



Distribution of known macrozooplankton abundance and biomass in the global ocean

R. Moriarty^{1,2,*}, E. T. Buitenhuis², C. Le Quéré^{1,2,**}, and M.-P. Gosselin^{1,***}

¹British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK

²School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

*Present address: School of Earth, Atmospheric and Environmental Sciences, University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL, UK

**Present address: Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

***Present address: Environment Agency, North West – Water Resources, Ghyll Mount, Gillan Way, Penrith 40 Business Park, Penrith CA11 9BP, UK

Correspondence to: R. Moriarty (r.moriarty@uea.ac.uk)

Received: 13 March 2012 – Published in Earth Syst. Sci. Data Discuss.: 20 April 2012

Revised: 11 June 2013 – Accepted: 18 June 2013 – Published: 12 July 2013

Abstract. Macrozooplankton are an important link between higher and lower trophic levels in the oceans. They serve as the primary food for fish, reptiles, birds and mammals in some regions, and play a role in the export of carbon from the surface to the intermediate and deep ocean. Little, however, is known of their global distribution and biomass. Here we compiled a dataset of macrozooplankton abundance and biomass observations for the global ocean from a collection of four datasets. We harmonise the data to common units, calculate additional carbon biomass where possible, and bin the dataset in a global 1×1 degree grid. This dataset is part of a wider effort to provide a global picture of carbon biomass data for key plankton functional types, in particular to support the development of marine ecosystem models. Over 387 700 abundance data and 1330 carbon biomass data have been collected from pre-existing datasets. A further 34 938 abundance data were converted to carbon biomass data using species-specific length frequencies or using species-specific abundance to carbon biomass data. Depth-integrated values are used to calculate known epipelagic macrozooplankton biomass concentrations and global biomass. Global macrozooplankton biomass, to a depth of 350 m, has a mean of $8.4 \mu\text{g CL}^{-1}$, median of $0.2 \mu\text{g CL}^{-1}$ and a standard deviation of $63.5 \mu\text{g CL}^{-1}$. The global annual average estimate of macrozooplankton biomass in the top 350 m, based on the median value, is 0.02 Pg C . There are, however, limitations on the dataset; abundance observations have good coverage except in the South Pacific mid-latitudes, but biomass observation coverage is only good at high latitudes. Biomass is restricted to data that is originally given in carbon or to data that can be converted from abundance to carbon. Carbon conversions from abundance are restricted by the lack of information on the size of the organism and/or the absence of taxonomic information. Distribution patterns of global macrozooplankton biomass and statistical information about biomass concentrations may be used to validate biogeochemical and plankton functional type models.

Macrozooplankton abundance and biomass dataset doi:10.1594/PANGAEA.777398.

1 Introduction

Global ocean biogeochemical models representing lower-trophic ecosystems have been widely used to study the interactions and feedbacks between climate and marine bio-

geochemistry. They have been applied to investigating the processes affecting atmospheric CO_2 concentration and climate (e.g. Lovenduski et al., 2008) and used to predict the possible effects of climate change on ecosystem structure, functioning and productivity. The most recent generation

of global biogeochemical models – called Dynamic Green Ocean Models (DGOMs) – represent marine ecosystems using multiple plankton functional types (PFTs; Le Quéré et al., 2005; Hood et al., 2006) and thus include basic trophic structure and ecosystem diversity. Independent datasets are required to validate the representation of each PFT in these models, in particular their distribution in various ocean regions, depth profiles and concentrations. Observational data on PFT-specific abundance and carbon biomass is sparse and highly variable, and consequently model validation is challenging. The lack of coherent global PFT-specific datasets is currently limiting the development of global models. The data presented in this paper are part of a wider community effort known as MARine Ecosystem DATA (MAREDAT), and cover data on a variety of major PFTs currently represented in marine ecosystem models (Buitenhuis et al., 2013).

MAREDAT is a collection of global biomass datasets gathered by marine ecosystem researchers. It contains data on the global distribution of a variety of the major PFTs currently represented in marine ecosystem models. These include bacteria, picophytoplankton, nitrogen fixers, calcifiers, dimethyl sulphide (DMS)-producers, silicifiers, microzooplankton, foraminifera, mesozooplankton (mostly copepods; Moriarty and O'Brien, 2013), pteropods and macrozooplankton. MAREDAT is part of the MARine Ecosystem Model Inter-comparison Project (MAREMIP) and is responsible for this compilation of observation-based global biomass datasets. The biomass data that populate MAREDAT are freely available for use in model evaluation and development, and to the scientific community as a whole (<http://maremip.uea.ac.uk/maredat.html>).

Macrozooplankton are found throughout the global ocean and epipelagic macrozooplankton in the sunlit portion. They include the holoplanktic and meroplanktic members of the thalicia, ctenophores, cnidaria, gastropoda, heteropoda, pteropoda, chaetognatha, polychaeta, amphipods, stomatopods, mysids, decapods, and euphausiids, among others, each taxon containing many species. Macrozooplankton are commonly classified by size but classification varies (see Sieburth et al., 1987; Schütt, 1892; Omori and Ikeda, 1984). Here we include all zooplankton whose adult size is greater than 2 mm (Le Quéré et al., 2005). They are separated from nekton (i.e. fish) as they do not have the locomotive capacity necessary to overcome ocean currents (Omori and Ikeda, 1984). Copepods are an exception to this definition. In order to try and avoid double counting of copepods, the biomass of this group have been included in a mesozooplankton/copepod-specific article (Moriarty and O'Brien, 2013; see Sect. 3.2.1).

Macrozooplankton are involved in an intricate trophic web, exerting a direct influence on all lower trophic levels through variation in species-specific feeding behaviours and prey preferences. They repackage autotrophic, heterotrophic and detrital material in the surface ocean. Part of this repackaged material moves to higher trophic levels through the pre-

dition on macrozooplankton (Deibel, 1998), making them an important link to higher trophic levels, especially fish. In some areas of the ocean they are a crucial link between lower and higher trophic levels, e.g. *Euphausia superba* in the Southern Ocean. Repackaged material may also be exported from the surface to the intermediate and deep ocean in the form of faecal pellets (Turner, 2002). In this manner macrozooplankton affect the cycling of carbon and nutrients in the ocean and have been included as a PFT in some DGOMs (Le Quéré et al., 2005, 2013).

PFT modelling uses a coarse division of the ecosystem, breaking the plankton into functional groups or plankton functional types (PFTs). PFT models treat macrozooplankton as a generalist group, linked by their common function in the removal of carbon from the sunlit waters of the global ocean to the deep ocean. Macrozooplankton demonstrate incredible diversity when it comes to feeding preferences, feeding on minute particles, bacteria, detritus, phytoplankton and other zooplankton. We presented macrozooplankton as a unique PFT compartment of the ecosystem, regardless of the position of individual species in the food web. This dataset is designed for application in the study of global ocean ecosystems from a plankton function or PFT perspective.

This paper presents a global synthesis of macrozooplankton abundance and biomass data, including full details of the biomass conversions used and the associated uncertainties. Patterns in the spatial and temporal distribution of epipelagic macrozooplankton abundance and biomass are examined. There is no one data centre responsible for the collection and synthesis of macrozooplankton abundance and biomass data. The synthesis presented here is a step towards gathering all available data to better understand the global distribution of macrozooplankton and the mechanisms controlling them.

2 Data and methods

2.1 Origin of data

The data compiled in this study have come from four existing databases and are summarised in Table 1a, b, c and d. Macrozooplankton data have been data mined from all databases with the exception of the rawKRILLBASE database, which already contained only macrozooplankton data.

The first two databases are from the Joint Global Ocean Flux Study: the Hawaii Ocean Time-Series (HOTS; Landry et al., 2001) and the Bermuda Atlantic Time-Series (BATS; Steinberg et al., 2012; Table 1a). These are long-term time-series stations representative of the oligotrophic subtropical gyres in the North Pacific and North Atlantic, respectively. Abundance and biomass data were collected by oblique net tows on monthly cruises at Station ALOHA (A Long-term Oligotrophic Habitat Assessment; 22°45' N, 158°00' W) for HOTS and at the BATS station (31°50' N, 64°10' W). The targeted maximum depth was between 150 and 200 m. At both locations data for two macrozooplankton size fractions

Table 1a. Summary of data points for original macrozooplankton abundance and biomass as gathered from JEGOFS long-term time-series datasets.

Dataset	Principal Investigator	No. of data		Depth range (m)	Reference
		Abundance	Biomass		
HOTS	Landry	272	1330	0–271	Landry et al. (2001)
BATS	Steinberg	0	1730*	0–306	Steinberg et al. (2012)

* Biomass values are originally in wet weight and dry weight.

2000–5000 μm and $>5000\mu\text{m}$ were collected. The data presented in this study were collected between 1994 and 2010; data collection is still ongoing at both sites. The HOTS dataset includes 272 abundance (individual (ind.) m^{-2}) and 1330 biomass (mg carbon (C) m^{-2}) data. For the purposes of this study, abundance data were converted from ind. m^{-2} to ind. L^{-1} by averaging the sampling depth. Biomass data were converted from mg C m^{-2} to μgCL^{-1} in a similar manner. The BATS dataset includes no abundance and 1730 biomass (mg wet mass m^{-3} and mg dry mass m^{-3}) data; it does not provide carbon biomass. BATS abundance data were converted to carbon biomass using a dry mass to carbon conversion (see Sect. 2.2).

