LIMNOLOGY and OCEANOGRAPHY: METHODS

ASLO

Limnol. Oceanogr.: Methods 21, 2023, 24–39

doi: 10.1002/lom3.10520

A RUpture-Based detection method for the Active mesopeLagIc Zone (RUBALIZ): A crucial step toward rigorous carbon budget assessments

Robin Fuchs,[1,2] **Chloé M. J. Baumas**,[1]* **Marc Garel**,[1] **David Nerini**,[1] **Frédéric A. C. Le Moigne**,[3] **Christian Tamburini**[1]

[1]Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO UM 110, Marseille, France
[2]Aix Marseille Univ, CNRS, I2M, Marseille, France
[3]LEMAR Laboratoire des Sciences de l'Environnement Marin, UMR6539, CNRS, UBO, IFREMER, IRD, Plouzané, Technopôle Brest-Iroise, France

Abstract

Determining mesopelagic organic carbon budgets is essential to characterize the ocean's role as a carbon dioxide sink. This is because the biological processes observed in the mesopelagic zone are crucial for understanding the biological carbon pump. Yet, field assessments of carbon budgets are often unbalanced with the carbon demand exceeding its supply. This underlines either methodological issues in the budget calculations or incomplete knowledge of the mesopelagic carbon cycling with potentially missing sources. Carbon budgets are built by partitioning the ocean into vertical depth zones. Vertical boundaries are conventionally defined between 200 and 1000 m depth or using various thresholds. Such approaches lack consistent methodology preventing robust comparison of mesopelagic carbon budget from region to region. Here, using a statistical rupture detection method applied to conductivity–temperature–depth (CTD)-cast variables (fluorescence, O_2 concentration, potential temperature, salinity, and density), we aim to provide independent estimates of mesopelagic boundaries. We demonstrate that the so-determined upper boundary is highly correlated with the knee points of the particulate organic carbon (POC) fluxes estimated by a power law and that over 90% of the POC flux attenuation occurs within our method boundaries. The identified zone therefore corresponds to the most active part of the conventional mesopelagic zone and we name it the "active mesopelagic zone" (AMZ). We find that the depths of the mesopelagic zone depend on the region considered. Our results demonstrate that the mesopelagic carbon budget discrepancy can vary up to four folds depending on the boundaries chosen and hence provide novel grounds to reassess existing and future mesopelagic carbon budgets.

In the euphotic zone of the ocean, phytoplankton convert carbon dioxide (CO_2) into biogenic carbon (C). A fraction of this biogenic C escapes the euphotic zone and crosses the mesopelagic zone of the ocean. The vertical export processes and the fate of organic carbon (OC) in the mesopelagic zone have received increased attention from the international community for the past decade and in particular in recent years (Buesseler and Boyd 2009; Robinson et al. 2010; Siegel et al. 2016; Martin et al. 2020). The mesopelagic zone harbors substantial fish resources (Irigoien et al. 2014) and above all, plays a key role in biogeochemical cycles, in particular in the downward pumping of biogenic carbon in the ocean. Mesopelagic organisms intercept about 90% of particulate organic carbon (POC) before it sinks deeper, and then respire CO_2 back into the water (Arístegui et al. 2005, 2009; Robinson et al. 2010; Costello and Breyer 2017). The mesopelagic zone is therefore a key component of the efficiency of the biological carbon pump (BCP), a crucial ecosystemic service being defined as the sum of all biological processes transporting C into the deep ocean (Eppley and Peterson 1979; Siegel et al. 2016; Le Moigne 2019).

Despite its paramount role in the BCP and thus in climate regulation, the mesopelagic zone, its global composition, and its ecology remain poorly known (Buesseler and Boyd 2009;

*Correspondence: chloe.baumas@mio.osupytheas.fr

#Robin Fuchs and Chloé M. J. Baumas contributed equally to this study.

Author Contribution Statement: C.B. conceived the idea and gathered the data. R.F. designed the rupture detection methodology. D.N. proposed and implemented the integration method of the PHP. R.F. and C.B. led the writing with significant contributions from all authors.

Additional Supporting Information may be found in the online version of this article.

Burd et al. 2010; Martin et al. 2020). The conventional sampling methods do not allow to gather representative data due to the vast size of the ocean, vertical heterogeneity, short temporal-scale research ship activities, hydrostatic pressure, and the avoidance tactics of metazoan (Robinson et al. 2010). In this respect, the lack of consensus concerning the boundaries of the mesopelagic zone is a stumbling block since the scientific community has failed to reconcile the mesopelagic C budget. Indeed, in most cases, measurements and estimates have shown a biological carbon demand often greater than the amount of POC exported (Reinthaler et al. 2006; Steinberg et al. 2008; Burd et al. 2010; Collins et al. 2015). In other words, the measured POC flux cannot support the measured metabolic C demand of prokaryotes and zooplankton altogether in the mesopelagic zone. In order to assess mesopelagic C budgets, C demand needs to be integrated over the whole mesopelagic zone, which by definition requires knowing its boundaries. Analyzing the work by Giering et al. (2014), it is worth noting that the choice of these boundaries significantly impacts the budget estimate, leading the balance toward a deficit, surplus, or a balanced C budget (see extended data fig. 5 in Giering et al. (2014)). In addition, the mesopelagic zone encompasses strong gradients in environmental conditions suggesting that the mesopelagic zone should not be considered as a homogeneous block toward the ocean. For these reasons, the mesopelagic zone boundaries need to be wisely and consistently defined before trying to provide interpretations about the mesopelagic C budget.

Similarly to Longhurst (2007), some studies have shown that the mesopelagic zone could be horizontally divided into 13–33 ecoregions by clustering physical or/and biological data (Proud et al. 2017; Reygondeau et al. 2018). However, concerning the vertical boundaries of the mesopelagic zone, fewer comprehensive data-based approaches have been proposed. In practice, this partition of the water column is often performed using fixed boundaries or thresholds. The mesopelagic zone is conventionally defined between 200 and 1000 m (Hedgpeth 1957). However, evidence begins to show that these boundaries can vary among oceanic biogeochemical provinces (Reygondeau et al. 2018), preventing accurate comparison between locations and studies. Besides, Buesseler et al. (2020) demonstrate that a fixed depths approach is not suitable for BCP efficiency assessment. Alternatively, criteria based on light and photosynthesis are often used (Lee et al. 2007). The upper boundary of the mesopelagic zone is then located where light is not sufficient for photosynthesis (between 0.1% and 1% of the surface photosynthetically active radiation [PAR] value). Yet, PAR-based approaches can only be implemented using CTD profiles acquired during the day and greatly depend on water turbidity (being therefore dependent on POC fluxes). Besides, they do not take into account the whole PAR profile but only two values: the surface value and the value at the limit depth of the euphotic zone. Other methods such as deep scattering layer (DSL) based on horizons where biomass-rich communities of zooplankton and fish stop during their daily migration have been proposed by Proud et al. (2017). These depths are readily detectable by echosounders but such methods require different measurements along the day and night as the depths of echos change, additional instruments, treatment skills, and time (e.g. Proud et al. 2015). The mesopelagic upper boundary can also be fixed below the mixed layer depth (MLD) (Giering et al. 2014; Belcher et al. 2016; Reygondeau et al. 2018). Two techniques exist to determine this depth, that is, the depth where the temperature was 0.5°C lower than surface temperature (Monterey and Levitus 1997) or the depth at which a change from the surface density of 0.125 kg m^{-3} has occurred (Levitus 1982). The main disadvantage of this method is the dependence of the result on the season and the chosen technique which provide significantly different results (Lukas and Lindstrom 1991). Instead of the euphotic zone, Owens et al. (2015) use the primary production zone (PPZ), considered as the zone between 0 m and the depth at which fluorescence reaches 10% of its highest value (Marra et al. 2014; Owens et al. 2015). Finally, these methods only provide an upper boundary for the mesopelagic zone but no lower boundary. Reygondeau et al. (2018) proposed to use the depth where the vertical POC fluxes gradient is sufficiently close to zero as a lower boundary. However, this approach determines the integration boundaries of the biogeochemical data using the biogeochemical data themselves. It hence presents an endogeneity problem for our purpose and could not be compared with the presented results.

Variables such as temperature, salinity, dissolved O_2 concentration, density, and fluorimetry data are well known and widely measured using sensors from CTD profiles, or casts, throughout the whole water column. Aside from their worldwide availability, these variables, among others, are considered to be significant ecological drivers and proxy measures of community structure or abundance (Sutton et al. 2017). These reasons make CTD profiles good candidates for moving toward a consistent and robust determination of the vertical boundaries of the mesopelagic zone.

In this study, we propose to use automatic rupture detection methods (Truong et al. 2020) applied to the CTD profiles to identify both the upper and lower boundaries of the mesopelagic zone. We name our approach RUBALIZ: a RUpture-Based detection method for the Active mesopeLagIc Zone. In essence, RUBALIZ provides boundaries that are independent of measurements of POC fluxes and/or C demand. Conversely, PAR-based methods are biased by particle load which is often used as a proxy for POC fluxes and can operate only during daytime, while the DSL method depends on migration activities. RUBALIZ can therefore be easily used for any cruise without taking care of the daytime or of the region.

In order to characterize the importance of boundary determination over the mesopelagic C budget, we present the associated integrated prokaryotic heterotrophic production (PHP) and POC flux based on the boundaries estimated by the different methods. To highlight the readiness of the proposed method, RUBALIZ has been applied to seven cruises that

occurred in the North Atlantic, the Mediterranean, the South Pacific, and the Arctic areas with contrasted stations in distinct oceanic biogeochemical provinces.

Material and procedures

In this study, data from seven cruises and 13 stations from distinct oceanic biogeochemical provinces were gathered (Table 1). These data include the potential temperature, salinity, dissolved O_2 concentration, density, and fluorimetry from CTD profiles, as well as PHP and POC fluxes. The CTD profiles were processed using the SeaDataProcess software. Only the downward CTD profiles have been used for all stations considered in this study.

Carbon fluxes

Sinking POC flux

The POC fluxes were measured using drifting sediment trap data, except for TONGA and DY032 for which they are derived from RESPIRE measurements (Boyd et al. 2015) and a Marine Snow Catcher (MSC) (Riley et al. 2012), respectively. They were communicated or already published elsewhere for this purpose (cruises and references indicated in Table 1). POC fluxes throughout the mesopelagic zone were calculated assuming that the data followed a power law as in Martin et al. (1987) for each of the 13 stations (Table 1). The knee points of the power law curves were estimated using the Unit Invariant Knee method (Christopoulos 2016).

