Supplementary Materials for Organic Carbon Burial in Global Continental Margin Sediments

Sandra Arndt[1,2,*], Dominik Hülse[3,*]

Markus Diesing[4], Sarah Paradis[5], Craig Smeaton[6]

[1]Universite libre de Bruxelles, Brussels, Belgium

[2]UiT The Arctic University of Norway, Tromsø, Norway

[3]MARUM University of Bremen, Bremen, Germany

[4]Geological Survey of Norway, Trondheim, Norway

[5]ETH Zurich, Zürich, Switzerland

[6]University of St Andrews, UK

[*]To whom correspondence should be addressed; E-mail: dhuelse@marum.de, sandra.arndt@ulb.be

S. Arndt and D. Hülse contributed equally and share the first authorship.

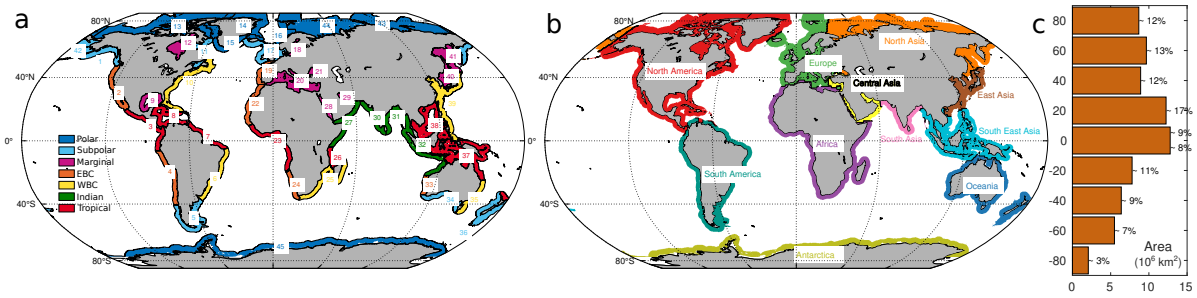
This PDF file includes:

Supplementary Methods

MARCATS region, RECCAP-2 segmentation, and Exclusive Economic Zones

We here define the continental margin according to the extended Margins and CATchment Segmentation (MARCATS) approach[1] that divides the global continental margin into 45 different segments. MARCATS aggregate the finer COSCAT (COastal Segmentation and related CATchments) segmentation of the global coastline[2] into larger geographical regions with broadly similar climatological, morphological, and oceanographic characteristics of the shelf sea. In addition, for a coarser segmentation, the 45 different MARCATS segments can be further aggregated into 7 classes (Eastern boundary currents, Western boundary currents, tropical margins, subpolar margins, polar margins, marginal seas, and Indian margins, ED Fig. 1a). The extended MARCATS region extends 300 km away from the shoreline, includes the global continental shelf and part of the global continental slope and rise, and covers a total global surface area of $\sim$74 million km^2. To represent estuaries, tidal wetlands and inland water bodies our resolution of 0.25° is too coarse and a more specific modelling approach would be needed. Therefore, these areas are excluded from our analysis. This definition of the continental margin is fully aligned with the major global budget synthesis efforts (e.g., RECCAP-1[3]; RECCAP-2[4]; Global Carbon Project syntheses[5]) and creates a zone of overlap with open ocean definitions that allows for the comparison of results at the dynamic interface between the continental margin and the open ocean.

In addition to the MARCATS segmentation, we use two other classifications to partition



Extended Data Figure 1: Maps of segmentations used to estimate organic carbon flux budgets: (a) MARCATS regions after[1,6], (b) RECCAP2 regions (adapted from[7]). (c) Latitudinal distribution of the continental margin area, aggregated in 20° bands.

our local estimates. First, we use eleven oceanic regions covering the entire globe that are based on the RECCAP-2 regions, produced to estimate land[4] and coastal ecosystem carbon fluxes[7]. We adapted the ten RECCAP-2 regions of[7] to extend 300 km away from the shoreline and included the continental margins around Antarctica to the segmentation (ED Fig. 1b). Second, we adopt the definition of a nation's Exclusive Economic Zone (EEZ) under the United Nations Convention on the Law of the Sea (UNCLOS) as the ocean area extending up to 200 nautical miles (approximately 370 kilometers) from a country's baseline – typically the low-water line along its coast – which is a maritime zone legally recognized under international law [8,9]. EEZs are used to calculate OC flux estimates on a national level, which may provide a useful, policy-relevant framework for greenhouse gas emission accounting and developing coastal management strategies.

General framework

We force the global reaction-transport model (RTM) OMEN-SED[10,11] on a 0.25° global grid with maps of organic carbon (OC) content, OC reactivity, sedimentation rate, sediment porosity, bioturbation depths and intensity in continental margin sediments to derive global maps of OC transfer fluxes through multiple defined sediment age horizons.

Reaction Transport Model

The one-dimensional reaction-transport model OMEN-SED solves the vertically resolved conservation equation for OC in porous media[12,13]:

$$\frac{\partial(1-\phi)OC}{\partial t} = \frac{\partial}{\partial z}\left(D_b(1-\phi)\frac{\partial OC}{\partial z}\right) - \frac{\partial(1-\phi)\omega OC}{\partial z} + (1-\omega)R_{OC} \qquad (1)$$

using an analytical approach that is based on ref.[10]. Here, OC (in $\frac{gC}{cm^3}$ dry sediment) is the concentration of organic carbon and ϕ is the porosity of the sediment. The term t is time, z denotes sediment depth, D_{b} (in $\frac{cm^2}{yr}$) is the bioturbation intensity, ω (in $\frac{cm}{yr}$) represents the sedimentation rate, and R_{OC} (in $\frac{gC}{cm^3 yr}$) the degradation rate of bulk OC.

