

# Satellite-based indicator of zooplankton distribution for global monitoring

Druon J.N., H  laou  t P., Beaugrand G., Fromentin J.M., Palialexis A., Hoepffner N.

## Supplementary information

### The Continuous Plankton Recorder

The Continuous Plankton Recorder (CPR) survey is among the few plankton monitoring programmes operated at a basin scale (another example being the California Cooperative Oceanic Fisheries Investigations [Calcofi]). The survey has monitored both phytoplankton and zooplankton monthly since the 1930s using an instrument (the CPR machine) that can be towed at high speed (about 15–20 knots) by merchant ships. This methodology has remained virtually unchanged since 1958, providing one of the most extensive ecological data sets available to the marine ecology community<sup>1–3</sup>. The length of the towing cable is calculated to produce a towing depth of about 10 m<sup>4</sup> at the operating speed of the vessel. Behind the ship, the upper layer of the water column is mixed due to the turbulent nature of the ship's wake and it is therefore accepted that the CPR samples the integrated first 10 m of the water column<sup>3</sup>. Water enters through the front 1.27 cm<sup>2</sup> aperture of the CPR machine and is filtered through a continuously moving band of silk (i.e. the filtering silk). Filtered plankton are subsequently covered by a second moving band of silk (i.e. the covering silk). This sandwich is automatically rolled onto the storage tank containing a dilute solution of borax-buffered formaldehyde (~4%) that fixes the plankton<sup>1</sup>. In the laboratory the silk is cut into samples, corresponding to approximately 10 nautical miles of sampling (or 18.52 km), before a microscopic procedure is carried out to ensure identification of the zooplankton to the highest practical level.

## Species inserted in the small and large mesozooplankton biomass indices

Two planktonic groups were created for the small (generally below 2 mm; 40 taxa) and large (generally above 2 mm; 91 taxa) mesozooplankton (Table SI 1 and Table SI 2), representing the most present species in the North Atlantic (Fig. SI 2). As mentioned in Richardson et al. 2006<sup>5</sup>, “Copepods identified usually represent Copepodite Stage V [(pre-adults)] and adults both because the CPR preferentially retains these larger copepods (Robertson, 1968<sup>6</sup>) and because they are easier to speciate than juveniles. In addition, females are generally more easily identified to species than are males. For example, females of three smaller species within the genus *Pleuromamma* are speciated (“*Pleuromamma borealis*”, “*Pleuromamma gracilis*”, “*Pleuromamma piseki*”), whereas their males are normally recorded as “*Pleuromamma* spp.”. Males of other genera such as *Euchaeta* are readily identifiable and are included with the females”.

**Table SI 1: List of the 40 relatively small taxa (generally below 2 mm) present in the whole dataset (n = 68,364) with associated average size ('CPR traverse' analysis) and relative contribution (species number multiplied by size in % of all species) for this size class.**

Taxon	Size (mm)	Contribution to index (%)	Taxon	Size (mm)	Contribution to index (%)
<i>Para-Pseudocalanus</i> spp.	1.45	27.3666	<i>Ctenocalanus vanus</i>	1.4	0.0320
<i>Acartia</i> spp.	1.4	19.9730	<i>Calocalanus</i> spp.	1	0.0199
<i>Centropages typicus</i>	1.4	17.8192	<i>Ctenocalanus</i> spp.	1.25	0.0136
<i>Temora longicornis</i>	1.25	11.1232	<i>Acartia danae</i>	1	0.0104
<i>Oithona</i> spp.	1.15	7.2734	<i>Euterpina acutifrons</i>	0.65	0.0097
<i>Pseudocalanus</i> spp.	1.45	6.0711	<i>Clytemnestridae</i>	1	0.0064
<i>Clausocalanus</i> spp.	1.25	3.3502	<i>Tortanus discaudatus</i>	2.25	0.0054
<i>Centropages hamatus</i>	1.15	1.5017	<i>Acartia longiremis</i>	0.95	0.0050
<i>Paracalanus</i> spp.	1	1.4137	<i>Lubbockia</i> spp.	2	0.0032
<i>Centropages</i> spp. (Unidentified)	1.6	1.0875	<i>Macrosetella gracilis</i>	1.35	0.0022
<i>Corycaeus</i> spp.	1.8	0.9310	<i>Microcalanus</i> spp.	0.85	0.0020
<i>Oncaea</i> spp.	0.95	0.7835	<i>Parapontella brevicornis</i>	1.45	0.0019
<i>Lucicutia</i> spp.	5.65	0.3844	<i>Temora turbinata</i>	1.45	0.0015
<i>Microsetella</i> spp.	0.55	0.1906	<i>Acartia negligens</i>	1.45	0.0015
<i>Isias clavipes</i>	1.5	0.1655	<i>Hemicyclops aberdonensis</i>	1.45	0.0015
<i>Scolecithricella</i> spp.	3.35	0.1638	<i>Pseudocalanus</i> adult Pacific	1.45	0.0008
<i>Temora stylifera</i>	1.65	0.1228	<i>Diaixis hibernica</i>	0.95	0.0005
<i>Centropages chierchiae traverse</i>	1.9	0.1066	<i>Acartia</i> spp. – Antarctic	1.4	0.0004
<i>Mecynocera clausi</i>	1.05	0.0533	<i>Pseudodiaptomus</i> spp.	1.3	0.0003
<i>Ctenocalanus vanus</i>	1.4	0.0320	<i>Acartia tonsa</i>	1.15	0.0003
<i>Calocalanus</i> spp.	1	0.0199	<i>Farranula</i> spp.	0.85	0.0002

