwidth, one that minimizes bias while maximizing variance information (bias-variance trade-off). Briefly, a best bandwidth is identified if the estimate of the pdf at any data point can also be

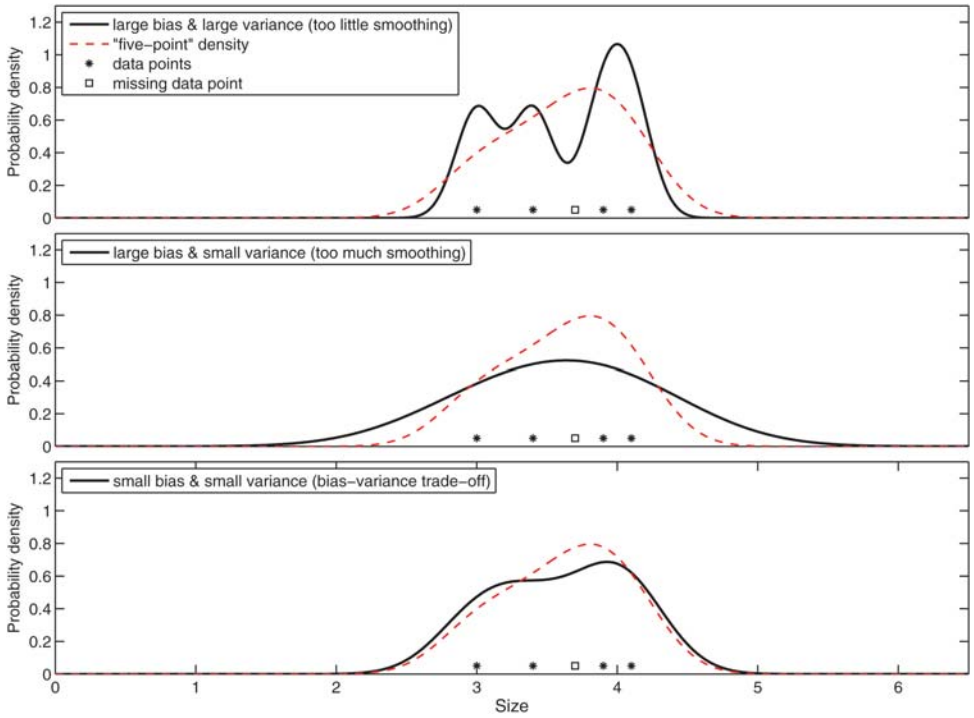

Fig. 2. Example of bandwidth selection. A reference density (dashed line) is described by five data points (asterisk and square symbols); our aim is to recover a density estimate (solid line) after removing one data point (square symbol). 'Bias' refers to the deviation between the dashed line and the solid line at the location of the missing point; 'variance' refers to the variability of the density estimates that might be gotten at that point, e.g., from different bootstrap estimates. See text for further explanation.

represented by a density estimate from neighbouring points. For example, if we remove one data point and if the bandwidth were too broad, the entire pdf would spread over too large a range, and the pdf estimate at the location of the missing point would be too small (biased estimate). On the other hand, if the bandwidth were too narrow, the pdf would consist of narrow spikes. The pdf values would have large variations within the size range of interest (too much variance). In the worst case, the pdf would become zero at the missing point. Figure 2 depicts an example of bandwidth selection where a reference density, which was described by five data points, has to be recovered by a density estimate from the four neighbouring data points. Optimum bandwidth selection provides reasonable density predictions for any data point removed.

A common approach to kernel density estimation is to specify a single bandwidth that remains constant throughout the analysis of a data set. Criteria exist for choosing an optimal constant bandwidth, mostly derived from practical considerations and based on normality assumptions. For example, assuming *a priori* that a final density estimate $\hat{f}$ can be expressed in terms of a normal density distribution facilitates the approximation of an optimal constant bandwidth (finding an asymptotic solution that describes the estimator's trade-off between variance and bias). Making such assumptions about unknown density distributions for bandwidth selection are often referred to as plug-in methods.

Pragmatic plug-in methods for choosing constant bandwidths have been shown to produce good results for size-structure analyses of fish larvae (Sanvicente-Añorve *et al.*, 2003) and for the estimation of size diversity (Quintana *et al.*, 2008). Our approach is similar, but we adopt an *ansatz*, discussed by Härdle *et al.* (Härdle *et al.*, 2004), who refined a rule-of-thumb bandwidth method originally suggested by Silverman (Silverman, 1986). Our bandwidth estimate h_j is inferred from the variance information of the normalized and log-transformed ESD observations, $s_{ij} = \ln(\mathrm{esd}_{ij}/s_0)$. To reduce the sensitivity of such an estimate to outliers, an inter-quartile range (R_j) was calculated by subtracting the upper 75% quartile from the lower 25% quartile:

$$R_j = s_{ij[0.75\mathrm{M}]} - s_{ij[0.25\mathrm{M}]} \quad (6)$$

Given the inter-quartile range R_j together with the standard deviation ($\hat{\sigma}_j$) of s_{ij}, the rule-of-thumb

bandwidth can then be calculated as:

$$h_j = 1.06 \cdot \min\left\{\hat{\sigma}_j, \frac{R_j}{1.34}\right\} M_j^{-1/5} \tag{7}$$

with M_j being the total number of data points for each plankton group. The factors 1.06 and 1.34 are derived from statistical considerations, assuming that the nonparametric density estimates resemble normal distributions (Härdle *et al.*, 2004). Numbers of data points for each plankton group, together with their observed size ranges, of this study are provided as Supplementary Material.

Fitting parametric distribution functions to plankton size density

For parametric density functions, we chose normal density functions on logarithmic scales to characterize the patterns suggested by the nonparametric density estimates. We hypothesize that each plankton group identified during measurement may actually consist of more than one species, or different clones if each resolved "type" represents a species. Plankton size distributions for each "type" are assumed to arise from a mixture of cells whose component populations are subject to different intrinsic growth parameters on the species level and are exposed to external variability due to micro-scale physics, nutrient availabilities and plankton interactions. As we are concerned with the dominant peaks of the densities, we restrict our analysis to resolving no more than two peaks. Therefore, the general form of the probability density function is bimodal (pdf_2), defined as an affine combination of two Gaussian functions on a logarithmic scale:

$$\begin{aligned} &pdf(s, q, a_1, b_1, a_2, b_2) \\ &\quad = (1-q)\cdot \mathcal{N}(s, a_1, b_1) + q \cdot \mathcal{N}(s, a_2, b_2) \\ &\quad = \frac{1}{\sqrt{2\pi}}\left[\frac{(1-q)}{b_1}\exp\left(-\frac{(s-a_1)^2}{2b_1^2}\right)\right. \\ &\quad \left. + \frac{q}{b_2}\exp\left(-\frac{(s-a_2)^2}{2b_2^2}\right)\right] \end{aligned} \tag{8}$$

The unimodal function (pdf_1) was derived simply by setting $q = 0$ in equation (8) and optimizing a_1 and b_1 only. The parameters (a_1, a_2 and b_1, b_2) are the means and variances of the normalized and log-transformed variable (s). The first mode of the pdf is characterized by its amplitude $(1 - q) \geq 0$ at location $s = a_1$ and its width b_1; similarly, the second mode is specified by its amplitude (q) at $s = a_2$ and width b_2.