The degradation of OC is simulated using a 500-G approximation of a reactive-continuum model (RCM,[11,14]). The advantage of this approximation is that it can be analytically solved for both the bioturbated as well as non-bioturbated zones of the sediment. The RCM describes the initial distribution of organic carbon compounds over the reactivity spectrum using a continuous Gamma distribution (Γ,[14]). We divide this initial distribution into 500 finite fractions, OC_i, each with its own first-order degradation rate constant, k_i. We divide the reactivity range, k = [10^{-15},

$10^{(-log(a)+2)}$], into 499 equal reactivity classes[15]. The least reactive class is given a degradation rate constant of $k_1 = 10^{-15}$ yr^{-1}, the most reactive class a constant of $k_{500} = 10^{-log(a)+2}$, and for the intermediate reactivity classes we calculate the degradation rate constant as $k_i = k_{j-1} + (k_j - k_{j-1})/2$.

The fraction of OC in the least reactive class is calculated based on the lower incomplete Gamma function:

$$f_1 = \int_0^{a10^{-15}} x^{\nu-1} \cdot exp(-x)dx. \tag{2}$$

The fraction of OC within the highest reactivity class is calculated based on the upper incomplete Gamma function:

$$f_{500} = \frac{\Gamma(\nu, ak_{500})}{\Gamma(\nu)}. \tag{3}$$

Fractions of OC in intermediate reactivity classes, $i \in [2, 499]$, are calculated as

$$f_i = \frac{\Gamma(\nu, 0, ak_{i+1}) - \Gamma(\nu, 0, ak_i)}{\Gamma(\nu)} \tag{4}$$

where a is the average initial age of the initial OC mixture, and ν is a dimensionless parameter defining the shape of the OC distribution. All fractions add up to unity.

The OC flux through a specific sediment depth (z), i.e., the respective transfer or burial flux, is calculated from the OC concentration at that depth according to

$$F(z) = -(1-\phi) \cdot \left(-D_b \cdot \frac{\partial OC(z)}{\partial z} + \omega \cdot OC(z) \right) \tag{5}$$

Determining the depth of a sediment age layer

To calculate the depth, z, of the respective sediment age layers below the depth of the bioturbated zone (i.e., target ages 100, 1000, and 5000 years) we consider the exponential decrease in porosity with depth and relate time, compaction length scale, and sedimentation rate to depth[12]. The porosity profile, ϕ, was calculated as

$$\phi(z) = \phi_0 \cdot exp(-c_0 \cdot z)$$

where ϕ_0 denotes the porosity at the sediment-water interface (SWI) and c_0 (per centimeter) stands for the compaction length scale, which characterizes how a given sediment type will compact under its own weight. We used a value of $c_0 = 0.5 \times 10^{-5}$ cm^{-1} for shelf sediments (i.e., seafloor depth $<$ 200m) and a value of $c_0 = 1.7 \times 10^{-6}$ cm^{-1} for sediments at depths $>$ 200m[16,17].

The age t of sediment at depth z is determined by integrating downcore sedimentation rate, adjusted for compaction:

$$t(z) = \int_0^z \frac{1}{\omega_s(z')dz'}$$

with $\omega_s(z')$ being the burial velocity, which itself depends on porosity and initial accumulation rate. For an exponential porosity law, this is typically expressed as:

$$\omega_s(z) = \omega_0 \frac{1-\phi_0}{1-\phi(z)}$$

where ω_0 is the sediment accumulation rate at the surface.

The Lambert W function, as implemented in MATLAB[18], is used for inverting expressions of the form $x = w \cdot exp(w)$[19], like the exponentials arising from the porosity equation. By

combining these expressions, the layer depth (z) is determined for each target age, and the OC transfer fluxes are subsequently simulated at the corresponding depths.

Model forcing

Sediment porosity

We spatially predict the porosity (ϕ_0) of global ocean surface sediments based on publicly available measurements and predictor variables with the quantile regression forests[?] machine learning algorithm. A global compilation of sediment porosity[20] (in %) was supplemented with records published on the PANGAEA database[21] between 2016 and 2022, i.e., after the publication of ref.[20]. Predictor variables (bathymetry and derived terrain attributes, primary productivity, chlorophyll, sea ice concentration, temperature, salinity, current speed, dissolved oxygen, iron, nitrate, silicate, and phosphate) were retrieved from Bio-ORACLE[22,23] and supplemented with global sediment thickness[24] and the distance to spreading ridges[25]. The predictor variables were submitted to a systematic variable selection process. The Boruta variable selection wrapper algorithm[26] identified important predictor variables. A subsequent de-correlation analysis[27] reduced the set of predictor variables to those with a variance inflation factor below 2.5. The final set of predictor variables was determined with forward feature selection[28] during model tuning.