The rawKRILLBASE (Atkinson et al., 2004; Table 1b and c) database is a Southern Ocean dataset of krill and salp numerical densities and length frequencies. It includes 8192 abundance (ind. m^{-2}) measurements of postlarval *Euphausia superba* and 9719 abundance measurements of pooled individuals of aggregate and solitary forms of salps (mainly *Salpa thompsoni* but also *Ihlea racovitzai*). The data spans the periods 1926–1951 and 1976–2003. All data are from random hauls or those at pre-fixed locations (hauls specifically targeted on krill or salp aggregations were excluded). All net hauls were oblique or vertical and were taken mainly in the summer months. Full details of this database, including net sampling details, are provided in Atkinson et al. (2004, 2008, 2009; including the supplementary method appendices of the first two). The krill and salp density data taken from the rawKRILLBASE are used in their raw form, i.e. the data owners have standardised these densities to a common sampling method but have stipulated that the standardised data may not be used as part of our study. For the purposes of this study abundance data were converted from ind. m^{-2} to ind. L^{-1} , using the depth over which the samples were gathered. All the rawKRILLBASE krill and salp data were converted to carbon biomass (see Sect. 2.2).

The Coastal & Oceanic Plankton Ecology, Production, & Observation Database (COPEPOD; O'Brien, 2005; Table 1d) is a global database and includes a myriad of macrozooplankton taxa: salps, doliolids, pyrosomes, ctenophores, cnidaria, scyphozoa, hydrozoa, anthozoa, pelagic molluscs, pelagic polychaetes, chaetognatha, amphipods, mysids, stomatopods, decapods, euphausiids, and appendicularia. The COPEPOD database is maintained by the United

States of America National Marine Fisheries Service (USA NMFS), and includes its own ecosystem survey and sampling programs, historical plankton data search and rescue work, institutional and project data, as well as individual submissions from researchers outside NMFS. In total there were 369 567 measurements of abundance within COPEPOD relating to macrozooplankton as defined above. Abundance measurements within COPEPOD were standardised to individuals per m^3 by NMFS. For the purposes of this study, abundance data were converted from ind. m^{-3} to ind. L^{-1} . It was possible to convert 15 297 abundance data to carbon biomass (see Sect. 2.2). COPEPOD is the only database considered here which included macrozooplankton abundance and biomass as a function of depth.

Where possible we have retained all taxonomic data associated with abundance and biomass values in the original raw datasets. While this is only directly important if we are converting abundance to biomass on a species-by-species level, we hope that this dataset will continue to be added to in terms of abundance, biomass and associated metadata and may be used for a wide variety of applications, a point of departure for research that requires more specific information than simple “macrozooplankton biomass” values.

2.2 Biomass conversion

The HOTS dataset contained biomass (mg C m^{-2}) data and it was only necessary to convert to common units and volume as described above. The BATS dataset contained both wet and dry mass biomass data. Dry mass data were converted using a dry mass to carbon conversion, $\text{C} = 0.36 \times \text{dry weight}$ (Madin et al., 2001, as documented in Steinberg et al., 2011).

For the rawKRILLBASE database the mean body mass of *Euphausia superba* was estimated as 140 mg dry mass ind. $^{-1}$ (A. Atkinson, personal communication, 2007) and was calculated from a length frequency distribution dataset containing 535 581 length measurements of *Euphausia superba* recovered from scientific hauls between October and April from 1926–1939 and 1976–2006 (Atkinson et al., 2009). A dry mass to carbon conversion of 0.45 (Pakhomov et al., 2002 and references therein) was used to convert dry mass to carbon per individual. The mean body mass, for both species of salp, was estimated at 120 mg dry mass ind. $^{-1}$ (Dubischar et al., 2006). A dry mass to carbon conversion of 0.2 (weighted

Table 1b. Summary of data points for *Euphausia superba* abundance in the Southern Ocean from the rawKRILLBASE dataset (Atkinson et al., 2004).

Principal Investigator	No. of abundance data	Depth range (m)	Country	Institute/Source/Reference
Loeb	1958	0–282	USA	Loeb et al. (2010) US Antarctic Marine Living Resources Field Season Reports
Atkinson	328	0–380	UK	BAS (unpublished data)
Atkinson	119	0–500	UK	Atkinson and Peck (1988)
Ward	59	0–200	UK	Ward et al. (2005)
Atkinson	55	0–400	UK	Ward et al. (2006) and BAS (unpublished data)
Chiba	25	0–268	Japan	Chiba et al. (1999) Japanese Antarctic Research Expedition
Chiba	35	0–150	Japan	Chiba et al. (2001) Japanese Antarctic Research Expedition
Pakhomov	1042	0–500	Russia	Pakhomov et al. (2002)
Hosie	66	0–200	Australia	Hosie et al. (2000) BROKE Survey, Australian Antarctic Division
Pakhomov	198	0–383	South Africa	Pakhomov et al. (2002)
Anadon	99	0–200	Spain	Anadon and Estrada (2002) FRUELA
Hosie	64	0–200	Australia	Hosie and Cochran (1994) Australian Antarctic Division
Hosie	83	0–200	Australia	Hosie et al. (1997) Australian Antarctic Division
Atkinson	1176	0–540	UK	Marr (1962); Atkinson et al. (2009); Foxtan (1966) Discovery Expeditions
Atkinson	8	0–875	Japan	Casareto and Nemoto (1986) SIBEX BIOMASS Programme
Ross and Quetin	631	0–460	USA	Ross et al. (2008) US Palmer LTER Program
Atkinson	125	0–300	Poland	Jazdzewski et al. (1982)
Siegel	1692	0–1200	Germany	Siegel (2005)
Siegel	117	0–2700		Siegel et al. (2004) CCAMLR
Siegel	147	0–213	Germany	Hunt et al. (2011)

average of solitary and aggregate forms) (Dubischar et al., 2006) was used to convert dry mass to carbon per individual.

For the COPEPOD database, only species-specific abundance data that had a corresponding species-specific carbon conversion were converted to biomass. Body mass data were compiled for a number of the species found in the COPEPOD dataset (for individual species conversions see Table 2). We used the macrozooplankton biomass conversion dataset database to generate the biomass data from species-specific abundance data. Adult mean body mass of the species was used for the conversion of abundance (ind. L^{-1}) data to carbon. Species-specific conversions were preferred even though it was only possible to convert a small fraction ($\sim 4\%$) of the dataset to carbon biomass.

Data were gridded using the original entries for latitude, longitude and month from all datasets. The mean depth of the sampling depth range of each macrozooplankton concen-

tration was used as sample depth. Macrozooplankton concentrations in ind. L^{-1} and $\mu\text{g C L}^{-1}$ were binned on the 4-dimensional World Ocean Atlas grid. This is a monthly grid with horizontal resolution of $1^\circ \times 1^\circ$ and 33 vertical levels resolved to 5 m in surface waters, increasing to 500 m from 2000 m downwards. Only data that are gridded in the top 350 m of the ocean is used for calculation of global epipelagic macrozooplankton annual average biomass but all available data (all depths) on macrozooplankton have been included in both ungridded and gridded datasets.

2.3 Quality control

Chauvenet's criterion for data rejection was used for the removal of high outliers from both abundance and biomass data (Glover et al., 2011; Buitenhuis et al., 2013). A normal distribution of the data is assumed and data are rejected when the

Table 1c. Summary of data points for *Salpa thompsoni* and *Ihlea racovitzai* abundance in the Southern Ocean from the rawKRILLBASE dataset (Atkinson et al., 2004).

Principal Investigator	No. of data Abundance	Depth range (m)	Country	Institute/Source/Reference
Loeb	1130	0–210	USA	Loeb et al. (2010) US Antarctic Marine Living Resources Field Season Reports
Atkinson	169	0–2236	Australia	Australian ANARE Research Notes
Atkinson	97	0–2500	UK	Atkinson and Peck (1988) and BAS (unpublished data)
Atkinson	409	0–2030	UK	BAS (unpublished data)
Atkinson	59	0–200	UK	Ward et al. (2005) and BAS (unpublished data)
Atkinson	55	0–400	UK	Ward et al. (2006) and BAS (unpublished data)
Chiba	24	0–268	Japan	Chiba et al. (1999) Japanese Antarctic Research Expedition
Chiba	44	0–150	Japan	Chiba et al. (2001) Japanese Antarctic Research Expedition
Pakhomov	66	0–1000	Russia	Pakhomov et al. (2002)
Hosie	66	0–200	Australia	Hosie et al. (2000) BROKE Survey, Australian Antarctic Division
Pakhomov	40	0–500	Germany	ANTARKTIS XIII5b
Pakhomov	9	0–394	South Africa	DEIMEC
Pakhomov	393	0–403	South Africa	Pakhomov et al. (2002)
Pakhomov	1423	0–1000	Russia	Pakhomov et al. (2002)
Anadon	99	0–200	Spain	Anadon and Estrada (2002) FRUELA
Hosie	147	0–200	Australia	Hosie and Cochran (1994) Australian Antarctic Division
Atkinson	2659	0–1300	UK	Foxtan (1966) Discovery Expeditions
Atkinson	13	0–1050	Japan	Casareto and Nemoto (1986) SIBEX BIOMASS Programme
Nishikawa	50	0–99	Japan	Nishikawa et al. (1995)
Atkinson	32	0–500	Japan	Nishikawa and Tsuda (2001)
Ross and Quetin	633	0–460	USA	Ross et al. (2008) Palmer LTER Program
Atkinson	96	0–300	Poland	Jazdzewski et al. (1982)
	63	0–245	Poland	Witek et al. (1985)
Siegel	1731	10–200	Germany	Siegel (2005)
Siegel	119	0–2700		Siegel et al. (2004) CCAMLR
Siegel	93	0–200	Germany	Hunt et al. (2011)

probability of deviation from the mean is less than $1/(2n)$, where n is the number of data points. Chauvenet's criterion could not be applied directly because the macrozooplankton abundance and biomass data were not normal distributed but instead ranged from low or undetectable concentrations, to very high value bloom events. Chauvenet's criterion was applied to all non-zero log-transformed data, which had a near normal distribution. Data with zero values for abundance and biomass are included in the dataset as they represent an absence of macrozooplankton, but could not be included in the log-transformed data.