Optimum parameter values were obtained by maximum likelihood estimation. Although the notion of comparing two continuous distributions (as it is the case for our nonparametric and the parametric densities) differs from a data-model comparison, an appropriate metric should reflect the number of data points. A perfect data-model comparison also demands the consideration of covariances between data points. The calculation of covariances between size measurements could possibly be achieved by incorporating the optimum bandwidth, defined before, as correlation length scale. Nevertheless, the proof of using the optimum bandwidth for computations of covariances requires an in-depth analysis that goes beyond the scope of this study. Furthermore, the optimum bandwidths differ, although slightly, between resample data sets of each plankton group. This means that an ideal data-model comparison would also require data-model fits for each resample data set. The average of the ensemble of parameters estimates for equation (8) would then describe a perfect fit. To keep computations moderate, we assumed a conditional probability that imposes independence between size measurements and describes deviations between the parametric function [equation (8)] and the nonparametric kernel density estimate [equation (5)] at points corresponding to ESDs in the original data sets. Our conditional probability can be defined as a likelihood (L):

$$\begin{aligned} &\mathrm{prob}(pdf_{est}(s_i)|q, a_1, b_1, a_2, b_2, pdf) = L \\ &\quad = \prod_{i=1}^{M} \frac{1}{\varepsilon_i\sqrt{2\pi}}\exp\left(-\frac{(pdf_{est}(s_i) - pdf(s_i))^2}{2\varepsilon_i^2}\right) \end{aligned} \tag{9}$$

This likelihood measures the fit to the nonparametric density estimate ($\mathrm{pdf}_{\mathrm{est}}$) at the points of measurement (s_i), given the parameter values (q, a_1, b_1, a_2, b_2) for the parametric function (pdf) [equation (8)]. The likelihood's variance information (ε_i^2, which is the uncertainty in our nonparametric density estimate) was calculated from the ensemble of density estimates, according to our bootstrapping approach. In practice, maximum likelihood estimates are obtained when the global minimum of the negative logarithm of the likelihood ($-\ln(L)$) is found. Rather than seeking for the minimum of the negative logarithm of the likelihood, we calculated the Akaike information criterion (AIC), which also accounts for the additional degree of freedom of the parametric function used to produce

the fit. In our case, the AIC was determined as follows:

$$AIC = 2 \cdot (-\ln(L) + P)$$
$$= -\sum_{i=1}^{M}\left(2\ln\left(\frac{1}{\varepsilon_i \cdot \sqrt{2\pi}}\right) - \frac{(pdf_{est}(s_i) - pdf(s_i))^2}{\varepsilon_i^2}\right)$$
$$+ 2P \qquad (10)$$

with M being the number of data points (Supplementary Material). P is the number of parameters needed for the parametric functions. We distinguished between AIC_1 with $P = 2$ for the unimodal- and AIC_2 with $P = 5$ for the bimodal function. Because of the first term in equation (10), the AIC can become negative. The lower the AIC value, the better the fit. The difference between AIC_2 and AIC_1 allows us to assess whether a bimodal function is an improvement compared to the unimodal pdf for a particular plankton group.

RESULTS

The IronEx II data contain 25 plankton groups, among which we distinguished 18 photoautotrophs and 7 heterotrophs. For simplicity, we show details only for those groups that were most abundant or experienced significant impacts from iron fertilization (Figs 3 and 4). All groups are, however, represented in the composite density estimates (Fig. 5). Density estimates for all 25 plankton groups are provided (with their standard deviations) as Supplementary Material. Tables II and III include information on reconstructing the density estimates with parametric functions [equation (8)], with the number of total cell concentration and parameter values, respectively. The last columns depict differences between AIC_2 and AIC_1 of the bimodal and the unimodal fits, respectively.

Nonparametric and parametric density estimates of plankton size

Phytoplankton

By comparing spectra from the fertilized patch to those from the control sites, we can assess the changes induced by iron addition. By examining the nonparametric density estimates on a logarithmic scale of ESD, we can identify three characteristic distributional patterns: symmetric unimodal, right-skewed and bimodal (two distinct peaks). Figure 3 shows the responses in size spectra for phytoplankton groups that were abundant within the fertilized patch and that were also observed at the control sites. Some large diatoms, such as *Rhizosolenia* and *Thalassiotrix* spp., were not found outside the patch (Supplementary Material).

Prymnesiophytes (Fig. 3A), the group that includes coccolithophorids, have a symmetric distribution on a logarithmic scale, and a unimodal pdf suffices to describe the density of the control (Table II). The dominant mode (the size with highest frequency of cells) is located between ESD = 2.8 μm ($a_1 = 1.03$) and ESD = 2.9 μm ($a_1 = 1.07$), with a small shift towards larger cell sizes when the fertilized patch is compared with the control. The parametric fit of the bimodal density function to the control density estimate yields $q = 0.04$ [equation (8)], so that a unimodal distribution can indeed be preferred over a bimodal fit. In contrast, a bimodal fit produces better results for the fertilized patch. Nonetheless, the parametric fit comprises a small contribution of the second mode ($q = 0.1$) to the overall density distribution, similar to the control. The apparent improvement is largely due to a better representation of the leptokurtic kurtosis that extends the density estimate up to sizes close to 5.9 μm. Apart from these details, the distribution shifts slightly towards larger sizes, which can be adequately described with the unimodal parametric function.

The size density changes for Pseudo-nitzschia (previously incorrectly reported as *Nitzschia* spp. in Landry *et al.*, 2000a) differ from those for prymnesiophytes (Fig. 3B). Instead of a slight shift towards larger cells, the distribution switches from a fairly symmetric unimodal density in the control samples to a bimodal distribution within the patch, with two distinct modes at ESD = 2.9 and 7.3 μm (with $a_1 = 1.05$ and $a_2 = 1.98$). The bimodality of the nonparametric estimates is well represented by the parametric fit, resolving the separation of the two peaks. The first mode (between ESD = 2 and 3 μm) is similar between control and patch. The second mode around 7 μm is detectable in the control samples, but extremely small. Subsequent analyses (data not shown) indicate that two modes are also well expressed in densities based on cell thickness and length rather than on ESD. Cells in the second mode thus seem morphologically similar to those in the first mode but are proportionally larger. The abundance increase of *Pseudo-nitzschia* spp. is most pronounced compared with other phytoplankton, and resolving the second mode in the parametric function is indeed crucial because it contains a significant biovolume response.

Autotrophic dinoflagellates show a clear bimodal distribution in both the control and the fertilized patch (Fig. 3C). However, the response to fertilization is different from the structural changes apparent for *Pseudo-nitzschia*. For dinoflagellates, cell concentration

