

# Preservation of organic carbon in marine sediments sustained by sorption and transformation processes

Received: 2 May 2023

Accepted: 29 October 2024

Published online: 3 January 2025

 Check for updates

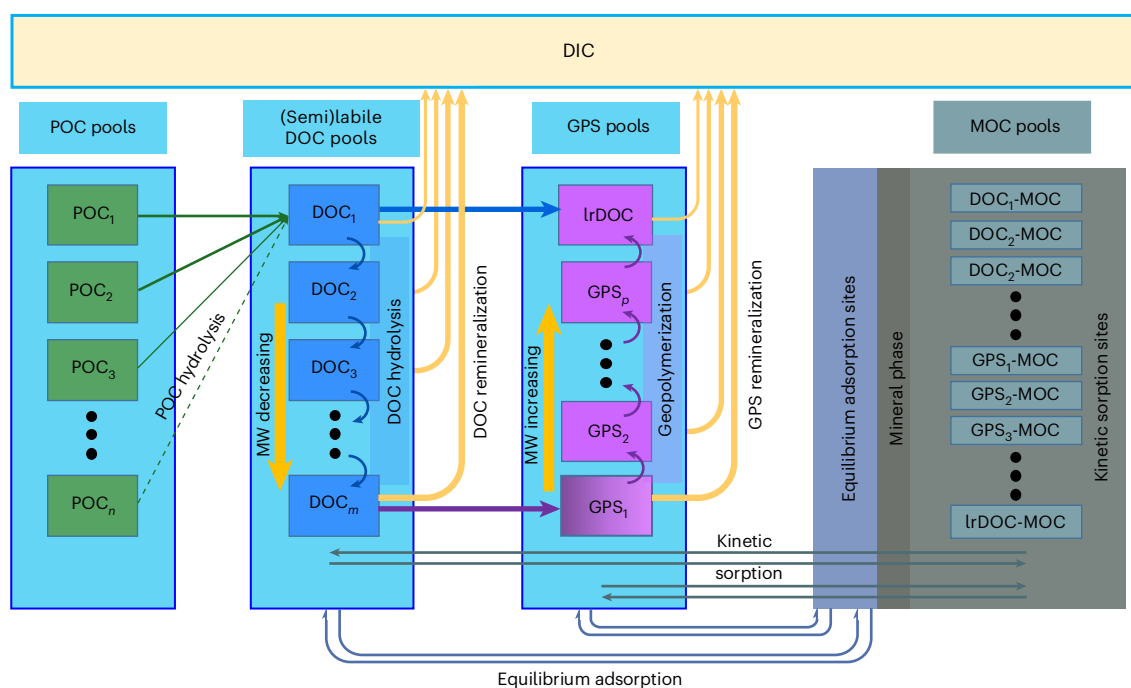
Peyman Babakhani<sup>1,2</sup>✉, Andrew W. Dale<sup>3</sup>, Clare Woulds<sup>4</sup>,  
Oliver W. Moore<sup>1,5</sup>, Ke-Qing Xiao<sup>1,6</sup>, Lisa Curti<sup>1</sup> & Caroline L. Peacock<sup>1</sup>

Controls on organic carbon preservation in marine sediments remain controversial but crucial for understanding past and future climate dynamics. Here we develop a conceptual-mathematical model to determine the key processes for the preservation of organic carbon. The model considers the major processes involved in the breakdown of organic carbon, including dissolved organic carbon hydrolysis, mixing, remineralization, mineral sorption and molecular transformation. This allows redefining of burial efficiency as preservation efficiency, which considers both particulate organic carbon and mineral-phase organic carbon. We show that preservation efficiency is almost three times higher than the conventionally defined burial efficiency and reconciles predictions with global field data. Kinetic sorption and transformation are the dominant controls on organic carbon preservation. We conclude that a synergistic effect between kinetic sorption and molecular transformation (geopolymerization) creates a mineral shuttle in which mineral-phase organic carbon is protected from remineralization in the surface sediment and released at depth. The results explain why transformed organic carbon persists over long timescales and increases with depth.