The mean \bar{x} and the standard deviation σ of the log-transformed data are calculated and used to calculate the critical value x_c . One half of $1/(2n)$ is used because the Chauvenet's criterion is a two-tailed test; however, we only rejected data on one tail, the high side. All log-transformed data with values higher than $\bar{x} + x_c$ are rejected (Luo et al., 2012).

2.4 Database formats

The original datasets are available as Excel files with full details of data sources and conversions. We also provide a

Table 1d. Summary of data points for macrozooplankton abundance from the COPEPOD dataset O'Brien.

COPEPOD Dataset	No. of abundance data	Depth (m)		Project	Institute	NODC accession no.
		Range	Mean			
ALMIRANTE SALDANHA Collection	785	0–274	72	–	Max-Planck-Institut Fuer Meteorologie	0000942
AtlantNIRO plankton	268	0–100	50	–	Atlantic Research Institute of Fishing Economy and Oceanography	9600039
BCF – POFI	454	0–260	55	Pacific Oceanic Fisheries Investigations	Maritime Regional Administration of Hydrometeorology	0051848
Biological Atlas of the Arctic Seas 2000: Plankton of the Barents and Kara Seas	30346	0–2828	89	–	–	0000283
BIOMAN	1888	0–100	43	BIOMass of ANchovy	–	9700075
Brodskii (1950)	312	0–3884	536	–	–	0064569
CHIU LIEN Collection	236	0–150	40	–	National Taiwan University Institute of Fishery Biology, Taipei	0000095, 0000097
CINECA I	65	0–600	184	Cooperative Investigations of the Northern-part of the Eastern Central Atlantic	Cent Nat pour L'Exploit des Ocg Bur Nat des Donn Ocg	9000076
CINECA II	2954	0–100	50	Cooperative Investigations of the Northern-part of the Eastern Central Atlantic	Mediterranean Marine Sorting Center	0000088
CINECA IV	696	0	0	CORiolis-INDONesia	Office de la Recherche Scientifique et Technique Outre Mer	0000527
CSK	52121	0–1023	56	Cooperative Study of the Kuroshio and adjacent regions	Fisheries Research and Development Agency, Republic of Korea Institute of Marine Research, Jakarta Kagoshima University Faculty of Fisheries Nagasaki University Fisheries Institute Tokyo University of Fisheries Kominato Marine Biological Lab, Awa-kominato Shimonoseki University of Fisheries Tokyo University, Ocean Research Institute Hokkaido University Faculty of Fisheries, Hakodate Hong Kong Fisheries Research Station, Hong Kong Fisheries Biology Unit, University of Singapore Marine Fisheries Laboratory, Department of Fisheries Philippines Fisheries Commission, Department of Agriculture and Natural Resources Japan Meteorological Agency Fisheries Research Institute, Malaysia Fisheries Research Institute, Republic of Korea Marine Fisheries Laboratory, Thailand Pacific Scientific Research Institute of Fisheries & Oceanography, USSR National Oceanography Institute, Vietnam SouthEast Asian Fisheries DEvelopment Center, Singapore	9500141
Drift Station Alpha	123	0–2000	401	International Geophysical Year	Scripps Institution of Oceanography	0000810
EASTROPAC	7336	0–277	101	Eastern Tropical Pacific 1967–1968	Smithsonian Oceanographic Sorting Center	9500089, 9500090
EASTROPIC	823	0–427	103	–	Maritime Regional Administration of Hydrometeorology – Rosgidromet	9700300, 9700074
EcoMon-SOOP (Gulf of Maine)	2012	10	10	MARine Resources Monitoring; Assessment & Prediction 1977–1987	Bermuda Container Line Ltd.	0051894
EcoMon-SOOP (Mid-Atlantic Bight)	2773	10	10	–	–	–
ECOSAR II	742	0–210	46	–	Federal University of Rio Grande	–
Finnish Baltic Sea Monitoring	2210	0–440	66	Helsinki Commission Baltic Monitoring	Finnish Environment Institute	–
GAVESHANI Collection	189	0–200	100	–	National Institute of Oceanography, Goa, India	0000941

Table 1d. Continued.

COPEPOD Dataset	No. of abundance data	Depth (m)		Project	Institute	NODC accession no.
		Range	Mean			
GILL Zooplankton Collection	10723	0–200	9	–	Woods Hole Oceanographic Institute & US Fish & Wildlife Service	9700101, 9700102, 9700103
Gulf of California 1983–1984	254	0–259	94	–	Hamburgische Schiffbauversuchsanstalt	0000911
HAKUHO MARU Collection	1003	0–2300	373	–	Japan Meteorological Agency	970003, 970005, 970007, 9700303
Historical data from the Japanese Oceanographic Data Center (JODC)	10167	0–300	98	–	–	9700311
IMECOCAL	13261	0–243	94	Investigaciones MEXicanas de la COriente del California	Hamburgische Schiffbauversuchsanstalt	0000911, 0000912, 0000913
IMR Norwegian Sea Survey	254	0–2680	121	–	Institute of Marine Research, Norway	0049894
INSTOP-6	2533	0–100	50	–	Mediterranean Marine Sorting Center	0000561
IIOE	91032	0–880	94	International Indian Ocean Expedition	–	9400059
IROP-4	797	0–100	50	–	Mediterranean Marine Sorting Center	0000561
JARE	19271	0–833	30	Japanese Antarctic Research Expedition	SAFHOS	0000039
JMA North Pacific Surveys	1795	0–999	73	–	–	0000051, 0000070, 0000398
Koyo Maru–Brazil	665	0–1300	68	–	Shimonoseki University of Fisheries	970003, 970005, 970007, 9700303
Koyo Maru Indian Ocean	75	0–199	41	–	–	–
Marion Dufrense	44	0–300	71	–	–	0000940
Minoda 1967	175	0–200	74	Office of US Naval Research	–	0000978
NEWP	2587	0–515	114	NorthEast Water Polynya project	Rosenstiel School of Marine & Atmospheric Sciences, University of Miami	9700074, 9700300
NMFS-PFEL Zooplankton	88	0–140	29	–	Institute Experimental of Meteorology, Obninsk, Russia	9800046
NODC Zooplankton Database	47366	–	–	–	–	Multiple
North Pacific Survey	3482	0–400	87	Cooperative Survey of the North Pacific US-Japan-Canada	Fisheries Research Board of Canada – Pacific Oceanographic Group, Nanaimo, British Columbia, USA & Japan Hydrographic Association – Marine Information Research Center	9700074, 9700300
Pacific Salmon Investigations	158	0–500	115	Pacific Salmon Investigations	Fish & Wildlife Service, Seattle, Washington, USA	9700101
Pearl Harbour 1946	5727	0–10	5	–	–	0051848
Pelagic Ecosystems of the Indian Ocean	6417	0–262	51	–	Institute of Biology of the Southern Seas	0001310
Pelagic Ecosystems of the Mediterranean	29654	0–1710	63	–	–	–
Pelagic Ecosystems of the Tropical Atlantic	38	0–2828	92	–	–	–
Pioneer Cruise 66	6835	0–600	185	–	University of Hawaii, Honolulu	9800165
SAHFOS-CPR Atlantic Ocean	891	10	10	–	SAHFOS	0000301

Table 1d. Continued.

