

MAREDAT: towards a world atlas of MARine Ecosystem DATa

Erik Buitenhuis, M. Vogt, R. Moriarty, N. Bednarsek, S.C. Doney, K. Leblanc, C. Le Quéré, Y.-W. Luo, C. O'Brien, T. O'Brien, et al.

▶ **To cite this version:**

Erik Buitenhuis, M. Vogt, R. Moriarty, N. Bednarsek, S.C. Doney, et al.. MAREDAT: towards a world atlas of MARine Ecosystem DATa. Earth System Science Data, 2013, 5, pp.227-239. 10.1594/PANGAEA.779970 . hal-00997771

HAL Id: hal-00997771
https://hal.science/hal-00997771

Submitted on 22 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Earth Syst. Sci. Data, 5, 227–239, 2013
www.earth-syst-sci-data.net/5/227/2013/
doi:10.5194/essd-5-227-2013
© Author(s) 2013. CC Attribution 3.0 License.



Open Access Earth System Science Data

MAREDAT: towards a world atlas of MARine Ecosystem DATa

E. T. Buitenhuis[1], M. Vogt[2], R. Moriarty[3], N. Bednaršek[4], S. C. Doney[5], K. Leblanc[6], C. Le Quéré[1], Y.-W. Luo[5], C. O'Brien[2], T. O'Brien[7], J. Peloquin[2], R. Schiebel[8], and C. Swan[2]

[1]Tyndall Centre for Climate Change Research and School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK
[2]Environmental Physics, Institute for Biogeochemistry and Pollutant Dynamics, ETH Zürich, Universitätsstrasse 16, 8092 Zürich, Switzerland
[3]School of Earth, Atmospheric and Environmental Sciences, University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL, UK
[4]NOAA Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115, USA
[5]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA
[6]Aix-Marseille Université, Université du Sud Toulon-Var, CNRS/INSU, IRD, MIO, UM110, 13288 Marseille, Cedex 09, France
[7]National Marine Fisheries Service, 1315 East–West Highway, Silver Spring, Maryland, USA
[8]LUNAM University, Angers University, UMR CNRS 6112 LPGN-BIAF, Laboratoire des Bio-Indicateurs Actuels et Fossiles, 2 Boulevard Lavoisier, 49045 Angers CEDEX 01, France

Correspondence to: E. Buitenhuis (tinyurl.com/contacterik)

Received: 13 November 2012 – Published in Earth Syst. Sci. Data Discuss.: 23 November 2012
Revised: 2 May 2013 – Accepted: 15 May 2013 – Published: 12 July 2013

Abstract. We present a summary of biomass data for 11 plankton functional types (PFTs) plus phytoplankton pigment data, compiled as part of the MARine Ecosystem biomass DATa (MAREDAT) initiative. The goal of the MAREDAT initiative is to provide, in due course, global gridded data products with coverage of all planktic components of the global ocean ecosystem. This special issue is the first step towards achieving this. The PFTs presented here include picophytoplankton, diazotrophs, coccolithophores, *Phaeocystis*, diatoms, picoheterotrophs, microzooplankton, foraminifers, mesozooplankton, pteropods and macrozooplankton. All variables have been gridded onto a World Ocean Atlas (WOA) grid ($1^\circ \times 1^\circ \times 33$ vertical levels $\times$ monthly climatologies). The results show that abundance is much better constrained than their carbon content/elemental composition, and coastal seas and other high productivity regions have much better coverage than the much larger volumes where biomass is relatively low. The data show that (1) the global total heterotrophic biomass (2.0–4.6 Pg C) is at least as high as the total autotrophic biomass (0.5–2.4 Pg C excluding nanophytoplankton and autotrophic dinoflagellates); (2) the biomass of zooplankton calcifiers (0.03–0.67 Pg C) is substantially higher than that of coccolithophores (0.001–0.03 Pg C); (3) patchiness of biomass distribution increases with organism size; and (4) although zooplankton biomass measurements below 200 m are rare, the limited measurements available suggest that *Bacteria* and *Archaea* are not the only important heterotrophs in the deep sea. More data will be needed to characterise ocean ecosystem functioning and associated biogeochemistry in the Southern Hemisphere and below 200 m. Future efforts to understand marine ecosystem composition and functioning will be helped both by further archiving of historical data and future sampling at new locations.

Microzooplankton database: doi:10.1594/PANGAEA.779970

All MAREDAT databases: http://www.pangaea.de/search?&q=maredat

Published by Copernicus Publications.

1 Introduction

The MARine Ecosystem Model Intercomparison Project (MAREMIP) was initiated in 2007 to facilitate communication, collaboration and the sharing of data and procedures, such as model evaluation techniques, between research groups developing Dynamic Green Ocean Models (DGOMs; Le Quéré et al., 2005). DGOMs are global ocean biogeochemical models that represent more than two plankton functional types (PFTs), thus including more ecological interactions than the unidirectional flow represented in Nutrient Phytoplankton Zooplankton Detritus (NPZD) models. After an exploratory phase, a kick-off meeting was held in 2009 (Le Quéré and Pésant, 2009). At this meeting it was decided to collectively synthesise existing biomass concentration measurements for the previously defined "key plankton functional types that need to be simulated explicitly to capture important biogeochemical processes in the ocean" (Le Quéré et al., 2005). This MARine Ecosystem biomass DATa (MAREDAT) special issue in Earth System Science Data is the result, with 11 papers on 11 PFTs and 1 paper on HPLC-based (high-performance liquid chromatography) phytoplankton pigments (http://www.earth-syst-sci-data.net/special_issue7.html). There are six papers relating to autotrophic groups: picophytoplankton (Buitenhuis et al., 2012a), diazotrophs (Luo et al., 2012), coccolithophores (O'Brien et al., 2013), *Phaeocystis* (Vogt et al., 2012), diatoms (Leblanc et al., 2012), and HPLC-based phytoplankton pigments (Peloquin et al., 2013). There are six papers relating to heterotrophic groups: picoheterotrophs (*Bacteria* and *Archaea*, Buitenhuis et al., 2012b), microzooplankton (here, we briefly reiterate and correct the microzooplankton biomass database that was recently published by Buitenhuis et al., 2010), planktic foraminifers (Schiebel and Movellan, 2012), mesozooplankton (Moriarty et al., 2013), pteropods (Bednaršek et al., 2012), and macrozooplankton (Moriarty et al., 2013). By this collaborative effort we are able to provide global databases for 9 out of the 10 PFTs that were proposed by Le Quéré et al. (2005).

The missing PFT is mixed phytoplankton, which is mostly made up of autotrophic dinoflagellates and nanophytoplankton other than coccolithophores and *Phaeocystis*. Nanophytoplankton are a taxonomically diverse group of algae, including prymnesiophytes, chlorophytes, and cryptophytes, which are not consistently treated as a distinct group in the literature. Where the term nanophytoplankton is used in the literature it often includes members of PFTs that have been included in the data products in this special issue. Therefore, we excluded this PFT from the current collection of data, even though nanophytoplankton represent a significant part of the phytoplankton biomass. Demarcation issues between nanophytoplankton and the other PFTs will need to be resolved before we can complete the abundance based atlas of phytoplankton biomass. Chemotaxonomic interpretation of the HPLC-based phytoplankton pigment database described in this special issue (Peloquin et al., 2013) offers one pathway toward resolving mixed phytoplankton biomass distribution. However, the most reliable way to prevent double counting and achieve a consistent dataset would be to measure the biomass and cell size of all distinct phytoplankton groups in the same samples in transects that cross all ocean basins.

A similar demarcation issue occurs for the zooplankton. The sum of micro-, meso-, and macrozooplankton should represent the total zooplankton population. However, although small foraminifers are microzooplankton, they tend not to be included in microscopic counts. Likewise, pteropods fall partly in both meso- and macrozooplankton size classes. In macrozooplankton studies, the focus is usually taxon specific, at a variety of levels between phylum and species. The sum of all relevant phyla is rarely accounted for, making an accurate assessment of the total macrozooplankton biomass difficult, and thus leading to a risk of underestimation, as opposed to the risk of double counting in the case of foraminifers and pteropods.