The preservation of organic carbon (OC) in marine sediments is critical to the global carbon and oxygen cycles and, thus, to Earth's climate and atmospheric composition<sup>1–4</sup>, the distribution of energy resources<sup>5,6</sup> and finding potential ocean-based mitigation strategies for the removal of excess atmospheric carbon dioxide that drives climate change<sup>7,8</sup>. The controls on carbon preservation, however, are currently unclear<sup>9–11</sup>. In the surface ocean, primary producers or phytoplankton take up atmospheric carbon dioxide to generate biomass, a fraction of which reaches the sediment as particulate OC (POC) and undergoes a series of complex degradation pathways, which may ultimately lead to either carbon remineralization or burial<sup>9,12</sup>. However, this paradigm neglects the role of dissolved OC (DOC) in

carbon preservation<sup>13,14</sup>. DOC, which is produced from the hydrolysis of POC, is a key intermediary in carbon cycling before OC is remineralized to inorganic carbon, for example, carbon dioxide. Present concentrations of DOC in marine sediment pore waters, however, may seem low<sup>15</sup>, leading to a lack of knowledge on its role in OC preservation and cycling. Nevertheless, experimental investigations and field measurements<sup>6,12,16–18</sup> suggest that DOC may accumulate over time as a result of its sorption to minerals forming mineral-phase OC (MOC) and/or its molecular transformation within sediments known as geopolymerization<sup>13,14,16,18–21</sup>, or its dilution in the water column after diffusing out from sediments<sup>22,23</sup>, and may thus substantially impact the carbon cycle on Earth<sup>24,25</sup>.

<sup>1</sup>School of Earth and Environment, University of Leeds, Leeds, UK. <sup>2</sup>Department of Civil Engineering and Management, University of Manchester, Manchester, UK. <sup>3</sup>GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany. <sup>4</sup>School of Geography, University of Leeds, Leeds, UK. <sup>5</sup>Department of Environment and Geography, University of York, York, UK. <sup>6</sup>Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China. ✉e-mail: [peyman.babakhani@manchester.ac.uk](mailto:peyman.babakhani@manchester.ac.uk)



**Fig. 1 | Conceptual model for DOC cycling in sediments.** The proposed conceptual model incorporates mechanisms of geopolymerization, equilibrium adsorption and kinetic sorption and a modified concept of hydrolysis that follows DOC cycling in the water column<sup>20</sup>. Model schematic nomenclature includes POC, GPS, IrDOC, various mineral-sorbed DOC ((semi)labile DOC-MOC, GPS-MOC and IrDOC-MOC), DIC and MW. All DOC, GPS and IrDOC pools can interact with minerals through equilibrium adsorption and kinetic sorption. In general, POC pools that originate from the water column can be hydrolysed

at any depth in the sediment or remain unhydrolysed. Their transport in the sediment is similar to the sediment solid minerals. MOC pools are transported similarly to sediment solid minerals and POC, but they originate from the net sorption of DOC, GPS and IrDOC to minerals and are further assumed to be unreactive unless the carbon is desorbed from minerals. Part of the POC, which is not hydrolysed at a given depth, and part of the MOC, which is not desorbed at that depth, are considered to be permanently buried.

So far, the contributions of these processes to OC preservation and, thus, carbon cycling have received relatively little attention and are poorly known. Moreover, the conventional concept of OC burial efficiency (BE) in sediments—an indicator of the potential to preserve carbon and quantify global budgets of carbon in modern and ancient sediments<sup>1,4,26–28</sup>—is conceptually incorrect if preservation of OC via DOC sorption and transformation is considerable and ignored. We redefine BE as preservation efficiency (PE), which includes both conventionally considered POC burial and MOC preservation.

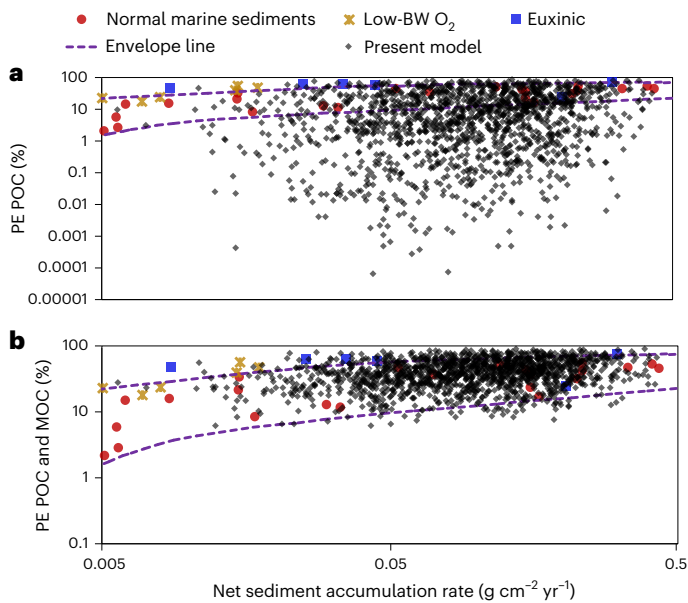
Here, we develop a mechanistic reaction-transport model (RTM) that considers the key processes of OC preservation in marine sediments via DOC cycling. After extensive validation, the model is used alongside Monte Carlo and artificial neural network (ANN) analyses to provide global insights into the role of the processes controlling carbon preservation in marine sediments and show where and how preservation occurs.