POC fluxes were estimated at the depths determined by the RUBALIZ method (*see* below). In the special case of the PEACETIME cruise, measured data were available from 200 to 1000 m. PEACETIME POC fluxes appeared to be constant throughout this zone, indicating that the major attenuation of interest likely occurred in shallower water. In order to obtain a proper integrable profile, the POC flux at 100 m depth was estimated using the method from Henson et al. (2011). This algorithm links export efficiency (*e-eff*) to sea surface temperature (SST): $e\text{-}eff = 0.23 \times e^{(-0.08 \times SST)}$. The *e-eff* is then multiplied by the primary production (PP) to estimate the exported F_{POC}. The POC flux was not available for the PS-1 station of the KN207-03 cruise.

Prokaryotic heterotrophic production

PHP was measured by incorporation of ^{3}H-Leucine as described in (Kirchman et al. 1985) and following two different protocols according to the different studies: (i) filtration on 0.2-μm 25-mm nitrocellulose filter and (ii) microcentrifugation techniques (see references in Table 1 for details). In brief, for both protocols, a volume of seawater samples was collected with a Niskin bottle and was incubated in the dark with 20 nM (saturating concentration) of ^{3}H-Leucine between 2 and 8 h according to the depth at in situ temperature. Then, samples were processed according to the protocols of the authors and counted with a scintillation counter as described by different authors in Table 1. For depths deeper or equal to 1000 m, samples were incubated with 10 nM (saturating concentration) of ^{3}H-Leucine for 10 h at in situ temperature. At the end of the experiment, incubations were stopped by adding formaldehyde, filtered on 0.2-μm 25-mm nitrocellulose, and counted with a scintillation counter. To calculate the PHP, we used an empirical conversion factor of 1.55 ng C pmol^{-1} of incorporated ^{3}H-Leu, assuming an isotopic dilution equal to 1, according to Simon and Azam (1989).

Prokaryotic respiration

Prokaryotic respiration (PR) was estimated from measured PHP and a prokaryotic growth efficiency (PGE) according to the equation from del Giorgio and Cole (1998). We use a fixed PGE of 7%, defined as the median of 32 values, measured or estimated from literature data that were computed using a conversion factor of 1.55 ng C pmol^{-1} Leu between 50 and 1000 m (Arístegui et al. 2005; Reinthaler et al. 2006; Baltar et al. 2010; Collins et al. 2015). The choice of the PGE as well as the conversion factor value are known to strongly impact the C budget (Burd et al. 2010; Giering and Evans 2022). However, asserting their impact is out of the scope of the present work and will be conducted in a dedicated study.

Prokaryotic carbon demand and C budget discrepancy

The prokaryotic carbon demand (PCD) was computed as the sum of PHP and PR. The C budget discrepancy, $\Delta_{POC,}$ was calculated using the boundaries determined by the RUBALIZ method and the following formula:

$$\Delta_{\mathbf{POC}} = \mathrm{POC}_{\mathrm{input}} - \mathrm{PCD},$$

with $\mathrm{POC}_{\mathrm{input}}$ being the POC flux available at the benchmark methods or RUBALIZ upper boundary and PCD = PHP + PR.

In order to characterize the impact of integration boundaries over the C budget discrepancy, the discrepancy obtained for each method in a given station was compared to the average discrepancy obtained by the methods in this station using *z*-scores:

$$z - \mathrm{score} = (\Delta_{\mathrm{POC}} - \mu)/\sigma,$$

where Δ_{POC} is the discrepancy of the method of interest, and μ and σ are the mean and standard error of the discrepancies obtained by all the methods. The *z*-scores are given in standard deviation (SD) to the discrepancy mean. The higher the *z*-score, the higher the method discrepancy is compared to the other methods in this station and conversely.

Mesopelagic boundaries detection and PHP integration

Integration boundaries: The RUBALIZ method

As explained in the introduction, RUBALIZ relies on routinely collected variables: potential temperature, salinity, dissolved O_2 concentration, density, and fluorimetry data to determine the boundaries of the mesopelagic zone. Density is determined by the salinity and potential temperature. However, the functional form relating these three quantities is highly complex (Roquet et al. 2015) and cannot be retrieved

Table 1. References and sources of the data used: cruise, geographical, and temporal information along with the database name. The associated Longhurst Province of each station was determined from Longhurst (1995). Depth ranges of the measurements are given for the POC fluxes and PHP data used. They correspond to the minimum and maximum sampling depths of both quantities. The number of collected PHP points is given in Table 2.

Cruise	Station	Region	Dates	Longhurst Provinces	CTD data	POC fluxes	PHP	Data depth ranges (m)
D341	PAP	North Atlantic	Jul–Aug 2009	NADR	BODC	Giering et al. (2014)	Giering et al. (2014)	50–1000
DY032	PAP	North Atlantic	Jun–Jul 2015	NADR	BODC	Belcher et al. (2016)	Baumas et al. (2021)	10–1200
KN207-01*	QL-1	North Atlantic	Apr–May 2012	NWCS	BCO-DMO	Collins et al. (2015)	Collins et al. (2015)	5–300
KN207-01*	QL-2	North Atlantic	Apr–May 2012	NASW	BCO-DMO	Collins et al. (2015)	Collins et al. (2015)	5–300
KN207-03	PS-1	North Atlantic	Jul 2012	NADR	BCO-DMO	Collins et al. (2015)	Collins et al. (2015)	5–150
KN207-03	PS-3&4	North Atlantic	Jul 2012	ARCT	BCO-DMO	Collins et al. (2015)	Collins et al. (2015)	5–300
MALINA	430	Arctic	Jul–Aug 2009	BPLR	SEANOE	Forest et al. (2013); Miquel et al. (2015)	Ortega-Retuerta et al. (2012)	3–210
MALINA	540	Arctic	Jul–Aug 2009	BPLR	SEANOE	Forest et al. (2013); Miquel et al. (2015)	Ortega-Retuerta et al. (2012)	3–210
MALINA	620	Arctic	Jul–Aug 2009	BPLR	SEANOE	Forest et al. (2013); Miquel et al. (2015)	Ortega-Retuerta et al. (2012)	3.5–210
PEACETIME	FAST	Mediterranean Sea	Jun 2017	MEDI	Guieu et al. (2020)	Guieu et al. (2020)	Marañón et al. (2021)	1–1000
PEACETIME	ION	Mediterranean Sea	Jun 2017	MEDI	Guieu et al. (2020)	Guieu et al. (2020)	Marañón et al. (2021)	5–1000
PEACETIME	TYRR	Mediterranean Sea	Jun 2017	MEDI	Guieu et al. (2020)	Guieu et al. (2020)	Marañón et al. (2021)	5–1000
TONGA	Station 8	South Pacific	Dec 2019	SPSG	Guieu and Bonnet (2019)	Bressac et al. (unpubl.)	Van Wambeke (unpubl.)	4–1000

*The potential temperature and density were not recorded during the KN207-01 cruise. Hence, the rupture detection was performed only using the salinity, dissolved O2 concentration, and fluorimetry profiles for the KN207-01 cruise.

by the rupture detection method. Taking into account the density signal therefore provides additional information and has an influence on the observed rupture (*see* Supporting Information Fig. S2). All CTD profile sources are indicated in Table 1.

The CTDs signals of the five variables were resampled using linear interpolation to have a value at each meter depth between the minimum and the maximum depths considered, $\underline{z}$ and $\overline{z}$, respectively. For each station, all CTD profiles were set to the same length and pulled together as a matrix y. y has $(\overline{z}-\underline{z})$ rows and a number of columns equal to five times the number of CTDs (each profile is made of five curves: the potential temperature, salinity, dissolved O_2 concentration, density, and fluorimetry). The rupture detection is performed over y and looks for common rupture points over all CTD and flux signals. In order for all CTD variables to be within the same magnitude, y was centered and reduced before performing the rupture detection. The number of CTD profiles available for each station is given in Supporting Information Table S1.

The rupture detection method was based on a kernelized mean change (Harchaoui and Cappe 2007; Truong et al. 2020). This was motivated by the fact that ruptures in the signal seemed more related to mean changes rather than changes in other statistical moments such as the variance. Besides, the kernelized mean change cost function did not make parametric assumptions about the shape of the statistical distribution of the data.

More formally, y was plunged into a reproducing kernel Hilbert space (rkhs) associated with a kernel function $k(.,.): R^d \times R^d \to R$ such that $k\left(y_z, y_{z'}\right) = \exp\left(-\gamma \left\| y_z - y_{z'} \right\|\right)$, with γ a positive bandwidth parameter. The mapping function between the original space and the rkhs is denoted by $\phi: R^d \to H$ and is implicitly defined by the requirement that:

$$< \phi(y_z) \mid \phi(y_{z'})_H > = k(y_z, y_{z'}) \text{and} \left\| \phi(y_z) \right\|_H^2 = k(y_z, y_z), \qquad (1)$$

for all embedded samples $(\phi(y_z), \phi(y_{z'})) \in R^d \times R^d$ and $\|.\|$ is the Euclidean norm.

Intuitively, the algorithm tries to split the full embedded signal $\{\phi(y_z)\}_{z \in [\underline{z}, \overline{z}]}$ into sub-signals $\{\phi(y_z)\}_{z \in [a,b], \underline{z} \le a \le b \le \overline{z}}$, such that each subpart of the signal is the closest to its mean and the farthest from the mean of the other subparts of the signal. In practice, this is captured by the following cost function c_{kernel} to minimize:

$$c_{\text{kernel}}(y_{a..b}) := \sum_{z=a}^{b} \left\| \phi(y_z) - \overline{\mu}_{a...b} \mid \right|_H^2,$$

where $y_{a..b}$ is the subsignal between depths a and b, $\overline{\mu}_{a..b}$ is the mean of the embedded subsignal $\{\phi(y_z)\}_{z \in [a,b], \underline{z} \le a \le b \le \overline{z}}$, and $\|.\|_H^2$ as defined in (1).

This cost function was minimized using a binary search method (Olshen et al. 2004), which determined an approximate minimum of the cost function using a sequence of twofold partitions of the signal.