Spatial auto-correlation in the training data is known to inflate estimates of predictive performance[29,30]. Our modelling framework accounts for this effect by generating spatially

separated folds for a ten-fold cross-validation[31] and estimates the area of applicability where the estimated cross-validation performance holds[32]. To assess model performance, we calculate the explained variance (R^2) and the root mean square error (RMSE).

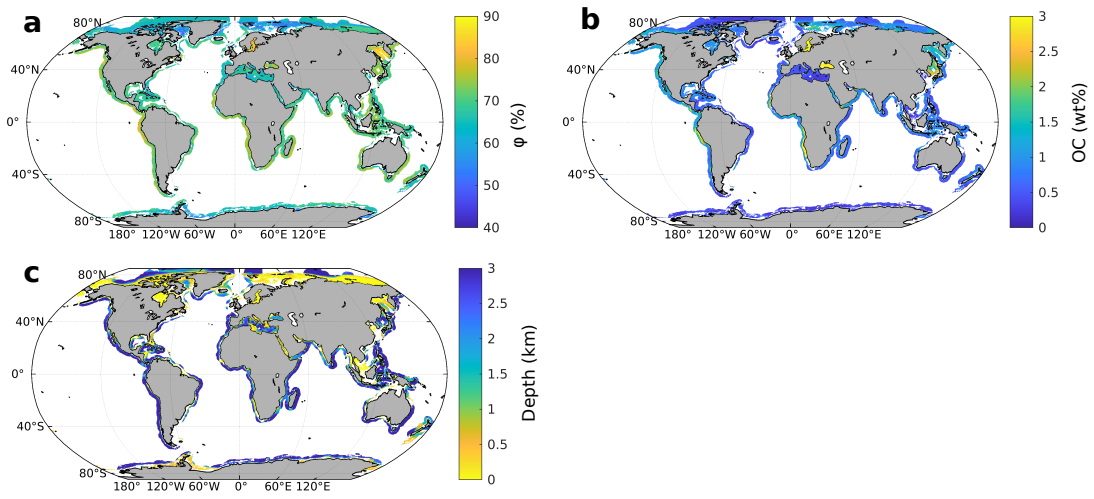
We then use the quantile regression forest (QRF) algorithm[?] to predict porosity spatially and to quantify the prediction uncertainty. QRF is a generalisation of the random forest algorithm[33], which aggregates the conditional mean from each tree in a forest to derive an ensemble prediction. QRF returns the whole conditional distribution of the response variable. This allows us to determine the median (q50) of the distribution and the underlying variability of the estimate by means of the 90% prediction interval ($PI90$), which is defined as:

$$PI90 = q_{95} - q_5; \tag{6}$$

with q_{95} and q_5 the 95^{th} and 5^{th} quantile of the conditional distribution, respectively. The complete analysis was carried out in R 4.1.3 statistical software. We use the median porosity distribution for our 'best-estimate' calculations and the q_{95} and q_5 quantiles to quantify conservative uncertainty estimates (see Section "Quantification of Model Uncertainty"). A map of the median porosity is shown in ED Fig. 2a.

Surface sediment OC content

We predict OC content of surface sediments with the same spatial modelling framework that was employed to model porosity. Measurements of OC content (in weight-%) have been retrieved



Extended Data Figure 2: Maps of model forcings and boundary conditions: Median surface sediment organic carbon (OC) content (a), porosity (ϕ, b), and seafloor depth (c).

from the MOSAIC database[34] and are limited to 0 - 1 cm sediment depth. Predictor variables include Bio-ORACLE data[22,23] (as above), predicted sediment porosity, sediment thickness[24] and partial pressure of CO_2 at the sea-air interface[35]. We, again, estimate the median, q_{95} and q_5 quantiles of the distribution (the median OC wt% is plotted in ED Fig. 2b) and use them to quantify conservative boundary condition uncertainties (see Section on "Quantification of Model Uncertainty").

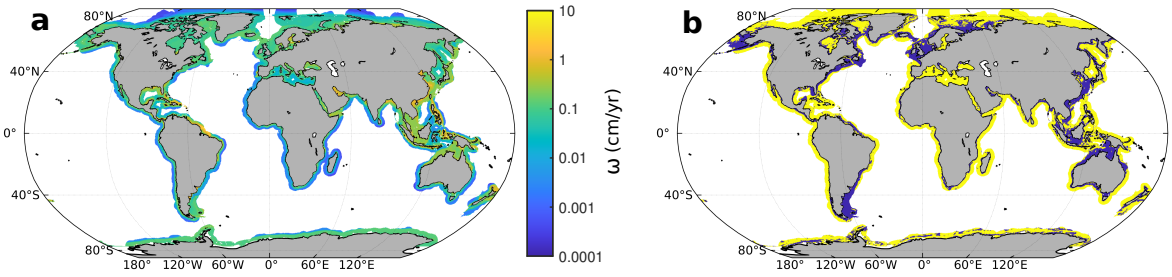
Sedimentation rate and non-depositional mask

Sedimentation rate estimates are very uncertain and have been shown to have a large effect on simulated benthic OC transfer efficiencies[15]. We use estimates from the k-nearest neighbor (k-NN) algorithm of Restreppo et al.[36], which follows a data-driven approach and uses para-

metrically, not geospatially, nearest observed data points to calculate probable sedimentation rates in areas with no data. Restreppo et al.[36] use known predictor features (e.g., water depth, distance from a river mouth, etc.) to produce the most probable sedimentation rate predictions on a 5-arc-minute global map. To our knowledge this is the only data product resolving the heterogeneity in sedimentation rates beyond simple water depth relationships. Predicted sedimentation rates decrease progressively with increasing distance from the coast and river mouths[36], a trend generally observed in the ocean (e.g., refs.[37,38]). We transformed the available map to our resolution (i.e., 0.25°) and restricted the data to the continental margin grid (ED Fig. 3a). Restreppo et al.[36] further provide standard deviations for their 'best' predictions which we use for a 'coherent-bias' stress test (see Section "Quantification of Model Uncertainty").