COPEPOD Dataset	No. of abundance data	Depth (m)		Project	Institute	NODC accession no.
		Range	Mean			
SSRF-312 North Pacific and Bering Sea Oceanography 1958	835	0–1710	113	–	Maritime Regional Administration of Hydrometeorology – Rosgidromet	9700074, 9700300
SSRF-377	504	0–212	45	–	–	–
SSRF-619	600	0–1291	40	–	Japan Hydrographic Association – Marine Information Research Center	9600088
Sub-arctic Frontal Zone Zooplankton	223	0–150	73	–	Alaska Fisheries Science Center	9700074
TASC	529	0–710	207	Trans-Atlantic Study of Calanus	Western Washington State College	0000566
USCG Chelan	617	0–25	11	–	Scripps Institution of Oceanography,	9500110
Volcano-7 Zooplankton	1013	0–1500	246	–	University of Rhode Island – Graduate School of Oceanography, Narragansett, Rhode Island, USA	9500081
WEBSEC	569	0–533	22	Western Bering Sea Ecological Cruise	–	9700074, 9700300
Zulfiquar Cruises	417	0–212	44	–	Meteorological Department, Pakistan	9400163

gridded NetCDF data file on the World Ocean Atlas grid ($1^\circ \times 1^\circ \times 33$ depths \times 12 months), which contains both carbon biomass for evaluation of ocean biogeochemical models and a number of other variables, including abundance, standard deviations and non-zero biomass (see Buitenhuis et al., 2013 for details). It should be noted that all figures presented here have been created using data from the Excel files (pre-gridding) in order to showcase the original datasets. Very similar figures can be created using the gridded NetCDF data file.

We have included abundance data in the dataset for a variety of reasons: (1) most macrozooplankton data are recorded in abundance terms and abundance data are often used in the calculation of biomass; (2) abundance may be used as an indication of where the animals are, and in what quantities, i.e. it may be used in a qualitative sense; (3) additions to both the abundance dataset and the conversion dataset will make it possible to convert more abundance data to biomass in the future; and (4) we have carefully separated the data collection and data processing steps so that the data would be easily adaptable for purposes other than biogeochemical model validation.

3 Results and discussions

3.1 Results of quality control

Both abundance and biomass have distributions close to normal after the data are log-transformed. After applying Chauvenet's criterion to the log-transformed abundance data, only 49 data points from the abundance dataset and 32 data points from the biomass dataset are rejected as outliers, their values being higher than the critical values for both datasets. The reasoning behind the rejection of the higher values is thus: if two hypothetical databases were constructed that were as

similar as possible to each other, but in which one is stochastically skewed with respect to the other because of an unrepresentative number of extremely high values, then those values are rejected to remove the skew. So while we think the data are values that reflect reality, we are consciously working to remove any skew. Abundance and biomass outliers originate from bloom taxon/species and occur in Northern and Southern Hemisphere spring/summer. These are most likely real values associated with a bloom rather than a problem with the sampling or lack of metadata available.

Sampling protocols, handling, preservation and measurement techniques have not been considered when removing outliers. Within the HOTS, BATS and rawKRILLBASE datasets, and the various projects within COPEPOD data, these variables are assumed consistent but are most likely not uniform across datasets and projects. Issues related to sampling such as the inherent variability of field populations (Landry et al., 2001), net mesh size, type of net, net avoidance, seasonal/diel vertical migrations, sample handling (e.g. sample splitting), size fractionation, and sample analysis – i.e. all sources of random sampling error – were considered to have a greater effect than the sampling issues across projects/datasets.

3.2 Data description

3.2.1 Abundance data

A total of 387 750 abundance data points (280 631 non-zero) between all four datasets cover the Indian Ocean, Barents Sea, Southern Ocean, west Atlantic, north eastern and north western Atlantic, the Caribbean, eastern equatorial Pacific, western North Pacific and Hawaii. There are few abundance data in the tropical and temperate south Pacific, and in the tropical north Atlantic (Fig. 1).

Table 2. Mean body mass values used to convert COPEPOD species abundance (ind. L⁻¹) to biomass (μmol CL⁻¹). Notes: *n* = number of data points per species, Min. = minimum body mass, Max. = maximum body mass, Stdev. = Standard deviation; all body mass units are μmol C. Data from Hirst (2003) and Moriarty (2009).

Group	Species	Body mass					
		<i>n</i>	Min.	Max.	Mean	Stdev.	% Stdev.
Ctenophore							
	<i>Bolionopsis infundibulum</i>	12	131	10 764	3896	4062	104
	<i>Pleurobrachia pileus</i>	31	13	571	166	181	109
Scyphozoa							
	<i>Aurelia aurita</i>	394	0	97 590	5728	11 724	204
	<i>Cyanea capillata</i>	25	114	135 159	10 283	28 330	276
Hydrozoa							
	<i>Aglantha digitale</i>	27	13	461	139	110	80
	<i>Eutonina indicans</i>	24	30	1013	311	230	74
	<i>Philalidium gregarium</i>	44	31	297	125	62	50
	<i>Sarsia princeps</i>	8	39	230	101	73	73
	<i>Sarsia tubulosa</i>	22	1.1	54	19	16	82
	<i>Stomotoca atra</i>	26	75	392	164	75	46
	<i>Agalma elegans</i>	7	0.5	4.6	1.6	1.4	89
Pelagic mollusc							
	<i>Clione limacina</i>	31	12	1819	198	357	180
	<i>Diacrea trispinosa</i>	1	293	293	293	–	–
	<i>Diphyes antarctica</i>	3	378	779	520	225	43
	<i>Limacina helicina</i>	5	13	25	17	4.9	28
Pelagic polychaete							
	<i>Tomopteris septentrionalis</i>	1	174	174	174	–	–
Chaetognath							
	<i>Parasagitta elegans</i>	530	0	187	54	44	82
	<i>Parasagitta enflata</i>	3	18	38	31	11	36
Amphipod							
	<i>Cyphocaris challengeri</i>	1	176	176	176	–	–
	<i>Hyperia galba</i>	5	139	401	258	106	41
	<i>Parathemisto japonica</i>	55	0.5	492	96	128	135
	<i>Themisto libellula</i>	1	80	80	80	–	–
	<i>Phronima sedentaria</i>	4	41	111	72	35	49
	<i>Themisto pacifica</i>	1	47	47	47	–	–
Mysid							
	<i>Acanthomysis pseudomacropsis</i>	2	89	124	106	25	23
Decapod							
	<i>Lucifer typus</i>	1	2.3	2	2.30	–	–
Euphausiid							
	<i>Euphausia krohnii</i>	2	8.2	9	8.60	1.0	11
	<i>Euphausia pacifica</i>	234	0.2	1674	195	186	96
	<i>Meganctiphanes norvegica</i>	2	1477	1477	1476	–	–
	<i>Thysanoessa inermis</i>	31	0.8	1495	373	377	101
	<i>Thysanoessa longipes</i>	1	321	321	321	–	–
	<i>Thysanoessa raschi</i>	5	51	309	166	108	65
	<i>Thysanoessa spinifera</i>	6	26	1330	619	548	89
Thaliacea							
	<i>Salpa fusiformis</i>	6	14.3	98	53.50	29.0	54
	<i>Salpa maxima</i>	10	1.3	23	9.00	8.0	88
	<i>Thalia democratica</i>	28	0.2	173	11.40	34.0	298
Appendicularia							
	<i>Fritillaria borealis sargassi</i>	8	0	0.02	0.01	0.01	100
	<i>Fritillaria haplostomai</i>	8	0	0.02	0.01	0.01	100
	<i>Oikopleura dioica</i>	151	0	5.5	0.9	1.42	158
	<i>Oikopleura longicauda</i>	18	0	1	0.20	0.0	–

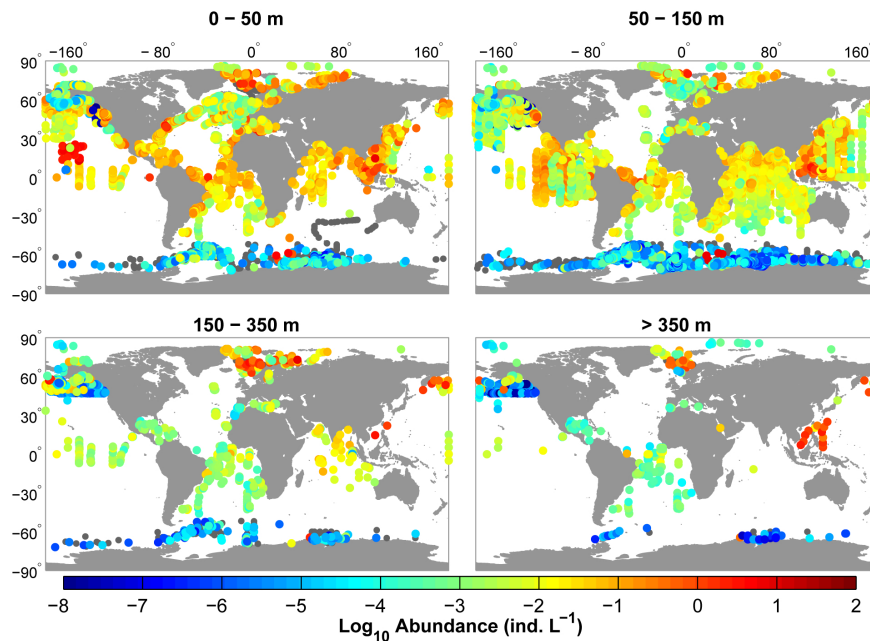


Figure 1. Global distribution of macrozooplankton abundance (ind. L⁻¹) at different depths. Grey points represent zero values.

The highest incidences of macrozooplankton abundance observations (Fig. 2a) are found at $\sim 60^\circ$ in both the Northern and Southern Hemispheres. The number of observations falls off to the higher latitudes, with no observations of abundance south of 75° S. While the distribution of observations at high latitudes is the same in both hemispheres, there is an asymmetry in the lower latitudes (50° S– 50° N) where the Southern Hemisphere has fewer observations, peaking at $\sim 30^\circ$ S.