The main hyperparameters to set in the method are γ the bandwidth parameter, $\underline{z}$ and $\overline{z}$. γ was set to the inverse of the median of the pairwise squared euclidean distances between all the samples of the full signal, following the heuristic given by Truong et al. 2020. As no consensus on the absolute upper and lower boundaries of the mesopelagic zone exists, several values of $\underline{z}$ and $\overline{z}$ could be specified to run the method. In order to assess the sensitivity of the approach to the choice of these two hyperparameters, we have estimated the boundaries for 10 equally spaced $\underline{z}$ and $\overline{z}$ values. The boundaries determined by the method correspond to the mean boundary values found for these 10 $\underline{z}$ and $\overline{z}$ values. The associated standard errors give an indication about the sensitivity of the results to the choice of $\underline{z}$ and $\overline{z}$. Thus, we have set $\underline{z}$ to 0 m and let $\overline{z}$ vary between 280 and 320 m to determine the upper boundary of the mesopelagic zone. To determine the lower boundary of the mesopelagic zone, the algorithm was run between the identified upper boundary and $\overline{z}$ varying between 1000 and 1300 m. The identified upper and lower boundaries are referred to as z_{upper} and z_{lower}, respectively.

A summary of the full rupture detection pipeline is given in Supporting Information Fig. S1.

PHP integration

The relationship between daily PHP flux and depth Z is commonly considered as a power law function of the form:

$$\text{PHP} = kZ^m, \qquad (2)$$

where k and m are parameters. When taking a log transformation so that setting $X = \ln(Z)$ and $Y = \ln(\text{PHP})$, model (2) can be re-expressed as

$$Y = b + aX,$$

where $b = \ln(k)$ and $a = m$. Estimation of parameters k and m is then achieved by linear regression using an observed sample $(x_i = \ln(z_i), y_i = \ln(\text{php}_i))$, $i = 1, \ldots, n$ of size n. However, the observation of the scatterplot of observations (x_i, y_i) (Fig. 1a) rather suggests that variables X and Y are connected through a piecewise linear model such that an estimate $\widehat{y}$ is expressed as

$$\widehat{y}(x) = \begin{cases} a_1 x + b_1 \text{if} x \le x_t, \\ a_2 x + b_2 \text{if} x > x_t, \end{cases}$$

under some continuity constraint of the form $\widehat{y}_1(x_t) = \widehat{y}_2(x_t)$.

Slope parameters a_1, a_2, intercept parameters b_1, b_2, and threshold parameter x_t are estimated when minimizing the sum of squares of the errors between data and model such that

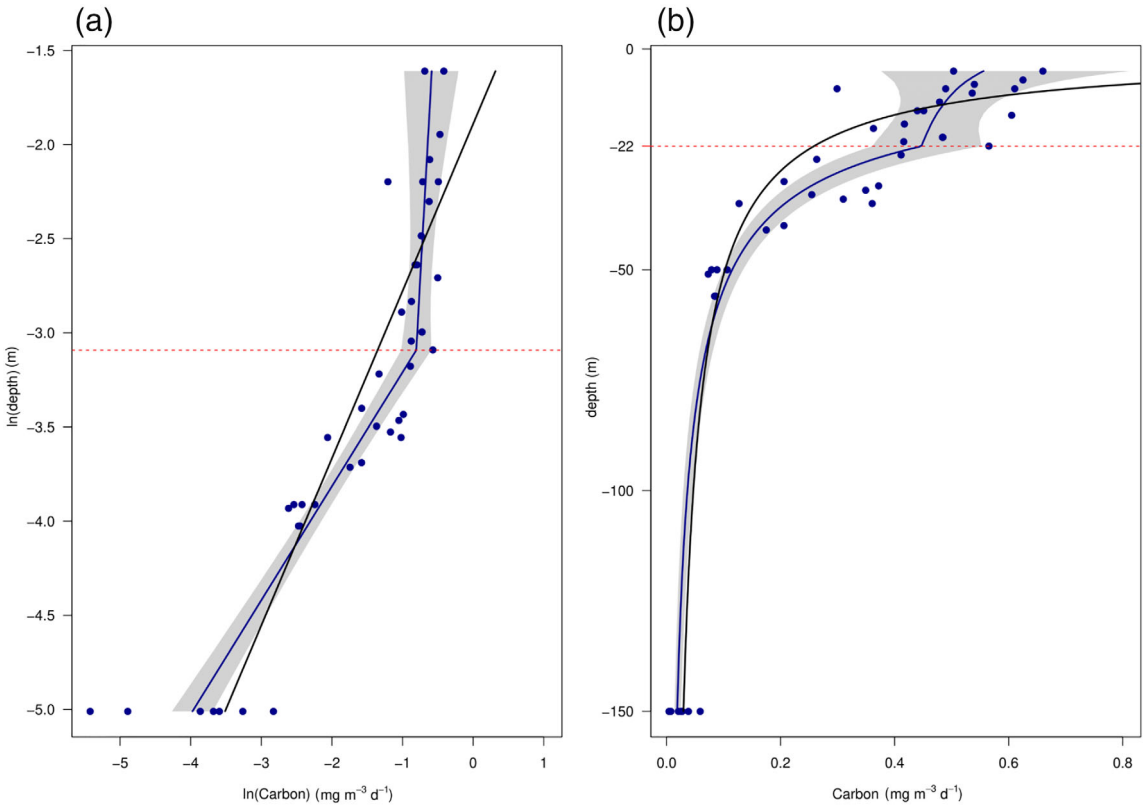

Fig. 1. Example of simple linear model and piecewise linear model fits on KN-207 03 PS3&4 data. The blue points are the observations, the black curve represents the simple linear fit, and the blue curve the piecewise fit. The red dashed line is the estimated threshold depth where the piecewise model changes and gray areas are 95% confidence intervals. The model fits are shown on the log-data (**a**) and on the original data (**b**).

$$\mathrm{SSE}(a_1, a_2, b_1, b_2, x_t) = \sum_{i=1}^{n} \left[y_i - \widehat{y}(x_i)\right]^2.$$

For a fixed value of x_t, the vector of parameters $\alpha = (b_1, a_1, a_2)'$ and the parameter b_2 are solution of the linear system given with

$$\begin{cases} \widehat{\alpha} = (\boldsymbol{X}\prime\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \\ \widehat{b_2} = (\widehat{a}_1 - \widehat{a}_2) \times x_t + \widehat{b}_1, \end{cases} \tag{3}$$

where $\mathbf{y} = (y_{(1)}, \ldots, y_{(n)})'$ is the vector of the observations y_i when the observations x_i have been sorted in ascending order. Matrix $\mathbf{X}$ is the $n \times 3$ design matrix such that

$$X = \begin{bmatrix} 1 & x_{(1)} & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_{(n_1)} & 0 \\ 1 & x_t & x_{(n_1+1)} - x_t \\ \vdots & \vdots & \vdots \\ 1 & x_t & x_{(n)} - x_t \end{bmatrix},$$

where $x_{(i)}$, $i = 1, \ldots, n$ is the sequence of the observations x_i sorted in ascending order, $1 \le n_1 \le n$ is the position of the last observation x_{n_1} such that $x_{n_1} \le x_t$. Only the search for the optimal value of x_t is achieved numerically (using any well-suited 1D root-finding algorithm) such that the solution is given by:

$$\widehat{x}_t = \underset{x_{(1)} \le x_t \le x_{(n)}}{\mathrm{argmin}}\ \mathrm{SSE}.$$

As shown in Eq. 3, the system owns five parameters but only 4° of freedom since the value of intercept b_2 is constrained by the continuity between both lines. The piecewise linear model is equivalent to a regression spline of degree 1 with one free knot (the threshold x_t).

Once an optimum threshold value $\widehat{x}_t$ has been found, the theory of linear regression provides tools for drawing confidence intervals for parameters a_1, a_2, b_1, and b_2. Under some normality assumptions of the residuals, the parameter α follows approximately a multivariate normal distribution with estimated mean $\widehat{\mu} = \left(\widehat{b}_1, \widehat{a}_1, \widehat{a}_2\right)'$ and estimated covariance matrix $\widehat{\Sigma} = \widehat{\sigma}^2 (X\prime X)^{-1}$ where $\widehat{\sigma}^2 = \frac{1}{n}\mathrm{SSE}\left(\widehat{a}_1, \widehat{a}_2, \widehat{b}_1, \widehat{b}_2, \widehat{x}_t\right)$ is the estimated variance of the residuals. The value of $\widehat{b}_2$ is deduced

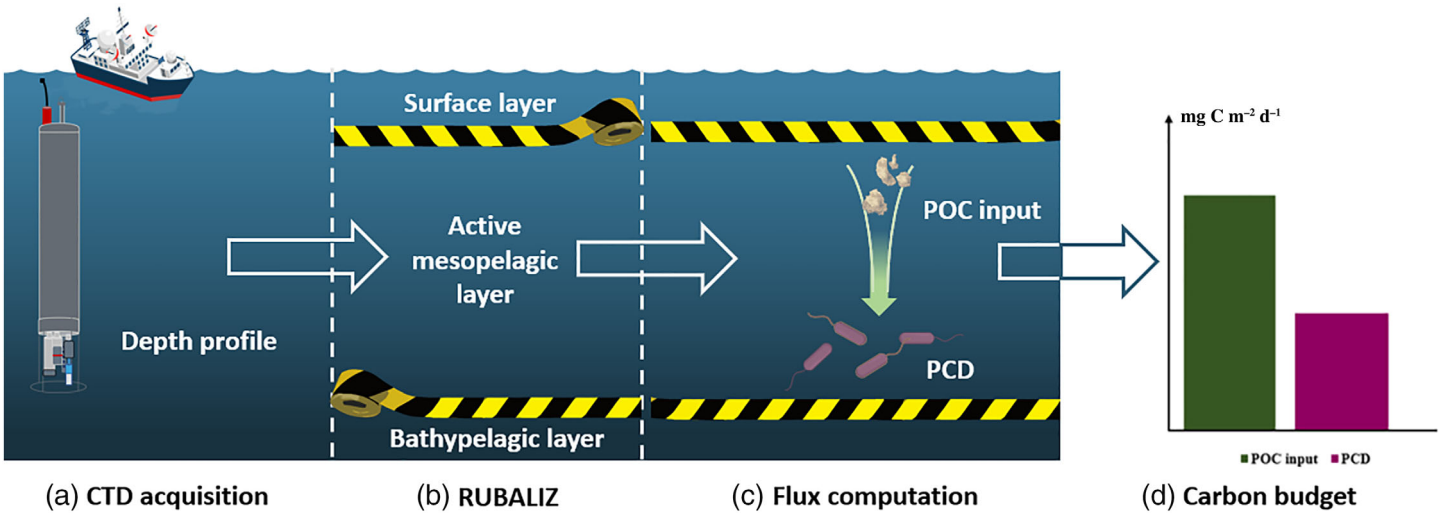

Fig. 2. Presentation of the RUBALIZ rupture detection pipeline. (**a**) Several potential temperature, salinity, dissolved O_2 concentration, density, and fluorimetry depth profiles are acquired. (**b**) RUBALIZ takes these five profiles and identifies the upper and lower boundaries of the active mesopelagic zone. (**c**) These boundaries are used to compute the gravitational POC flux input to the active mesopelagic zone (AMZ), integrate the PCD profiles and provide C budgets (**d**).

from Eq. 3. A 95% confidence interval can then be computed for the piecewise linear model and plotted into the original system of coordinates (Fig. 1b). The confidence interval lengths in the linear regression case were three times bigger than in the piecewise regression case (not shown).