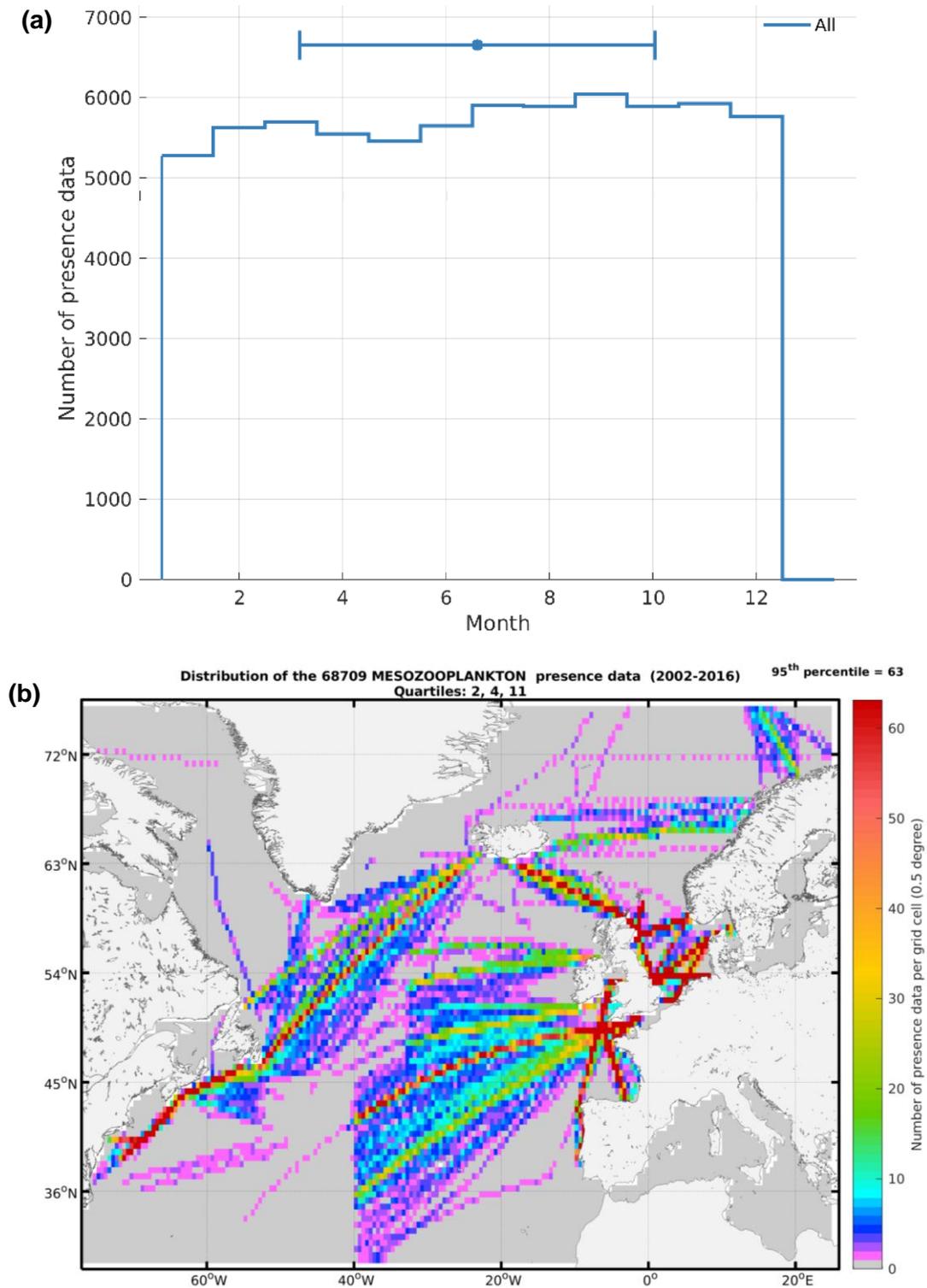
**Table SI 2: List of the 91 relatively large taxa (generally above 2 mm) present in the whole dataset (n = 68,364) with associated average size ('CPR eye-count' analysis) and relative contribution (species number multiplied by size in % of all species) for this size class.**