## Conceptualizing carbon cycling and preservation in sediments

Our conceptual model for carbon cycling and preservation in sediments begins with the hydrolysis of several discrete POC fractions (POC<sub>1</sub>, POC<sub>2</sub> and so on) to a single DOC pool (DOC<sub>1</sub>) of high molecular weight (MW) (Fig. 1). Continuing sequential hydrolysis of DOC<sub>1</sub> produces DOC with increasingly lower MW (DOC<sub>2</sub>, DOC<sub>3</sub> and so on) and higher reactivity (remineralization rate), analogous to the recent paradigm for DOC cycling in the water column<sup>20</sup>. The sequential hydrolysis approach is markedly different from the previous hypothesis (Fig. 1), where each discrete POC fraction produces a corresponding DOC pool with similar reactivity to its parent POC pool<sup>14</sup>. To represent the production of least-reactive DOC (IrDOC), we include the direct accumulation of freshly hydrolysed but intrinsically undegradable DOC

via a direct transfer of DOC<sub>1</sub> to IrDOC<sup>14</sup> (Fig. 1). We further consider the geopolymerization of low-MW DOC (DOC<sub>m</sub>) to higher-MW molecules herein referred to as geopolymerized substances (GPS-DOC or simply GPS)<sup>21,29</sup> with increasingly lower reactivity (GPS<sub>1</sub>, GPS<sub>2</sub> and so on), which eventually contribute to the IrDOC pool (Fig. 1). Our IrDOC pool, thus, comprises both intrinsically undegradable DOC and aged molecularly transformed GPS. Reactivities of DOC pools are selected on the basis of their lifetimes in the water column and are categorized as labile and semi-labile (DOC<sub>1</sub>, DOC<sub>2</sub> and so on, grouped as (semi)labile-DOC, with a lifetime of ~9 h to ~1.5 years), mid-reactive (GPS<sub>1</sub>, GPS<sub>2</sub> and so on, with a lifetime of ~20 years) and least-reactive (IrDOC, with a lifetime of ~16,000 years)<sup>24</sup>. To represent the sorption of DOC to minerals, we include a two-site description of ‘equilibrium adsorption’<sup>16,18,19</sup> and ‘kinetic sorption/desorption’. Kinetic sorption/desorption, which is a general description of sorption/desorption processes that are not in equilibrium, mostly aims to represent occlusion within minerals, co-precipitation and/or aggregation with minerals, and reverse processes that result in the desorption of carbon from the mineral matrix<sup>17,18,30</sup>. The net result of the kinetic sorption–desorption leads to the formation of MOC<sup>31–33</sup> pools, including (semi)labile DOC-MOC, GPS-MOC and IrDOC-MOC (Fig. 1). These MOC pools are then transported with minerals similar to POC pools that are mixed with minerals, although MOC is different from POC by origin, that is, POC originates from the water column whereas MOC is formed in the sediments via kinetic sorption of DOC to minerals.

We incorporate our conceptual model (Fig. 1) into a vertically resolved RTM for marine sediments that couples transport processes (for example, sediment burial velocity before compaction, and bioturbation mixing)<sup>34</sup> and biogeochemical reactions (for example, DOC remineralization). We execute the validated RTM in a Monte Carlo approach (>1,000 simulations) using input parameters that are varied randomly



**Fig. 2 | Comparing model-generated PE with literature data. a, b,** Data generated for 1,450 model runs in a Monte Carlo approach (transparent black dots) for the conventional approach to calculate PE of POC that considers only POC (a) and for the newly defined PE that considers both POC and MOC (b). These are compared with field data from previous studies<sup>12,36</sup>. The spread of model data is derived from a normal distribution of the net sediment accumulation rate data observed in the global grid datasets (Supplementary Figs. 1 and 2). The envelope line (dashed line) represents the general boundaries of the spread of data identified in previous studies<sup>13,41</sup>. Low BW O<sub>2</sub> stands for low bottom water oxygen concentration.