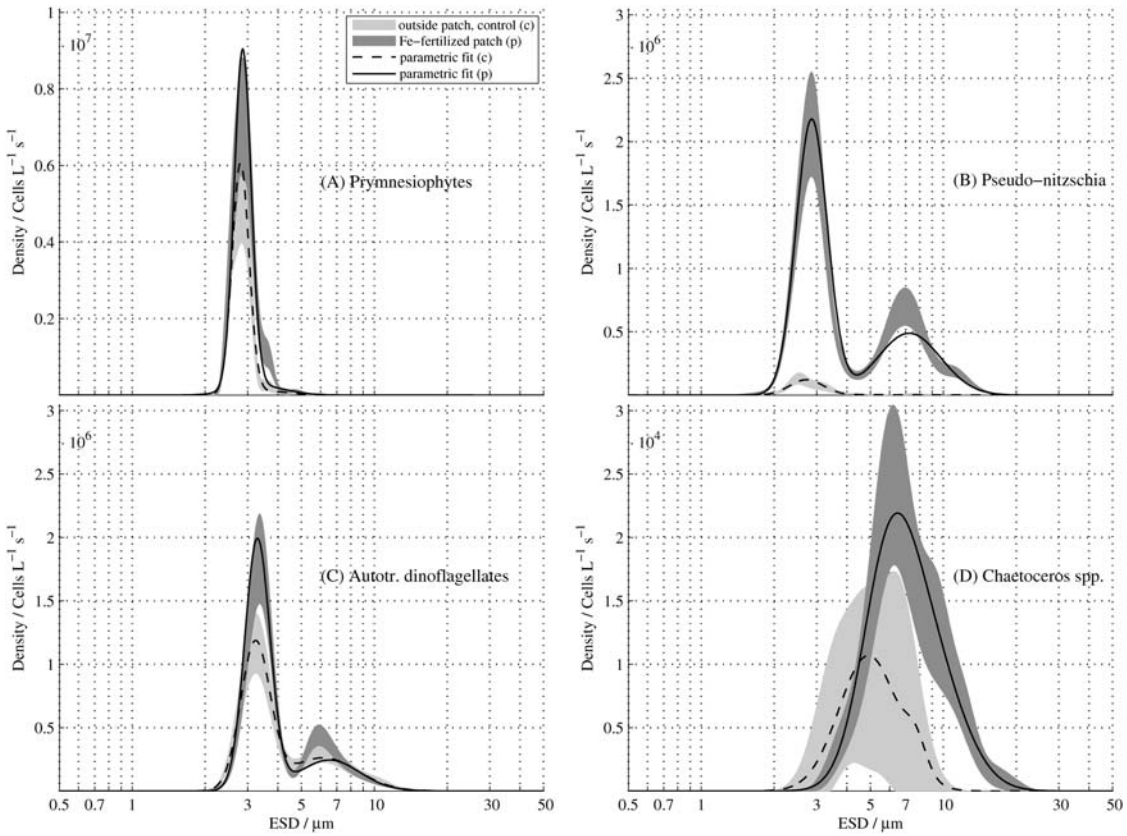

Fig. 3. Cell concentration densities of the four dominant taxonomic phytoplankton groups during the IronEx II experiment. (**A**) Prymnesiophytes, (**B**) *Pseudo-nitzschia*, (**C**) photo-autotrophic dinoflagellates, and (**D**) *Chaetoceros* spp. Shaded areas include nonparametric density estimates of the mean and the 95% confidence limits (± 2 standard deviations). The horizontal axis of equivalent spherical diameter (ESD) is on a logarithmic scale (base 10). ESD can be back-transformed from the normalized, logarithmic variable ($s = \ln(\mathrm{ESD}/1.0\ \mu\mathrm{m})$) to a linear scale according to $\mathrm{ESD} = \exp(s)$.

density increases significantly within the first mode, between 2.8 and 3.5 µm ESD (with $a_1 = 1.17$ and 1.19 for control and patch, respectively). The response in the size range between 5 and 10 µm (second mode) is less expressed. With a_2 between 1.19 and 1.89, the parametric fits do show two distinct modes, but the amplitudes of the first mode are better resolved both inside and outside the patch. As with *Pseudo-nitzschia*, some limitation exists in applying the bimodal function, to match the second mode's amplitude of both nonparametric densities. This may possibly be attributed to an overestimate of the width of the second mode ($b_2 = 0.29$). The second mode seems wider than the first mode and may include a third mode at the upper end of the distributions for *Pseudo-nitzschia* and photoautotrophic dinoflagellates. Only if b_2 were smaller could the amplitudes of the function's second mode fit the density estimates while resolving the amplitudes of the first mode equally well.

The fertilization response of *Chaetoceros* spp. combines most of the characteristics described before, with a shift of the first mode together with an accentuation of the second (Fig. 3D). These structural changes can be inferred from the optimum values of the parametric fits. According to small differences in AIC and the better performance of the unimodal fit in the control, a function with a single mode may suffice to describe the shift toward larger cells. The dominant first mode moves from ESD = 5 to 6 µm. This shift is exaggerated in the parametric density estimate of the unimodal function. The nonparametric density of the iron-patch suggests the existence of a third and perhaps even a fourth mode for larger ESD. Again, as in the parametric fits of photoautotrophic dinoflagellates and *Pseudo-nitzschia*, the second mode of *Chaetoceros* is wide enough ($b_2 = 0.34$) to fit the kurtosis at the upper tail and thus resolves the right-side skewness of the nonparametric density estimate. Such density estimates are particularly challenging, since their skewed distributions and their kurtosis are not ideally represented by either unimodal or bimodal functions on log-scale.

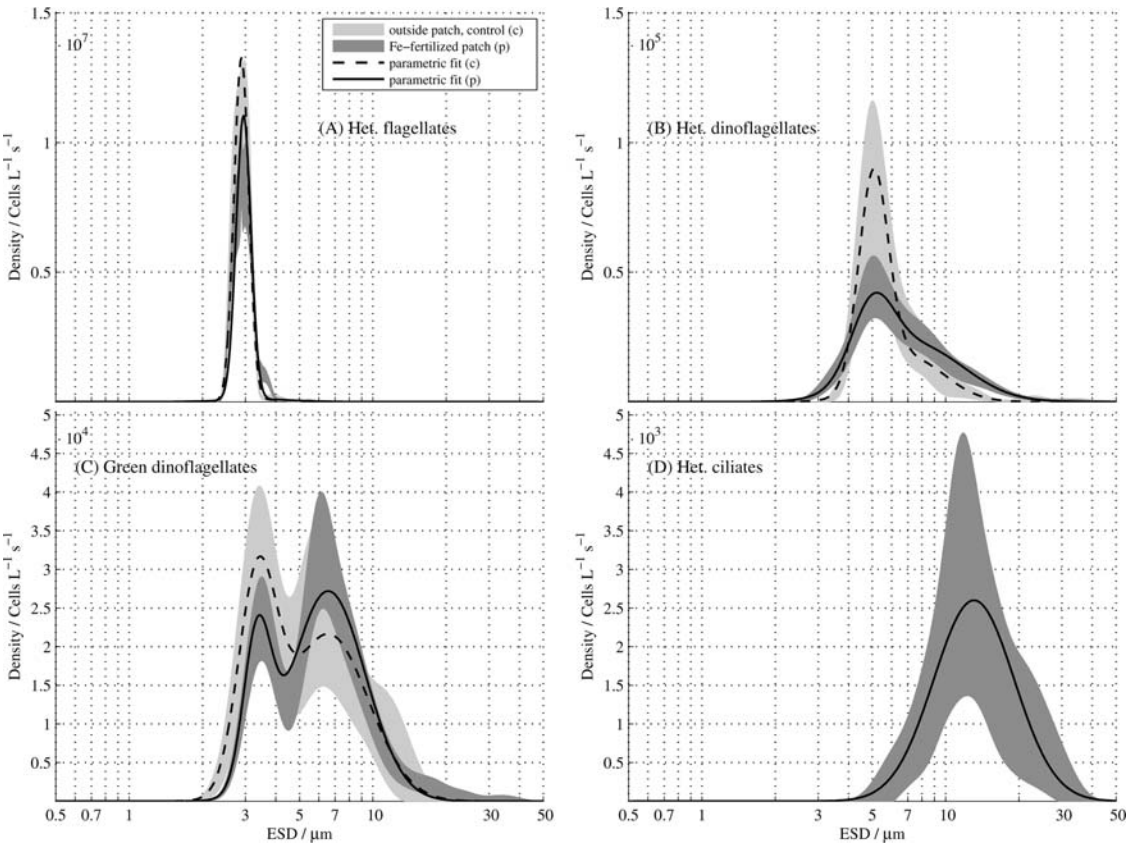

Fig. 4. Cell concentration densities of microzooplankton during the IronEx II experiment. (**A**) Heterotrophic nano-flagellates, (**B**) heterotrophic dinoflagellates, (**C**) heterotrophic, green dinoflagellates with endosymbionts, and (**D**) ciliates. Shaded areas include nonparametric density estimates of the mean and the 95% confidence limits (± 2 standard deviations). The horizontal axis of equivalent spherical diameter (ESD) is on a base 10 logarithmic scale. ESD can be back-transformed from the normalized, logarithmic variable ($s = \ln(\mathrm{ESD}/1.0\ \mu\mathrm{m})$) to a linear scale according to $\mathrm{ESD} = \exp(s)$.

Heterotrophs

Heterotrophs also showed three different response patterns in size densities that are similar to those of the phytoplankton. In contrast, there are major differences between heterotrophic density spectra and photoautotrophic density spectra in the response of cell concentration to iron fertilization. Heterotrophic cells are decreased within the fertilized patch for sizes smaller than 5 µm. Cells with ESDs larger than 5 µm were either equally abundant inside and outside the patch or increased after iron fertilization. Flagellates, the smallest heterotrophic nano-plankton observed, exhibited symmetric unimodal densities both inside and outside the fertilized patch (Fig. 4A). The parametric unimodal density function fits both nonparametric estimates well, as indicated by a positive difference between the AIC (Table III). The size peak of the flagellates is close to that of the prymnesiophytes, at approximately ESD = 2.9 µm. Although both fill a similar narrow niche in ESD, the abundance of heterotrophic flagellates decreased markedly at the same times that prymnesiophytes increased, suggesting that grazing on heterotrophic flagellates could be preferred to grazing on prymnesiophytes. Alternatively, both groups could have experienced similar grazing pressure, but the specific growth rate of the prymnesiophytes may have exceeded their loss to grazing.