Taxon	Size (mm)	Contribution to index (%)	Taxon	Size (mm)	Contribution to index (%)
<i>Calanus finmarchicus</i>	3.7	65.09250	<i>Candacia pachydactyla</i>	2.45	0.00853
<i>Calanus helgolandicus</i>	2.7	17.33127	<i>Candacia bipinnata</i>	2.55	0.00845
<i>Metridia lucens</i>	2.75	5.35962	<i>Alteutha</i> spp.	0.9	0.00799
<i>Paraeuchaeta norvegica</i>	8.25	3.84413	<i>Euchaeta media</i>	3.95	0.00716
<i>Calanus hyperboreus</i>	7.65	1.01623	<i>Eucalanus hyalinus</i>	5.75	0.00686
<i>Euchaetidae</i>	6.85	0.96420	<i>Miracia efferata</i>	1.7	0.00677
<i>Pleuromamma borealis</i>	1.95	0.89215	<i>Centropages violaceus</i>	2	0.00639
<i>Paraeuchaeta hebes</i>	3.15	0.70059	<i>Rhincalanus cornutus</i>	3.3	0.00630
<i>Pleuromamma gracilis</i>	2.05	0.51751	<i>Aetideus armatus</i>	1.8	0.00528
<i>Nannocalanus minor</i>	1.85	0.45241	<i>Undeuchaeta</i> spp.	4.75	0.00499
<i>Candacia armata</i>	2.45	0.41125	<i>Paracandacia bispinosa</i>	1.8	0.00438
<i>Calanus glacialis</i>	4.55	0.35729	<i>Undeuchaeta major</i>	4.75	0.00431
<i>Pleuromamma abdominalis</i>	3.45	0.32992	<i>Labidocera</i> spp.	2.9	0.00374
<i>Metridia longa</i>	3.05	0.32463	<i>Heterorhabdus</i> spp.	3.1	0.00370
<i>Pleuromamma robusta</i>	3.4	0.31022	<i>Subeucalanus monachus</i>	2.3	0.00335
<i>Pleuromamma piseki</i>	2.05	0.27878	<i>Paraeuchaeta</i> spp.	6.8	0.00276
<i>Neocalanus gracilis</i>	3	0.20123	<i>Parathalestris croni</i>	2	0.00229
<i>Euchaeta acuta</i>	4	0.18701	<i>Candacia curta</i>	2.2	0.00184
<i>Euchirella rostrata</i>	3.05	0.15824	<i>Paracandacia</i> spp.	1.85	0.00137
<i>Calanoides carinatus</i>	2.8	0.15021	<i>Pseudodiaptomus</i> spp.	1.3	0.00130
<i>Centropages chierchiae</i>	1.9	0.14811	<i>Neocalanus robustior</i>	3.75	0.00125
<i>Pleuromamma xiphias</i>	4.85	0.09949	<i>Aetideus giesbrechti</i>	1.65	0.00102
<i>Undeuchaeta plumosa</i>	3.75	0.09527	<i>Paraeuchaeta gracilis</i>	6.05	0.00101
<i>Euchaeta marina</i>	3.1	0.08600	<i>Scolecithrix</i> spp.	3.5	0.00092
<i>Sapphirina</i> spp.	4.5	0.07997	<i>Pontellina plumata</i>	1.45	0.00083
<i>Rhincalanus nasutus</i>	4.4	0.07662	<i>Euchirella messinensis</i>	4.5	0.00075
<i>Candacia</i> spp.	3.15	0.07048	<i>Candacia longimana</i>	3.15	0.00075
<i>Candacia ethiopica</i>	2.5	0.04091	<i>Scaphocalanus echinatus</i>	1.95	0.00074
<i>Undinula vulgaris</i>	2.55	0.03917	<i>Monstrilloida</i>	3.25	0.00070
<i>Scolecithrix danae</i>	2.1	0.03376	<i>Scolecithrix bradyi</i>	1.3	0.00065
<i>Subeucalanus crassus</i>	3.5	0.03056	<i>Subeucalanus pileatus</i>	2.1	0.00040
<i>Centropages bradyi</i>	1.9	0.02837	<i>Scaphocalanus</i> spp.	3.35	0.00040
<i>Mesocalanus tenuicornis</i>	2.45	0.02279	<i>Haloptilus spiniceps</i>	4.05	0.00029
<i>Metridia</i> spp. (V-VI)	5.9	0.02167	<i>Subeucalanus mucronatus</i>	3	0.00029
<i>Labidocera wollastoni</i>	2.3	0.01860	<i>Heterorhabdus oikoumenikis</i>	2.8	0.00027
<i>Anomalocera patersoni</i>	3.3	0.01551	<i>Aetideus</i> spp.	2.15	0.00026
<i>Euchirella</i> spp.	5.35	0.01289	<i>Aetideus acutus</i>	1.6	0.00011
<i>Labidocera aestiva</i>	2.4	0.01265	<i>Augaptilus</i> spp.	4.25	0.00010
<i>Corycaeus speciosus</i>	1.55	0.01191	<i>Haloptilus acutifrons</i>	3.4	0.00008
<i>Urocorycaeus</i> spp.	2.15	0.01174	<i>Euchirella amoena</i>	3.35	0.00008
<i>Heterorhabdus norvegicus</i>	3.6	0.00953	<i>Heterostylites longicornis</i>	3.3	0.00008

## Supplementary information: Productive fronts and mesozooplankton dynamics

<i>Paracandacia simplex</i>	1.95	0.00944	<i>Pontellopsis regalis</i>	3.05	0.00007
<i>Neocalanus</i> spp.	6	0.00930	<i>Spinocalanus</i> spp.	2.95	0.00007
<i>Heterorhabdus papilliger</i>	2.15	0.00923	<i>Heterorhabdus abyssalis</i>	2.85	0.00007
<i>Labidocera acutifrons</i>	4	0.00897	<i>Haloptilus longicornis</i>	1.9	0.00005
<i>Copilia</i> spp.	5.2	0.00868			

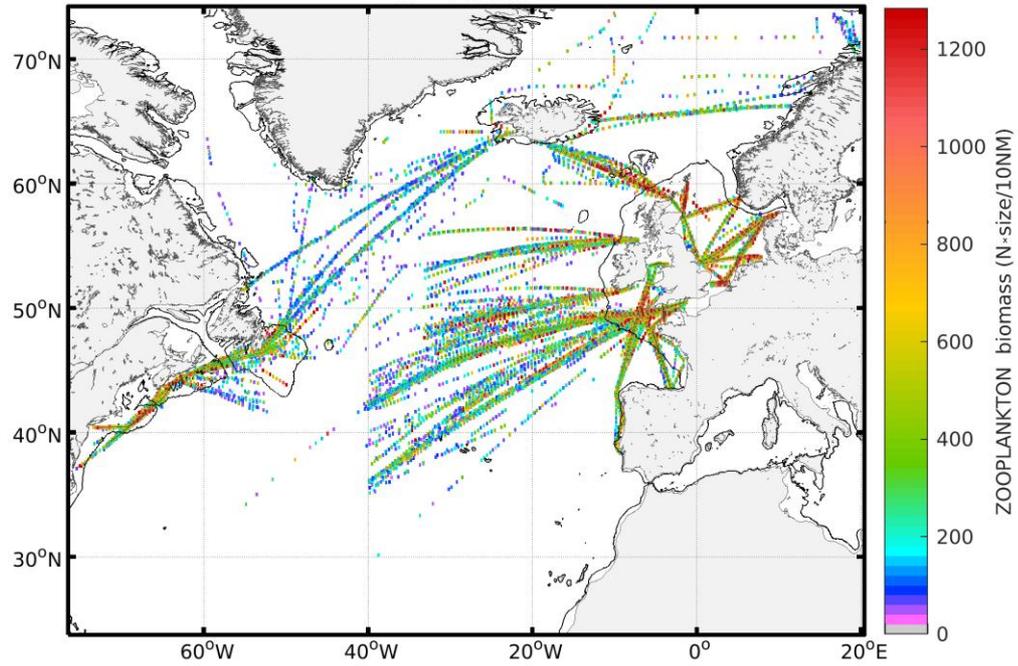
### Distribution of the mesozooplankton presence data (2002–2016)