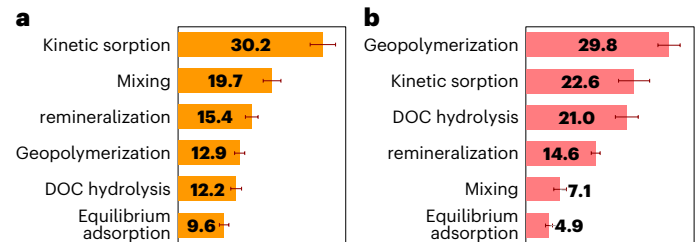
within globally relevant ranges (Supplementary Table 1) based on statistical distributions taken from six field-modelling datasets (Supplementary Figs. 1 and 2) and ten previous studies (Supplementary Table 2) to examine the broader role of different OC preservation processes (Supplementary Fig. 3, stage 1).

We then use the RTM-generated dataset to train an ANN<sup>35</sup> and determine the importance of processes that control OC preservation (Supplementary Fig. 3, stage 2). Our model allows both MOC formation in the solid-phase OC and POC burial at depth to be included in the definition of PE in contrast to the conventional approach where only POC burial is considered in the model<sup>27,28</sup>. The equations for the conventional and the newly defined PE (equations (6) and (7)) and other details of our model development, testing and execution are given in Supplementary Sections 1 and 2. We use PE to evaluate the overall model performance against global data. We further validate our mathematical model and other aspects of our model as summarized in Methods and fully described in Supplementary Section 2, Supplementary Figs. 4–12 and Supplementary Tables 3–7.

## Model evaluation

The incorporation of MOC into models of PE is compared with the classical approach<sup>28,36</sup> that considers only POC in Fig. 2 (refs. 12,36). The PE of POC alone cannot explain the observed trend and envelope of PE values versus sediment accumulation rate (Fig. 2a, dashed lines)<sup>12,36</sup>. Only when the preservation of MOC is considered along with POC does the vast majority of model output data fall inside the envelope (Fig. 2b).

Many different factors have been proposed to explain OC preservation in marine sediments<sup>312</sup>. The ability of our model to predict available global PE data indicates that consideration of DOC cycling and related processes, especially sorption and geopolymerization, is critical to predicting PE. This is remarkable because no model parameters were tuned in these simulations (Fig. 2). Our approach of combining RTM, ANN and Monte Carlo allows the process-based understanding of



**Fig. 3 | The relative importance of different processes. a, b,** The relative importance (%) of six processes to PE when MOC is considered in addition to POC (a) and to preservation rates for MOC (b). The six processes are DOC hydrolysis, DOC remineralization, mixing, equilibrium adsorption, kinetic sorption and geopolymerization. The newly defined PE is given by equation (7). The preservation rates for MOC are shown as the rate of MOC formation, which is the sum of net kinetic sorption rates integrated at the depth of 1 m ( $\mu\text{mol cm}^{-2} \text{yr}^{-1}$ ) for DOC, GPS and IrDOC. The importance of each process is obtained on the basis of the maximum sensitivity of the parameters categorized for each process. The categorization is presented in Supplementary Table 9. Each bar is the mean of 1,000 executions of the process importance analysis, and the error bars represent the 95% confidence interval. Details of sampling in the Monte Carlo method for process importance analysis are provided in Supplementary Sections 1.3–1.5 and in previous studies<sup>35</sup>.