Dinoflagellates are the second most abundant group among the heterotrophs. After iron fertilization, their total cell concentration decreased slightly, but larger cells (ESD $\geq$ 8 µm) became more abundant at the same time (Fig. 4B). The size spectrum does not reveal distinct peaks but is right-skewed. The density peak is located around ESD = 5 µm, which is captured by the unimodal parameter fit. The AICs suggest a preference for the bimodal parametric fits, because they better resolve the right-skewness of the distributions. It is noteworthy that the dinoflagellate density maxima almost coincide with the local minimum (depression) seen in the size densities of photoautotrophic dinoflagellates and *Pseudo-nitzschia*.

The green dinoflagellates, although containing pigments from prokaryotic endosymbionts, are treated

Table II: Best parameter estimates for dominant phytoplankton groups, given unimodal and bimodal parametric probability density functions (pdf_1 and pdf_2)

Photo-autotrophic plankton	Location	Cell concentration	Parameter estimates							pdf comparison
			pdf_1		pdf_2					
		C^{tot}/Cells L^{-1}	a_1	b_1	q	a_1	b_1	a_2	b_2	$\Delta AIC = (AIC_2 - AIC_1)$
Prymnesiophytes	Patch	1.93×10^6	1.07	0.1	0.08	1.05	0.08	1.2	0.25	-1.2×10^3
	Control	1.26×10^6	1.03	0.08	0.04	1.03	0.08	1.3	0.21	-1.6×10^3
Pseudo-nitzschia	Patch	1.19×10^6	1.04	0.14	0.31	1.05	0.15	1.98	0.3	-2.4×10^5
	Control	5.29×10^4	1.02	0.18	0.09	1	0.16	1.6	0.58	-1.8×10^2
Photo-autotroph. dinoflagellates	Patch	7.73×10^5	1.2	0.12	0.23	1.19	0.12	1.87	0.29	-3.2×10^5
	Control	6.20×10^5	1.19	0.14	0.35	1.17	0.14	1.81	0.33	-5.0×10^4
Chaetoceros spp.	Patch	1.85×10^4	1.96	0.34	0.56	1.77	0.25	2.08	0.34	-3.0×10^2
	Control	8.07×10^3	1.61	0.28	0.07	1.59	0.28	2.04	0.1	4.8×10^0

Total cell concentrations are equal to the integral of the nonparametric densities of the original data. Differences in Akaike information criterion (unimodal, AIC_1; bimodal, AIC_2) compare the goodness of the parametric fits for each plankton group.

Table III: Best parameter estimates for dominant zooplankton groups, given unimodal and bimodal parametric probability density functions (pdf_1 and pdf_2)

Heterotrophic plankton	Location	Cell concentration	Parameter estimates							pdf comparison
			pdf_1		pdf_2					
		C^{tot}/Cells L^{-1}	a_1	b_1	q	a_1	b_1	a_2	b_2	$\Delta AIC = (AIC_2 - AIC_1)$
(Nano-) flagellates	Patch	2.00×10^6	1.09	0.08	0.04	1.08	0.07	1.22	0.33	1.8×10^3
	Control	2.37×10^6	1.06	0.06	0.02	1.03	0.07	1.37	0.25	-1.3×10^4
Dinoflagellates	Patch	3.82×10^4	1.92	0.42	0.64	1.61	0.2	2.03	0.45	-3.2×10^3
	Control	4.20×10^4	1.68	0.18	0.33	1.62	0.14	1.96	0.33	-3.0×10^2
Green dinoflagellates	Patch	3.02×10^4	1.52	0.28	0.79	1.21	0.13	1.88	0.35	-1.2×10^4
	Control	3.27×10^4	1.52	0.32	0.63	1.21	0.18	1.88	0.38	-3.3×10^2
Ciliates	Patch	2.48×10^3	2.57	0.38	0.17	2.46	0.32	3.07	0.26	-1.8×10^0

Total cell concentrations are equal to the integral of the nonparametric densities of the original data. Differences in Akaike information criterion (unimodal, AIC_1; bimodal, AIC_2) compare the goodness of the parametric fits for each plankton group.

here as heterotrophs. Both density estimates (patch and control) clearly show two modes, whose peaks are identical with those of the photoautotrophic dinoflagellates (Fig. 4C). The green dinoflagellates are less abundant than their photoautotrophic counterparts. Notably, their responses to iron fertilization can be interpreted as mirror images: cell abundance of green dinoflagellates decreases within the first peak of smaller sizes, while the second mode hardly changes; the opposite response is observed for the photoautotrophic dinoflagellates. The fit of the bimodal function to the nonparametric densities proved superior to a unimodal function, which can be inferred from the negative AIC values. As previously seen, the local density minimum occurs within the size range of 4 and 5 μm ESD.

Figure 4D depicts heterotrophic ciliates in the larger size range, where they appear in small numbers. Density estimates could not be made outside the fertilized patch because there were too few data points with independent length and width measurements. Consequently, we could only infer a density estimate from observations made within the fertilized patch. The increase in abundance represents a large fraction of the biovolume increase of the microbial community. An approximately 3-fold increase in ciliate biovolume is reported in Landry *et al.* (Landry *et al.*, 2000a), who suggested a substantial increase in grazing on smaller plankton following fertilization. A unimodal density function with $a_1 = 2.57$ and $b_1 = 0.38$ (peak at ESD = 13 μm) suffices to describe the nonparametric density estimate. No improvements are achieved when optimizing the additional three parameters of the bimodal function.

Composite density spectra

In Figure 5, we depict the combined nonparametric density estimates for all plankton groups, distinguishing photoautotrophs from heterotrophs. These composite density spectra cover a wide size range and represent the response of the microplankton community of the

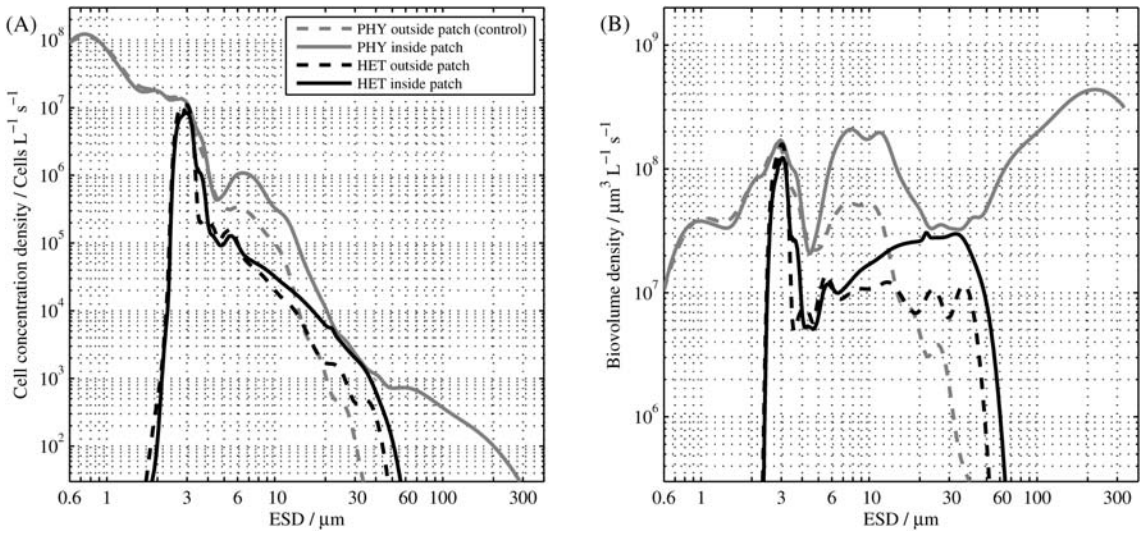

Fig. 5. Composite size-spectra: Plankton densities of the photo-autotrophic and heterotrophic plankton communities were obtained by summing corresponding mean nonparametric density estimates, respectively. (**A**) Cell concentration densities inside (solid lines) and outside (dashed lines) the iron fertilized patch. (**B**) Corresponding densities of biovolume (as indicator for biomass) inside (solid lines) and outside (dashed lines) the fertilized patch.