Once every piecewise linear model has been fitted, computation of integrated PHP fluxes along depth is achieved using the explicit formulation for the integral

$$I_s = \frac{\exp\left(\widehat{b}_1\right)}{\widehat{a}_1+1}\left(z_t^{\widehat{a}_1+1} - z_{\text{upper}}^{\widehat{a}_1+1}\right) + \frac{\exp\left(\widehat{b}_2\right)}{\widehat{a}_2+1}\left(z_{\text{lower}}^{\widehat{a}_2+1} - z_t^{\widehat{a}_2+1}\right),$$

where $\widehat{a}_1, \widehat{a}_2, \widehat{b}_1, \widehat{b}_2,$ and $\widehat{z}_t = \exp(\widehat{x}_t)$ are estimated parameters from piecewise regression of data $(z_i,\ \text{php}_i)$, $i = 1, \ldots, n$ sampled on cruise s. As these parameters are associated with a 95% confidence interval, it is also possible to appreciate the uncertainty of the estimations of integrated carbon fluxes.

Benchmarks methods

Finally, the benchmark approaches, namely the approaches based on the PAR values (Ez0.1 and Ez1), on the Mixed Layer Depth or PPZ, enabled to determine the beginning of the mesopelagic zone but not its end. Hence, in order to revise C budget estimations, we have set the end of the mesopelagic zone for the benchmark approaches to 1000 m deep in agreement with the literature conventional value. Conversely, the lower boundary of the mesopelagic zone was properly determined by the method in the case of RUBALIZ.

General approach summarized

The general approach of the paper is summarized in Fig. 2 and the rupture detection itself in Supporting Information Fig. S1. The code and data to reproduce the results are available at https://github.com/RobeeF/rubaliz_paper and the DOI associated specifically with the RUBALIZ package is: 10.5281/zenodo.6840898. The RUBALIZ package can be downloaded for stand-alone use at https://github.com/RobeeF/rubaliz/

Results

In this section, we first compare the RUBALIZ approach to existing methods. Then, we show that the zone identified by RUBALIZ matches the biogeochemically active part of the mesopelagic zone, which is the zone of interest in C budget assessments. Finally, the estimated boundaries are used to integrate biogeochemical data and we compute the related C budgets.

Assessment of the approach

RUBALIZ boundaries

The method determined the upper and lower boundaries of the mesopelagic zone for the 13 locations considered. An illustration of the boundaries found for DY032 PAP, MALINA 430, PEACETIME ION, and KN207-01 QL-2 along with the associated variables of a CTD-cast is shown in Fig. 3. The profiles were not smoothed before performing the rupture detection and presented small amplitude fluctuations that did not influence the boundaries found. The upper boundaries identified by RUBALIZ were located right below the fluorescence peaks and below the significant O_2 variations, that is, at 127, 79, 118, and 191 m deep for DY032 PAP, MALINA 430, PEACETIME ION, and KN207-01 QL-2, respectively. This result is confirmed by the sensitivity analysis reported in Supporting Information Fig. S2, which showed that the main variables driving the upper boundary estimation are O_2 and

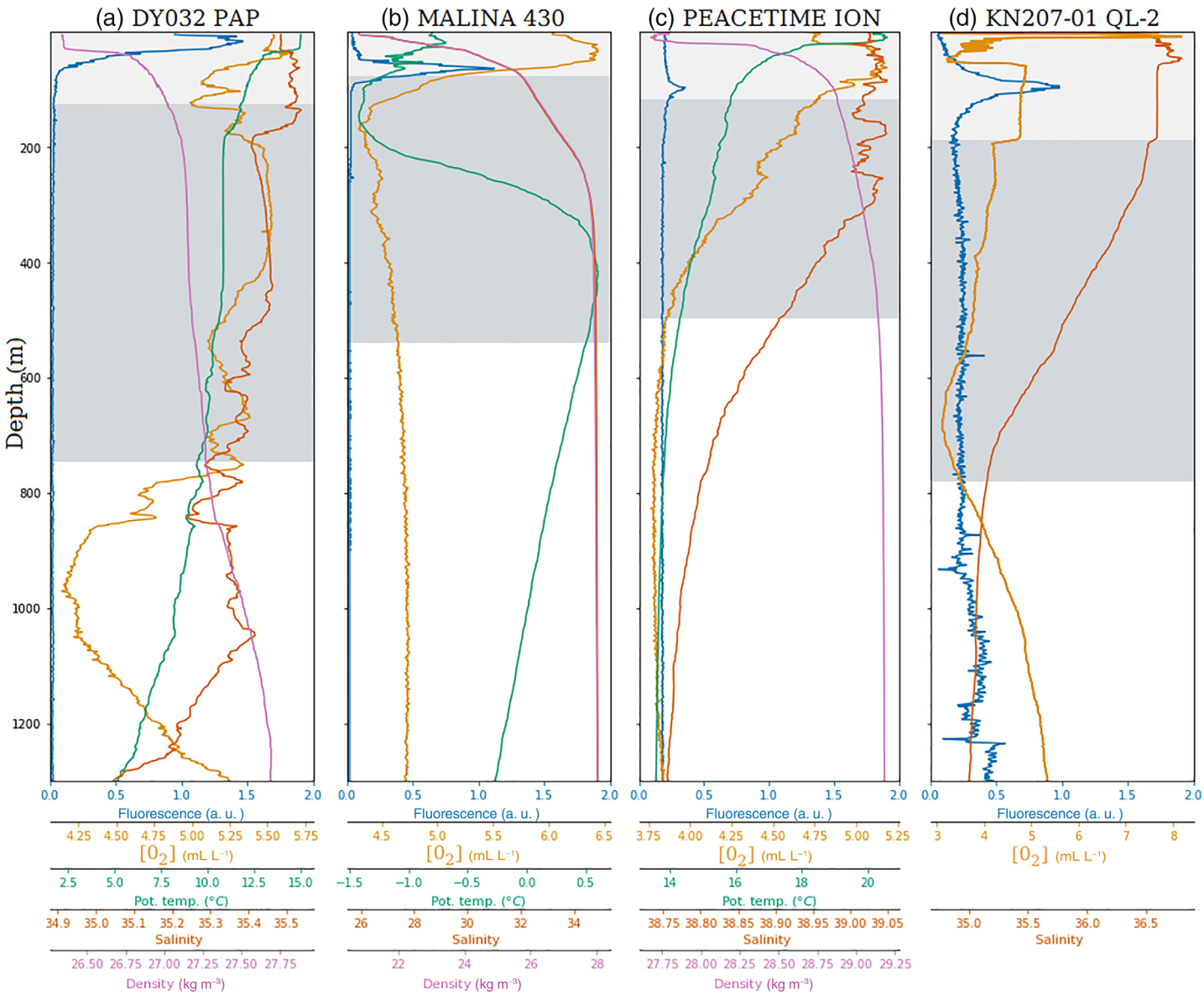

Fig. 3. Illustration of the RUBALIZ boundaries of the mesopelagic zone (shaded in grey) along with one CTD signal for (**a**) PAP DY032, (**b**) MALINA 430, (**c**) PEACETIME ION and (**d**) KN207-01 QL-2. The fluorescence curve is in blue, the O_2 concentration in light orange, the potential temperature in green, the salinity in dark orange, and the density in pink. The curves are the raw curves and were not smoothed beforehand.

fluorescence. The upper boundaries of the other stations presented comparable values (e.g., 135 m for D341 PAP and 153 m for Tonga Station 8).

The lower boundaries were estimated between 490 m (KN207-03 PS-1) and 783 m (KN207-01 QL-2) (*see* Supporting Information Table S1). As presented in Fig. 2, these boundaries reflected abrupt changes in most variables (PAP), which were located below an inflection point in some profiles (e.g., O_2 at KN207-01 QL2, or potential temperature at Station 430), or at a slope rupture (e.g., O_2 signal at PEACETIME ION). The sensitivity analysis (Supporting Information Fig. S2) highlighted the prime importance of the O_2 signal in the lower boundary determination, followed by the salinity and potential temperature. The upper boundaries were more precisely estimated than the lower boundaries (Supporting Information Table S1; Fig. S4).