The temporal distribution of macrozooplankton abundance presented here covers 84 yr from 1926 to 2010 (Fig. 2b). Sampling did not occur between 1939 and 1950, with the exception of 1946; most observations were collected towards the late 1960s. There are far fewer observations between 2005 and 2010 (Fig. 2b), either because not all data have been archived within a data repository or in a database or there have been a decline in this type of sampling activity.

The mean sampling depth of macrozooplankton abundance presented here is 85 m (± 90 m standard deviation) and ranges from the surface to ~ 2500 m (Fig. 2d). Most data are concentrated in the top 500 to 1000 m. Macrozooplankton may be found throughout the water column; species of macrozooplankton that live in the epipelagic, or sunlit surface waters of the ocean, are usually found in the top 350 m. The mean depth of sampling is well suited for investigating the concentrations of epipelagic macrozooplankton, the main focus of this study. Macrozooplankton abundance data have been collected in all months of the year in both the Northern and Southern Hemispheres (Fig. 2e and f).

The majority of the data in the Southern Ocean belongs to the three species *Euphausia superba*, *Salpa thompsoni* and

Ihlea racovitzai (data from the rawKRILLBASE dataset), whereas the majority of the data for the remainder of the global ocean is representative of the whole macrozooplankton community. This accounts partially for the difference in abundances in the northern and equatorial latitudes and those of the Southern Ocean. It is difficult to quantify the proportion of total global macrozooplankton abundance that is made up of the three species mentioned above. These species have distributions in the Southern Hemisphere but are predominant in the Antarctic waters of the Southern Ocean. They are not cosmopolitan in the global ocean. This is discussed further in Sect. 3.3.2.

There is potential ambiguity about whether large copepods are included in meso- or macrozooplankton sampling. Although we have used a cut-off of 2 mm adult body size for other taxonomic groups, previous work on mesozooplankton has used cut-off sizes between 5 and 30 mm to delimit mesozooplankton (Buitenhuis et al., 2006; see supplementary table 3). To prevent double counting with the MAREDAT mesozooplankton database (Moriarty and O'Brien, 2013), we have excluded copepod species that were available in the COPEPOD database. However, in the HOT and BATS databases, we only had access to the total macrozooplankton biomass data, which did include copepods greater in size than 2 mm. Large copepods can avoid nets with a small mesh size, such as are used for sampling small copepods (typically $200\ \mu\text{m} < \text{mesh size} < 330\ \mu\text{m}$; Harris et al., 2000; Moriarty and O'Brien, 2013), but this under sampling has not been comprehensively quantified. We were therefore unable

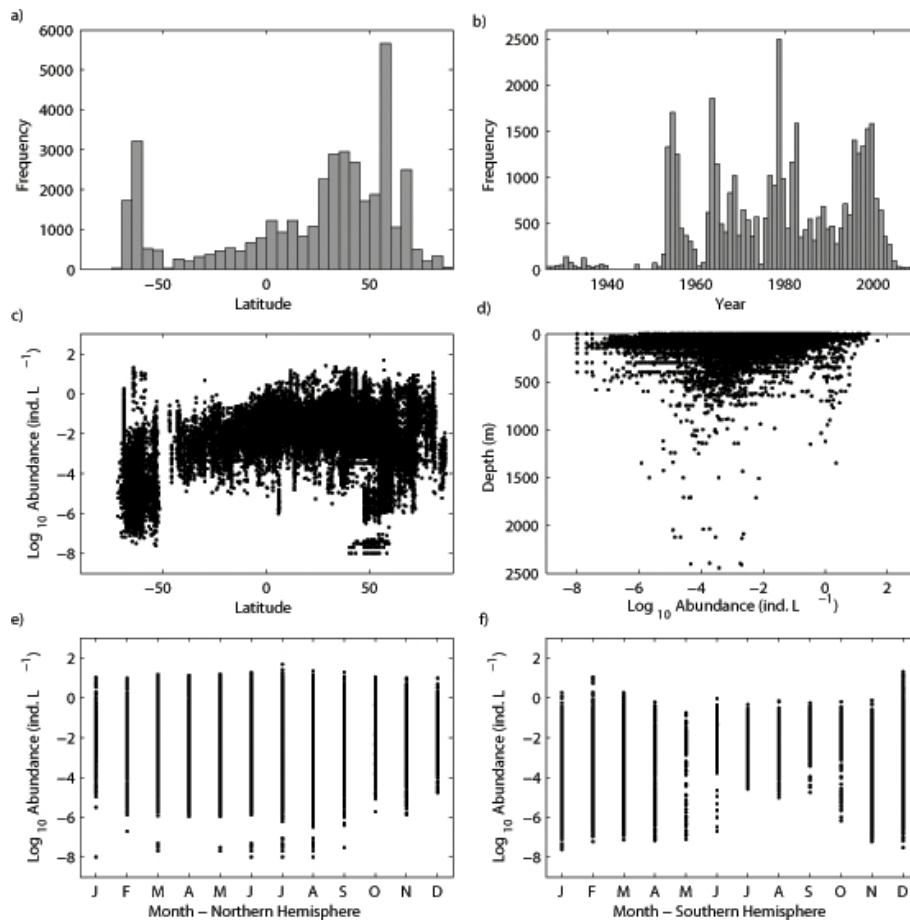


Figure 2. Description of macrozooplankton abundance: (a) latitudinal distribution of observations, (b) yearly distribution of observations, (c) latitudinal distribution of abundance, (d) depth distribution of abundance, (e) monthly abundance distribution in the Northern and (f) Southern Hemispheres.

to estimate whether there is double counting or a gap between the mesozooplankton and macrozooplankton datasets.

3.2.2 Biomass data

A total of 36 268 biomass data points (28 104 non-zero) (Fig. 3) between all four datasets leaves much of the global ocean uncharacterised, with the obvious exception of the Southern Ocean, Barents Sea, Hawaii and Bermuda. All rawKRILLBASE and BATS data were converted to biomass, along with the majority of HOTS data. Only ~ 15 000 COPEPOD abundance data were converted, which is why biomass has much less spatial coverage than abundance. In vast areas of the global ocean there is little or no information on biomass. This is a direct result of only converting abundance to biomass using species-specific conversions. This approach was necessary as bulk conversions of abundance to biomass as yet are not sophisticated enough to account for many of the variables, e.g. region, season, life history, food concentration and food quality, that are important to the amount of carbon in any particular individual or species.

Carbon values are a much more useful measurement than abundance data; however there are no published generic relations for the conversion of macrozooplankton abundance to biomass. This type of conversion has not been included in the analysis as the conversion factors are too general, and the large deviation of the bulk conversions from the species-specific conversions show the former would severely distort the results.

Efforts to assemble a comprehensive listing of conversions for macrozooplankton by groups such as the ICES Working Group on Zooplankton Ecology have been ongoing for years. The scientists involved, experts in the field, find this effort overwhelming, and the differences due to regions, seasons and life stage (length to body composition) make the equations hugely variable. A blanket global conversion, without a better conversion estimate, is not the best way to convert abundance to biomass. Without valid conversion equations from abundance (number of macrozooplankton per sample) to biomass (mass of biomass to sample), there is a need for length frequency data, mass data and carbon data. In the

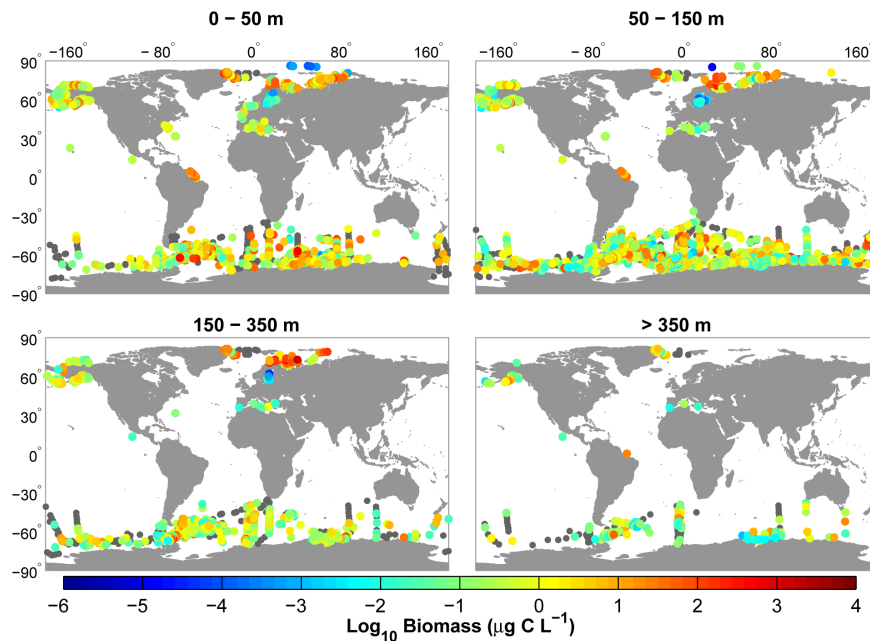


Figure 3. Global distribution of macrozooplankton biomass (μgCL^{-1}) at different depths. Grey points represent zero values.

majority of cases these data are not available for macrozooplankton species or the entire macrozooplankton size class or cohort of species.