eastern equatorial Pacific to the IronEx II fertilization. Phytoplankton cell densities (Fig. 5A) show that the patch and control distributions diverge about ESD = 5 µm, with patch abundance rising to a distinct peak around 6.5 µm. Cells in the picophytoplankton size range (<2 µm) hardly change in abundance, but the density slightly declines (Supplementary Material), most likely due to stimulated grazing pressure of microzooplankton in the fertilized patch (Landry *et al.*, 2000b). Cells in the size range of 3–5 µm show no noticeable density difference between the patch and control, but sizes larger than ESD = 5 µm in the patch clearly exceed the control concentrations. Heterotrophic dinoflagellates and ciliates seem to benefit from the increased abundance of prey larger than 5 µm, and a significant increase in microzooplankton cell density is evident at ESD = 7 µm. We interpret this cautiously as indicating an approximate predator–prey size ratio of 1.4 (i.e. 7:5 µm), which is roughly consistent with the ~1:1 ratio expected for dinoflagellate predators (Hansen *et al.*, 1994). It is interesting to see how cell densities of heterotrophic dinoflagellates and ciliates converge around ESD = 15 µm, with ciliates reaching the approximate 2×10^3 cells L^{-1} abundance of dinoflagellates. Large diatoms, mainly *Rhizosolenia* and *Thalassiotrix* spp., appear in significant numbers only within the fertilized patch. These species, with ESDs larger than 30 µm, likely escape microzooplankton grazing. Although large diatoms appear only in small numbers, their contribution to biovolume is substantial (Fig. 5B).

DISCUSSION

Application of kernel density estimator for the analysis of plankton size

Nonparametric kernel density estimation is a widely accepted statistical method in many research fields, including astronomy (e.g. Chen, 1996; Carpenter, 2000), economics (e.g. Aït-Sahalia and Lo, 1998; Herzfeld and Weiss, 2007) and biomedical research (e.g. Aitchison and Aitken, 1976; Gasser *et al*, 1984). In marine and aquatic sciences, however, kernel density estimation has seldom been applied. Sanvicente-Añorve *et al.* (Sanvicente-Añorve *et al.*, 2003) were first to promote the method for investigations of length-frequency distributions of fish larvae. Their motivation was to allow better characterization of skewness and multimodality in their distributions, which can be difficult to see in scattered, discontinuous histograms. Quintana *et al.* (Quintana *et al.*, 2008) provided in-depth comparison of parametric and nonparametric density estimation methods and assumptions for deriving a robust size-diversity model based on the Shannon index. They found nonparametric methods, kernel density estimators in particular, to be superior to fitting parametric functions to size and weight data. Similar to our composite density estimates (Fig. 5), they found distributional patterns that can hardly be represented by a single generic parametric function, like a Pareto-based biomass-size density (Vidondo *et al.*, 1997). Quintana *et al.* (Quintana *et al.*, 2008) also stressed that most parametric

functions that represent the entire plankton community size-structure have no true ecological meaning, apart from exponential-shaped pdfs.

A study by Havlicek and Carpenter (Havlicek and Carpenter, 2001) compared plankton size-structure patterns among lakes, constructing size distributions of phytoplankton, zooplankton and fish by summing over a series of Gaussian density functions. Although they referred to their procedure as kernel density estimation, it differed from the conventional approach by imposing parametric pdfs. In our study, we followed Sanvincente-Añorve *et al.* (Sanvincente-Añorve *et al.*, 2003) and Quintana *et al.* (Quintana *et al.*, 2008) in using kernel density estimation as a tool to derive nonparametric density estimates. Our analyses are therefore fundamentally different from those of Havlicek and Carpenter (Havlicek and Carpenter, 2001) with respect to prior assumptions. Havlicek and Carpenter (Havlicek and Carpenter, 2001) assumed that plankton size densities were normally distributed on a linear scale and that each bandwidth of their density function was equivalent to the standard deviation of size measurements of each plankton group. From our analysis, we learned that such *a priori* assumptions are not justified because asymmetries are apparent within a taxonomic plankton group. In contrast to Havlicek and Carpenter (Havlicek and Carpenter, 2001), we avoided prior assumptions of plankton size distributions by using nonparametric size densities of individual plankton groups, which were then added together to yield composite densities of heterotrophs and photoautotrophs. Such an approach was outlined by Quintana *et al.* (Quintana *et al.*, 2008) to improve the investigation of size diversity. Subsequent analyses of our composite densities may therefore help address the issue of size-diversity measures.

In our study, we imposed constant bandwidths on each plankton group analysed. We also tested an approach where bandwidths of variable size were determined (data not shown) by applying the algorithm for adaptive bandwidth selection of Brewer (Brewer, 2000). This algorithm relies on Bayesian statistical considerations in that it explicitly considers conditional dependencies between neighbouring kernels. A data point-specific bandwidth was obtained for every kernel (of each s_{ij}) after multiplying the constant bandwidth of equation (7) by a correction factor. Hence, the local (or data-point-specific) correction factor induced bandwidths smaller than the constant one [equation (7)] in size ranges where data points are close together, thus allowing better density resolution. In contrast, the correction factor leads to larger bandwidths for data points that are widespread. We found that adaptive bandwidth selection has the benefit of achieving higher resolution within certain size ranges, but can be disproportionately affected by outliers at the upper and lower bounds of size-density spectra. Large bandwidths were obtained for minimum and maximum ESDs of individual plankton groups, which artificially enhanced the appearance of leptokurtic kurtosis in our densities. In addition, our resampling strategy tended to mask most of the benefit of higher resolution, while the outlier effects remained. Algorithms for adaptive bandwidth selection are intuitively appealing, but we conclude that a more detailed methodological analysis would be needed to specify conditions where the benefits of adaptive bandwidth selection prevail. Given the large numbers of data points in our study, we restricted ourselves to constant bandwidths. Optimal constant bandwidths were much easier to calculate, which was particularly desirable when used in conjunction with the bootstrapping approach.

Choosing and fitting parametric size-density functions

Through visual inspection of the nonparametric density estimates on logarithmic scales of ESD, we identified three distinct distributional patterns: symmetric unimodal, right skewed and bimodal. Using a unimodal normal distribution function to fit the symmetrical patterns seemed an obvious choice. Additional degrees of freedom were introduced when imposing a bimodal function, which were expected to allow better representation of right-skewed and bimodal size densities. In addition, we learned from our analysis that estimating the three additional parameters (q, a_2, b_2) sometimes also resolves kurtosis better (e.g. in heterotrophic dinoflagellates), which is indicated by small estimates of q and larger values of b_2 than of b_1. This better representation of kurtosis is thus a side effect of fitting the bimodal parametric function to some density estimates. The existence of leptokurtic kurtosis, for example, in the distributions of heterotrophic- and green dinoflagellates, appears to be robust. Right-skewed and leptokurtic distributions of natural nano- and microplankton communities were similarly expressed in the data analysed by Sabetta *et al.* (Sabetta *et al.*, 2005), who proposed using Gamma functions to represent these distributional patterns.

Campbell and Yentsch (Campbell and Yentsch, 1989) related synchronous and asynchronous cell growth to specific size-distribution patterns. Interestingly, their computations revealed skewed and bimodal densities for an asynchronous cell population, where cell growth is restricted to a short period in the middle of the cellular

life cycle. When deriving parametric size density functions for plankton, it is worthwhile considering causal links between population growth, life cycles and variations of size density. Similarly, distribution functions that combine cell cycle information with "random" Gaussian noise were discussed in Armstrong (Armstrong, 2003a, b). In the modelling study of Litchman *et al.* (Litchman *et al.*, 2009), evolutionary stable strategy sizes of plankton were investigated in response to varying pulse periods of nitrogen supply. For intermediate pulse periods (between 30 and 40 days) of nitrogen supply, their model yields coexistence of small and larger cells, which generates bimodal size distribution patterns. According to these model predictions, bimodal patterns can be sustained simply by balancing the allometric dependencies of maximum growth rate and nutrient-storage ability. These model findings allow for a bottom-up interpretation of the bimodalities seen in plankton size densities.