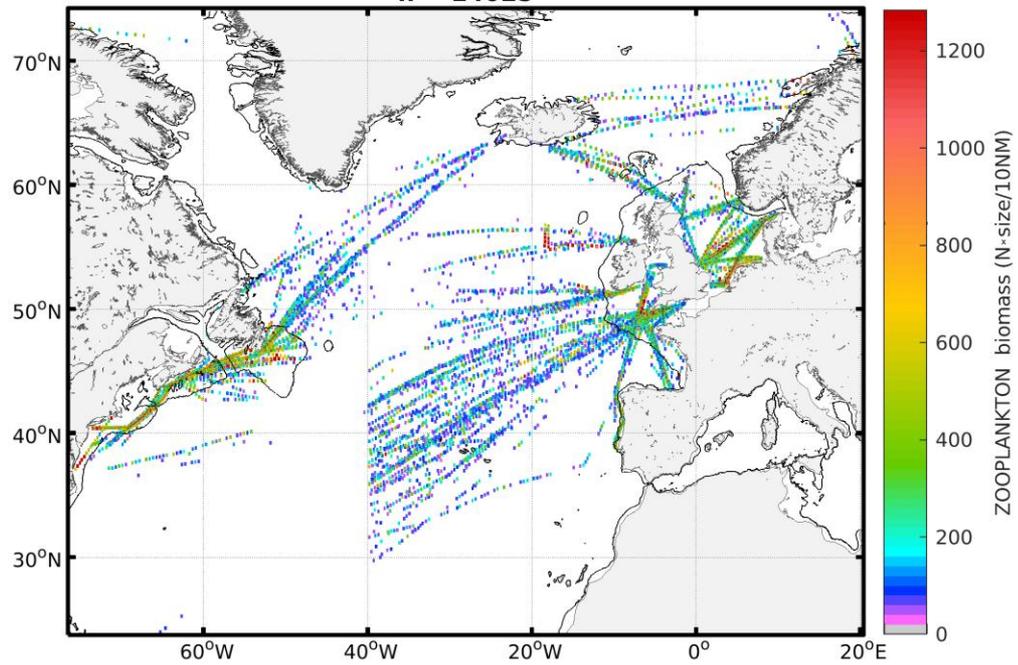
**Figure SI 1** (a) Monthly and (b) spatial distributions of the CPR sampling from 2002 to 2016 (including absence data,  $n = 68,709$ ). Only positive abundance data ( $n = 54,282$ ) were used in the analysis.

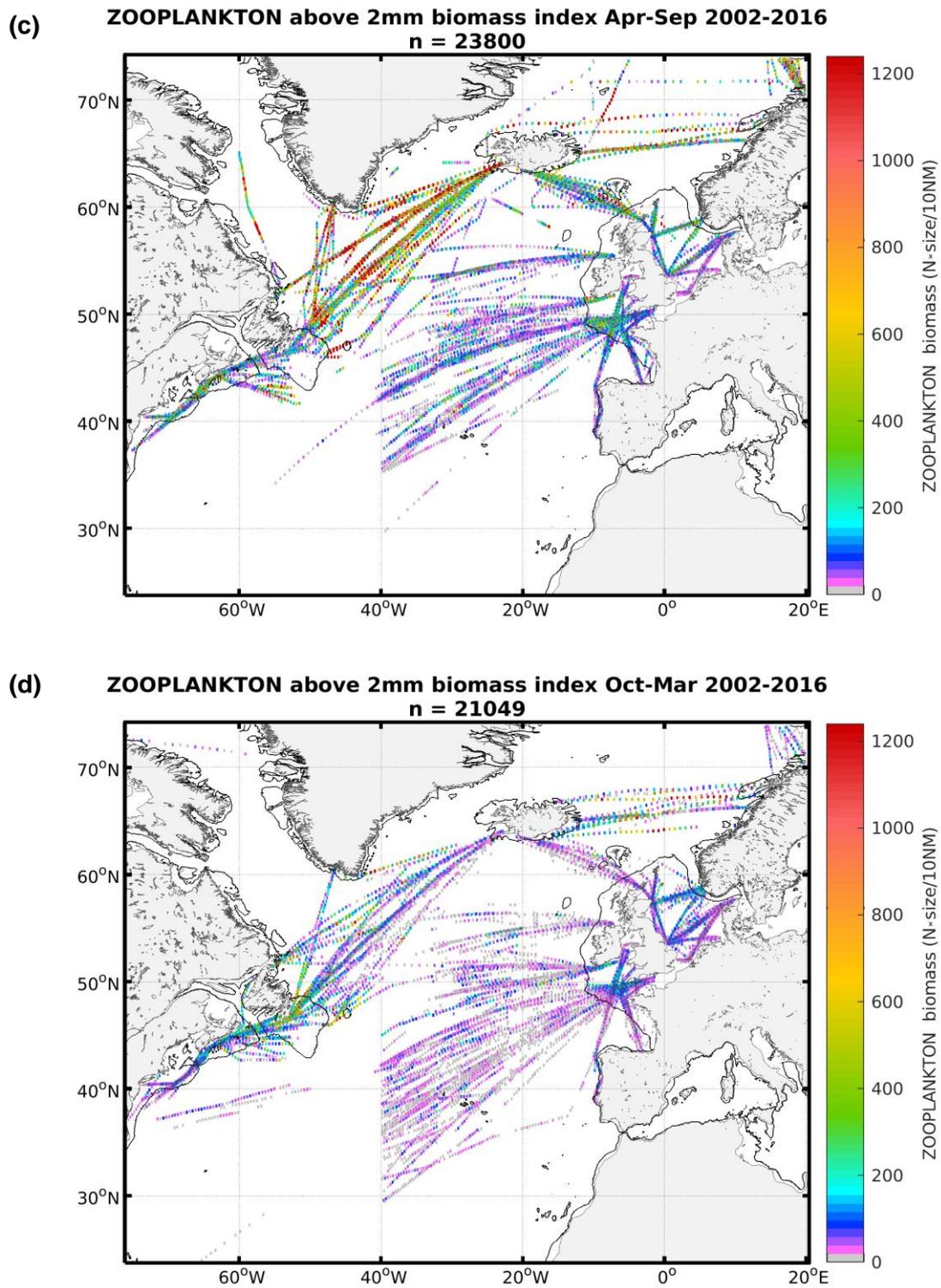
### Original mesozooplankton biomass index by size class