Latitudinal distribution of biomass echoes that of the abundance data, with peaks in observations at $\sim 60^\circ$ in both hemispheres but with fewer in the north (Fig. 4a). There are greater gaps in the biomass data, with few data between 40° S and 20° N. The temporal distribution of macrozooplankton biomass also echoes largely what has been said above for the abundance data. There are, however, larger gaps in the distribution, with no data between 1939 and 1950, and few observations between the late 1950s and early 1970s (Fig. 4b). In the mid 1970s biomass observations increase and remain higher, only occasionally dropping down to pre-1975 values. Much work was done on the chemical composition of macrozooplankton species in the mid 1970s to early 1990s, which may be one explanation for the increase in biomass data associated with the end of the 20th century.

The mean sampling depth of macrozooplankton biomass is 88 m (± 104 m standard deviation) and ranges from the surface to ~ 2500 m (Fig. 4c and d). Macrozooplankton biomass data have been collected in all months of the year in the Northern Hemisphere but there are no biomass data associated with the winter months (July, August and September) in the Southern Hemisphere (Fig. 4e and f).

3.3 Global estimates

Here, we use the gridded dataset to determine the depth integrated global values for macrozooplankton abundance and biomass in the top 350 m. Ninety-three percent of total abun-

dance and ninety-nine percent of total biomass are found in the top 350 m of the global ocean. We have specifically chosen data gridded in the top 350 m of the ocean for use in these calculations of global epipelagic macrozooplankton annual average abundance and biomass as this is where macrozooplankton are usually found and because we have considerable coverage down to that depth (Figs. 2d and 4d).

3.3.1 Abundance

Global abundance to a depth of 350 m has a mean of $0.018 \text{ ind. L}^{-1}$, a median of $0.0006 \text{ ind. L}^{-1}$ and a standard deviation of 0.12 ind. L^{-1} (Table 3a). The fact that the means are much higher than the median and that the standard deviation is high indicates that very high concentrations of abundance are occasionally observed (Luo et al., 2012). Abundance data show no clear latitudinal patterns. Mean and median latitudinal abundance values north of 15° S are of a similar range whereas the latitudinal abundances to the south have lower means and much lower medians. Differences in the number of samples (n) make it difficult to fully decipher if there are any broad latitudinal patterns in abundance concentrations.

3.3.2 Biomass

Global macrozooplankton biomass to a depth of 350 m has a mean of $8.4 \mu\text{gCL}^{-1}$, a median of $0.2 \mu\text{gCL}^{-1}$ and a standard deviation of $63.5 \mu\text{gCL}^{-1}$ (Table 3b). Again, as in the abundance data, there are differences in the order of magnitude, within the mean, and within the median values between

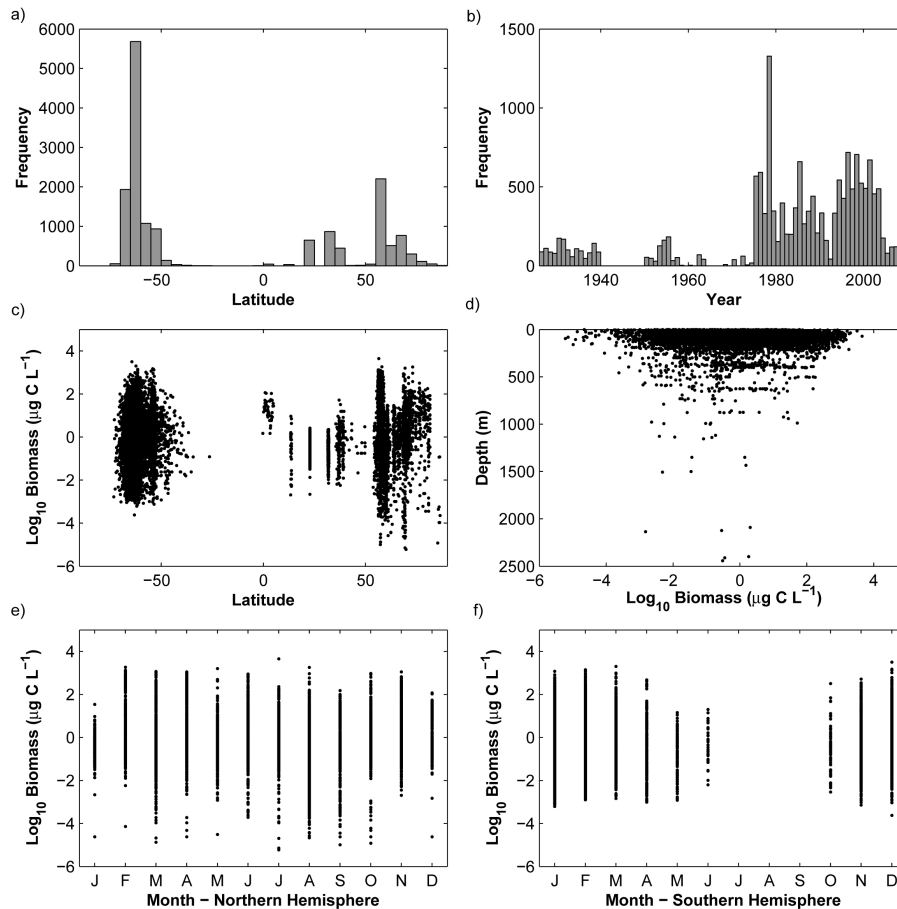


Figure 4. Description of macrozooplankton biomass: (a) latitudinal distribution of observations, (b) yearly distribution of observations, (c) latitudinal distribution of biomass, (d) depth distribution of biomass, (e) monthly biomass distribution in the Northern and (f) Southern Hemispheres. There are no data for biomass in the Southern Hemisphere winter months (July, August, September).

Table 3a. Global and latitudinal band values for the gridded macrozooplankton abundance data.

Latitude	Abundance (ind. L ⁻¹)					
	<i>n</i>	Min.	Max.	Mean	Median	± std.
Global	21 293	9.19×10^{-9}	5.11	0.018	0.0006	0.123
90–40° N	7216	1.00×10^{-8}	5.11	0.034	0.0017	0.179
40–15° N	3537	1.00×10^{-8}	2.67	0.023	0.0030	0.122
15° N–15° S	3547	2.39×10^{-7}	3.70	0.012	0.0015	0.095
15–40° S	1039	4.06×10^{-7}	0.07	0.001	0.0002	0.004
40–90° S	5954	9.19×10^{-9}	0.83	0.002	2.82×10^{-6}	0.023

latitudinal bands. Median biomass values to the south of 15° S are lower than their northern counterparts.

In Table 3b there is a difference of two orders of magnitude between the biomass values for 40 to 90° N and 40 to 90° S. This may be explained by differences in the type of data in the datasets associated with each of these regions. The rawKRILLBASE data, three species, are most of the data in the Southern Ocean, while the COPEPOD data, representa-

tive of the entire macrozooplankton community, are found throughout the global ocean. The rawKRILLBASE data are composed of three species that are the predominant macrozooplankton species in the Southern Ocean and can make up 90 % of the biomass (Witek et al., 1985). Depending on temporal and spatial scales, *Euphausia superba* and *Salpa thompsoni/Ihlea racovitzai* are estimated to account for between 30–90 % of the biomass in the Southern Ocean. Both

Table 3b. Global and latitudinal band values for the gridded macrozooplankton biomass data.

Latitude	Biomass (μgCL^{-1})					
	<i>n</i>	Min.	Max.	Mean	Median	\pm std.
Global	8146	6.00×10^{-6}	3967	8.38	0.15	63.46
90–40° N	2147	6.00×10^{-6}	3967	16.62	0.42	114.39
40–15° N	270	0.0033	13.32	0.76	0.26	1.27
15° N–15° S	42	0.0026	116.8	18.06	10.23	25.96
15–40° S	44	1.60×10^{-2}	4.58	0.29	1.00×10^{-7}	0.88
40–90° S	5643	2.20×10^{-4}	582.3	5.60	0.08	28.14

species have bloom capabilities and patchy distributions and relatively high biomass in the Southern Ocean.

The median value, $0.2 \mu\text{gCL}^{-1}$, for global epipelagic macrozooplankton biomass to a depth of 350 m has been used to estimate an annual average epipelagic macrozooplankton biomass of 0.02 Pg C. There are two reasons why we have picked 350 m for this calculation: (1) biomass data are more evenly distributed at this depth, and (2) below this depth macrozooplankton have different metabolic rates. PFT models with a macrozooplankton component usually only consider macrozooplankton in the epipelagic surface ocean and this estimate has been tailored to this end. A number of caveats accompany this estimate of annual average epipelagic macrozooplankton biomass: (1) the data are not uniformly distributed spatially or temporally because some areas are not covered and because there is a slight bias in the biomass data against winter values in the Southern Hemisphere; and (2) data are not proportionally distributed between the various biomes of the global ocean. We have used the median value of the epipelagic data to calculate the annual average of epipelagic macrozooplankton biomass (the value for macrozooplankton from all depths is given in Table 2). We have chosen the median as an appropriate value to base the annual average biomass of epipelagic macrozooplankton on as it indicates a midpoint value rather than a value skewed by the occasional very high value, as indicated by the high values for the mean and standard deviation.