An alternative top-down interpretation of deviations from normality can be based on predator impacts: densities might be unimodal in the absence of grazing, with deviations from normality imposed by extensive grazing within a particular size range. This perspective is inspired by the existence of a persistent depression in densities within the iron-fertilized patch between 4 and 5 μm ESD (Figs 3 and 4). Furthermore, disproportionate grazing losses of cells, in the 4–5 μm size range, could explain the bimodal structures depicted for *Pseudo-nitzschia* (also autotrophic dinoflagellates and heterotophic green dinoflagellates), as well as the densities of *Chaetoceros* and heterotrophic dinoflagellates. Rather than a "grazing induced" separation into two modes, the asymmetry of the unimodal, but right-skewed, size densities could be due to disproportionate grazing on smaller cells on the left-hand side of the distributions.

Figure 6 depicts bimodal densities of photoautotrophic and heterotrophic plankton density estimates together with parametric fits that consider a size preference for grazing (φ) and a corresponding width for the grazing impact (κ), which remain constant. The distribution of size-selective grazing acts on all plankton groups that fall within the size range of interest, with only the magnitude of the grazing impact varying among the different plankton groups. The parametric function used can be interpreted as a unimodal density function (pdf_1) that is multiplied by a grazing filter (ϕ): $\hat{f} = \mathrm{pdf}_1 \cdot \phi$, with $\phi = A/(\kappa \cdot \sqrt{2\pi}) \cdot [1 - B \cdot \exp\{(s-\varphi)^2/\kappa^2\}]$. The parameter A relates the amplitude of the unimodal pdf_1 to the maximum grazing impact (B). Such parametric functions cannot be fit to represent the individual densities within the confidence limits. However, the general characteristics of the parametric fits provide support for a top-down explanation of bimodality in the size range between 2 and 20 μm ESD. The qualitatively good, but statistically non-significant fits of this parametric approach could result from morphological differences of plankton prey that remain unresolved when normalizing to ESD. To resolve this issue, we may need a dynamical modelling approach, where convolutions of a grazing kernel function with individual (unimodal) size densities of plankton groups or species are explicitly considered, similar to the procedure discussed by Armstrong (Armstrong, 2003a, b).

Resolving details in plankton size structure and implications for ecological modelling

With the recognized importance of organism size as an organizing trait for plankton community structure, trait-based modelling approaches (Litchman and Klausmeier, 2008) seem likely to improve descriptions of plankton interactions. Specification of trade-offs among different traits (like size) is important and can link ecosystem function to biogeochemical pathways of organic matter flux (e.g. Verdy *et al.*, 2009). For example, trade-offs between nutrient uptake affinity and predation defence (Merico *et al.*, 2009) could be expressed in terms of size-dependencies, including aspects of cellular or metabolic size scaling (e.g. Irwin *et al.*, 2006; Litchman *et al.*, 2007) and adaptation of mean or optimum cell-size (Jiang *et al.*, 2005; Yoshiyama and Klausmeier, 2008). Given these theoretical prospects together with new modelling approaches, the availability of high-resolution data of plankton community structure (in terms of size and taxonomic differentiations; Armstrong, 2003b) is of crucial importance. The degree to which details can be resolved depends primarily on the quantity and quality of available data. To produce these data, rigorous procedures for statistical treatment of size measurements are needed. Plankton size data are usually available as histograms that vary in bin width and point of origin. Analysing size data from different investigators or experiments is difficult, as one needs to combine a series of differently binned data. If continuous densities of plankton size are derived using kernel estimation methods, combining them is straightforward. With the bootstrapping approach, as applied in our study, it should be possible to reanalyse historical binned size data and provide continuous density estimates together with confidence limits. This procedure would facilitate inter-comparisons among different data sets and support investigations of species interactions while also providing constraints for size-based plankton models.

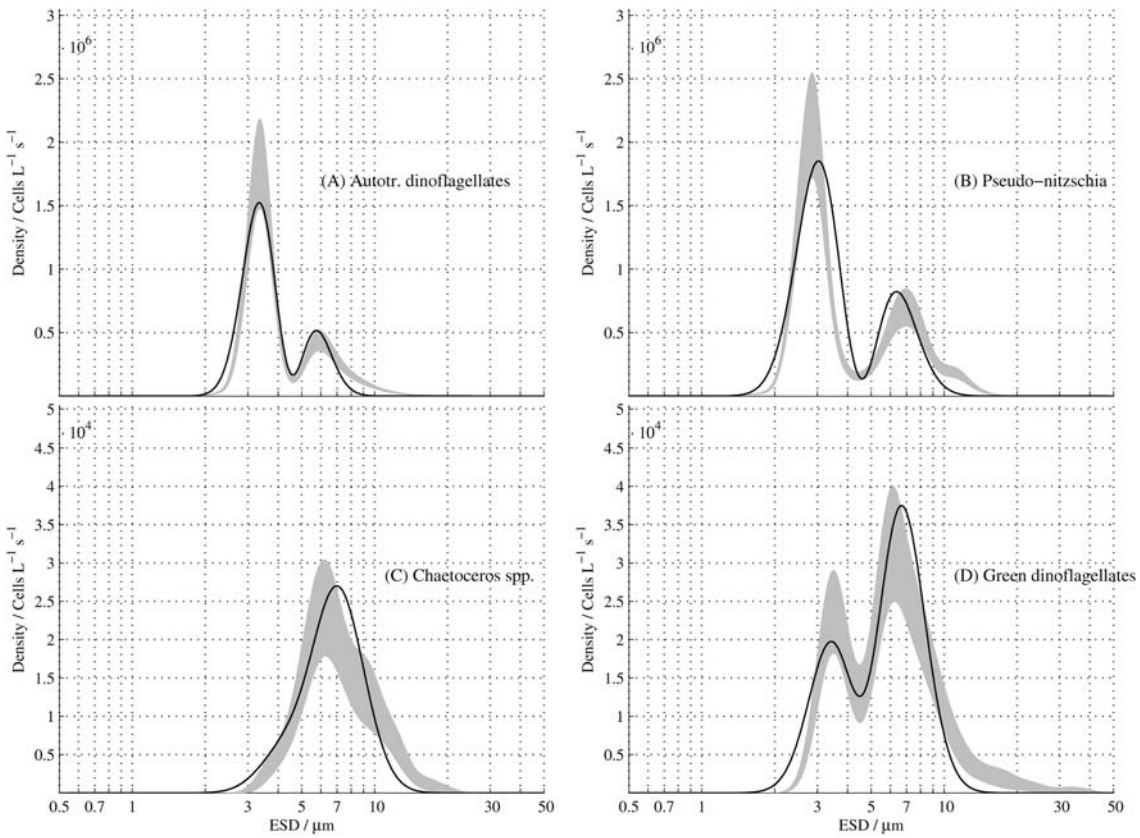

Fig. 6. Nonparametric density estimates and their 95% confidence limits (shaded). Parametric fit with a unimodel density function (pdf_1) multiplied by a grazing distribution ϕ): $\hat{f} = pdf_1 \cdot \phi$,, with $\phi = A/(\kappa \cdot \sqrt{2\pi}) \cdot [1 - B \cdot \exp\{(s-\varphi)^2/\kappa^2\}]$. The grazing distribution consists of a preference size for grazing ($\phi = 1.52$) and a corresponding width for the grazing impact ($\kappa = 0.88$), which remain constant for all plankton groups: (**A**) Photoautotrophic dinoflagellates (with $a_1 = 1.44$, $b_1 = 0.2$, $A = 130$, $B = 0.998$), (**B**) *Pseudo-nitzscha* ($a_1 = 1.44$, $b_1 = 0.27$, $A = 90$, $B = 0.998$), (**C**) *Chaetoceros* spp. ($a_1 = 1.75$, $b_1 = 0.27$, $A = 30$, $B = 0.96$) and (**D**) green dinoflagellates ($a_1 = 1.59$, $b_1 = 0.27$, $A = 65$, $B = 0.99$).