In addition to the 9 PFTs mentioned above, we include data on two groups of zooplankton calcifiers, the calcite producing planktic foraminifers (Schiebel and Movellan, 2012) and the pteropods (Bednaršek et al., 2012), which include both shelled aragonite-producing species and naked species. These data should be valuable in complementing research into phytoplankton calcifiers, i.e. coccolithophores, and in addressing the biogeochemical cycling of calcium carbonate, and thus alkalinity and atmospheric CO_2. The diazotroph dataset contains both biomass estimates and nitrogen fixation rate data, which are useful to evaluate the ecological roles of diazotrophs or to quantify marine nitrogen fixation (Luo et al., 2013).

Since 1994, the World Ocean Atlas (WOA) has been synthesising interpolated global gridded climatological datasets of physical and chemical parameters (temperature, oxygen, nutrients, etc.). Initially, these datasets were annual averages, but increasingly they cover seasonal variations on a monthly basis in the surface ocean. WOA provides datasets that achieve data coverage across all grid points by interpolation. These data products are used extensively by global ocean modellers for initialisation and/or evaluation (e.g. Doney et al., 2009), in combination with biogeochemical datasets such as dissolved inorganic carbon, alkalinity, pCO_2, and DMS, which have been synthesised through initiatives like MAREDAT (e.g. the CARINA special issue in ESSD http://www.earth-syst-sci-data.net/special_issue2.html). The goal of MAREDAT is to provide, in due course, similar data products and coverage for all planktic components of the global ocean ecosystem, and this special issue is the first step towards achieving this. For ease of use we have gridded the biological variables onto the same grid as used for the WOA ($1^\circ \times 1^\circ \times 33$ vertical levels $\times$ monthly climatologies). Because of the large seasonal variability in biological components over most of the ocean, we have chosen to provide monthly files from the start. At this point, there is not yet

Table 1. Database characteristics.

		doi:10.1594/ PANGAEA.	Raw data		Chauvenet's criteria		Number of data on WOA grid	Number of 10° boxes with data[a]
			Total number	Excluded outliers	Standard deviation	Critical value (*zc*)		
Autotrophs	Picophytoplankton	777385	40 946	0	0.79	4.36	10 747	147
	Diazotrophs	774851	3849	1[b]	1.59[b]	3.66[b]	2280	110
	Coccolithophores	785092	11 503	1	0.92	4.03	5162	167
	Phaeocystis	779101	3526	0	1.28	3.72	392	63
	Diatoms	777384	91 704	151	1.15	4.54	3852	158
	Pigments	793246	35 634	ND	ND	ND	10 880	174
Heterotrophs	Picoheterotrophs	779142	39 766	0	0.44	4.37	9284	84
	Microzooplankton	779970	4282	0	0.91	3.67	2029	121
	Foraminifers	777386	1057	0	1.05	3.49	745	25
	Mesozooplankton	785501	153 163	0	1.13	4.65	42 245	458
	Pteropods	777387	15 134	0	1.52	4.13	7151	248
	Macrozooplankton	777398	36 237	32	1.93	4.29	8147	187

[a] Out of 539 ocean boxes.
[b] The Chauvenet's criterion for the diazotroph database was calculated separately for the different data sources, but here we calculate one value for the standard deviation and *zc* for comparison to the other databases.

enough information to furnish filled (interpolated) datasets. Hence, we did not interpolate the data but produced datasets with missing values. Our aim in bringing together these data has been to (1) stimulate research on observation-based improvements in our knowledge of the ecological and biogeochemical functioning of the ocean and (2) provide in situ-based data constraints for numerical models and satellite algorithms that distinguish multiple plankton groups.

2 Data

Both the raw data files and the gridded files that were compiled for the 11 PFTs and the phytoplankton pigments are publicly available at the PANGAEA Data Publisher for Earth and Environmental Science (http://www.pangaea.de/search?&q=maredat) and the MAREDAT website (http://www.maredat.info). The latter also includes a technical corner with the gridding program, QC script, etc. Each of the 12 datasets can be cited with the DOIs reported in Table 1. The raw data include longitude, latitude, depth, date, abundance, biomass conversion factors, biomass, references and quality control flags. The pigment database contains pigment concentrations instead of abundance, biomass and biomass conversion factors. The *Phaeocystis* database contains biomass information with and without an estimate of the gelatinous mucus surrounding colonial cells. The diazotroph database includes biomass information based on cell counts and on *nif*H gene counts. The biomass concentrations of coccolithophores and foraminifers are organic carbon only, while the pteropod biomass concentrations include both organic and inorganic carbon. In this paper, we only use the *Phaeocystis* biomass without mucus, and the diazotroph biomass based on cell counts.

Furthermore, we provide a database of microzooplankton distribution that was recently published by Buitenhuis et al. (2010). Some errors in the previous version (C. Stock, personal communication, 2011) have been corrected ($n = 682$), and we also took this opportunity to grid this dataset in the same way as the others (doi:10.1594/PANGAEA.779970).

2.1 Quality control by Chauvenet's criterion

In all contributing PFT papers, Chauvenet's criterion has been used to exclude very high values from the gridded databases (Glover et al., 2011). While all data had already been quality controlled by the contributing researchers, there was still a risk of over-representation of high values from (1) studies that specifically targeted the occurrence/bloom of the relevant PFT, (2) high productivity coastal sites or (3) errors in the reported units. Chauvenet's criterion assumes that the data has a normal distribution and rejects data on the presumption that if a set of n measurements were carried out twice, outliers would be excluded at $1/(2n)$ probability of occurrence, thus preventing any bias between the outlier rejection in the two sets of measurements. The critical value occurs at $p = 1 - 1/2n$. None of the datasets had a normal distribution, therefore Chauvenet's criterion was performed on log-transformed data. Log-transformation of biomass values meant the exclusion of zero values. Zero and very low values for biomass are a true reflection of the ecology of the oceans. Zero abundance or biomass is often treated as a lack of measurement (i.e. represented by a blank entry rather than a zero concentration), even though it supplies useful information about the absence of an organism or PFT. This means that zero values are usually under-represented, especially in the deep sea. Therefore, a one-sided Chauvenet's criterion was applied to identify only the high value outliers. For the

smallest dataset (foraminifers, $n = 1057$ non-zero observations) the critical value was 3.50 times the standard deviation from the average, while for the largest dataset (mesozooplankton, $n = 153\,163$) it was 4.65 times the standard deviation. The pigment database was subjected to a different set of quality control procedures as described by Peloquin et al. (2013).

2.2 Gridding

In order to arrive at a consistent collection of gridded data products for all abundance based PFTs, they were all gridded with the same program, and include the number of observations, average abundance, average biomass, median abundance, median biomass, standard deviation of abundance, and standard deviation of biomass, both for the total datasets and for the non-zero observations. In some datasets, some of this information was excluded for methodological reasons: there is no abundance for mesozooplankton, and the databases for picoheterotrophs, diatoms and mesozooplankton contain no zero observations. All PFT and pigment data were gridded on a $1° \times 1°$ horizontal grid, with grid box centres from 179.5° W to 179.5° E and 89.5° S to 89.5° N. The vertical axis also follows the WOA spacing, centred on the 33 depths: 0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2500, 3000, 3500, 4000, 4500, 5000, and 5500 m. The time axis uses a climatological year with 12 months. By using a climatological year we implicitly ignore any temporal trend in the datasets, some of which span several decades, but at the present coverage this seems justified.

2.3 Patchiness

We use the following formula as a mathematical representation of the patchiness of the horizontal distribution of the PFTs summed over all depths:

$$P = 1/\overline{B} \times \sqrt{(\Sigma((L - B)^2)/(n - 1))} \quad (1)$$

where P is patchiness, $\overline{B}$ is mean biomass, L is the average within 10° longitude and latitude of the individual observations B, and n is the number of observations.