The approach of Restreppo et al.[36] does not account for erosion and non-deposition, which prevail over large swathes of highly dynamic coastal zones and continental shelves[39,40]. To account for areas where sedimentation rates over longer time scales (years to millennia) are effectively zero, we assume that coarse-grained siliciclastic sediments (sand, gravel and coarser) are indicative of erosional and non-depositional environments[41]. To identify these areas, we predict the spatial distribution of biogenic, transitional, fine-grained siliciclastic and coarse-grained siliciclastic sediments by utilising sediment data from refs.[41,42] and predictor variables from Bio-ORACLE[22,23] as well as predicted surface OC and porosity. The modelling framework is based on ref.[43] but updated to include a spatial ten-fold cross validation[31] and an estimate of the area of applicability[32]. The output of biogenic, transitional, fine-grained siliciclastic and

coarse-grained siliciclastic sediments is converted to a binary mask with 0 indicating erosion and non-deposition (coarse-grained siliciclastic sediments) and 1 indicating deposition (biogenic, transitional and fine-grained siliciclastic sediments, ED Fig. 3b). The mask is finally multiplied with the sedimentation rates as obtained by the approaches described above.



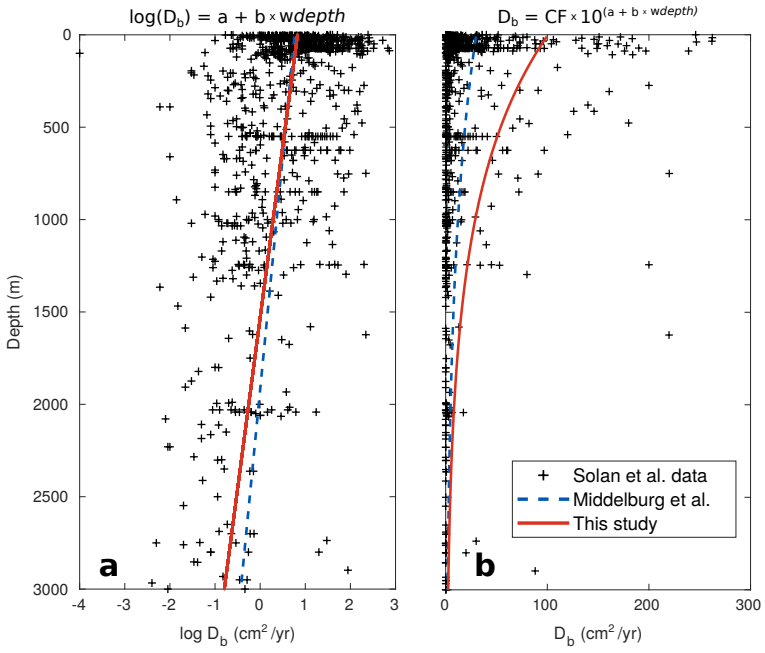
Extended Data Figure 3: Maps of sedimentation rate (ω, a) after Restreppo et al.[36] and the non-depositional mask (b) used to force the diagenetic model, OMEN-SED. A blue colour in (b) indicates areas of erosion and non-deposition, i.e., areas that are excluded in our budget analysis.

Model parameters

Bioturbation

To predict the bioturbation intensity coefficient, D_b (in cm^2/yr), we derive an empirical equation based on the data of Solan et al.[44]. Therefore, we restrict the Solan et al.[44] data to our continental margin region of interest (N=918) and follow the approach of Middelburg et al.[45] – which is, we use an exponential relationship with water depth (d) as an independent variable of the general form $D_b = CF \cdot 10^{(a+b \cdot d)}$, with $CF = exp(2.65s^2)$ being a correction factor to account for the

systematic bias due to skewness and s^2 the variance of the distribution.



Extended Data Figure 4: Bioturbation intensity (a: on log-scale, b: linear-scale) vs. water depth plots. Symbols indicate the data from Solan et al.[44], lines are the regression curves (red-solid is equation (7) and blue-dashed is equation (8)) using the form indicated above the subplots.