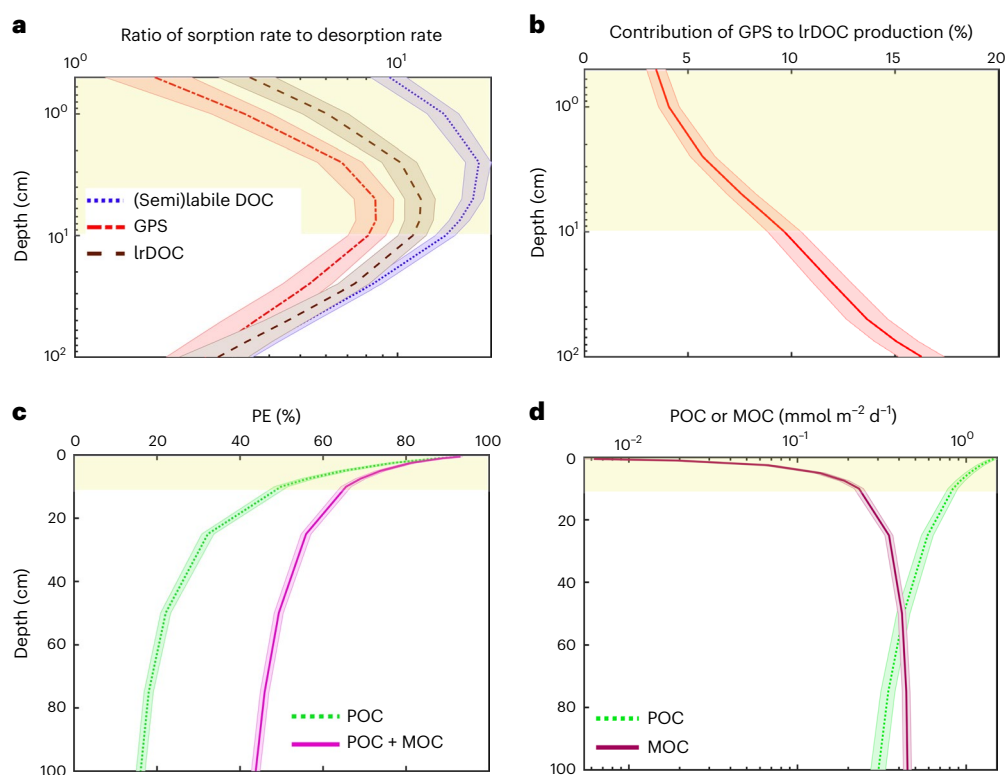
carbon dynamics to be translated into global understanding without the necessity for model fitting to specific sites that may introduce uncertainties related to site-specific conditions.

Our model definition of POC hydrolysis is similar to that considered as POC remineralization elsewhere<sup>9,27,28</sup>, and thus, our estimation for PE of POC is similar to the conventional approach<sup>28</sup>, which also underestimates field PE with estimated global values that are less than 10% (ref. 28). Our model, however, additionally tracks DOC hydrolysed from POC and its sorption to minerals, which eventually leads to a much better prediction of PE field data compared with the conventional approach. The mean values for all Monte Carlo model runs ( $n = 1,450$ ) are  $16.1 \pm 1\%$  and  $43.8 \pm 1\%$  for PE of POC and POC + MOC, respectively.

## The role of different processes in carbon preservation

We quantify the importance of six model processes in controlling key indicators of carbon preservation, including PE and MOC formation rates (or DOC sorption rates). Three processes are traditionally understood to be important for carbon preservation<sup>37</sup>, namely, DOC hydrolysis to increasingly lower-MW DOC, DOC remineralization and sediment mixing by fauna. The importance of the other three processes is disputed or poorly understood, namely, kinetic sorption, equilibrium adsorption and geopolymerization (Supplementary Tables 8 and 9).

The results reveal that kinetic sorption is the most important process for PE with a relative importance of  $30.2 \pm 3\%$  among the six processes, followed by mixing ( $19.7 \pm 2\%$ ), remineralization ( $15.4 \pm 1\%$ ), geopolymerization ( $12.9 \pm 1\%$ ), DOC hydrolysis ( $12.2 \pm 1\%$ ) and equilibrium adsorption ( $9.6 \pm 1\%$ ) (Fig. 3a). In general, processes that control POC burial flux and carbon turnover at depth, such as mixing and remineralization, are more important for PE than geopolymerization and DOC hydrolysis. We also investigate the most important processes for the preservation rate of DOC-MOC species (Supplementary Fig. 13) that are averaged in Fig. 3b. These show that geopolymerization ( $29.8 \pm 2\%$  relative to all six processes) is the most important process for DOC-MOC preservation, followed by kinetic sorption ( $22.6 \pm 3\%$ ), DOC hydrolysis ( $21.0 \pm 2\%$ ), remineralization ( $14.6 \pm 1\%$ ), mixing ( $7.1 \pm 1\%$ ) and equilibrium adsorption ( $4.9 \pm 1\%$ ) (Fig. 3b). The highest importance of the geopolymerization process in controlling DOC-MOC preservation adds weight to the existence of a synergic effect of geopolymerization with DOC sorption, in which geopolymerization renders DOC less reactive and sorption provides extra protection from microbial remineralization<sup>18,21</sup>.