3 Results

3.1 Microzooplankton biomass database

The microzooplankton biomass database contains 4044 georeferenced data points. The gridded database contains 2029 data points (Fig. 1, Table 1). Data from the Northern Hemisphere makes up 64 % of the database, data from the top 225 m makes up 96 % of the database, and data from the spring and summer months makes up 63 % of the data. The average biomass is $7.0 \pm 15.3\,\mu g\,C\,L^{-1}$ with a median

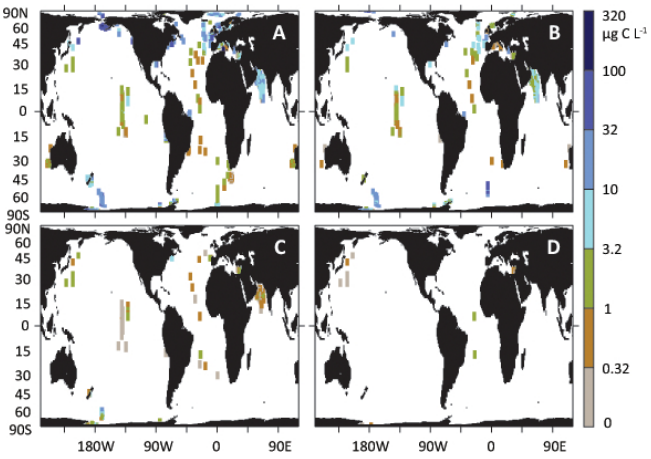


Figure 1. Microzooplankton biomass [$\mu g\,C\,L^{-1}$]. **(A)** 0–40 m, **(B)** 40–112.5 m, **(C)** 112.5–225 m, **(D)** > 225 m.

of $1.8\,\mu g\,C\,L^{-1}$. Biomass is considerably lower in the tropics ($3.4 \pm 7.5\,\mu g\,C\,L^{-1}$) than in the Northern Temperate (23–67° N, $7.9 \pm 18.7\,\mu g\,C\,L^{-1}$) and Southern Temperate Zones ($8.4 \pm 13.2\,\mu g\,C\,L^{-1}$), and increases further towards the Arctic ($9.7 \pm 9.9\mu g\,C\,L^{-1}$) and Antarctic ($16.1 \pm 25.5\,\mu g\,C\,L^{-1}$).

Compared to the microzooplankton database published by Buitenhuis et al. (2010), the corrected database contains both lower and higher values. The mean of the whole dataset has decreased by almost a factor 3, while the median has decreased by 10 %.

3.2 Comparison of PFT biomass concentrations

The number of data points that are available for each PFT differs by two orders of magnitude (Table 1). The larger databases have been built on earlier data synthesis efforts (O'Brien et al., 2002; Vaulot et al., unpublished data; Gosselin et al., unpublished data). We intend to maintain this group effort and extend the databases in future. By making both the raw and gridded databases publicly available, we also hope to encourage other researchers to publish extended versions. The horizontal distribution of each PFT at selected depths is presented in the contributing papers. Here, we compare the global average vertical profiles of the 11 PFTs and phytoplankton pigments (Fig. 2) and the zonal average sections (depth versus latitude) of the 11 PFTs in the top 200 m (Fig. 3; for the zonal average sections of pigments see figure 8 in Peloquin et al., 2013).

3.2.1 Patchiness

We calculated the horizontal patchiness of each PFT. There is a clear increase in patchiness with organism size (Fig. 4), although the relationship is different for autotrophs (patchiness $= 2.14 \times \log(\text{size}) + 0.98$, $n = 5$, $p = 0.04$) and heterotrophs ($\log(\text{patchiness}) = 0.25 \times \log(\text{size}) - 0.35$, $n = 6$, $p = 0.04$).

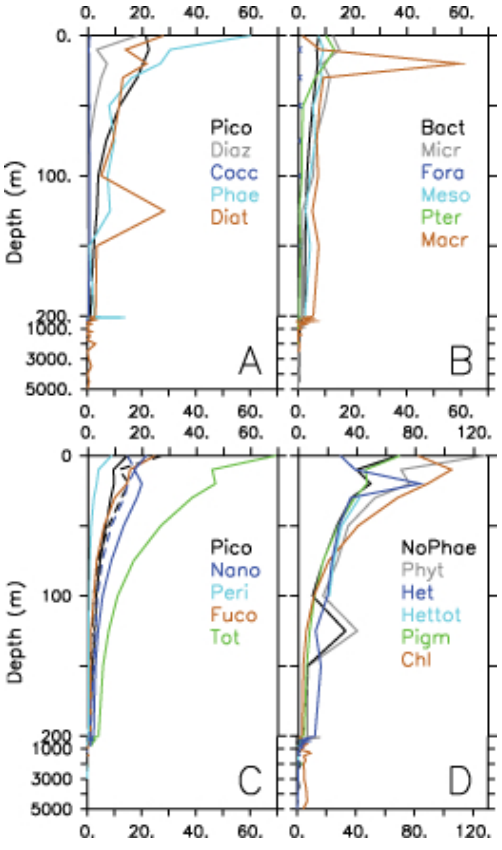


Figure 2. Depth profiles of average PFT biomass [$\mu g C L^{-1}$]. **(A)** Autotrophs from abundance. Black, "Pico": picophytoplankton; grey, "Diaz": diazotrophs; blue, "Cocc": coccolithophores; cyan, "Phae": *Phaeocystis*; brown, "Diat": diatoms. **(B)** Heterotrophs from abundance. Black, "Bact": picoheterotrophs; grey, "Micr": microzooplankton; blue crosses, "Fora": foraminifers; cyan, "Meso": mesozooplankton; green, "Pter": pteropods; brown, "Macr": macrozooplankton. **(C)** Autotrophs from pigments. Black, "Pico": prokaryotes from chlorophyll *b* and zeaxanthin; blue, "Nano": nanophytoplankton from 19'-hexanoyloxyfucoxanthin, 19'-butanoyloxyfucoxanthin, and alloxanthin; cyan, "Peri": dinoflagellates from peridinin; brown, "Fuco": diatoms from fucoxanthin; green, "Tot": total; dashed black: prokaryotes from sum of abundance based diazotroph and cyanobacteria; dashed blue: sum of abundance based biomass of coccolithophores and picoeukaryotes. **(D)** Black, "NoPha": sum of available phytoplankton without and grey, "Phyt": with *Phaeocystis*; blue, "Het": sum of heterotrophs without and cyan, "Hettot": with foraminifers and pteropods; green, "Pigm": total phytoplankton from pigments; Brown, "Chl": WOA 2005 chlorophyll × model C : Chl ratio. See text for details.

This increased patchiness may reflect a change from small K-selected organisms, which tend to form a constant background biomass, to large r-selected organisms (MacArthur and Wilson, 1967), which go through bloom and bust cycles. For the zooplankton it may in addition reflect a tendency for larger organisms to show coordinated vertical migration and/or swarming.

3.2.2 Autotrophic biomass

The highest phytoplankton biomass recorded in our datasets is found for *Phaeocystis* (Fig. 2a). As can be seen in figure 7 of Vogt et al. (2012), there is considerable sampling bias towards coastal waters, where dense *Phaeocystis* blooms can be a regular occurrence, and under which conditions a mix of colonial and single cells are often found. Under non-bloom conditions, *Phaeocystis* is mostly found as single cells, which cannot be distinguished from other nanophytoplankton with standard microscopic protocols, and therefore such measurements of background numbers of single *Phaeocystis* cells are almost absent from the database.

The next most abundant phytoplankton PFTs are picophytoplankton and diatoms (Fig. 2a), which is consistent with accepted wisdom about the importance of these two groups. However, dominance of picophytoplankton in the low latitudes and diatoms in the temperate latitudes has been suggested (Alvain et al., 2005; Uitz et al., 2006), while our plots of zonal average biomass show very little latitudinal differences in these groups (Fig. 3a, e). Dominance of picophytoplankton in low latitudes can be consistent with a homogenous latitudinal distribution because of the increase in biomass from low to temperate latitudes (Fig. 3f), but for diatoms there is a real discrepancy. The diatom biomass profile (Fig. 2a) has a distinct peak at 125 m. This peak is caused by a single observation of $7210\,\mu g C L^{-1}$, which was measured in a massive accumulation of mat forming *Rhizosolenia*, a feature which has been regularly observed in various oligotrophic environments (Shipe et al., 1999).

The average biomass of coccolithophores is surprisingly low (Fig. 2a), though it is almost constant down to 150 m, which is consistent with previous studies that showed the importance of coccolithophores in the lower euphotic layer (Cortés et al., 2001; Haidar and Thierstein, 2001). Coccolithophore biomass is lower even than diazotroph biomass, although the latter have hardly been measured at high latitudes (Fig. 3b), where they are thought not to occur (Carpenter, 1983).