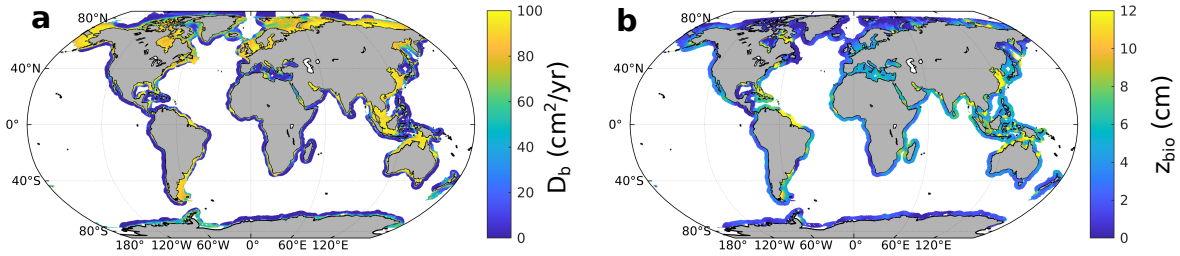
Our empirical predictive equation for D_b based on the data of Solan et al.[44] is then:

$$D_b = 14.9998 \cdot 10^{0.8268 - 0.0005401 \cdot d} \tag{7}$$

We also tested the classic empirical relationship of Middelburg et al.[45]:

$$D_b^M = 5.2 \cdot 10^{0.7624 - 0.0003972 \cdot d} \tag{8}$$

which was based on N=132 global data points. Extended Datta Fig. 4 compares both approaches with the observations compiled by Solan et al.[44]. Both approaches lead to very similar OC burial estimates (not shown), which was expected as simulated OC burial rates are less sensitive to variations in bioturbation intensity[15]. Extended Datta Fig. 5a shows the corresponding bioturbation intensity map calculated from equation (7).



Extended Data Figure 5: Maps of bioturbation intensity parameter (a) and mixed layer depth (b) used to force the diagenetic model, OMEN-SED. Bioturbation intensity, D_b, is estimated from equation (7) based on the data of Solan et al.[44] (a). The mixed layer depth map (b) has been interpolated from the data compiled by Song et al.[42]. TODO: calculate the area-weighted mean and std, and add this to the figures.

Traditionally, it was believed that globally, the extent of the mixed layer depth is largely restricted to the upper 10 cm of the sediments and largely independent of water depth and sedimentation rate[46,47]. However, more recent observations show that highly dynamic regions such as large estuaries have thicker mixed layer depths than most oceanic sediments[42]. To account for this variability, we interpolated the data compiled by Song et al.[42] on our global map of the continental margins (see ED Fig. 5b) using the MATLAB function griddata. We tested vari-

ous interpolation methods (i.e., 'linear', 'natural', 'cubic'). Because all lead to similar results we used the triangulation-based 'natural' neighbour interpolation, which represents an efficient tradeoff between the 'linear' and 'cubic' methods. We also tested a globally fixed mixed layer depth of 5.75 cm[47] and 9.7 cm[46] – again leading to very similar results (not shown), since OC burial rates are less sensitive to variations in mixed layer depth[15].

OC Reactivity

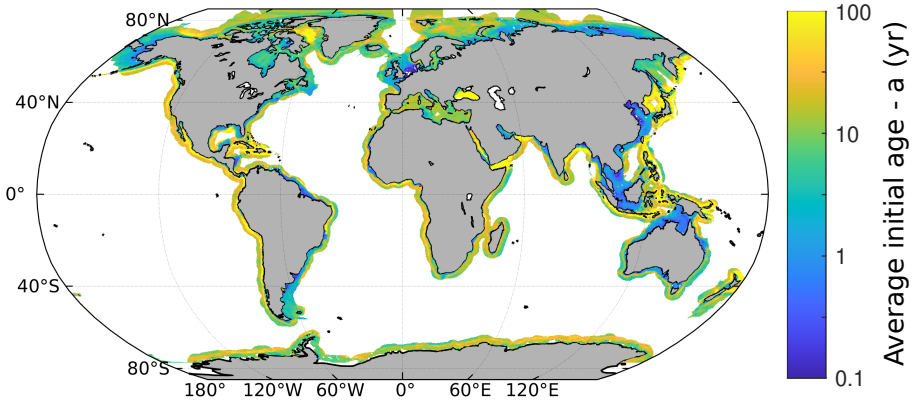
We build on the approach of Pika et al.[48], who inversely determined organic matter reactivity parameters from benthic diffusive oxygen uptake (DOU) measurements at 392 drill sites and derived a predictive equation for the reactivity parameter a. Global maps of a were then generated by applying this equation to extrapolated DOU maps from Jørgensen et al.[49] and Seiter et al.[50]. Here, we refine the approach of Pika et al.[48] by more directly linking benthic DOU rates to OM reactivity and configuring OMEN-SED explicitly for each continental margin grid cell. Rather than deriving a predictive equation for the RCM parameter a from a limited set of DOU observations, we use the global DOU map of Jørgensen et al.[49] to inversely calculate RCM parameter a. This enables the model to reproduce the local DOU rate in every grid cell.

For the global mapping of benthic DOU rates, Jørgensen et al.[49] developed a multicomponent linear regression model that relates measured DOU rates (mmol O_2 m^{-2} d^{-1}) to water depth (m) and sea-surface net primary production (NPP, g C m^{-2} y^{-1}). This predictive relation-

ship was combined with a gridded global dataset of water depth [51] and annual mean sea-surface NPP (see ref.[49] for details) to generate a global map of benthic DOU rates at 0.1° resolution, which we regridded to 0.25°. Missing DOU values in continental margin grids were filled with the mean of adjacent cells; if no neighbors were available, the global mean predicted for the corresponding depth interval was assigned (see Table 4 in ref.[49]).