Comparison with benchmark methods

Figure 4 depicts the mesopelagic vertical boundaries established by the RUBALIZ approach with regards to existing approaches, namely 1% and 0.1% PAR (Ez1 and Ez0.1), MLD computed on temperature or density, the PPZ, and the usual fixed 200–1000 m boundaries. Regardless of the method, the upper boundary was always set shallower than the standard

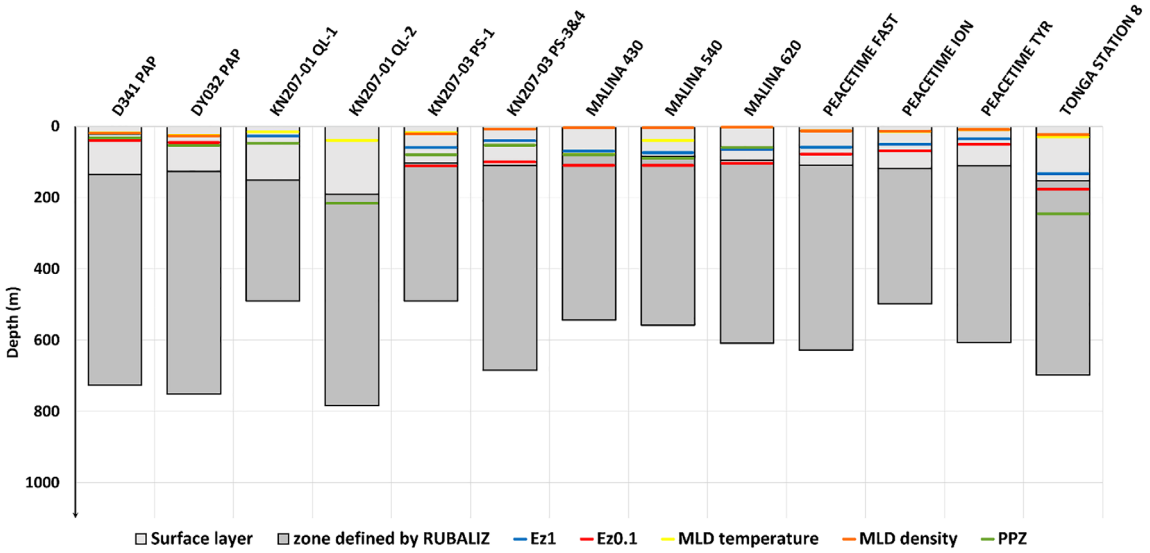

Fig. 4. Comparison of the upper boundaries found by each method and presentation of the lower boundary found by RUBALIZ. The missing bars are due to inoperant methods at a given station (e.g., unavailable data, the variable threshold used by the method did not exist).

200 m value, except for the PPZs of Station QL-2 (KN207-01) and Station 8 (TONGA) with 216 and 246 m, respectively. The RUBALIZ upper boundary was generally deeper than the upper boundary of the other methods (PAP for both cruises, QL-1, and the three PEACETIME stations) or equivalent to Ez0.1 and PPZ (KN207-03 stations, MALINA stations). In all cases, the shallowest depth appeared to be determined by MLD temperature or MLD density and the deepest by RUBALIZ, PPZ, or Ez0.1 (Fig. 4). The upper boundaries often present the same general depth ordering, from the shallowest to the deepest: MLD density, MLD temperature, EZ1, PPZ, EZ0.1, and RUBALIZ.

Concerning the lower boundary, RUBALIZ shallowest results corresponded to the PS-1 station (490 m deep), the deepest to the QL-2 station (783 m deep), and a mean depth of 621 m for all 13 stations. Therefore, these lower boundaries were always shallower than the 1000 m classically used to define the end of the mesopelagic zone.

RUBALIZ targets the active mesopelagic zone

The vertical boundaries were determined by RUBALIZ using the five physical CTD profiles independently from the POC and C demand fluxes. However, we highlight an important result: the identified upper and lower boundaries are closely linked to the major attenuation of the POC flux. First, the onset of the POC flux attenuation begins at the RUBALIZ upper boundary as indicated in Fig. 5. Indeed, Fig. 5 shows a 1 : 1 line between the POC curve knee points and the RUBALIZ upper boundary ($R^2 = 0.94$) with an average spread of $\pm$ 31 m. This indicates that the onset of the POC flux attenuation begins at the RUBALIZ upper boundary. Second, below the RUBALIZ lower boundary the POC flux attenuation is limited. Since POC fluxes are represented by a power law, the difference between the maximum and minimum values at the related depths quantifies the attenuation of the POC flux. We calculated the attenuation within the RUBALIZ boundary and compared it to the attenuation between the lower RUBALIZ boundary and the 1000 m depth. We found that over 90% of the POC flux is attenuated within the RUBALIZ boundaries.

Regardless of seasons and locations, the area bounded by the upper and lower boundaries of the RUBALIZ always appears to be located near the maximum attenuation and in the vicinity of the depth where the POC flux attenuation strongly slows down. The boundaries determined on the physical conditions are hence consistent with the patterns observed on the biological fluxes. As a result, we propose to call the "active mesopelagic zone" (AMZ), the zone determined by RUBALIZ, and this denomination will be used in the sequel.

C budget assessment

Integrating biogeochemical rates data

The active mesopelagic zone boundaries presented above for each station were used to integrate PHP fluxes and construct C budgets. The cruises presenting the highest integrated PHP were TONGA (54.16 mg C m^{-2} d^{-1}), PEACETIME (26.87 mg C m^{-2} d^{-1} on average), and DY032 (24.82 mg C m^{-2} d^{-1}). The different stations of a given cruise presented analogous PHP except for PEACETIME FAST ($\sim$ 2 times higher than the two other stations) and MALINA Station 620 ($\sim$ 10 times higher than the two other stations). The R^2, which informs about how well the estimated relationship described the data, was higher than 0.62 for all stations, except for station QL-2, with a mean of 0.85 (*see* Table 2). The best estimations were performed for TONGA and PEACETIME ($R^2 \geq 0.92$). The largest confidence intervals with respect to the

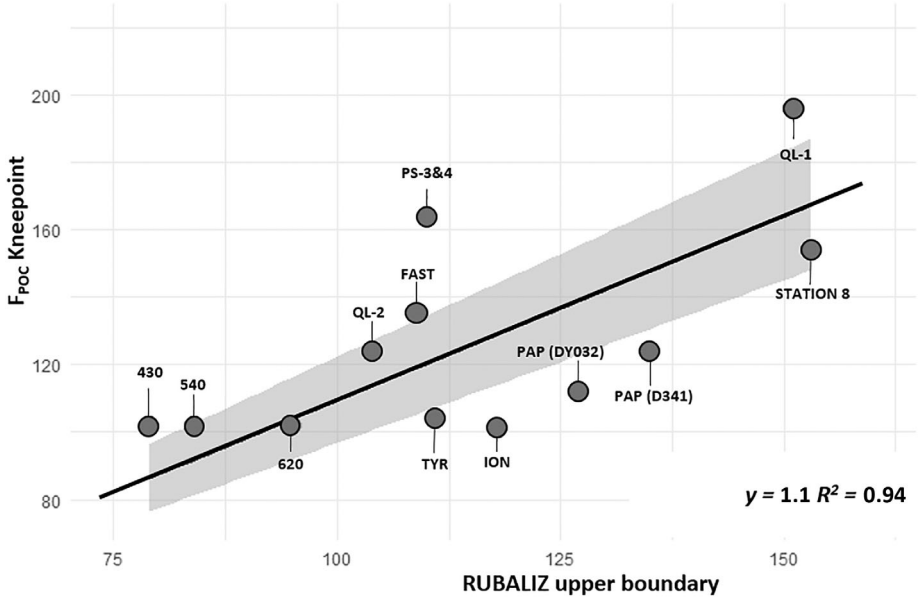

Fig. 5. Linear relationship between the boundaries detected by RUBALIZ and biogeochemistry data (symbolized by the knee points of the respective POC fluxes estimated by a power law). The intercept coefficient was not significant and the *p*-value of the slope coefficient was 6.63×10^{-08}. The shaded area corresponds to the 95% confidence interval.

estimated PHP were due either to a low R^2 (QL-2) or to a limited number of points (MALINA), but the confidence interval size remained inferior or equal to the estimated PHP for all stations.

Assessing active mesopelagic zone C budget

The C budget discrepancy, Δ_{POC}, that is, the difference between gravitational sinking POC_{input} and PCD, was negative except in the KN207-03 PS-3&4 station for PPZ, Ez1, MLD temperature, and density methods (Supporting Information Fig. S3). This implies that POC gravitational input is not sufficient to satisfy the PCD in most cases regardless of the method. The estimated discrepancy depends significantly on the boundary determination method used and cruise with a mean of -325.10 ± 283.13, a maximum of -1099.09, and a minimum of 108.05 mg C m^{-2} d^{-1} both obtained by MLD density estimation by TONGA Station 8 and KN207-03 PS-3&4, respectively. Figure 6 presents the *z*-score per station associated with all the benchmark methods and RUBALIZ. For all negative carbon discrepancies, the higher the *z*-score is, the less negative the discrepancy is. Thus, using RUBALIZ boundaries reduces the C budget discrepancy compared to the other methods. In the special case of KN207-03 PS-3&4, RUBALIZ found a slightly more negative discrepancy compared to the

Table 2. Estimated integrated PHP fluxes using the detected boundaries of RUBALIZ.

cruise	station	Active mesopelagic zone boundaries	PHP estimated (mg C m^{-2} d^{-1})	PHP Confidence Interval (mg C m^{-2} d^{-1})	R^2	Number of points
D341	PAP	(135; 726)	15.26	(11.27; 21.23)	0.82	16
DY032	PAP	(127; 751)	24.82	(21.45; 28.80)	0.89	82
KN207-01	QL-1	(151; 490)	11.41	(9.17; 14.34)	0.70	39
KN207-01	QL-2	(191; 783)	6.85	(4.89; 11.49)	0.42	24
KN207-03	PS-1	(104; 490)	13.99	(12.34; 16.03)	0.80	28
KN207-03	PS-3&4	(110; 684)	13.37	(11.40; 15.93)	0.88	42
MALINA	430	(79; 544)	3.54	(2.36; 5.99)	0.62	6
MALINA	540	(84; 558)	3.22	(2.46; 4.52)	0.93	6
MALINA	620	(95; 609)	30.99	(22.73; 44.70)	0.98	6
PEACETIME	FAST	(109; 628)	38.55	(31.71; 46.85)	0.92	54
PEACETIME	ION	(118; 498)	18.47	(16.52; 20.80)	0.96	31
PEACETIME	TYR	(111; 606)	23.60	(19.73; 28.38)	0.95	25
TONGA	Station 8	(153; 702)	54.16	(45.88; 64.12)	0.96	14