3.2.3 Comparison of autotrophic biomass with pigment distributions

Recent syntheses suggest that virtually no pigment can be unequivocally assigned to quantify one marine algal type or species, since most of these pigments are shared across multiple phytoplankton taxa (Higgins et al., 2011). For this reason, statistical methods employing multiple pigment to chlorophyll ratios are required to adequately resolve the algal community composition (e.g. Mackey et al., 1996; Van den Meersche et al., 2008). Nonetheless, a preliminary analysis of the basin-scale distribution of a few key diagnostic pigments from the global phytoplankton pigment database (Peloquin et al., 2013) permits a rough comparison with some of the autotrophic biomass distributions from the

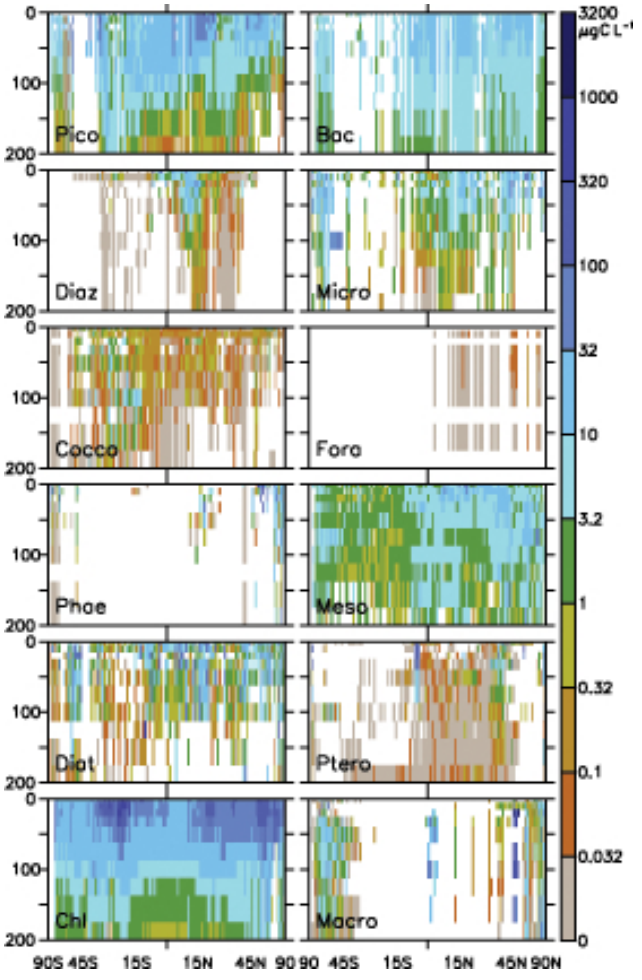


Figure 3. Zonal average PFT biomass [$\mu g C L^{-1}$]. (Pico) Picophytoplankton, (Diaz) diazotrophs, (Cocco) coccolithophores, (Phae) *Phaeocystis*, (Diat) diatoms, (Chl) WOA 2005 chlorophyll × PlankTOM5.3 C : Chl ratio, (Bac) picoheterotrophs, (Micro) microzooplankton, (Fora) foraminifers, (Meso) mesozooplankton, (Ptero) pteropods, and (Macro) macrozooplankton.

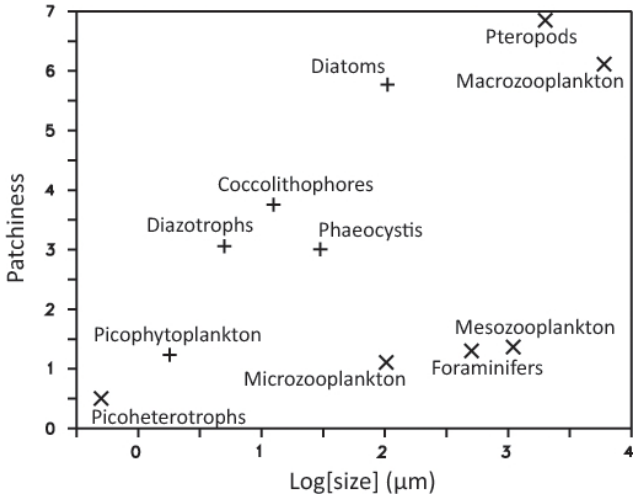


Figure 4. Patchiness (see Sect. 2.3) as a function of log (nominal organism size in µm). Plusses: autotrophs; crosses: heterotrophs.

abundance databases. For example, zeaxanthin, roughly indicative of the presence of cyanobacteria, exhibited a maximum concentration around the equatorial region, and is an order of magnitude lower poleward of 50° (figures 8l and 9l in Peloquin et al., 2013). This is consistent with the biomass data of *Prochlorococcus* and *Synechococcus* (figure 6a, b in Buitenhuis et al., 2012a) and the observed lack of diazotroph biomass at high latitudes (figure 3b, Luo et al., 2012). In addition, divinyl chlorophyll *a* is a strong biomarker for the presence of *Prochlorococcus*. Concentration of this pigment in the subtropics occurs at slightly deeper depths in the Southern Hemisphere than in the Northern Hemisphere (figure 9f in Peloquin et al., 2013), mirroring those patterns observed in the *Prochlorococcus* biomass distribution (figure 6a in Buitenhuis et al., 2012a). Fucoxanthin, a widely prevalent pigment among diatom species, exhibits maxima at high latitudes, particularly in the Southern Hemisphere (figures 8g and 9g in Peloquin et al., 2013). However, Fig. 3e indicates much less meridional variability in diatom abundance, which further highlights the need for incorporating the complexity of multiple pigment ratios when assessing phytoplankton distributions.

In order to also compare the depth profiles of biomass from the abundance and pigment databases, we first use pigment concentrations to calculate the contribution of different size classes of the phytoplankton to the overall chlorophyll concentration using the conversion factors of Uitz et al. (2006), and then convert chlorophyll to biomass using the C : Chl ratios of the PlankTOM5.3 model (Buitenhuis et al., 2013), which uses photosynthetic parameters synthesised by Geider et al. (1997). Because the C : Chl ratio of diatoms is about a third of the other two phytoplankton groups in the model, we use one profile of C : Chl ratios as a function of depth for diatoms and another profile for all other phytoplankton. The resulting pigment based biomass concentrations for diatoms and prokaryotes agree quite well with the abundance based biomass concentrations (Fig. 2a, c). The pigment based biomass for nanophytoplankton is close to the abundance based biomass of coccolithophores plus picoeukaryotes. This is rather surprising given the taxonomic diversity of larger eukaryotes that have not been included in the biomass from abundance, and the exclusion of *Phaeocystis* biomass. The pigment based biomass of dinoflagellates is estimated at roughly a third of the diatom biomass.

A thorough chemotaxonomic analysis of the global pigment database will reveal patterns in phytoplankton community structure on finer scales, and potentially contribute missing information on autotrophic dinoflagellates and nanophytoplankton. In concert with MAREDAT biomass development and analysis, analysis of global pigment distributions

Table 2. PFT characteristics.

		Approximate diameter [μm]		PFT biomass ($\mu g C L^{-1}$) in top 200 m[b]					PFT global biomass [Pg C][c]	
		Min	Max	Min	Max	Mean	Median	Std	Min	Max
Autotrophs	Picophytoplankton	0.6	3	0	449	12.7	5.47	22.1	0.28	0.64
	Diazotrophs	0.6	2000[a]	0	676	5.5	0.03	27.4	0.008	0.12
	Coccolithophores	5	20	0	83	0.4	0.05	2.4	0.001	0.03
	Phaeocystis	5	2000[a]	0	1248	28.3	2.18	96.0	0.11	0.71
	Diatoms	10	200	0	4155	16.5	1.74	104.7	0.10	0.94
Heterotrophs	Picoheterotrophs	0.4	0.6	0.2	58	8.1	6.64	6.0	1.00	1.10
	Microzooplankton	5	200	0	304	9.3	3.12	17.1	0.48	0.73
	Foraminifers	100	1000	0	0.4	0.03	0.008	0.05	0.0009	0.002
	Mesozooplankton	200	2000	0	345	6.0	2.77	10.6	0.33	0.59
	Pteropods	1000	3000	0	878	2.9	0.01	25.4	0.026	0.67
	Macrozooplankton	2000	10 000	0	3967	9.3	0.21	67.7	0.22	1.52

[a] Lower end of size range is for individual cells, higher end for colonies.
[b] Calculated from gridded databases.
[c] Lower estimate using median depth profiles, higher estimate using mean depth profiles; see Sect. 3.2.5 for details.

will help further guide the representation of all PFTs in ecosystem models.