To inversely estimate parameter a from this DOU map, we configured OMEN-SED with our new set of spatially explicit boundary conditions, including sediment porosity, surface sediment OC content, sedimentation rate, and bioturbation intensity and depth. Parameter a was iteratively adjusted until the simulated DOU rate matched with the local DOU rate from Jørgensen et al.[49], within a 5% tolerance.

ED Fig. 6 shows the resulting global distribution of the RCM parameter a along continental margins. Low OC reactivities (high a values, $> 10-100$ yr; green–yellow) dominate Eastern boundary upwelling systems (e.g., Peru–Chile, California, West Africa, Namibia), the margins of the Indian Ocean, as well as higher latitude shelves and Arctic/Antarctic margins. Regions with high reactivity (low a values, < 1 yr; blue–purple) are more localized, often around tropical/subtropical margins (e.g., parts of Southeast Asia, equatorial West Africa, Amazon outflow), semi-enclosed, marginal seas, and shallower environments closer to the coast. We note that our results indicating low organic carbon reactivities in eastern boundary systems should be interpreted with caution, as these regions often experience low-oxygen bottom waters, where DOU



Extended Data Figure 6: Map of organic carbon (OC) reactivity or average initial age (i.e., parameter a) using sediment rate estimates based on the approaches of Restreppo et al.[36]. For all other model forcings and boundary conditions (i.e., OC content, porosity, bioturbation parameters) we used our 'default' estimates as decribed in the text.

may not be a reliable predictor of organic matter degradation reactivity or rates.

Quantification of Model Uncertainty

Parameters and boundary conditions used to drive the diagenetic model are a primary and large source of uncertainty for our estimates. We quantify uncertainties resulting from three of the most important model parameters and boundary conditions[15]: surface sediment OC content, sediment porosity and sedimentation rate. Because uncertainties in these boundary conditions are large, we simulate conservative upper and lower estimates based on the 5^{th} and 95^{th} quantile of the conditional distributions for sediment OC content and porosity, as well as the local standard deviation for sedimentation rate based on the machine learning estimates of ref.[36]. The

available data on biotubation depth[42] and efficiency[44] lack uncertainty estimates. Moreover, because geological OC tranfer efficiency is largely insensitive to bioturbation parameters[15], we exclude these parameters from our uncertainty analysis. We also excluded the reactivity parameter a, as it is a derived quantity that depends on sedimentation rate, organic carbon content, porosity, and DOU rates.

Uncertainty related to surface sediment OC and porosity

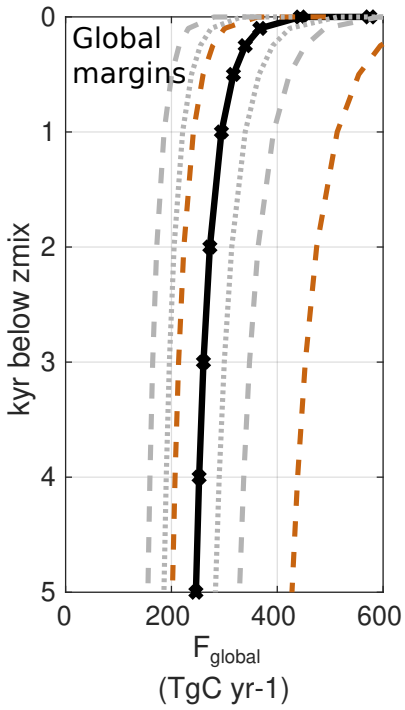
Typical surface sediment OC density distributions are unimodal but asymmetric, with a steep decline below the median and a heavier right tail (not shown). To capture the skewness and obtain a parsimonious representation for sampling, we approximate the distribution with two components: For values below the median, the density is relatively flat and drops sharply around the q5. A uniform distribution between q5 and the median provides a simple yet conservative representation of this range without overweighting rare low-end values. For values above the median, the distribution resembles a bell-shaped curve with a longer tail. We therefore approximate it with a normal distribution, using a standard deviation of 0.25 $\times$ (q95 - median) to ensure that the spread reflects the observed variability in the upper tail while keeping the bulk of the probability mass close to the median. This piecewise approach balances realism and simplicity: it reproduces the observed asymmetry of the distribution while avoiding overparameterization, and it ties the assumed spread directly to robust quantile statistics rather than being sensitive to extreme values.

For each grid cell, based on the piecewise approximation of the distribution (uniform below the median, normal above the median), we sample 10 surface sediment OC values below the median to represent the lower half of the distribution. Each sampled value is used as a boundary condition in a separate diagenetic model run. The mean of these 10 simulations provides the lower uncertainty estimate related to surface sediment OC content of our model results. We apply the same approach to the upper half of the distribution by sampling 10 values above the median, yielding an upper uncertainty estimate. Uncertainty estimates for global organic carbon fluxes across all sediment ages, associated with surface sediment OC content, are shown as dashed grey lines in Extended Data Figure 7.

By sampling 10 values from each half of the distribution and averaging the resulting model runs, we avoid overemphasizing extreme values and reduce sensitivity to outliers. This approach does not attempt to reproduce the full distributional complexity, but instead provides a robust and transparent bracketing of possible outcomes that is consistent with the observed variability in the data. In this way, the bounds should be interpreted as conservative estimates of uncertainty, rather than formal confidence intervals.