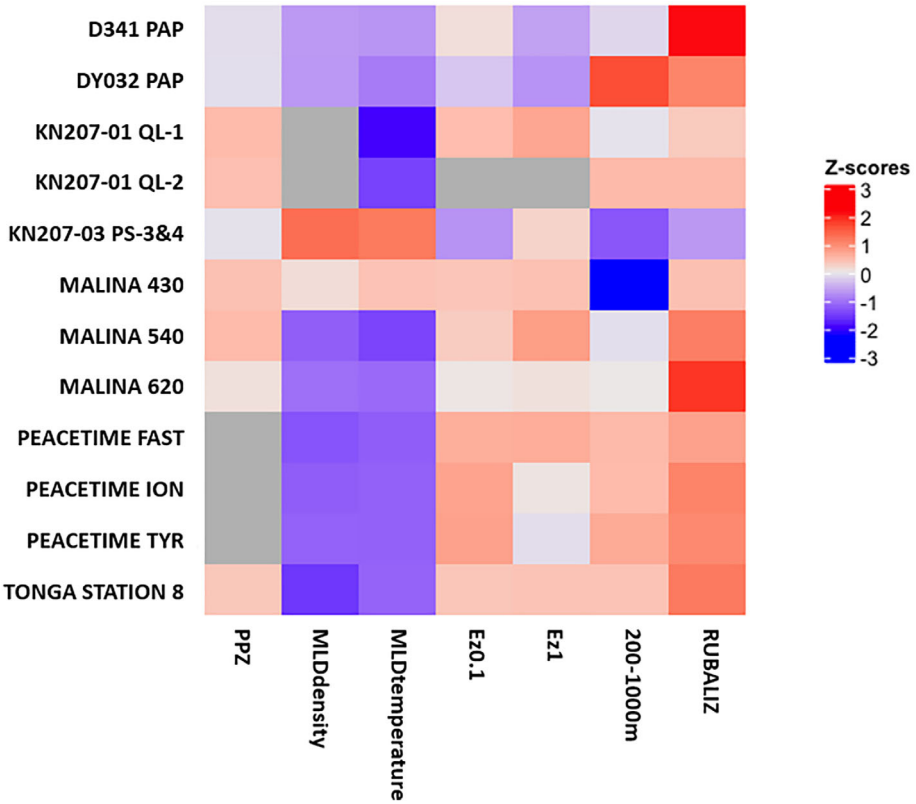

Fig. 6. *Z*-scores per station of the carbon budget discrepancy for all methods. The *z*-score is the number of standard deviations separating a raw score from the mean. The gray cells correspond to stations for which a given method could not determine an upper boundary. The POC flux of KN207-03 PS-1 was not available and the associated *z*-score was not represented here.

other methods. Compared to the usual 200–1000 m boundaries, the other benchmark methods estimated more pronounced negative discrepancies, especially the MLD-based methods (Supporting Information Fig. S3).

Discussion

A robust methodology for boundary determinations

The mesopelagic zone is the scene of the highest attenuation rate of POC flux, a key point for assessing carbon sequestration across the ocean (Robinson et al. 2010). However, to date, no easy-to-use and universal method exist to define the boundaries of the mesopelagic zone in a meaningful and consistent way.

The customary definition of the mesopelagic zone between 200 and 1000 m depth (since Hedgpeth 1957) is practical from a theoretical point of view but not relevant to compare studies of different biogeochemical provinces. Indeed, recent research (e.g. Reygondeau et al. 2018) has challenged this view and demonstrated the variability in time and space of these vertical boundaries.

Here, we propose to revisit the mesopelagic boundary determination by taking into account the vertical variability of five variables well known to characterize the water column: fluorescence, potential temperature, salinity, density, and O_2 concentration (Sprintall and Cronin 2001; Lavigne et al. 2015). Fluorescence and [O_2] signals are not by default collected by the CTDs while they have a strong influence on the boundary estimations (Supporting Information Table S2; Fig. S2). Yet, in biogeochemistry studies, which is the field for which RUBALIZ was designed for, collecting [O_2] and fluorescence signals is common and can be performed by a simple request to the dedicated technical staff.

The complete vertical profiles of these five variables were used all together contrary to existing methods that define a threshold operating on a single variable. As demonstrated in our sensitivity analysis (Supporting Information Fig. S2), all five variables participated in the determination of the boundaries, whereas the benchmark approaches were based on a single variable. Furthermore, using the whole profiles and a nonparametric mean-change kernel rather than a single threshold, makes RUBALIZ less sensitive to outlier points frequent in in situ data, and robust to missing profiles as during the KN207-01 cruise. The rupture detection approach also makes RUBALIZ robust to useless/noisy profiles as shown in Supporting Information Fig. S5. Finally, the general trends shared by different CTD casts at a given station were captured without being influenced by cast-specific background noise.

The choice of the depth intervals on which the upper and lower boundaries were determined ($\underline{z}$ and $\overline{z}$) constituted one of the main limits of our approach. However, the low estimation variances, especially for the upper boundary, underlined that this choice was not the main source of variability in our estimates and that the lower boundary was more difficult to estimate. Other rupture criteria than the kernelized mean change, such as Gaussian process change point models or least absolute deviation methods, were implemented (results not shown) but focused more on local features of the profiles rather than on the changes in trends and inflection points.

Analysis of the ruptures found

The upper boundary delimited by RUBALIZ was located deeper than the ones provided by the benchmark models but shallower than the 200 m boundary. Conversely, the RUBALIZ lower boundaries were all located above 1000 m, certainly denoting that most of the mesopelagic remineralization occurs before 1000 m (Robinson et al. 2010). We have calculated that over 90% of the POC flux is attenuated within the RUBALIZ boundaries. The RUBALIZ boundaries were determined using exogenous CTD physical data, which enabled to separate the zone boundary determination problem from the integration of biological fluxes problem. Yet, the so-determined upper boundaries matched POC flux attenuation knee points (Fig. 5), and the lowest boundaries delimit the end of the maximum POC fluxes attenuation zone, indicating that the five physical CTD variables used shared common information with the biological POC flux and motivated the denomination of "active mesopelagic zone" (AMZ). The link between the active mesopelagic zone and biogeochemical processes could be explained by the influence of environmental variables on how prokaryotes degrade POC. Indeed, prokaryotes diversity (DeLong et al. 2006; Ghiglione et al. 2008; Severin et al. 2016; Garel et al. 2019; Sebastián et al. 2021), growth efficiency (del Giorgio and Cole 1998; Nagata et al. 2010) or even gene expression (Bergauer et al. 2018) are known to be dynamic according to physical variables. Most of these processes are still poorly understood and the associated data are scarce (Burd et al. 2010; Robinson et al. 2010; Baumas et al. 2021; Giering and Evans 2022). However, the close link between environmental physical variables and prokaryotic activities could, together, strongly drive how POC flux is attenuated.

RUBALIZ as a ready-to-use shipboard tool

Given the link existing between POC flux and RUBALIZ boundaries, the present method could be a useful tool to adapt the sampling strategy during seagoing cruises. From an operational perspective, given the cost, manpower, specific equipment, low sea state, and post-analysis efforts required to use sediment traps and get POC fluxes from them (McDonnell et al. 2015), RUBALIZ could help to optimize sediment trap position and deployments at sea. From our results, we recommend using at least one CTD cast down to 1300 m to fully resolve the active mesopelagic zone. Indeed, as shown in Supporting Information Fig. S4 in the PEACETIME FAST case, RUBALIZ provided reliable estimates from the first acquired CTD cast, the spread with the final estimation being less than 1 m and 35 m for the upper and lower boundaries, respectively. Hence, the physical and biological sampling strategy can be designed at the very beginning of a station occupation, when only a few CTDs are available. Similarly, after strong weather events, hydrography could be significantly modified (Lavigne et al. 2015) and RUBALIZ could be used to rapidly adapt the sampling strategy.

C budget and perspectives

PHP data are usually integrated by trapezoidal rule (Reinthaler et al. 2006; Gazeau et al. 2021) or using a power law (Giering et al. 2014). Here, we showed that using a piecewise model with a single node on the log-data provided a better fit to the data, by increasing the R^2 of the fit and decreasing the confidence interval of the PHP fluxes estimated (*see* Supporting Information Fig. S6). This model could be applied as soon as at least three PHP samples are collected, the most influencing PHP points being the ones located at the very surface, near the node, and after the RUBALIZ lower boundary.

The PHP fluxes were integrated using the boundaries identified by the benchmark methods and RUBALIZ at each station to compute C budgets. Our results emphasized that regardless of the method used to determine the integration boundaries, the C budget discrepancies were systematically negative (except for KN207-03 PS-3&4). This problem implying that the estimated POC gravitational input is not sufficient to satisfy the estimated PCD has been an issue for several decades (Burd et al. 2010). However, we show that RUBALIZ significantly reduced this discrepancy (Fig. 6), and thus provided a solid basis for comparison between mesopelagic C budgets from different regions and seasons.

Figure 7 presents a first attempt to consistently compare 13 C budgets to one another. By comparing these various C budgets from contrasting regions and seasons all together, we conclude that this discrepancy remains a widespread feature of the ocean and that additional research is still needed to resolve this issue. RUBALIZ is only a first step and a better estimation of C fluxes in the mesopelagic zone still requires further investigation about (i) the validity of PGE used to estimate PR (Burd et al. 2010) and of the CF Leu/C used to convert leucine incorporation into PHP (Giering and Evans 2022); (ii) the role of attached to sinking particles prokaryotes which are not included here as we only use free-living PHP data from Niskin (Baumas et al. 2021); (iii) the high-pressure effect as is it now proved that pressure can have an important effect on prokaryotic activities and diversity, especially at depths below 200 m (Garel et al. 2019; Tamburini et al. 2021); (iv) the additional C sources such as from particles