3.2.4 Heterotrophs

The largest zooplankton biomass is found for macrozooplankton, up to 59 $\mu g C L^{-1}$ at 20 m depth (Fig. 2b). Macrozooplankton include some species that swarm, which could explain this sharp biomass peak, but it is also possible that some sampling bias such as proposed for the high diatom concentrations at 125 m depth occurred at a biomass concentration that was not quite high enough to be excluded by Chauvenet's criterion. Also, as noted above, the biomass distribution of all PFTs tends to become more patchy with increasing organism size (Fig. 4), thus increasing the errors around the means. The biomass of picoheterotrophs, mesozooplankton, microzooplankton and pteropods is fairly similar in the top 100 m, around 6 $\mu g C L^{-1}$ (Fig. 2b), while below 100 m the microzooplankton biomass is about half that of the picoheterotrophs and mesozooplankton, and pteropod biomass about a tenth, although there are no mesozooplankton data below 500 m. The biomass of planktic foraminifers is low, 0.04 $\mu g C L^{-1}$ at 10 m depth, and decreases to less than 0.01 below 300 m. At present, the database of foraminifers is limited to the Northern Hemisphere (Fig. 3i).

The numbers of observations drop dramatically below 1000 m for the zooplankton. Picoheterotrophs are the only PFT for which there are still around 100 observations at each level down to 3000 m. It is difficult to make definite statements about the ecology of the deep sea with so little information, but the data suggest that there are non-negligible concentrations of zooplankton in the deep sea: up to 6 $\mu g C L^{-1}$ macrozooplankton in the mesopelagic and 0.19 $\mu g C L^{-1}$ microzooplankton below 2750 m, compared to 0.36 $\mu g L^{-1}$ picoheterotrophs. These biomass concentrations are low relative to the surface biomass, but the volume of the deep sea is much larger, suggesting that zooplankton could make a substantial contribution to global ocean biogeochemical cycles in the deep sea as well. Picoheterotrophs have received much more attention with respect to their role in the biogeochemical cycling of organic matter in the deep ocean, but the ecology and biogeochemistry of the deep sea with a significant zooplankton contribution could be quite different from what we expect for a picoheterotroph-dominated deep sea.

3.2.5 Global PFT biomass inventories

The integrated global biomass inventories of the 11 PFTs are presented in Table 2. A low estimate was calculated by multiplying the median biomass at each depth by the volume of water at that depth, and a high estimate by using the mean concentrations at each depth. We vertically interpolated biomass between depths without observations. The main uncertainties in the biomass estimates are the uncertainties around conversion from abundance to biomass, a lack of coverage (which is due to both lack of samples and to only partial availability of those samples that have been taken), and also the tendency to sample near the coast or where a PFT is thought to occur or even bloom. Because of the latter sampling bias, the maximum global biomass inventories are likely to be overestimates. The conversion factors are probably the main source of uncertainty in all the datasets, while the importance of lack of coverage or of sampling bias appear to vary between datasets, e.g. for *Phaeocystis* there is a substantial sampling bias towards the coast, as noted above, which appears to explain the high horizontally averaged profile (Fig. 2a). For bacteria it was shown that a potential bias

towards the coast (in this case due to an observed change in the conversion factor rather than over-representation of sampling) was unimportant (Buitenhuis et al., 2012b).

We compare our phytoplankton biomass concentrations to the WOA 2005 total chlorophyll measurements. Without conversion to carbon, this equates to 18 to 30 Tg Chl. We calculated the total phytoplankton biomass by multiplying the WOA 2005 chlorophyll concentration by the C : Chl ratio profile of the PlankTOM5.3 model (Buitenhuis et al., 2013, Fig. 2d). The sum of our phytoplankton PFT biomass inventories is between 0.7 and 1.6 times the total phytoplankton derived from WOA chlorophyll down to 250 m (Fig. 2d), despite the fact that we did not include a large part of the nanophytoplankton nor the autotrophic dinoflagellates. One of the reasons for the overestimate at the surface is probably a strong sampling bias in the *Phaeocystis* dataset towards high values, as suggested above. Dividing our phytoplankton carbon without *Phaeocystis* by WOA chlorophyll results in C : Chl ratios of between 44 and 99 down to 250 m, except at 125 m where diatom biomass is unexpectedly high. While these C : Chl ratios are reasonable, they can be quite variable between taxonomic groups, even at a given light intensity. Therefore, these C : Chl ratios do not constitute a very stringent test, and a biomass similar to the picophytoplankton depth profile could be added to the total global mean biomass at any depth and still result in reasonable C : Chl ratios.

For the phytoplankton both the biomass and the number of observations decrease rapidly below 225 m (< 1–3 % of the data). For diatoms, and in particular for *Phaeocystis*, we suspect an additional possible bias because there are a few high values at depth, which show a considerable departure from the expected decrease of biomass with depth. These high concentrations below the euphotic zone are likely to be sinking cells after a bloom in the upper ocean rather than viable populations. However, recent samples from a depth range of 2000–4000 m that were taken during the Malaspina 2010 expedition confirm the ubiquitous presence of morphologically well-preserved, living phytoplankton cells of different taxa in the deep ocean (Agustí et al., ASLO conference presentation, 2012). The latter results suggest the existence of a far more efficient biological pump than previously thought, or the presence of physiological mechanisms that preserve cells at high pressures and low temperatures. We have integrated *Phaeocystis* biomass only to 225 m, and diatoms to 550 m.

For the picoheterotrophs, estimating global biomass from the depth profile seems to give reasonable results since there are still a fair number of observations and the biomass in the deep sea is fairly homogenous. However, the uncertainty about abundance to biomass conversion applies to picoheterotrophs as well. For the mesozooplankton the biomass has only been integrated where there was data down to 500 m, and for the pteropods the biomass has only been integrated down to 1050 m (with only 4 observations between 1050 and 2000 m).

3.2.6 Comparison of autotrophic and heterotrophic PFT biomass

In order to place PFT biomass in a wider context, total heterotrophic biomass is examined in relation to total autotrophic biomass in the global ocean (Fig. 5a). The median concentration for each group is considered. Lower biomass in the autotrophic component of the ecosystem reflects the higher turnover and metabolic costs of these small organisms (Odum, 1971). With higher turnover and metabolic costs, a low standing crop with high productivity can supply higher trophic levels with the energy that is then stored in their biomass because of lower turnover and metabolic costs. The global biomass data for each plankton group (Fig. 5a) do not show a blunt food pyramid as is typically found in terrestrial ecosystems, but instead confirm the high $H : A$ (heterotroph : autotroph) ratio of around 1 (see also Fig. 2d) that was previously found by Gasol et al. (1997).

In general, the coastal biomass concentrations are higher than in the open ocean (Fig. 5b, c), as expected. Foraminifers and picophytoplankton are the exceptions, with lower concentrations in the coastal ocean. For foraminifers this has been reported previously (Retailleau et al., 2009). Coccolithophores show a lower mean but higher median in the coastal ocean. Judging from the standard deviation associated with many of the PFTs (Table 2, Fig. 5), caution is required not to over-interpret these data. In accordance with the results of Gasol et al. (1997), we find that metazoan biomass possibly increases more than phytoplankton in the coastal ocean, but in contrast to them we find that microzooplankton also increase at least as much as phytoplankton. The bacterial abundance is the same in the coastal and open ocean, which is consistent with a decrease relative to phytoplankton in Gasol et al. (1997), but more work is needed to confirm whether the carbon content of bacteria is higher in the coastal ocean (see Buitenhuis et al., 2012b).