3.4 Effects of conversion factors

The limited availability of carbon conversion data is one of the major limitations of this dataset. A general conversion from dry mass to carbon mass was used for the BATS data. It was not thought to be appropriate to apply a general carbon conversion of this type to the abundance data (see above; see Mizdalski, 1988) because macrozooplankton span a range of diverse phyla, and there is huge variety within and between species; temperature, food quality, and life stage all affect the chemical composition and size of the organism. As a result the body mass of macrozooplankton spans at least 8 orders of magnitude. The uncertainty in biomass is greatest when there is no indication of body mass, body length or life stage.

There is difficulty in assessing the error on the global biomass values stated above. Only one dataset, HOTS, gives an indication of the combined sampling and conversion error (standard deviation) of 25 %. The standard deviation associated with the BATS and rawKRILLBASE dataset biomass conversion values are 17 % and 5 %, respectively. Conversion errors associated with the species-specific biomass conversions within COPEPOD range from 11 % (standard deviation as a percentage of the mean body mass) to 298 % (Table 2). The wide range in standard deviation associated with the COPEPOD species-specific conversions shows the wide range of body masses within macrozooplankton species.

4 Conclusions and recommendations

The global biomass of macrozooplankton is estimated and presented here alongside partial coverage for macrozooplankton abundance and biomass distribution. This work is presented as a first step towards a quantitative analysis of global distribution of macrozooplankton biomass. From the present dataset we estimate a biomass median of $0.2 \mu\text{gCL}^{-1}$ (= 0.02 Pg C annual average epipelagic macrozooplankton biomass) and a standard deviation of $63.5 \mu\text{gCL}^{-1}$. The global, latitudinal and depth estimates of biomass concentrations will be useful for understanding ocean biogeochemistry, and for evaluating global models that include macrozooplankton. Species level abundance data will be useful for understanding biodiversity, both globally and regionally, and will be of interest to researchers outside PFT and biogeochemical modelling. Although the dataset is not yet fully comprehensive in terms of taxonomic data or temporal and spatial distributions, it is a foundation upon which a comprehensive dataset can be based. The present database can act as a nucleus for a fully comprehensive dataset of macrozooplankton biodiversity in the global ocean, which will justify further details in their representation in models, e.g. inclusion of a separate representation for herbivorous and carnivorous macrozooplankton.

There is a requirement for the provision of guidelines for macrozooplankton abundance and biomass data and meta-data collection. These guidelines may be consulted during planning stages of research cruises and supplementary data

may be considered for collection. If detailed data on a taxonomic level, life stage, size, and chemical composition, was gathered this would augment the number of biomass data points from the COPEPOD dataset. Detailed carbon information on a species level along with environmental data would be incredibly useful. This would expand the supplementary conversion dataset to aid the accurate conversion of abundance data, length, wet, and dry mass data to carbon; a currency valued by the modelling community.

For the first time macrozooplankton data from national data centres have been collected in the original datasets, and synthesised in the gridded dataset, to create a macrozooplankton data product. The original datasets preserve all metadata that was received, including taxonomic information, although this information was not always detailed to species level. The gridded data includes amongst others abundance and biomass values. Both the original and gridded datasets will be of interest to researchers across biological oceanography and biogeochemical and PFT modelling. The taxonomic, abundance, conversion and biomass data may be extracted for a variety of uses. Although at present there are more biomass data at high latitudes, there are at least some data at low latitudes as well, so that the data can be used at both regional and global scales. Le Quéré et al. (2013) have shown the importance of macrozooplankton to the functioning of the lower trophic level in the ocean ecosystem and associated biogeochemistry, so we look forward to a wider interest in this group of organisms over the coming years.

Apart from COPEPOD, HOTS and BATS databases no national data centres are yet in a position to facilitate the provision of macrozooplankton data. Central data repositories are relatively new and time is required to gather, assess and supply accurate data and metadata. Communication between biogeochemical modellers, data managers and experimentalists is continually improving (Le Quéré and Pesant, 2009) and there is an ever increasing interest to combine expertise from the modelling and experimentalist communities to produce and share the data products necessary to parameterise and validate marine ecosystem models.

Acknowledgements. We thank Angus Atkinson and collaborators for their permission to use and reproduce the rawKRILLBASE database (Atkinson et al., 2004) and for discussions regarding length frequency distributions and body size of *Euphausia superba*, *Salpa thompsoni* and *Ihleia racovitzai*. We thank Todd O'Brien for compiling and collating macrozooplankton abundances in Coastal & Oceanic Plankton Ecology, Production & Observation Database (COPEPOD) (O'Brien, 2005). We thank Michael Landry and collaborators for permission to use and reproduce the macrozooplankton component of the Hawaii Ocean Time-Series (HOTS) dataset (Landry et al., 2001). We thank Deborah Steinberg and collaborators for permission to use and reproduce the macrozooplankton component of the Bermuda Atlantic Time-Series (BATS) dataset (Steinberg et al., 2012). Our most profound thanks go to all the people who participated in the collection, identification and analysis of

the original zooplankton samples; without them and their efforts this work would not have been possible.

We thank Clare Enright for technical assistance, Stéphane Pesant for his support, and the members of the Dynamic Green Ocean Project for their input. We thank the European Union for funding to RM (FAASIS project MEST-CT-2004-514159) and M-PG (Eurocean project) and UK NERC for funding to ETB (MARQUEST project NE/C516079/1).

Edited by: D. Carlson

Reviewed by: three anonymous referees

References

- Anadon, R. and Estrada, M.: The FRUELA cruises. A carbon flux study in productive areas of the Antarctic Peninsula (December 1995–February 1996), *Deep-Sea Res. Part II*, 49, 567–583, doi:10.1016/s0967-0645(01)00112-6, 2002.
- Atkinson, A. and Peck, J. M.: A summer-winter comparison of zooplankton in the oceanic area around South Georgia, *Polar Biol.*, 8, 463–473, doi:10.1007/bf00264723, 1988.
- Atkinson, A., Siegel, V., Pakhomov, E., and Rothery, P.: Long-term decline in krill stock and increase in salps within the Southern Ocean, *Nature*, 432, 100–103, doi:10.1038/nature02996, 2004.
- Atkinson, A., Siegel, V., Pakhomov, E. A., Rothery, P., Loeb, V., Ross, R. M., Quetin, L. B., Schmidt, K., Fretwell, P., Murphy, E. J., Tarling, G. A., and Fleming, A. H.: Oceanic circumpolar habitats of Antarctic krill, *Mar. Ecol.-Prog. Ser.*, 362, 1–23, 2008.
- Atkinson, A., Siegel, V., Pakhomov, E. A., Jessopp, M. J., and Loeb, V.: A re-appraisal of the total biomass and annual production of Antarctic krill, *Deep-Sea Res. Pt. I*, 56, 727–740, doi:10.1016/j.dsr.2008.12.007, 2009.
- Buitenhuis, E. T., Le Quéré, C., Aumont, O., Beaugrand, G., Bunker, A., Hirst, A., Ikeda, T., O'Brien, T., Piontkovski, S., and Straile, D.: Biogeochemical fluxes through mesozooplankton, *Global Biogeochem. Cy.*, 20, GB2003, doi:10.1029/2005GB002511, 2006.
- Buitenhuis, E. T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S. C., Leblanc, K., Le Quéré, C., Luo, Y.-W., O'Brien, C., O'Brien, T., Peloquin, J., Schiebel, R., and Swan, C.: MAREDAT: towards a world atlas of MARine Ecosystem DATA, *Earth Syst. Sci. Data*, 5, 227–239, doi:10.5194/essd-5-227-2013, 2013.
- Casareto, B. E. and Nemoto, T.: Salps of the Southern Ocean (Australian sector) during the 1983–84 summer, with special reference to the species *Salpa thompsoni*, *Foxtton 1961*, *Memoirs of National Institute of Polar Research*, 221–239 1986.
- Chiba, S., Ishimaru, T., Hosie, G. W., and Wright, S. W.: Population structure change of *Salpa thompsoni* from austral mid-summer to autumn, *Polar Biol.*, 22, 341–349, doi:10.1007/s003000050427, 1999.
- Chiba, S., Ishimaru, T., Hosie, G. W., and Fukuchi, M.: Spatio-temporal variability of zooplankton community structure off east Antarctica (90 to 160 degrees E), *Mar. Ecol.-Prog. Ser.*, 216, 95–108, doi:10.3354/meps216095, 2001.
- Deibel, D.: Feeding and metabolism in Appendicularia, in: *The Biology of Pelagic Tunicates*, edited by: Bone, Q., Oxford University Press, New York, 139–150, 1998.
- Dubischar, C. D., Pakhomov, E. A., and Bathmann, U. V.: The tunicate *Salpa thompsoni* ecology in the Southern Ocean II. Prox-