The response of plankton community structure to iron addition during IronEx II can be explained by a combination of bottom-up and top-down effects (Armstrong, 1994; Price *et al.*, 1994). The natural plankton community of the eastern equatorial Pacific is subject to a tight control inherent to the microbial loop, where growth, grazing and remineralization are strongly coupled (Pomeroy, 1974; Azam *et al.*, 1983). The bloom develops as a disturbance of the balance between utilization (growth uptake) of the limiting trace element, iron and its regeneration via microzooplankton predation. The composite community size distributions reveal a narrow depression between ESD = 4 and 5 µm, which is close to the peaks in densities of *Chaetoceros* and heterotrophic dinoflagellates outside of the fertilized patch. We speculate that the natural zooplankton community outside the fertilized patch maintains a grazing preference for cells in this size range. Cell concentrations in this range declined after iron addition, whereas abundance of slightly larger cells increased, producing a shift in the peak of *Chaetoceros* towards larger cells and enhancing the skewness in the distribution of heterotrophic dinoflagellates. At the same time, cell concentrations of *Pseudo-nitzschia* and autotrophic dinoflagellates increased; but since their size variance includes larger cells that could escape extensive grazing, their size densities expressed a clear bimodality.

From our analysis, it appears that the 4–5 µm ESD size range may represent a sensitive link of predator–prey interaction. At approximately 5-µm ESD, there is a step-like increase in densities of larger cells in the fertilized patch relative to the control (Fig. 5). The slight increase in abundance of photoautotrophic cells smaller than 4 µm ESD has little impact on biovolume, and thus on biomass. The sensitive size ranges in the proximity of $\geq$5 µm and >40 µm ESD can also be interpreted in terms of the "loophole" concept (Bakun and Broad, 2003; Irigoien *et al.*, 2005), in which phytoplankton growth exceeds predation, allowing escape from the tightly linked predator–prey interactions of the microbial food web. However, the loophole concept does not adequately account for bottom-up, allometric size-scaling

effects discussed above. Assuming that mineralization of organic matter and the regeneration of organic-iron compounds regulate the stability and efficiency of the microbial loop, a characteristic size-density pattern of the plankton community is established, as observed at the control stations. The recent model of Yoshiyama and Klausmeier (Yoshiyama and Klausmeier, 2008) makes the point, for example, that size dependencies extend to the level of the molecules of dissolved organic matter that serves as resources for microbial consumers. Ultimately, the balance between bottom-up and top-down size-dependent processes needs to be explored using systems models that relate directly to the patterns of size variability observed in natural systems, both in their base states and in response to perturbations such as iron fertilization. For such modelling efforts, rigorous methods to develop, analyse and synthesize size-based assessments of plankton community structure are necessary.

SUMMARY AND CONCLUSIONS

Using cell counts and size data from IronEx II, we derived nonparametric characterizations of size-frequency distributions for several photoautotrophic and heterotrophic plankton groups. Symmetric nonparametric density estimates were described well by unimodal parametric density functions. Asymmetries, such as right-skewed and bimodal densities, as seen for example in the density spectra of *Pseudo-nitzschia* and *Chaetoceros*, could be resolved using bimodal parametric functions. Size-based analysis of the nonparametric densities revealed a significant depression among photoautotrophic and heterotrophic plankton groups in the range of 4–5 μm ESD. Since this characteristic is expressed for both autotrophs and heterotrophs, it seems more likely to be associated with size-selective grazing than with bottom-up effects on primary producers. By imposing parametric functions that represent a size-specific grazing filter, we found qualitatively good agreement to the bimodal nonparametric densities.

The issue of deriving parametric functions for predator–prey relationships needs to be addressed further. As seen for the size densities of individual plankton groups, the critical size range 4–5 μm ESD appears relatively invariant inside and outside the fertilized patch, while the biovolume of smaller cells is altered little by iron addition. Those plankton groups or species with cells larger than 5-μm ESD yield the major fraction of net biovolume increase after fertilization, with heterotrophs showing a size-shifted response and an implicit mean predator:prey size ratio suggestive of grazing by dinoflagellates.

In this study, we have shown that kernel density estimation is well suited to representing important aspects of plankton community structure. The smoothed nonparametric densities are ideal for subsequent analyses of spectra to obtain size functions for growth rates and grazing mortalities, as suggested for example by Zhou (Zhou, 2006). Since most density estimates yield symmetric distributions outside the fertilized patch, we conclude that the undisturbed microbial plankton community can be well described by unimodal Gaussian functions. Asymmetries are clearly expressed after fertilization. Only by means of dynamical modelling can we fully resolve the issues of bottom-up and top-down control of the plankton community (Armstrong, 1994; Price *et al.*, 1994).

In addition to their use in modelling, kernel density estimates of plankton size can contribute to intercomparisons among independent studies of plankton community structure and dynamics on different spatial and temporal scales. Such activities are currently hindered by the multiplicity of schemes and scales for size binning in histograms. Use of the bootstrapping approach, in conjunction with the kernel density estimation, provides a statistical analysis procedure that can be utilized to derive continuous size-densities of plankton from historical, binned data.

SUPPLEMENTARY DATA

Supplementary data can be found online at http://plankt.oxfordjournals.org.

ACKNOWLEDGEMENTS

We thank Sara Tanner and Susan Brown for their efforts and care in producing the plankton community data set for IronEx II. We also thank Carsten Lemmen for assistance during the implementation of C-code into MATLAB Mex-files. We are thankful for comments by Darcy Taniguchi. We greatly appreciate supportive and constructive comments made by four anonymous reviewers.

FUNDING

This research was conducted as part of the NSF grants OCE 03-24666, OCE 99-11765 and OCE 04-17616. The reanalysis was partially funded by the Helmholtz Association (Germany) PACES programme 'Polar regions and Coasts in a changing Earth System'.

REFERENCES

Aitchison, J. and Aitken, C.G.C. (1976) Multivariate Binary Discrimination by the Kernel Method. *Biometrika*, **63**, 413–420.

Aït-Sahalia, Y. and Lo, A. W. (1998) Nonparametric estimation of state-price densities implicit in financial asset prices. *J. Finance*, **53**, 499–547.

Armstrong, R. A. (2003a) A hybrid spectral representation of phytoplankton growth and zooplankton response: the "control rod" model of plankton interaction. *Deep-Sea Res. II*, **50**, 2895–2916.

Armstrong, R. A. (2003b) Representing biogeochemical diversity and size spectra in ecosystem models of the ocean carbon cycle. In Canham, C. D., Cole, J. and Laurenroth, W. K. (eds.), *Models in Ecosystem Science*. Princeton University Press, Princeton, NJ, pp. 254–271.

Armstrong, R. A. (1994) Grazing limitation and nutrient limitation in marine ecosystems: steady state solutions of an ecosystem model with multiple food chains. *Limnol. Oceanogr.*, **39**, 597–608.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A. and Thingstad, F. (1983) The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol. Prog. Ser.*, **10**, 257–263.

Azam, F. and Malfatti, F. (2007) Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, **5**, 782–791. doi:10.1038/nrmicro1747.

Bakun, A. and Broad, K. (2003) Environmental 'loopholes' and fish population dynamics: comparative pattern recognition with focus on El Niño effects in the Pacific. *Fish. Oceanogr.*, **12**, 458–473.

Blanchot, J., Andre, J.-M., Navarette, C. *et al.* (2001) Picophytoplankton in the equatorial Pacific: vertical distributions in the warm pool and in the high nutrient low chlorophyll conditions. *Deep-Sea Res. I*, **48**, 297–314.

Brewer, M. (2000) A Bayesian model for local smoothing in kernel density estimation. *Stat. Comput.*, **10**, 299–309.

Calbet, A. and Landry, M. R. (2004) Phytoplankton growth, microzooplankton grazing and carbon cycling in marine systems. *Limnol. Oceanogr.*, **49**, 51–57.

Campbell, J. W. and Yentsch, C. M. (1989) Variance Within Populations, I: Homogeneous Phytoplankton Theoretical Framework for Interpreting Histograms. *Cytometry*, **10**, 587–595.

Carpenter, J. M. (2000) 2mass observations of the perseus, orion a, orion b, and monoceros r2 molecular clouds. *Astron. J.*, **120**, 3139–3161.

Cavender-Bares, K. K., Rinaldo, A. and Chisholm, S. W. (2001) Microbial size spectra from natural and nutrient enriched ecosystems. *Limnol. Oceanogr.*, **46**, 778–789.