4 Discussion

At a time when we, the human species, are subjecting the biosphere to unprecedented rates of change, we know very little at the global scale of the baseline functioning of the biosphere. There is a large gap between the detailed but anecdotal information that is available about the physiology of individual species under particular in situ or controlled laboratory conditions, and what we can say about the functioning of the biosphere as a whole. In ocean ecosystem modelling we have only begun to address this gap in the last decade. The present effort at data synthesis on the global biomass distribution of most of the PFTs representing the lower trophic levels of the ocean ecosystem is the first attempt at comprehensively addressing this gap. It is necessarily preliminary and raises more questions than it answers. These questions indicate that despite the scarcity of data for most groups we still know much more about the abundance of organisms than

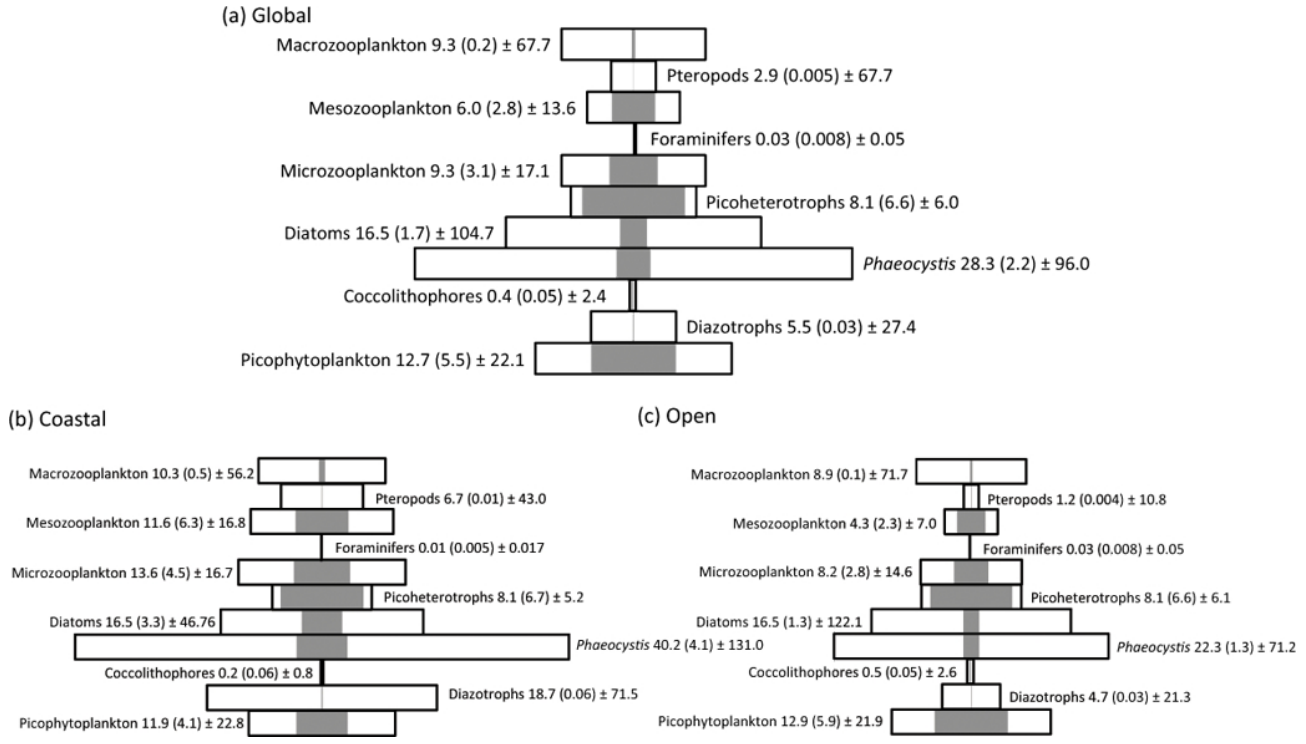


Figure 5. Trophic pyramid of autotrophic and heterotrophic PFTs. Mean (black outline) and median (grey fill, values in brackets) biomass ($\mu g\,C\,L^{-1}$) in the top 200 m for each of the PFTs presented in the MAREDAT special issue. Standard deviation is listed but not shown in the graph. **(a)** Global ocean (for full details see Table 2, **(b)** coastal and **(c)** open ocean.

about their carbon content/elemental composition, and much more about places where organisms are abundant than about the much larger volumes of global ocean where biomass is relatively low.

Information becomes even more anecdotal, and understanding more tenuous, when looking at time scales longer than a few seasons. For large regions of the global ocean, systematic changes at an interannual scale (or longer) are largely unknown for most PFTs. Extending our perspective to the longer term and geological time scales, MAREDAT data may lead to a better understanding of both environmental and climate change, and the rate of CO_2 increase experienced over not only glacial-interglacial cycles but also since the beginning of industrialisation. Over time intervals of millions of years, distribution and size of planktic foraminifers (Schmidt et al., 2004) and coccolithophores (Henderiks and Pagani, 2007) have been affected by climate change, and this might also be true for other PFTs. The MAREDAT approach is a step towards a more complete, qualitative and quantitative understanding of PFTs and associated feedbacks with past and future environmental and, in particular, climate change at a global scale and over long intervals of time.

Currently, at least two large-scale programs are collecting plankton samples that will increase our knowledge on plankton community composition and species diversity in the global oceans. The Tara Oceans expedition (http://oceans.taraexpeditions.org) is collecting plankton samples from 2009 to 2013 for a depth range of 0 to 200 m, including samples from all the major ocean basins. The Malaspina 2010 expedition (http://www.expedicionmalaspina.es) completed its journey in 2011, and collected samples from the deep oceans, at depths down to several thousand metres (Laursen, 2011). The latter expedition will furnish us with new data to increase our understanding of deep ocean ecosystems, while the former will increase data coverage in the upper layers of the ocean.

The ranges of global biomass inventories we calculated from the profiles of median and mean biomass in several cases span more than an order of magnitude. Despite this uncertainty they indicate that heterotrophic PFTs are at least as abundant as autotrophic PFTs (Table 2; Figs. 2d, 5), even if we account for the risk of double counting foraminifers as microzooplankton and pteropods as macrozooplankton in the sum of the zooplankton biomass. Within the uncertainty in the data this is in agreement with Gasol et al. (1997), who estimated that the open ocean biomass of phytoplankton, total zooplankton and picoheterotrophs is roughly the same. We have compared our estimates of the sum of the available total phytoplankton biomass estimates with in situ chlorophyll. We find the sum of phytoplankton biomass is higher than biomass estimated from chlorophyll. All these databases are biased towards oversampling of the coastal ocean. The

coastal ocean represents 5 % of the ocean area, but the phytoplankton biomass in the coastal ocean represents between 7 % (diazotrophs) and 34 % (Phaeocystis and diatoms) of the respective databases, while the WOA 2005 in situ chlorophyll represents 20 % of that database. In future updates of the databases we intend to diminish this bias by including additional observational data and improving the cell to carbon conversion algorithms, in particular for those PFTs with a large range of morphotypes and cell sizes. This will allow a better determination of total phytoplankton biomass in the global ocean.

Plankton are governed by physical, chemical and biological processes that occur on a vast variety of temporal and spatial scales. This is one reason why large datasets are needed to reliably estimate the global biomass distribution of any PFT. Physical processes can drive patchiness in PFTs on scales from millimetres to thousands of kilometres (Pinel-Alloul, 1993; Folt and Burns, 1999). Over fine scales, millimetres to tens of metres, biological processes are often more important. For the picoheterotrophs, the estimated global biomass using the median and mean concentrations is very similar, indicating that the horizontal variation in biomass makes only a small contribution to the uncertainty in the estimated global biomass. This range does not include the uncertainty from the respective conversion factors from abundance to biomass. For the autotrophic PFTs, horizontal variability is larger (Fig. 4), and the difference between the median and the mean is more than an order of magnitude for three PFTs (Table 2). For diazotrophs this variability is compounded by lack of spatial coverage, but for coccolithophores and pteropods there is also large variability despite a better spatial coverage (Table 1, Fig. 3). For the zooplankton PFTs, individual behaviour, e.g. mating, predator avoidance and searching for food (Folt and Burns, 1999), and variables such as food concentration, swimming behaviour (Pinel-Alloul, 1993) and species interactions (Mackas et al., 1985) are important examples of small-scale biological processes that affect patchiness. Growth rates for a number of groups within the macrozooplankton, particularly the gelatinous members, salps, ctenophores, cnidarians and appendicularians, are higher than those of mesozooplankton copepods (Hirst and Bunker, 2003). The ability of these groups to "bloom" or swarm, through a combination of high grazing rates, growth rates and life history, when food concentration or other environmental factors are favourable, means that macrozooplankton may reach high biomass concentrations in areas where they amass, resulting in a spatially heterogeneous distribution.