- imate and elemental composition, *Mar. Biol.*, 149, 625–632, doi:10.1007/s00227-005-0226-8, 2006.
- Foxton, P.: The distribution and life-history of *Salpa thompsoni* Foxton with observations on a related species, *Salpa gerlachei* Foxton, Discovery Report, 1–116, 1966.
- Glover, D. M., Jenkins, W. J., and Doney, S. C.: *Modeling Methods for Marine Science*, Cambridge University Press, Cambridge, 588 pp., 2011.
- Harris, R., Weibe, P., Lenz, J., Skjoldal, H.-R., and Huntley, M.: *ICES Zooplankton Methodology Manual*, Academic Press, London, 2000.
- Hirst, A. G., Roff, J. C., and Lampitt, R. S.: A synthesis of growth rates in marine epipelagic invertebrate zooplankton, *Adv. Mar. Biol.*, 44, 1–142, 2003.
- Hood, R. R., Laws, E. A., Armstrong, R. A., Bates, N. R., Brown, C. W., Carlson, C. A., Chai, F., Doney, S. C., Falkowski, P. G., Feely, R. A., Friedrichs, M. A. M., Landry, M. R., Moore, J. K., Nelson, D. M., Richardson, T. L., Salihoglu, B., Schartau, M., Toole, D. A., and Wiggert, J. D.: Pelagic functional group modeling: Progress, challenges and prospects, *Deep-Sea Res. Pt. II*, 53, 459–512, 2006.
- Hosie, G. W. and Cochran, T. G.: Mesoscale distribution patterns of macrozooplankton communities in Prydz Bay, Antarctica January to February 1991, *Mar. Ecol.-Prog. Ser.*, 106, 21–39, doi:10.3354/meps106021, 1994.
- Hosie, G. W., Cochran, T. G., Pauly, T., Beaumont, K. L., Wright, S. W., and Kitchener, J. A.: Zooplankton community structure of Prydz Bay, Antarctica, January–February 1993 (18th Symposium on Polar Biology), *Proceedings of the NIPR Symposium on Polar Biology*, 10, 90–133, 1997.
- Hosie, G. W., Schultz, M. B., Kitchener, J. A., Cochran, T. G., and Richards, K.: Macrozooplankton community structure off East Antarctica (80–150 degrees E) during the Austral summer of 1995/1996, *Deep-Sea Res. Pt. II*, 47, 2437–2463, doi:10.1016/s0967-0645(00)00031-x, 2000.
- Hunt, B. P. V., Pakhomov, E. A., Siegel, V., Strass, V., Cisewski, B., and Bathmann, U.: The seasonal cycle of the Lazarev Sea macrozooplankton community and a potential shift to top-down trophic control in winter, *Deep-Sea Res. Pt. II*, 58, 1662–1676, doi:10.1016/j.dsr2.2010.11.016, 2011.
- Jazdzewski, K., Kittel, W., and Lotocki, K.: Zooplankton studies in the southern Drake Passage and in the Bransfield Strait during the austral summer (BIOMAS-FIBEX, February–March 1981), *Pol. Polar Res.*, 3, 203–242, 1982.
- Landry, M. R., Al-Mutairi, H., Selph, K. E., Christensen, S., and Nunnery, S.: Seasonal patterns of mesozooplankton abundance and biomass at Station ALOHA, *Deep-Sea Res. Pt. II*, 48, 2037–2061, doi:10.1016/s0967-0645(00)00172-7, 2001.
- Le Quéré, C. and Pesant, S.: Plankton Functional Types in a new generation of biogeochemical models, *EOS: Transactions American Geophysical Union*, 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.
- Le Quéré, C., Buitenhuis, E. T., Moriarty, R., Vogt, M., Sailley, S., Chollet, S., Stephens, N., Enright, C., Franklin, D., Larsen, S., Legendre, L., Platt, T., Rivkin, R. B., and Sathyendranath, S.: Role of plankton functional diversity for marine ecosystem services, in preparation, 2013.
- Loeb, V., Hofmann, E. E., Klinck, J. M., and Holm-Hansen, O.: Hydrographic control of the marine ecosystem in the South Shetland-Elephant Island and Bransfield Strait region, *Deep-Sea Res. Pt. II*, 57, 519–542, doi:10.1016/j.dsr2.2009.10.004, 2010.
- Lovenduski, N. S., Gruber, N., and Doney, S. C.: Toward a mechanistic understanding of the decadal trends in the Southern Ocean carbon sink, *Global Biogeochem. Cy.*, 22, Gb3016, doi:10.1029/2007gb003139, 2008.
- 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., Moisaner, 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.
- Madin, L. P., Horgan, E. F., and Steinberg, D. K.: Zooplankton at the Bermuda Atlantic Time-series Study (BATS) station: diel, seasonal and interannual variation in biomass, 1994–1998, *Deep-Sea Res. Pt. II*, 48, 2063–2082, doi:10.1016/s0967-0645(00)00171-5, 2001.
- Marr, J. W. S.: The natural history and geography of the Antarctic krill (*Euphausia superba* Dana), *Discovery Report*, 32, 33–464, 1962.
- Mizdalski, E.: Weight and length data of zooplankton in the Weddell Sea Antarctica in Austral Spring 1988 ANT V-3, *Berichte zur Polarforschung*, 1–72, 1988.
- Moriarty, R.: Respiration rates in epipelagic macrozooplankton: a dataset, *PANGAEA*, 2009.
- Moriarty, R. and O'Brien, T. D.: Distribution of mesozooplankton biomass in the global ocean, *Earth Syst. Sci. Data*, 5, 45–55, doi:10.5194/essd-5-45-2013, 2013.
- Nishikawa, J. and Tsuda, A.: Diel vertical migration of the tunicate *Salpa thompsoni* in the Southern Ocean during summer, *Polar Biol.*, 24, 299–302, doi:10.1007/s003000100227, 2001.
- Nishikawa, J., Naganobu, M., Ichii, T., Ishii, H., Terazaki, M., and Kawaguchi, K.: Distribution of salps near the South Shetland Islands During Austral Summer, 1990–1991 with special reference to krill distribution, *Polar Biol.*, 15, 31–39, 1995.
- O'Brien, T. D.: COPEPOD: A global plankton database, US Dep. Commerce, NOAA Technical Memorandum, NMFS-F/SPO-73, 136, 2005.
- Omori, M. and Ikeda, T.: *Methods in marine zooplankton ecology*, Wiley-Interscience Publications, John Wiley & Sons, Japan, 331 pp., 1984.
- Pakhomov, E. A., Froneman, P. W., and Perissinotto, R.: Salp/krill interactions in the Southern Ocean: spatial segregation and implications for the carbon flux, *Deep-Sea Res. Pt. II*, 49, 1881–1907,

- doi:10.1016/s0967-0645(02)00017-6, 2002.
- Ross, R. M., Quetin, L. B., Martinson, D. G., Iannuzzi, R. A., Stamerjohn, S. E., and Smith, R. C.: Palmer LTER: Patterns of distribution of five dominant zooplankton species in the epipelagic zone west of the Antarctic Peninsula, 1993–2004, *Deep-Sea Res. Pt. II*, 55, 2086–2105, 2008.
- Schütt, F.: *Analytische Planktonstudien*, Lipsius and Tischer, Kiel, 117 pp., 1892.
- Sieburth, J. M., Smetacek, V., and Lenz, J.: Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fraction, *Limnol. Oceanogr.*, 23, 1256–1263, 1987.
- Siegel, V.: Distribution and population dynamics of *Euphausia superba*: summary of recent findings, *Polar Biol.*, 29, 1–22, doi:10.1007/s00300-005-0058-5, 2005.
- Siegel, V., Kawaguchi, S., Ward, P., Litvinov, F., Sushin, V., Loeb, V., and Watkins, J.: Krill demography and large-scale distribution in the southwest Atlantic during January/February 2000, *Deep-Sea Res. Pt. II*, 51, 1253–1273, 2004.
- Steinberg, D. K., Lomas, M. W., and Cope, J. S.: Long-term increase in mesozooplankton biomass in the Sargasso Sea: Linkage to climate and implications for food web dynamics and biogeochemical cycling, *Global Biogeochem. Cy.*, 26, GB1004, doi:10.1029/2010GB004026, 2012.
- Turner, J. T.: Zooplankton fecal pellets, marine snow and sinking phytoplankton blooms, *Aquat. Microb. Ecol.*, 27, 57–102, 2002.
- Ward, P., Shreeve, R., Whitehouse, M., Korb, B., Atkinson, A., Meredith, M., Pond, D., Watkins, J., Goss, C., and Cunningham, N.: Phyto- and zooplankton community structure and production around South Georgia (Southern Ocean) during Summer 2001/02, *Deep-Sea Res. Pt. I*, 52, 421–441, doi:10.1016/j.dsr.2004.10.003, 2005.
- Ward, P., Shreeve, R., Atkinson, A., Korb, B., Whitehouse, M., Thorpe, S., Pond, D., and Cunningham, N.: Plankton community structure and variability in the Scotia Sea: austral summer 2003, *Mar. Ecol.-Prog. Ser.*, 309, 75–91, doi:10.3354/meps309075, 2006.
- Witek, Z., Kittel, W., Czykieta, H., Zmijewska, M. I., and Presler, E.: Macrozooplankton in the southern Drake Passage and in the Bransfield Strait Antarctica during BIOMASS-SIBEX Dec. 1983–Jan. 1984, *Pol. Polar Res.*, 6, 95–116, 1985.