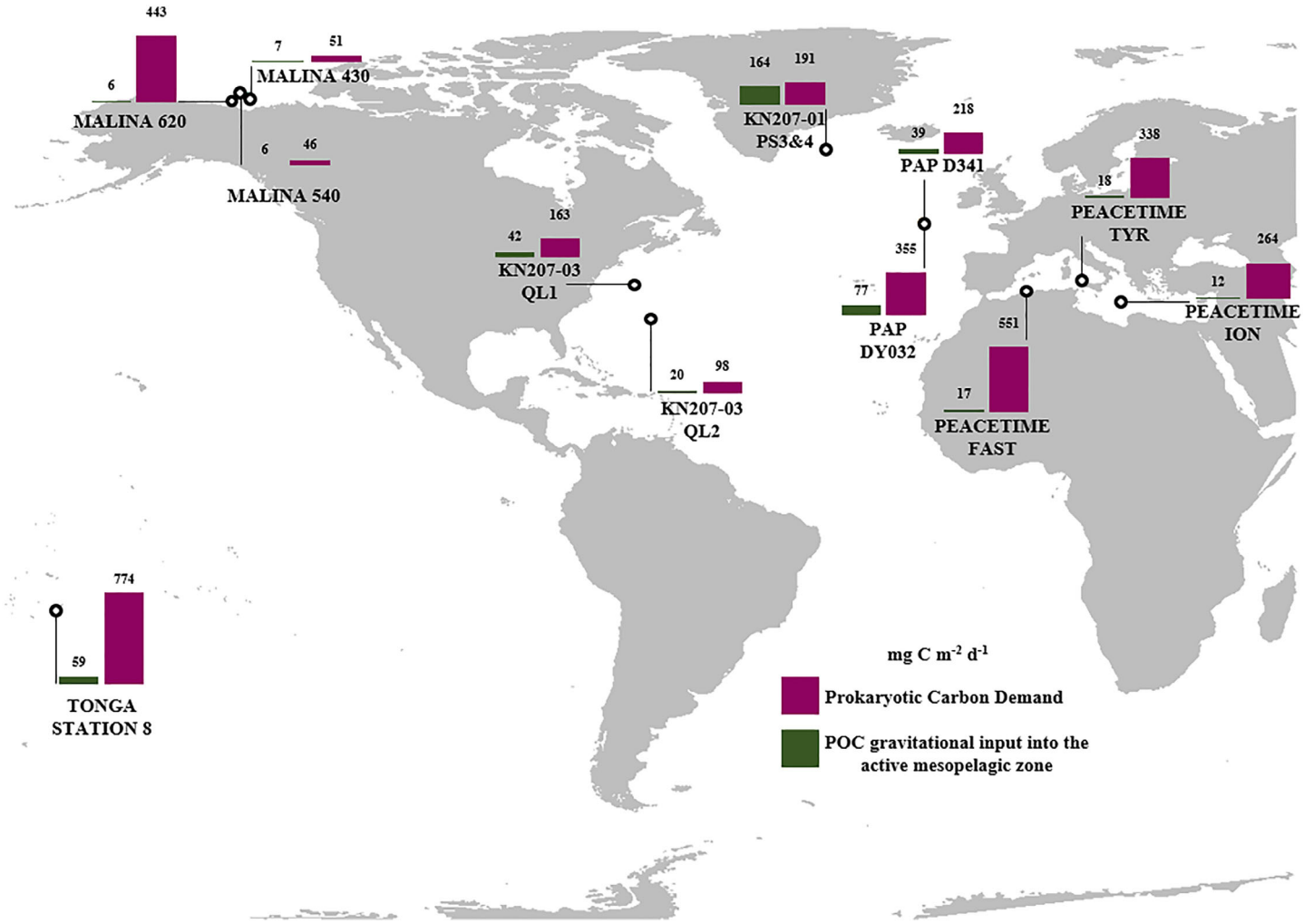

Fig. 7. RUBALIZ derived C budget of the active mesopelagic zone (AMZ). Pink bars represent the PCD and green bars the POC_{input}. The POC flux of KN207-03 PS-1 was not available and the associated C budget was not represented here.

injection pump or nychthemeral migrations of zooplankton or micronekton (Steinberg and Landry 2017; Aumont et al. 2018; Boyd et al. 2019); and (v) assessing the contribution of chemolithoautotrophs as a new source of OC in the dark ocean (Herndl and Reinthaler 2013). The research field aiming to decipher the C cycle in the mesopelagic zone would benefit from a worldwide effort in mapping the RUBALIZ active mesopelagic zone across regions and seasons. In that sense, autonomous and semi-autonomous platforms such as Argo floats data covering the global ocean could be used to understand how the active mesopelagic zone is varying and as a second step how to model it to predict how POC sequestration may evolve in the future. Finally, RUBALIZ, in addition to a precise sampling strategy directly on board, provides a first step toward a world mapping in Longhurst et al. style of the active mesopelagic zone. Such large-scale mapping could in turn be linked to particle flux composition, prokaryotic diversity and activities, zooplankton ecology, and POC degradation processes in order to set a new regionalization of the BCP efficiency in response to changing ocean dynamics.

Data availability statement

Data are available at https://github.com/RobeeF/rubaliz_paper.

References

Arístegui, J., C. M. Duarte, J. M. Gasol, and L. Alonso-Sáez. 2005. Active mesopelagic prokaryotes support high respiration in the subtropical northeast Atlantic Ocean. Geophys. Res. Lett. **32**: 1–4. doi:10.1029/2004GL021863

Arístegui, J., J. M. Gasol, C. M. Duarte, and G. J. Herndld. 2009. Microbial oceanography of the dark ocean's pelagic realm. Limnol. Oceanogr. **54**: 1501–1529. doi:10.4319/lo.2009.54.5.1501

Aumont, O., O. Maury, S. Lefort, and L. Bopp. 2018. Evaluating the potential impacts of the diurnal vertical migration by marine organisms on marine biogeochemistry. Global Biogeochem. Cycl. **32**: 1622–1643. doi:10.1029/2018GB005886

Baltar, F., J. Arístegui, E. Sintes, J. M. Gasol, T. Reinthaler, and G. J. Herndl. 2010. Significance of non-sinking particulate organic carbon and dark CO_2 fixation to heterotrophic carbon demand in the mesopelagic northeast Atlantic. Geophys. Res. Lett. **37**: 1–6. doi:10.1029/2010GL043105

Baumas, C. M. J., and others. 2021. Mesopelagic microbial carbon production correlates with diversity across different marine particle fractions. ISME J. **15**: 1695–1708. doi:10.1038/s41396-020-00880-z

Belcher, A., M. Iversen, S. Giering, V. Riou, S. A. Henson, L. Berline, L. Guilloux, and R. Sanders. 2016. Depth-resolved particle-associated microbial respiration in the northeast Atlantic. Biogeosciences **13**: 4927–4943. doi:10.5194/bg-13-4927-2016

Bergauer, K., A. Fernandez-Guerra, J. A. L. Garcia, R. R. Sprenger, R. Stepanauskas, M. G. Pachiadaki, O. N. Jensen, and G. J. Herndl. 2018. Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. Proc. Natl. Acad. Sci. USA **115**: E400–E408. doi:10.1073/pnas.1708779115

Boyd, P. W., A. McDonnell, J. Valdez, D. LeFevre, and M. P. Gall. 2015. RESPIRE: An in situ particle interceptor to conduct particle remineralization and microbial dynamics studies in the oceans' Twilight Zone. Limnol. Oceanogr. Methods **13**: 494–508. doi:10.1002/lom3.10043

Boyd, P. W., H. Claustre, M. Levy, D. A. Siegel, and T. Weber. 2019. Multi-faceted particle pumps drive carbon sequestration in the ocean. Nature **568**: 327–335. doi:10.1038/s41586-019-1098-2

Buesseler, K. O., and P. W. Boyd. 2009. Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. Limnol. Oceanogr. **54**: 1210–1232. doi:10.4319/lo.2009.54.4.1210

Buesseler, K. O., P. W. Boyd, E. E. Black, and D. A. Siegel. 2020. Metrics that matter for assessing the ocean biological carbon pump. Proc. Natl. Acad. Sci. **117**: 9679–9687. doi:10.1073/pnas.1918114117

Burd, A. B., and others. 2010. Assessing the apparent imbalance between geochemical and biochemical indicators of meso- and bathypelagic biological activity: What the @$#! is wrong with present calculations of carbon budgets? Deep. Res. Part II Top. Stud. Oceanogr. **57**: 1557–1571. doi:10.1016/j.dsr2.2010.02.022

Christopoulos, D. 2016. Introducing Unit Invariant Knee (UIK) as an objective choice for elbow point in multivariate data analysis techniques. SSRN Electron. J.: 1–7. doi:10.2139/ssrn.3043076

Collins, J. R., B. R. Edwards, K. Thamatrakoln, J. E. Ossolinski, G. R. DiTullio, K. D. Bidle, S. C. Doney, and B. A. S. Van Mooy. 2015. The multiple fates of sinking particles in the North Atlantic Ocean. Global Biogeochem. Cycl. **29**: 1471–1494. doi:10.1002/2014GB005037

Costello, M. J., and S. Breyer. 2017. Ocean depths: The mesopelagic and implications for global warming. Curr. Biol. **27**: R36–R38. doi:10.1016/j.cub.2016.11.042

del Giorgio, P. A., and J. J. Cole. 1998. Bacterial growth efficiency in natural aquatic systems. Annu. Rev. Ecol. Syst. **29**: 503–541. doi:10.1146/annurev.ecolsys.29.1.503

DeLong, E. F., and others. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science **311**: 496–503. doi:10.1126/science.1120250

Eppley, R. W., and B. J. Peterson. 1979. Particulate organic matter flux and planktonic new production in the deep ocean. Nature **282**: 677–680. doi:10.1038/282677a0

Forest, A., and others. 2013. Ecosystem function and particle flux dynamics across the Mackenzie Shelf (Beaufort Sea, Arctic Ocean): An integrative analysis of spatial variability and biophysical forcings. Biogeosciences **10**: 2833–2866. doi:10.5194/bg-10-2833-2013

Garel, M., P. Bonin, S. Martini, S. Guasco, M. Roumagnac, N. Bhairy, F. Armougom, and C. Tamburini. 2019. Pressure-retaining sampler and high-pressure systems to study deep-sea microbes under in situ conditions. Front. Microbiol. **10**: 453. doi:10.3389/FMICB.2019.00453

Gazeau, F., and others. 2021. Impact of dust addition on the metabolism of Mediterranean plankton communities and carbon export under present and future conditions of pH and temperature. Biogeosciences **18**: 5423–5446. doi:10.5194/bg-18-5423-2021

Ghiglione, J. F., and others. 2008. Role of environmental factors for the vertical distribution (0–1000 m) of marine bacterial communities in the NW Mediterranean Sea. Biogeosciences **5**: 1751–1764. doi:10.5194/bg-5-1751-2008

Giering, S. L. C., and others. 2014. Reconciliation of the carbon budget in the ocean's twilight zone. Nature **507**: 480–483. doi:10.1038/nature13123

Giering, S. L. C., and C. Evans. 2022. Overestimation of prokaryotic production by leucine incorporation—and how to avoid it. Limnol. Oceanogr. **67**: 1–13. doi:10.1002/lno.12032

Guieu, C., and S. Bonnet. 2019. TONGA cruise. RV L'Atalante. doi:10.17600/18000884

Guieu, C., and others. 2020. BIOGEOCHEMICAL dataset collected during the PEACETIME cruise. SEANOE. doi:10.17882/75747

Harchaoui, Z., and O. Cappe. 2007. Retrospective mutiple change-point estimation with kernels, p. 768–772. *In* 2007 IEEE/SP 14th Workshop on Statistical Signal Processing. IEEE. doi:10.1109/SSP.2007.4301363

Hedgpeth, J. W. 1957. Classification of marine environments. *In* Treatise on marine ecology and paleoecology. Geological Society of America.