We show that the patchiness of PFT biomass increases with size. In terms of database synthesis, this means that more data would be needed for larger organisms before we can reliably generate interpolated datasets of biomass. In terms of model representation, it suggests that it will be more difficult to accurately represent individual data points, and that it will only be possible to represent the mean state or to build stochastic representations. It should be noted that by using point measurements (usually representative of $< 1\,m^2$) and binning these into a horizontal grid spacing of 1°, there is potential for distorting the results. On the one hand, using gridded data means that the calculation cannot detect variations in patchiness that are smaller than 1°. On the other hand, there are not enough data points in each grid point to obtain a true reflection of the average biomass in that grid cell, and therefore fine scale variability could potentially be ascribed to a larger scale. There is no indication that the latter invalidates the increase in patchiness with size that we find, since if it did then patchiness should be significantly higher for smaller datasets, which is not the case ($p = 0.5$, $n = 11$).

In addition to the depth and coastal ocean bias already noted for the phytoplankton, all of the heterotroph databases also have a depth bias (62–96 % of samples at < 225 m) and all but one have a coastal bias (13–28 % of samples in the coast). The exception is foraminifers (2 %). In addition, 9 of the databases show a substantial bias towards samples in the Northern Hemisphere (64–100 % of the databases). The exceptions are pigments and coccolithophores, while macrozooplankton show a bias towards the Southern Hemisphere (70 %). The databases also tend to show some bias towards spring and summer (56–86 % of the databases).

Qualitative information on the presence or absence of specific autotrophic groups from MAREDAT may help guide the a priori selection of pigment ratios that is essential for quantifying algal type abundance through Bayesian-type analyses (Mackey et al., 1996; Van der Meersche et al., 2008). Conversely, the biomass databases from MAREDAT will be instrumental in evaluating those quantitative results, which may help bridge some of the information gaps in phytoplankton distributions.

In its first iteration it is difficult to judge the overall quality of the MAREDAT dataset. The data gathering effort has been a massive success (see Table 1). The data has been quality controlled and each dataset has documented the associated uncertainties. There is consistency across the datasets in this respect, and ensuring this consistency is one of the major achievements of MAREDAT (see Table 1).

5 Summary

This introduction to MAREDAT and the data, results and discussion presented here represents what was at one point an almost impossible goal for many of those involved in this project. We are now in a position to make the first data based estimates of global annual average biomass for 11 PFTs and to validate biogeochemical models. PFT modelling is a relatively new development in ocean biogeochemical modelling, and global models embedded in general circulation models (GCMs) have only been available for the last 10 yr (Aumont et al., 2003). Improved validation of these models will also support climate change research, helping early detection

of impacts and feedbacks from current climate change, and building better models for projections into the future.

The MAREDAT data will also be fundamental for the investigation of global plankton biogeography and ecology. In this initial analysis, we investigate zonal and vertical patterns of plankton biomass, their relative contribution to global plankton biomass, and the relationship of body size with patchiness. We find that (1) the global total heterotrophic biomass is at least as high as the total autotrophic biomass, (2) the biomass of zooplankton calcifiers is higher than that of coccolithophores, (3) patchiness of biomass distribution increases with organism size, and (4) although zooplankton biomass measurements below 200 m are rare, the limited measurements available suggest that *Bacteria* and *Archaea* are not the only heterotrophs in the deep sea. Future studies will investigate patterns of plankton diversity, ecological niches, and habitat structure (Luo et al., 2013) and dynamics.

The MAREDAT initiative has been accepted by many within the community as a timely, important and "enormously useful" contribution to biological oceanography and biogeochemistry. Similar efforts have been suggested to collect other marine ecological data, such as phytoplankton traits (e.g. Barton et al., 2013). The general weaknesses, e.g. fundamental lack of open water data and lack of elemental information for a variety of taxa, have also been recognised. Another major weakness of the initiative was that it was not possible to include some existing published and unpublished datasets, either because they were not publicly available, due to a lack of time for data retrieval, or because the response rates of the data owners to our data call were low. However, we have now set a precedent within the community for publishing this type of data and giving full accreditation to those involved in gathering it. Hopefully this will encourage data holders to contribute data in the future. The authors involved in this special issue have agreed that datasets and documentation will be updated for the next release of MAREDAT, planned for 2015. The experience gained from this MAREDAT iteration and the subsequent analysis of plankton ecology based on this data will help inform MAREDAT2015.

Overall, this special issue has brought together abundance and biomass data for most of the PFTs in the lower trophic levels of the ocean ecosystem, as well as a global database of HPLC-based phytoplankton pigments. In several cases the biomass databases represent the first of their kind to cover all ocean basins, and in the remaining cases they are substantially larger than what has been available so far. In all, we have brought together 436 887 biomass measurements in 12 databases. The gridded data products have been provided for the ocean ecosystem modelling community. The raw data, including references and metadata, are also publicly available at http://www.pangaea.de/search?&q=maredat.

Acknowledgements. We thank Marie-Pierre Gosselin for collecting the first versions of several PFT databases, funded by EUR-OCEANS. We thank the guest editors Stéphane Pesant and Walker O. Smith, the ESSD editorial team Dave Carlson and Hans Pfeiffenberger, and the PANGAEA team. We acknowledge funding from the EU (to ETB, EMBRACE, contract 282672, to COB, Greencycles, contract 238366, to RM, FAASIS, MEST-CT-2004-514159), ETH Zürich (to MV), and the National Science Foundation (to SCD and YL, NSF-EF-0424599).

Edited by: W. Smith

References

Agustí, S.: Ubiquitous presence of healthy phytoplankton cells in the deep ocean, ASLO ocean sciences meeting, 2012.

Alvain, S., Moulin, C., Dandonneau, Y., and Breon, F. M.: Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery, Deep-Sea Res. Pt. I, 52, 1989–2004, 2005.

Aumont, O., Maier-Reimer, E., Blain, S., and Monfray, P.: An ecosystem model of the global ocean including Fe, Si, P co-limitations, Global Biogeochem. Cy., 17, 1060, doi:10.1029/2001GB001745, 2003.

Barton, A. D., Pershing, A. J., Litchman, E., Record, N. R., Edwards, K. F., Finkel, Z. V., Kiorboe, T., and Ward, B. A.: The biogeography of marine plankton traits, Ecol. Lett., 16, 522–534, 2013.

Bednaršek, N., Možina, J., Vogt, M., O'Brien, C., and Tarling, G. A.: The global distribution of pteropods and their contribution to carbonate and carbon biomass in the modern ocean, Earth Syst. Sci. Data, 4, 167–186, doi:10.5194/essd-4-167-2012, 2012.

Buitenhuis, E. T., Rivkin, R. B., Sailley, S., and Le Quere, C.: Biogeochemical fluxes through microzooplankton, Global Biogeochem. Cy., 24, GB4015, doi:10.1029/2009GB003601, 2010.

Buitenhuis, E. T., Li, W. K. W., Vaulot, D., Lomas, M. W., Landry, M. R., Partensky, F., Karl, D. M., Ulloa, O., Campbell, L., Jacquet, S., Lantoine, F., Chavez, F., Macias, D., Gosselin, M., and McManus, G. B.: Picophytoplankton biomass distribution in the global ocean, Earth Syst. Sci. Data, 4, 37–46, doi:10.5194/essd-4-37-2012, 2012a.

Buitenhuis, E. T., Li, W. K. W., Lomas, M. W., Karl, D. M., Landry, M. R., and Jacquet, S.: Picoheterotroph (Bacteria and Archaea) biomass distribution in the global ocean, Earth Syst. Sci. Data, 4, 101–106, doi:10.5194/essd-4-101-2012, 2012b.

Buitenhuis, E. T., Hashioka, T. and Le Quéré, C.: Combined constraints on global ocean primary production using observations and models, Global Biogeochem. Cy., doi:10.1029/2012GB004516, submitted, 2013.

Carpenter, E. J.: Physiology and Ecology of Marine Planktonic Oscillatoria (Trichodesmium), Mar. Biol. Lett., 4, 69–85, 1983.

Cortes, M. Y., Bollmann, J., and Thierstein, H. R.: Coccolithophore ecology at the HOT station ALOHA, Hawaii, Deep-Sea Res. Pt. II, 48, 1957–1981, 2001.

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., Mahowald, N., Glover, D. M., and Takahashi, T.: Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data, J. Marine Syst., 76, 95–112, 2009.

Folt, C. L. and Burns, C. W.: Biological drivers of zooplankton patchiness, Trends Ecol. Evol., 14, 300–305, 1999.

Gasol, J. M., del Giorgio, P. A., and Duarte, C. M.: Biomass distribution in marine planktonic communities, Limnol. Oceanogr., 42, 1353–1363, 1997.