Since the typical density distribution for porosity looks similar but are left-skewed, we use an analogues, piecewise way to approximate the distribution, which is using a normal distribution with a standard deviation of 0.25 $\times$ (median - q5) for values below the median and a uniform distribution between the median and q95. Uncertainty estimates for global organic

carbon fluxes across all sediment ages, associated with sediment porosity, are shown as dotted grey lines in Extended Data Figure 7.



Extended Data Figure 7: Model uncertainties in simulated global organic carbon (OC) burial flux as a function of burial age for continental margin sediments. The black solid line shows the best estimate; dashed brown lines indicate uncertainties associated with sediment accumulation rates; dashed grey lines represent uncertainties related to surface sediment OC content; and the grey dotted line shows uncertainties arising from porosity estimates.

<u>'Coherent-bias' stress test for sedimentation rate</u>

Sedimentation rate is a key parameter because it influences both sediment transport and the predicted organic carbon reactivities and fluxes, and therefore has a major impact on our analysis.

Uncertainties in local sedimentation rates are large[36] and the distributions are strongly skewed and positive-only. In such cases, the reported standard deviations by Restreppo et al.[36] are likely dominated by a few high-end values and can exaggerate the plausible spread, while also implying negative values on the low side that are physically impossible. Because only the mean (μ) and standard deviation (σ) per grid cell are reported by Restreppo et al.[36], we therefore apply a reduced factor of $0.25 \times \sigma$ as a pragmatic shrinkage to the uncertainty band. By scaling σ down, we acknowledge that σ is not a good symmetric uncertainty descriptor for a positive-skewed variable. The $\pm 0.25 \times \sigma$ band is a pragmatic shrinkage and still proportional to the reported variability, while avoiding letting a handful of extreme tails dominate the uncertainty analysis if the full $\pm\sigma$ range is applied. In addition, we impose the global minimum of the mean values ($\mu_{\min}$) as a hard lower bound, so that $\mu - 0.25 \times \sigma$ can never fall below $\mu_{\min}$. This ensures that perturbations remain physically realistic. Our choice acknowledges the limitations of (μ, σ) as descriptors of a non-normal variable, while still scaling perturbations to the reported variability. Importantly, this band (dashed brown line in Fig. 1a and ED Fig. 7) is not used as a probabilistic interval, but rather as a controlled 'coherent-bias' stress test of systematic shifts in the sedimentation field. Stochastic perturbations within the reported $\pm\sigma$ range produced negligible model responses, indicating that diagenetic outputs are much more sensitive to systematic bias than to local random noise. Future sensitivity analysis should therefore prioritize improving sedimentation rate estimates and their local distributions, so that uncertainties can be quantified more rigorously.

1. Roobaert, A. *et al.* The Spatiotemporal Dynamics of the Sources and Sinks of CO2 in the Global Coastal Ocean. *Global Biogeochemical Cycles* **33**, 1693–1714 (2019).

2. Meybeck, M., Dürr, H. H. & Vörösmarty, C. J. Global coastal segmentation and its river catchment contributors: A new look at land-ocean linkage. *Global Biogeochemical Cycles* **20** (2006).

3. Enting, I. G., Rayner, P. J. & Ciais, P. Carbon Cycle Uncertainty in REgional Carbon Cycle Assessment and Processes (RECCAP). *Biogeosciences* **9**, 2889–2904 (2012).

4. Ciais, P. *et al.* Definitions and methods to estimate regional land carbon fluxes for the second phase of the REgional Carbon Cycle Assessment and Processes Project (RECCAP-2). *Geoscientific Model Development* **15**, 1289–1316 (2022).

5. Friedlingstein, P. *et al.* Global Carbon Budget 2024. *Earth System Science Data* **17**, 965–1039 (2025).

6. Laruelle, G. G. *et al.* Global multi-scale segmentation of continental and coastal waters from the watersheds to the continental margins. *Hydrology and Earth System Sciences* **17**, 2029–2051 (2013).

7. Rosentreter, J. A. *et al.* Coastal vegetation and estuaries are collectively a greenhouse gas sink. *Nature Climate Change* **13** (2023).

8. Flanders Marine Institute. Maritime boundaries geodatabase: Maritime boundaries and exclusive economic zones (200nm), version 11 (2019).

9. Nordquist, M. *United Nations Convention on the Law of the Sea 1982, Volume VII: A Commentary* (BRILL, 2011).

10. Hülse, D., Arndt, S., Daines, S., Regnier, P. & Ridgwell, A. OMEN-SED 1.0: a novel, numerically efficient organic matter sediment diagenesis module for coupling to Earth system models. *Geoscientific Model Development* **11**, 2649–2689 (2018).

11. Pika, P., Hülse, D. & Arndt, S. OMEN-SED(-RCM) (v1.1): a pseudo-reactive continuum representation of organic matter degradation dynamics for OMEN-SED. *Geoscientific Model Development* **14**, 7155–7174 (2021).

12. Berner, R. A. *Early Diagenesis: A Theoretical Approach* (Princeton University Press, Princeton, 1980).

13. Boudreau, B. P. *Diagenetic models and their implementation*, vol. 505 (Springer, Berlin, 1997).

14. Boudreau, B. P. & Ruddick, B. R. On a reactive continuum representation of organic matter diagenesis (1991).

15. Bradley, J. A., Hülse, D., LaRowe, D. E. & Arndt, S. Transfer efficiency of organic carbon in marine sediments. *Nature Communications* **13**, 7297 (2022).