(a) ZOOPLANKTON below 2mm biomass index Apr-Sep 2002-2016  
n = 23065



(b) ZOOPLANKTON below 2mm biomass index Oct-Mar 2002-2016  
n = 14025



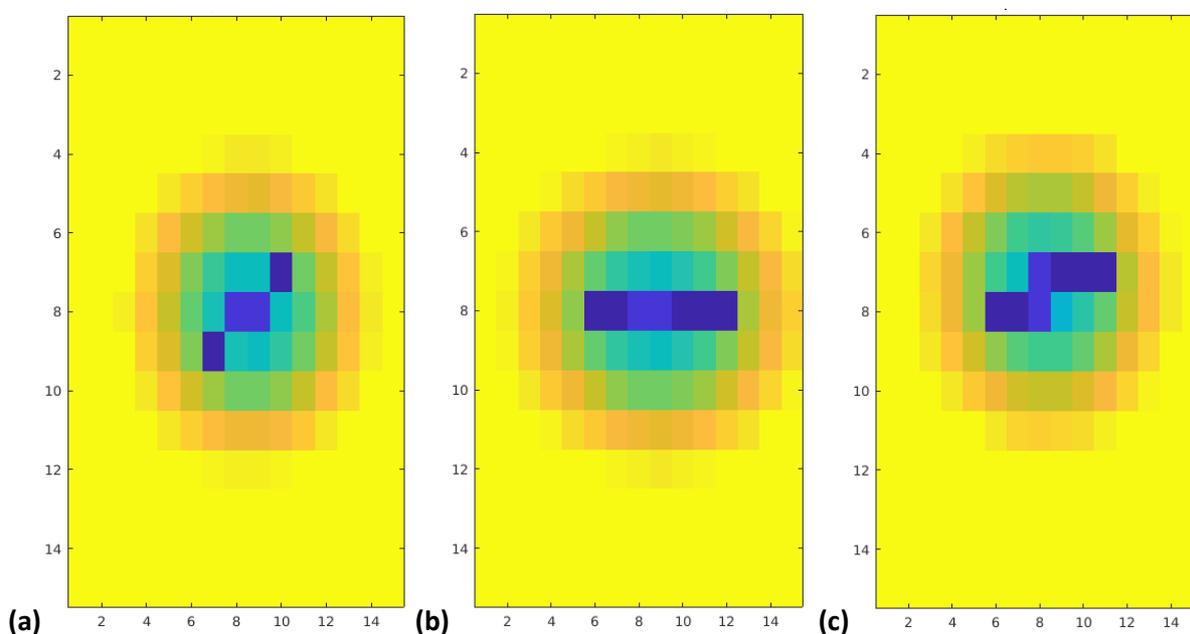


**Figure SI 2** Geographical mean index of mesozooplankton biomass (positive values only) in the North Atlantic **(a, c)** for the high (April-September, upper map;  $n = 23,800$ ) and **(b, d)** the low (October-March, lower map;  $n = 21,049$ ) seasons and the **(a, b)** small and **(c, d)** large mesozooplankton observations (generally below and above 2 mm size) collected between 2002 and 2016 by the Continuous Plankton Recorder survey (mean relative biomass in cells of  $0.3^\circ$  by  $0.3^\circ$ ). The biomass index is expressed in number per mean size and 10 nautical miles' sample. The maximum value of the colour bar was set to the 95th percentile of the high season (1,225 and 1,230 for the small and large mesozooplankton, respectively).

### Diel vertical migration of mesozooplankton

Diel vertical migration (DVM) is an almost universal feature of zooplankton communities, especially in areas south of the oceanic Polar Front<sup>7</sup>. The main observed DVM pattern is an increase in zooplankton abundance/biomass in the subsurface of the water column during the night followed by a diminution at daylight. The pattern is commonly explained to reduce the risk of mortality by predation. The principal pieces of evidence supporting this hypothesis are (i) the normal DVM is more pronounced for heavily pigmented taxa that are more easily perceived by visual predators<sup>8</sup> and (ii) migrating zooplankton tends to be near the surface only when illumination levels are low<sup>8</sup>.

### Spatial selection of biomass and environmental data matchups



**Figure SI 3** Examples of matchups (cells in dark blue) between samples of mesozooplankton biomass from the Continuous Plankton Recorder survey of 10 nautical miles length of direction **(a)** 64, **(b)** 90 and **(c)** 70 degrees (central position centred in the central cell of graphs – in position 8, 8) and the selected grid cells of the habitat index ( $1/24^\circ$  resolution).