Chavez, F. P., Buck, K. R., Coale, K. H. *et al.* . (1991) Growth rates, grazing, sinking, and iron limitation of equatorial Pacific phytoplankton. *Limnol. Oceanogr.*, **36**, 1816–1833.

Chen, B. (1996) A non-parameter global method of interpreting star count data. *Astronomy & Astrophysics Supplement Series*, **118**, 181–189.

Coale, K. H., Johnson, K. S., Fitzwater, S. E. *et al.* (1996) A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization in the equatorial Pacific. *Nature*, **383**, 495–501.

Efron, B. and Tibshirani, R. J. (1993) *An introduction to the bootstrap*. Chapman & Hall, 436 pp.

Gasser, Th., Müller, H. G., Köhler, W. *et al.* (1984) Nonparametric regression analysis of growth curves. *Ann. Stat.*, **12**, 210–229.

Gaedke, U. (1993) Ecosystem analysis based on biomass size distributions: a case study of a plankton community in a large lake. *Limnol. Oceanogr.*, **38**, 112–127.

Härdle, W., Müller, M., Sperlich, S. *et al.* (2004) *Nonparametric and Semiparametric Models*. Springer Series in Statistics, Springer-Verlag, pp. 340.

Hansen, B., Bjørnsen, P. K. and Hansen, P. J. (1994) The size ratio between planktonic predators and their prey. *Limnol. Oceanogr.*, **39**, 395–403.

Havlicek, T. D. and Carpenter, S. R. (2001) Pelagic species size distributions in lakes: Are they discontinuous? *Limnol. Oceanogr.*, **46/5**, 1021–1033.

Herzfeld, T. and Weiss, C. (2007) Corruption clubs: empirical evidence from kernel density estimates. *Appl. Econ.*, **39/12**, 1565–1572.

Irigoien, X., Flynn, K. J. and Harris, R. P. (2005) Phytoplankton blooms: a 'loophole' in microzooplankton grazing impact? *J. Plankton Res.*, **27/4**, 313–321.

Irwin, A. J., Finkel, Z. V., Schofield, O. M. E. *et al.* (2006) Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. *J. Plankton Res.*, **28/5**, 459–471.

Jiang, L., Schofield, O. M. E. and Falkowski, P. G. (2005) Adaptive evolution of phytoplankton cell size. *Am. Nat.*, **166/4**, 496–505.

Kerr, S. R. (1974) Theory of size distribution in ecological communities. *J. Fish. Res. Board Can*, **31**, 1859–1862.

Kerr, S. R. (1979) Prey availability, metaphoetesis, and the size structures of lake trout stocks. *Invest. Pesq.*, **43**, 187–198.

Landry, M. R., Ondrusek, M. E., Tanner, S. J. *et al.* (2000a) Biological response to iron fertilization in the eastern equatorial Pacific (IronEx II). I. Microplankton community abundances and biomass. *Mar. Ecol. Prog. Ser.*, **201**, 27–42.

Landry, M. R., Constantinou, J., Latasa, M. *et al.* (2000b) Biological response to iron fertilization in the eastern equatorial Pacific (IronEx II). III. Dynamics of phytoplankton growth and microzooplankton grazing. *Mar. Ecol. Prog. Ser.*, **201**, 57–72.

Litchman, E. and Klausmeier, C. A. (2008) Trait-based community ecology of phytoplankton. *Ann. Rev. Ecol. Evol. Syst.*, **39**, 615–639.

Litchman, E., Klausmeier, C. A., Schofield, O. M. *et al.* (2007) The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. *Ecol. Lett.*, **10**, 1170–1182.

Litchman, E., Klausmeier, C. A. and Yoshiyama, K. (2009) Contrasting size evolution in marine and freshwater diatoms. *Proc. Natl Acad. Sci.*, **106**, 2665–2670.

Merico, A., Bruggeman, J. and Wirtz, K. (2009) A trait-based approach for downscaling complexity in plankton ecosystem models. *Ecol. Model.*, **220**, 3001–3010.

Moloney, C. L. and Field, J. G. (1989) General allometric equations for rates of nutrient uptake, ingestion, and respiration in plankton organisms. *Limnol. Oceanogr.*, **34**, 1290–1299.

Parzen, E. (1962) On estimation of a probability function and mode. *Ann. Math. Stat.*, **33/3**, 1065–1076.

Platt, T. and Denman, K. (1977) Organisation in the pelagic ecosystem. *Helgoland Mar. Res.*, **30**, 575–581.

Price, N. M., Ahner, B. A. and Morel, F. M. M. (1994) The equatorial Pacific Ocean: grazer-controlled phytoplankton populations in an iron-limited system. *Limnol. Oceanogr.*, **39**, 520–534.

Pomeroy, L. R. (1974) The ocean's food web, A changing paradigm. *Bioscience*, **24**, 499–504.

Quintana, X. D., Brucet, S., Boix, D. *et al.* (2008) A nonparametric method for the measurement of size diversity with emphasis on data standardization. *Limnol. Oceanogr. Methods*, **6**, 75–86.

Reul, A., Rodriguez, J., Blanco, J. M. *et al.* (2006) Control of microplankton size structure in contrasting water columns of the Celtic Sea. *J. Plankton Res.*, **28**, 449–457.

Rinaldo, A., Maritan, A., Cavender-Bares, K. K. *et al.* (2002) Cross-scale ecological dynamics and microbial size spectra in marine ecosystems. *Proc. R. Soc. Lond. Ser. B*, **269**, 2051–2059.

Rodriguez, J. and Mullin, M. M. (1986) Relation between biomass and body weight of plankton in a steady oceanic ecosystem. *Limnol. Oceanogr.*, **31**, 361–370.

Rykaczewski, R. R. and Checkley, D. M. (2008) Influence of ocean winds on the pelagic ecosystem in, upwelling regions. *Proc. Natl Acad. Sci. USA*, **105**, 1965–1970.

Sabetta, L., Fiocca, A., Margheriti, L. *et al.* (2005) Body size-abundance distributions of nano- and micro-phytoplankton guilds in coastal marine ecosystems. *Estuarine, Coastal and Shelf Science*, **63**, 645–663.

San Martin, E. S., Harris, R. P. and Irigoien, X. (2006a) Latitudinal variation in plankton size spectra in the Atlantic Ocean. *Deep-Sea Res. II*, **53**, 1560–1572.

San Martin, E. S., Irigoien, X., Harris, R. P. *et al.* (2006b) Variation in the transfer of energy in marine plankton along a productivity gradient in the Atlantic Ocean. *Limnol. Oceanogr.*, **51**, 2084–2091.

Sanvicente-Añorve, L., Salgado-Ugarte, I. and Castillo-Rivera, M. (2003) The use of kernel density estimators to analyze length-frequency distributions of fish larvae. In Browman, H. I. and Skiftesvik, A. B. (eds), *The Big Fish Bang: Proceedings of the 26th Annual Larval Fish Conference*. Bergen, Norway, pp. 419–430.

Sheldon, R. W., Prakash, A. and Sutcliffe, W. H. Jr. (1972) Size distribution of particles in the ocean. *Limnol. Oceanogr.*, **17**, 327–340.

Silverman, B. W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Verdy, A., Follows, M. and Flierl, G. (2009) Optimal phytoplankton cell size in an allometric model. *Mar. Ecol. Prog. Ser.*, **379**, 1–12.

Vidondo, B., Prairie, Y. T., Blanco, J. M. *et al.* (1997) Some Aspects of the Analysis of Size Spectra in Aquatic Ecology. *Limnol. Oceanogr.*, **42**, 184–192.

Wegman, E. J. (1972) Nonparametric probability density estimation density estimation methods. *J. Stat. Comput. Simulat.*, **1**, 225–245.

Yoshiyama, K. and Klausmeier, C. A. (2008) Optimal cell size for resource uptake in fluids: a new facet of resource competition. *Am. Nat.*, **171**, 59–70.

Zhou, M. (2006) What determines the slope of a plankton biomass spectrum? *J. Plankton Res.*, **28**, 437–448.

Zhou, M. and Huntley, M. E. (1997) Population dynamics theory of plankton based on biomass spectra. *Mar. Ecol. Prog Ser.*, **159**, 61–73.