Henson, S. A., R. Sanders, E. Madsen, P. J. Morris, F. Le Moigne, and G. D. Quartly. 2011. A reduced estimate of the strength of the ocean's biological carbon pump. Geophysical Research Letters. **38**(4): 1–5. doi:10.1029/2011gl046735

Herndl, G., and T. Reinthaler. 2013. Microbial control of the dark end of the biological pump. Nat. Geosci. **6**: 718–724. doi:10.1038/ngeo1921

Irigoien, X., and others. 2014. Large mesopelagic fishes biomass and trophic efficiency in the open ocean. Nat. Commun. **5**: 3271. doi:10.1038/ncomms4271

Kirchman, D., E. K'nees, and R. Hodson. 1985. Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. Appl. Environ. Microbiol. **49**: 599–607. doi:10.1128/aem.49.3.599-607.1985

Lavigne, H., F. D'Ortenzio, M. Ribera D'Alcalà, H. Claustre, R. Sauzède, and M. Gacic. 2015. On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea: A basin-scale and seasonal approach. Biogeosciences **12**: 5021–5039. doi:10.5194/bg-12-5021-2015

Le Moigne, F. A. C. 2019. Pathways of organic carbon downward transport by the oceanic biological carbon pump. Front. Mar. Sci. **6**: 1–8. doi:10.3389/fmars.2019.00634

Lee, Z., A. Weidemann, J. Kindle, R. Arnone, K. L. Carder, and C. Davis. 2007. Euphotic zone depth: Its derivation and implication to ocean-color remote sensing. J. Geophys. Res. **112**: C03009. doi:10.1029/2006JC003802

Levitus, S. 1982. Climatological atlas of the world ocean. US Department of Commerce, National Oceanic and Atmospheric Administration. p. 173.

Longhurst, A. 1995. Seasonal cycles of pelagic production and consumption. Prog. Oceanogr. **36**: 77–167. doi:10.1016/0079-6611(95)00015-1

Longhurst, A. R. 2007. Ecological geography of the sea. Amsterdam: Elsevier p. 542.

Lukas, R., and E. Lindstrom. 1991. The mixed layer of the western equatorial Pacific Ocean. J. Geophys. Res. **96**: 3343. doi:10.1029/90JC01951

Marañón, E., and others. 2021. Deep maxima of phytoplankton biomass, primary production and bacterial production in the Mediterranean Sea. Biogeosciences **18**: 1749–1767. doi:10.5194/bg-18-1749-2021

Marra, J. F., V. P. Lance, R. D. Vaillancourt, and B. R. Hargreaves. 2014. Resolving the ocean's euphotic zone. Deep. Res. Part I Oceanogr. Res. Pap. **83**: 45–50. doi:10.1016/j.dsr.2013.09.005

Martin, J. H., G. A. Knauer, D. M. Karl, and W. W. Broenkow. 1987. VERTEX: carbon cycling in the northeast Pacific. Deep Sea Res. Part A. Oceanogr. Res. Pap. **34**: 267–285. doi:10.1016/0198-0149(87)90086-0

Martin, A., and others. 2020. The oceans' twilight zone must be studied now, before it is too late. Nature **580**: 26–28. doi:10.1038/d41586-020-00915-7

McDonnell, A. M. P., and others. 2015. The oceanographic toolbox for the collection of sinking and suspended marine particles. Prog. Oceanogr. **133**: 17–31. doi:10.1016/j.pocean.2015.01.007

Miquel, J.-C., B. Gasser, J. Martín, C. Marec, M. Babin, L. Fortier, and A. Forest. 2015. Downward particle flux and carbon export in the Beaufort Sea, Arctic Ocean; the role of zooplankton. Biogeosciences **12**: 5103–5117. doi:10.5194/bg-12-5103-2015

Monterey, G., and S. Levitus. 1997. Seasonal variability of mixed layer depth for the world ocean. NOAA ATLAS, NESDIS,14, p. 96.

Nagata, T., and others. 2010. Emerging concepts on microbial processes in the bathypelagic ocean—Ecology, biogeochemistry, and genomics. Deep Sea Res. Part II Top. Stud. Oceanogr. **57**: 1519–1536. doi:10.1016/j.dsr2.2010.02.019

Olshen, A. B., E. S. Venkatraman, R. Lucito, and M. Wigler. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics **5**: 557–572. doi:10.1093/biostatistics/kxh008

Ortega-Retuerta, E., and others. 2012. Carbon fluxes in the Canadian Arctic: Patterns and drivers of bacterial abundance, production and respiration on the Beaufort Sea margin. Biogeosciences **9**: 3679–3692. doi:10.5194/BG-9-3679-2012

Owens, S. A., S. Pike, and K. O. Buesseler. 2015. Thorium-234 as a tracer of particle dynamics and upper ocean export in the Atlantic Ocean. Deep Sea Res. Part II Top. Stud. Oceanogr. **116**: 42–59. doi:10.1016/j.dsr2.2014.11.010

Proud, R., M. J. Cox, S. Wotherspoon, and A. S. Brierley. 2015. A method for identifying sound scattering layers and extracting key characteristics. Methods Ecol. Evol. **6**: 1190–1198. doi:10.1111/2041-210X.12396

Proud, R., M. J. Cox, and A. S. Brierley. 2017. Biogeography of the global ocean's mesopelagic zone. Curr. Biol. **27**: 113–119. doi:10.1016/j.cub.2016.11.003

Reinthaler, T., H. van Aken, C. Veth, J. Arístegui, C. Robinson, P. J. L. B. Williams, P. Lebaron, and G. J. Herndl. 2006. Prokaryotic respiration and production in the meso- and bathypelagic realm of the eastern and western North Atlantic basin. Limnol. Oceanogr. **51**: 1262–1273. doi:10.4319/lo.2006.51.3.1262

Reygondeau, G., and others. 2018. Global biogeochemical provinces of the mesopelagic zone. J. Biogeogr. **45**: 500–514. doi:10.1111/jbi.13149

Riley, J. S., R. Sanders, C. Marsay, F. A. C. Le Moigne, E. P. Achterberg, and A. J. Poulton. 2012. The relative contribution of fast and slow sinking particles to ocean carbon export. Global Biogeochem. Cycl. **26**: 1–10. doi:10.1029/2011GB004085

Robinson, C., and others. 2010. Mesopelagic zone ecology and biogeochemistry—A synthesis. Deep Sea Res. Part II Top. Stud. Oceanogr. **57**: 1504–1518. doi:10.1016/j.dsr2.2010.02.018

Roquet, F., G. Madec, T. J. McDougall, and P. M. Barker. 2015. Accurate polynomial expressions for the density and specific volume of seawater using the TEOS-10 standard. Ocean Model. **90**: 29–43. doi:10.1016/j.ocemod.2015.04.002

Sebastián, M., E. Ortega-Retuerta, L. Gómez-Consarnau, M. Zamanillo, M. Álvarez, J. Arístegui, and J. M. Gasol. 2021. Environmental gradients and physical barriers drive the basin-wide spatial structuring of Mediterranean Sea and adjacent eastern Atlantic Ocean prokaryotic communities. Limnol. Oceanogr. **1–19**: 4077–4095. doi:10.1002/lno.11944

Severin, T., and others. 2016. Impact of an intense water column mixing (0–1500 m) on prokaryotic diversity and activities during an open-ocean convection event in the NW Mediterranean Sea. Environ. Microbiol. **00**: 4378–4390. doi:10.1111/1462-2920.13324

Siegel, D. A., and others. 2016. Prediction of the export and fate of global ocean net primary production: The EXPORTS Science Plan. Front. Mar. Sci. **3**: 22. doi:10.3389/fmars.2016.00022

Simon, M., and F. Azam. 1989. Protein content and protein synthesis rates of planktonic marine bacteria. Mar. Ecol. Prog. Ser. **51**: 201–213. doi:10.3354/meps051201

Sprintall, J., and M. F. Cronin. 2001. Upper ocean vertical structure, p. 3120–3128. *In* Encyclopedia of ocean sciences. Elsevier.

Steinberg, D. K., B. A. S. Van Mooy, K. O. Buesseler, P. W. Boyd, T. Kobari, and D. M. Karl. 2008. Bacterial vs. zooplankton control of sinking particle flux in the ocean's twilight zone. Limnol. Oceanogr. **53**: 1327–1338. doi:10.4319/lo.2008.53.4.1327

Steinberg, D. K., and M. R. Landry. 2017. Zooplankton and the ocean carbon cycle. Annu. Rev. Mar. Sci. **9**: 413–444. doi:10.1146/ANNUREV-MARINE-010814-015924

Sutton, T. T., and others. 2017. A global biogeographic classification of the mesopelagic zone. Deep Sea Res. Part I Oceanogr. Res. Pap. **126**: 85–102. doi:10.1016/j.dsr.2017.05.006

Tamburini, C., and others. 2021. Increasing hydrostatic pressure impacts the prokaryotic diversity during *Emiliania huxleyi* aggregates degradation. Water **13**: 2616. doi:10.3390/w13192616

Truong, C., L. Oudre, and N. Vayatis. 2020. Selective review of offline change point detection methods. Signal Process. **167**: 107299. doi:10.1016/j.sigpro.2019.107299

Acknowledgments

The authors wish to express their gratitude to D. Lefèvre, F. D'ortenzio, G. Reygondeau, and F. Van Wambeke for stimulating and informative discussions. The authors warmly thank J.R. Collins, E.O. Retuerta, P. Massicotte, A. Martin, M. Bressac, F. Van Wambeke, F. Gazeau, T. Blasco, and N. Bhairy for sharing data. The authors thank the SAM platform from the Mediterranean institute of oceanography as well as the PNIO-DT INSU for their technical expertise and facilities. This work was supported by the French National program LEFE (Les Enveloppes Fluides et l'Environnement), through the PARTY project (remineralization des PARticules marines et Transfert vers les abYsses) awarded to FACLM. This manuscript is a contribution of the ANR APERO (project number ANR-21-CE01-0027).

Submitted 10 April 2022
Revised 29 August 2022
Accepted 23 September 2022

Associate editor: Gordon T. Taylor