## Cluster analysis method and illustrative abiotic variables

Cluster analysis is a suitable method for identifying homogenous groups of objects or 'clusters' (here, mesozooplankton habitat suitability), regardless of their respective number. A cluster analysis is well suited to identify habitats that are marginally represented and may otherwise be interpreted as outliers by other statistical methods. The number of clusters was chosen as the minimum number that favours interpretation. We used 3-day mean values of CHL and gradCHL to increase the number of matchups assuming that variability is low at  $\pm 1$  day from the sampling date and at a spatial resolution of  $1/24^\circ$ . CHL and gradCHL datasets were log-transformed prior to use, due to the wide variability in their values, and were normalised by the mean and standard deviation prior to performing the cluster analysis.

K-means clustering<sup>9</sup> based on a Euclidean distance was used to estimate the association of data points between clusters and to minimise the within-cluster sum of squared errors. In k-means clustering, the number of clusters  $k$  was first chosen (here the minimum number that favours interpretation) and the cluster centres were initialised randomly. Each data point was then assigned to the closest cluster based on a selected distance measure (similarity) and updated cluster centre. At each iteration step, the new cluster centres were computed as the mean vectors of the assigned data points. These two steps, data point assignment and cluster centre update, were repeated until the cluster centres did not change any more or until a sufficient number of iterations were performed. Matlab's k-means function was used with 500 iterations/restarts and the Euclidian distance setting. The z-score transformation<sup>10</sup> was performed before clustering, in which each data variable was normalised to zero mean and unit variance to guarantee that each selected variable had equal influence on the minimisation of the within-cluster sum of squares objective function<sup>10</sup>.

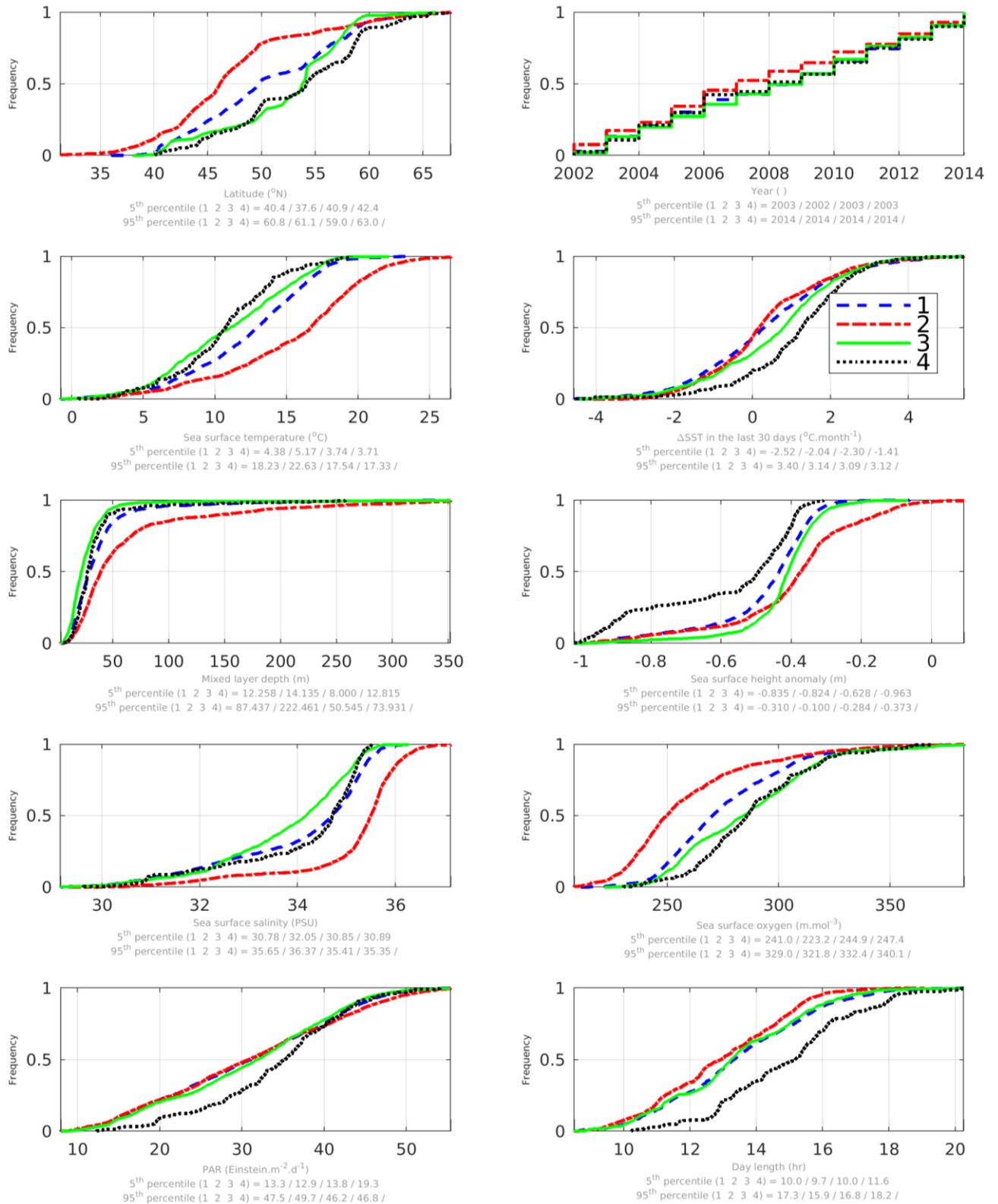
The cluster analysis (Fig. 2) was applied with non-influential (illustrative) variables to reveal the coherence of the results on mesozooplankton habitat suitability with respect to the abiotic conditions. Abiotic variables were extracted from the EU Copernicus Marine Environment Monitoring Service (see description below). The cumulative distribution of illustrative abiotic variables (Fig. SI 4) highlights that the red cluster, corresponding to the lowest values of the index of mesozooplankton biomass, tends to occur mostly in surface waters characterised by warmer temperature, being more deeply-stratified, in further southern areas, with high levels of sea surface height anomaly (SSHa) and salinity (high evaporation), and low oxygen levels compared with the other clusters. These traits indeed characterise relatively poorly productive areas. On the contrary, the cluster of the highest mesozooplankton biomass index (black) tends to occur in generally recently stratified waters

of higher latitude, in lower SSHa and temperature levels and longer day length and higher oxygen levels, which are typical traits of productive waters. The green and blue clusters (medium level of mesozooplankton biomass) generally correspond to intermediate levels of abiotic conditions except for salinity and mixed layer depth where the green cluster shows the lowest levels likely in relation to high precipitation and recently formed mixed layer depth (spring bloom).