16. Hantschel, T. & Kauerauf, A. I. *Fundamentals of Basin and Petroleum Systems Modeling* (Springer Science & Business Media, 2009).

17. Wallmann, K. *et al.* The Global Inventory of Methane Hydrate in Marine Sediments: A Theoretical Approach. *Energies* **5** (2012).

18. Moler, C. The Lambert W Function. https://blogs.mathworks.com/cleve/2013/09/02/the-lambert-w-function/ (2013).

19. Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J. & Knuth, D. E. On the LambertW function. *Advances in Computational Mathematics* **5**, 329–359 (1996).

20. Martin, K. M., Wood, W. T. & Becker, J. J. A global prediction of seafloor sediment porosity using machine learning. *Geophysical Research Letters* **42**, 10,640–10,646 (2015).

21. Diepenbroek, M. *et al.* PANGAEA—an information system for environmental sciences. *Computers & Geosciences* **28**, 1201–1210 (2002).

22. Tyberghein, L. *et al.* Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography* **21**, 272–281 (2012).

23. Assis, J. *et al.* Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography* **27**, 277–284 (2018).

24. Straume, E. O. *et al.* GlobSed: Updated Total Sediment Thickness in the World's Oceans. *Geochemistry, Geophysics, Geosystems* **20**, 1756–1772 (2019).

25. Lee, T., Wood, W. & Phrampus, B. Global derived datasets for use in k-nn machine learning prediction of global seafloor total organic carbon (version 1)[data set]. *Zenodo. doi* **10** (2018).

26. Kursa, M. B. & Rudnicki, W. R. Feature Selection with the Boruta Package. *Journal of Statistical Software* **36**, 1–13 (2010).

27. Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. Where is positional uncertainty a problem for species distribution modelling? *Ecography* **37**, 191–203 (2014).

28. Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. & Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software* **101**, 1–9 (2018).

29. Millard, K. & Richardson, M. On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing* **7**, 8489–8515 (2015).

30. Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).

31. Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Arroita, G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution* **10**, 225–232 (2019).

32. Meyer, H. & Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution* **12**, 1620–1633 (2021).

33. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).

34. Paradis, S. *et al.* The Modern Ocean Sediment Archive and Inventory of Carbon (MOSAIC): version 2.0. *Earth System Science Data* **15**, 4105–4125 (2023).

35. Global Ocean Biogeochemistry Hindcast.

36. Restreppo, G. A., Wood, W. T. & Phrampus, B. J. Oceanic sediment accumulation rates predicted via machine learning algorithm: towards sediment characterization on a global scale. *Geo-Marine Letters* **40**, 755–763 (2020).

37. Leithold, E. L., Blair, N. E. & Wegmann, K. W. Source-to-sink sedimentary systems and global carbon burial: A river runs through it. *Earth-Science Reviews* **153**, 30–42 (2016).

38. Duarte, C. M. Reviews and syntheses: Hidden forests, the role of vegetated coastal habitats in the ocean carbon budget. *Biogeosciences* **14**, 301–310 (2017).

39. Burdige, D. J. Preservation of Organic Matter in Marine Sediments: Controls, Mechanisms, and an Imbalance in Sediment Organic Carbon Budgets? *Chemical Reviews* **107**, 467–485 (2007).

40. Emery, K. O. Relict Sediments on Continental Shelves of World1. *AAPG Bulletin* **52**, 445–464 (1968).

41. Dutkiewicz, A. & Müller, D. Seafloor sediments in the world's ocean (2021).

42. Song, S. *et al.* A global assessment of the mixed layer in coastal sediments and implications for carbon storage. *Nature Communications* **13** (2022).

43. Diesing, M. Deep-sea sediments of the global ocean. *Earth System Science Data* **12**, 3367–3381 (2020).

44. Solan, M. *et al.* Worldwide measurements of bioturbation intensity, ventilation rate, and the mixing depth of marine sediments. *Scientific Data* **6**, 58 (2019).

45. Middelburg, J. J., Soetaert, K. & Herman, P. M. Empirical relationships for use in global diagenetic models. *Deep Sea Research Part I: Oceanographic Research Papers* **44**, 327–344 (1997).

46. Boudreau, B. P. Mean mixed depth of sediments: The wherefore and the why. *Limnology and Oceanography* **43**, 524–526 (1998).

47. Teal, L., Bulling, M., Parker, E. & Solan, M. Global patterns of bioturbation intensity and mixed depth of marine soft sediments. *Aquatic Biology* **2**, 207–218 (2010).

48. Pika, P. A., Hülse, D., Eglinton, T. I. & Arndt, S. Regional and Global Patterns of Apparent Organic Matter Reactivity in Marine Sediments. *Global Biogeochemical Cycles* **37**, e2022GB007636 (2023).

49. Jørgensen, B. B., Wenzhöfer, F., Egger, M. & Glud, R. N. Sediment oxygen consumption: Role in the global marine carbon cycle. *Earth-Science Reviews* **228**, 103987 (2022).

50. Seiter, K., Hensen, C. & Zabel, M. Benthic carbon mineralization on a global scale. *Global Biogeochemical Cycles* **19** (2005).

51. Amante, C. & Eakins, B. W. Etopo1 arc-minute global relief model: procedures, data sources and analysis (2009).