Geider, R. J., MacIntyre, H. L., and Kana, T. M.: Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature, Mar. Ecol.-Progr. Ser., 148, 187–200, 1997.

Glover, D. M., Jenkins, W. J., and Doney, S. C.: Modeling Methods for Marine Science, Cambridge University Press, Cambridge, UK, 2011.

Haidar, A. T. and Thierstein, H. R.: Coccolithophore dynamics off Bermuda (N. Atlantic), Deep-Sea Res. Pt. II, 48, 1925–1956, 2001.

Henderiks, J. and Pagani, M.: Refining ancient carbon dioxide estimates: Significance of coccolithophore cell size for alkenone-based pCO_2 records, Paleoceanography, 22, PA3202, doi:10.1029/2006PA001399, 2007.

Higgins, H., Wright, S., and Schluter, L.: Quantitative interpretation of chemotaxonomic pigment data, in: Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography, edited by: Roy, S., Llewellyn, C., Egeland, E., and Johnsen, G., Cambridge University Press, 2011.

Hirst, A. G. and Bunker, A. J.: Growth of marine planktonic copepods: global rates and patterns in relation to chlorophyll a, temperature, and body weight, Limnol. Oceanogr., 48, 1988–2010, 2003.

Laursen, L.: Malaspina expedition: starting with a splash, Nature News Blog, http://blogs.nature.com/news/2011/02/malaspina_expedition_starting.html, 2011.

Leblanc, K., Arístegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., Breton, E., Cornet, V., Gibson, J., Gosselin, M.-P., Kopczynska, E., Marshall, H., Peloquin, J., Piontkovski, S., Poulton, A. J., Quéguiner, B., Schiebel, R., Shipe, R., Stefels, J., van Leeuwe, M. A., Varela, M., Widdicombe, C., and Yallop, M.: A global diatom database – abundance, biovolume and biomass in the world ocean, Earth Syst. Sci. Data, 4, 149–165, doi:10.5194/essd-4-149-2012, 2012.

Le Quéré, C. and Pésant, S.: Plankton Functional Types in a new generation of biogeochemical models, EOS, 90, 30–31, 2009.

Le Quéré, C., Harrison, S. P., Prentice, I. C., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Da Cunha, L. C., Geider, R., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., and Wolf-Gladrow, D.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Glob. Change Biol., 11, 2016–2040, 2005.

Luo, Y.-W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., Bonnet, S., Boström, K. H., Böttjer, D., Capone, D. G., Carpenter, E. J., Chen, Y. L., Church, M. J., Dore, J. E., Falcón, L. I., Fernández, A., Foster, R. A., Furuya, K., Gómez, F., Gundersen, K., Hynes, A. M., Karl, D. M., Kitajima, S., Langlois, R. J., LaRoche, J., Letelier, R. M., Marañón, E., McGillicuddy Jr., D. J., Moisander, P. H., Moore, C. M., Mouriño-Carballido, B., Mulholland, M. R., Needoba, J. A., Orcutt, K. M., Poulton, A. J., Rahav, E., Raimbault, P., Rees, A. P., Riemann, L., Shiozaki, T., Subramaniam, A., Tyrrell, T., Turk-Kubo, K. A., Varela, M., Villareal, T. A., Webb, E. A., White, A. E., Wu, J., and Zehr, J. P.: Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates, Earth Syst. Sci. Data, 4, 47–73, doi:10.5194/essd-4-47-2012, 2012.

Luo, Y.-W., Lima, I. D., Karl, D. M., and Doney, S. C.: Data-based assessment of environmental controls on global marine nitrogen fixation, Biogeosciences Discuss., 10, 7367–7412, doi:10.5194/bgd-10-7367-2013, 2013.

MacArthur, R. H. and Wilson, E. O.: The theory of island biogeography, Princeton University Press, 1967.

Mackas, D. L., Denman, K. L., and Abbott, M. R.: Plankton patchiness: biology in the physical vernacular, B. Mar. Sci., 37, 652–674, 1985.

Mackey, M. D., Mackey, D. J., Higgins, H. W., and Wright, S. W.: CHEMTAX – A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton, Mar. Ecol.-Prog. Ser., 144, 265–283, 1996.

Moriarty, R., Buitenhuis, E. T., Le Quéré, C., and Gosselin, M.-P.: Distribution of known macrozooplankton abundance and biomass in the global ocean, Earth Syst. Sci. Data, 5, 241–257, doi:10.5194/essd-5-241-2013, 2013.

O'Brien, C. J., Peloquin, J. A., Vogt, M., Heinle, M., Gruber, N., Ajani, P., Andruleit, H., Arístegui, J., Beaufort, L., Estrada, M., Karentz, D., Kopczyn'ska, E., Lee, R., Poulton, A. J., Pritchard, T., and Widdicombe, C.: Global marine plankton functional type biomass distributions: coccolithophores, Earth Syst. Sci. Data, 5, 259–276, doi:10.5194/essd-5-259-2013, 2013.

O'Brien, T. D., Conkright, M. E., Boyer, T. P., Stephens, C., Antonov, J. I., Locarnini, R. A., and Garcia, H. E.: World Ocean Atlas 2001, Volume 5: Plankton, NOAA Atlas NESDIS 53, edited by: Levitus, S., US Government Printing Office, Washington, DC, 95 pp., available at: http://www.nodc.noaa.gov/OC5/indpub.html#pub2002, 2002.

Odum, E. P.: Fundamentals of Ecology, W. B. Saunders, Philadelphia, 574 pp., 1971.

Peloquin, J., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., Uitz, J., Barlow, R., Behrenfeld, M., Bidigare, R., Dierssen, H., Ditullio, G., Fernandez, E., Gallienne, C., Gibb, S., Goericke, R., Harding, L., Head, E., Holligan, P., Hooker, S., Karl, D., Landry, M., Letelier, R., Llewellyn, C. A., Lomas, M., Lucas, M., Mannino, A., Marty, J.-C., Mitchell, B. G., Muller-Karger, F., Nelson, N., O'Brien, C., Prezelin, B., Repeta, D., Jr. Smith, W. O., Smythe-Wright, D., Stumpf, R., Subramaniam, A., Suzuki, K., Trees, C., Vernet, M., Wasmund, N., and Wright, S.: The MAREDAT global database of high performance liquid chromatography marine pigment measurements, Earth Syst. Sci. Data, 5, 109–123, doi:10.5194/essd-5-109-2013, 2013.

Pinel-Alloul, B.: Spatial Heterogeneity as a Multiscale Characteristic of Zooplankton Community, 2nd International Joint Conference on Limnology and Oceanography, Evian, France, 17–42, 1993.

Retailleau, S., Howa, H., Schiebel, R., Lombard, F., Eynaud, F., Schmidt, S., Jorissen, F., and Labeyrie, L.: Planktic foraminiferal production along an offshore-onshore transect in the southeastern Bay of Biscay, Cont. Shelf Res., 29, 1123–1135, doi:10.1016/j.csr.2008.12.021, 2009.

Schiebel, R. and Movellan, A.: First-order estimate of the planktic foraminifer biomass in the modern ocean, Earth Syst. Sci. Data, 4, 75–89, doi:10.5194/essd-4-75-2012, 2012.

Schmidt, D. N., Thierstein, H. R., Bollmann, J., and Schiebel, R.: Abiotic forcing of plankton evolution in the Cenozoic, Science, 303, 207–210, 2004.

Shipe, R. F., Brzezinski, M., Pilskaln, C. H., and Villareal, T. A.: *Rhizosolenia* mats: An overlooked source of silica production in the open sea, Limnol. Oceanogr., 44, 1282–1292, 1999.

Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, J. Geophys. Res.-Oceans, 111, C08005, doi:10.1029/2005JC003207, 2006.

Van den Meersche, K., Soetaert, K., and Middelburg, J. J.: A Bayesian compositional estimator for microbial taxonomy based on biomarkers, Limnol. Oceanogr.-Meth., 6, 190–199, 2008.

Vogt, M., O'Brien, C., Peloquin, J., Schoemann, V., Breton, E., Estrada, M., Gibson, J., Karentz, D., Van Leeuwe, M. A., Stefels, J., Widdicombe, C., and Peperzak, L.: Global marine plankton functional type biomass distributions: Phaeocystis spp., Earth Syst. Sci. Data, 4, 107–120, doi:10.5194/essd-4-107-2012, 2012.