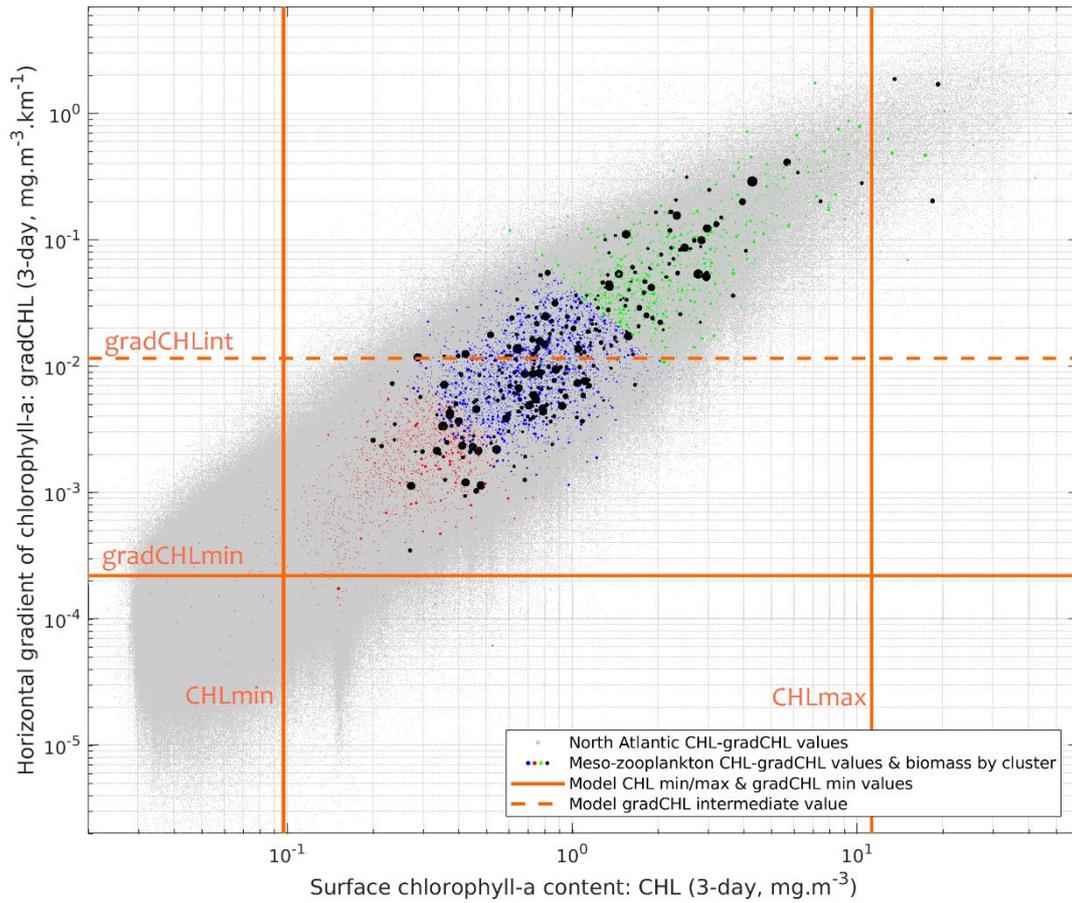
The comparison of the relative influence of the photosynthetically active radiation (PAR) at sea surface with the day length on each cluster reveals that day length discriminates the high mesozooplankton biomass cluster (black) better. This is probably because surface phytoplankton productivity is generally not limited by the solar radiation near the sea surface, while day length is likely to be a major limitation of primary productivity. Finally, all clusters show a relatively regular sampling from year to year.

Data for the abiotic variables (i.e. SSHa (m), SST ( $^{\circ}$ C), SSS (psu), MLD (m) and  $O_2$  ( $\text{mmol m}^{-3}$ ), see figure legend of Fig. SI 4) are extracted from the global ocean model (1997–2014) provided by the EU Copernicus Marine Environment Monitoring Service (<http://marine.copernicus.eu/>). Monthly mean data are extracted from the global model (Glorys2V3) with a  $1/4^{\circ}$  horizontal resolution and 75 unevenly spaced vertical levels. This ocean model includes a variational data assimilation scheme for vertical temperature and salinity profiles and satellite sea level anomalies. These monthly data are interpolated onto the MODIS-Aqua grid ( $1/24^{\circ}$  resolution) and then linearly interpolated to obtain daily values that match the sampling day. This month-to-day interpolation step is assumed to produce suitable estimates of the seasonal changes that define habitat. SST and SSS values are calculated from the upper model layer (ca. 3 m) and are thus considered to be representative of the mixed layer. MLD is defined as the maximum of the vertical density gradient. PAR data from the MODIS-Aqua sensor were extracted from the NASA portal (<https://oceancolour.gsfc.nasa.gov/cgi/l3>).

Supplementary information: Productive fronts and mesozooplankton dynamics

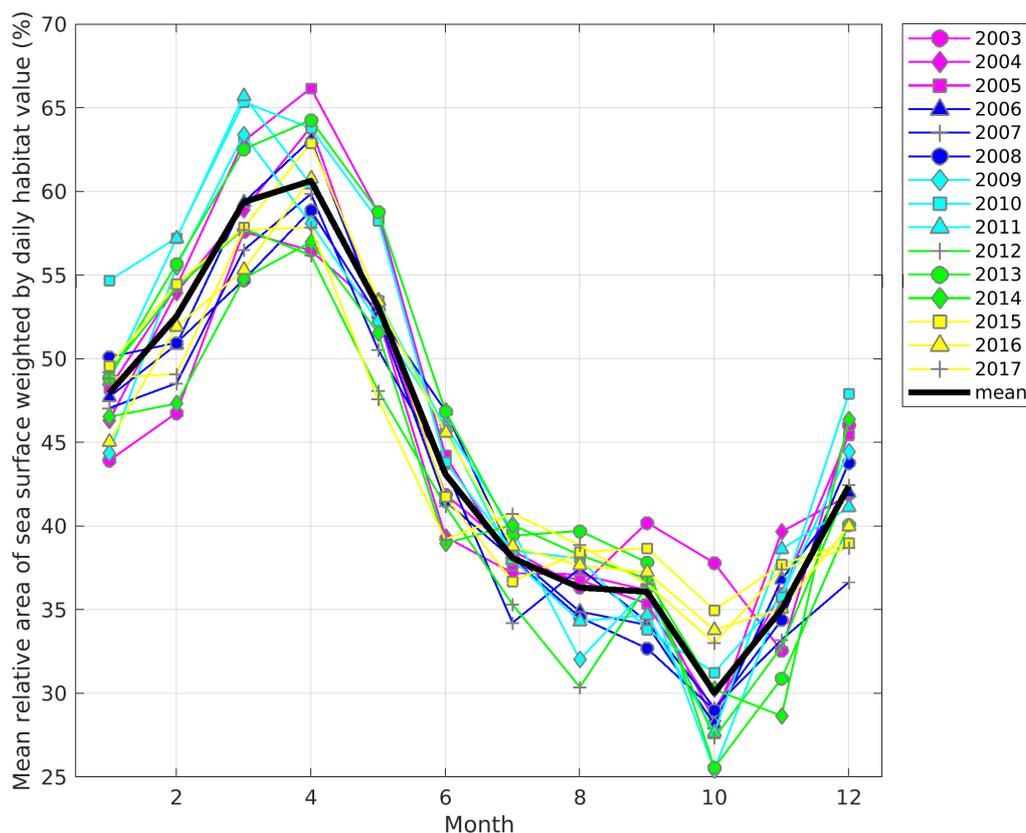


**Figure SI 4** Cumulative distributions for the illustrative abiotic variables as output of the main cluster analysis (Fig. 2 of the main publication for details). The illustrative variables are: latitude, year, sea surface temperature (SST), SST increase over the last 30 days ( $\Delta$ SST), mixed layer depth (MLD), sea surface height anomaly (SSHa), sea surface salinity (SSS), sea surface oxygen ( $O_2$ ), photosynthetically active radiation (PAR) and day length.



**Figure SI 5** Scatter plot (logarithmic scale) of chlorophyll-a content (CHL) versus horizontal gradient of chlorophyll-a (gradCHL) for the entire North Atlantic (grey dots). Mesozooplankton biomass index is superimposed (the size of the dots being proportional to mesozooplankton biomass) with colours according to the clustering (see Fig. 2). The suitable range of CHL, the minimum value of gradCHL (orange solid lines) and the intermediate value of gradCHL (orange dashed line) are superimposed (see Fig. 3).

### Seasonal variability of mesozooplankton habitat



**Figure SI 6** Monthly habitat size of mesozooplankton in the North Atlantic Ocean for the period 2003–2017 (values relative to the surface ocean area). The mean seasonal habitat is shown (black solid line).

### References

1. Batten, S. D. *et al.* CPR sampling: the technical background, materials and methods, consistency and comparability. *Prog. Oceanogr.* **58**, 193–215 (2003).
2. Reid, P. C., Colebrook, J. M., Matthews, J. B. L. & Aiken, J. The Continuous Plankton Recorder: concepts and history, from plankton indicator to undulating recorders. *Prog. Oceanogr.* **58**, 117–173 (2003).
3. Warner, A. J. & Hays, G. C. Sampling by the continuous plankton recorder survey. *Prog. Oceanogr.* **34**, 237–256 (1994).
4. Hélaouët, P., Beaugrand, G. & Reygondeau, G. Reliability of spatial and temporal patterns of *C. finmarchicus* inferred from the CPR survey. *J. Mar. Syst.* **153**, 18–24 (2016).
5. Richardson, A. J. *et al.* Using continuous plankton recorder data. *Prog. Oceanogr.* **68**, 27–74 (2006).
6. Robertson, A. continuous plankton recorder: a method for studying the biomass of calanoid copepods. *Bull. Mar. Ecol.* (1968).
7. Beaugrand, G., Ibañez, F. & Lindley, J. A. Geographical distribution and seasonal and diel changes in the diversity of calanoid copepods in the North Atlantic and North Sea. *Mar. Ecol. Prog. Ser.* **219**, 189–203 (2001).
8. Hays, G. C., Proctor, C. A., John, A. W. G. & Warner, A. J. Interspecific differences in the diel vertical migration of marine copepods: the implications of size, color, and morphology. *Limnol. Oceanogr.* **39**, 1621–1629 (1994).

9. MacQueen, J. Some methods for classification and analysis of multivariate observations. in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* **1**, 14 (1967).
10. Berthold, M., Borgelt, C. & Höppner, F. *Guide to intelligent data analysis*. **42**, (Springer, pp. 394, 2010).