**Fig. 4 | Depth profiles obtained from Monte Carlo modelling.** **a**, The ratio of the kinetic sorption rates averaged for 1,450 runs of the Monte Carlo modelling to the desorption rates also averaged for 1,450 model runs at different depths. **b**, The percentage contribution of the final GPS pool to the total IrDOC production (Fig. 1) at different depths. These contributions are averaged for 1,450 runs of the Monte Carlo modelling. The mathematical equations used to produce these plots are presented in Supplementary Section 3. **c**, The PE (%) for only POC and for both POC and MOC together, demonstrating that MOC increases

calculated PE by a factor of 2.7 at 1 m depth compared with the conventional approach considering only POC. **d**, POC fluxes and MOC preservation rate. MOC content exceeds POC at a depth of ~50 cm and still continues to rise below this depth. All values have been averaged for 1,450 Monte Carlo model runs. The shaded areas of the curves represent 95% confidence intervals obtained from Monte Carlo model runs. The light-yellow-shaded region represents the mixed layer depth (top 10 cm).

Previous considerations of DOC sorption in marine sediments have been limited to a simple equilibrium surface adsorption process expressed using a partition coefficient ( $K_d$ ), whether in a modelling framework<sup>38,39</sup> or in field and experimental investigations<sup>16,19</sup>. Others have argued that equilibrium adsorption cannot be responsible for the preservation of intrinsically labile organic compounds and some type of irreversible sorption and/or change in OC reactivity might be responsible<sup>33,40,41</sup>. Proposed patterns, such as monolayer adsorption of DOC onto mineral surfaces<sup>40,42</sup>, still rely on the assumption of mere surface-mediated adsorption and have been criticized for not being able to describe the observed preservation in various ocean settings<sup>33,41,43</sup>. Our consideration of several concurrent processes shows that, while carbon preservation indicators are still not insensitive to the equilibrium adsorption process, their control by this process is the least among the six processes considered (Fig. 3a,b). While equilibrium adsorption can reduce the DOC concentration that is bioavailable in pore fluids, thereby retarding DOC degradation<sup>18,44</sup>, DOC at equilibrium sites can be instantaneously desorbed if the DOC concentrations in the pore water decrease (for example, as a result of microbial degradation or kinetic sorption)<sup>33</sup>. Kinetically controlled mineral DOC protection mechanisms such as co-precipitation, occlusion and aggregation, however, can protect DOC over much longer timescales<sup>31–33</sup>. Our analysis reveals that the kinetic sorption process is the most important factor in controlling PE (Fig. 3a) and the second most important in controlling averaged MOC preservation rates (Fig. 3b).

Recently, it has been shown that geopolymerization in the form of a Maillard-type condensation reaction through catalysis by dissolved or particulate iron and manganese is a crucial process for benthic DOC

cycling<sup>21</sup>. Yet, quantification of the geopolymerization process and its importance compared with other processes of carbon preservation has so far remained difficult<sup>10</sup>. Our analysis shows that geopolymerization is equally, or more, important for PE compared with DOC hydrolysis (Fig. 3a). Furthermore, on average, geopolymerization is the most important factor for the preservation rate of DOC species among all six processes considered (Fig. 3b).

### How sorption and geopolymerization control OC preservation

We use sediment depth profiles (Fig. 4 and Supplementary Section 3) obtained from the Monte Carlo simulations to provide a broader insight into how different processes control OC preservation. We observe that the mixed layer acts as a shuttle for different DOC pools by protecting them from exposure to oxygen, nutrients and microbial enzymes and, consequently, limiting their rapid remineralization in the mixed layer and delivering them to greater depths (Fig. 4a). We also investigate the pathways of IrDOC production, which show that geopolymerization contributes 16.3% to IrDOC formation (Fig. 4b). These are fully discussed in Supplementary Section 4.

We further observe that PE can vary with the sediment depth horizon or age horizon<sup>28</sup>. The modelled profiles of PE versus depth in sediment considering only POC as well as POC plus MOC (Fig. 4c) agree with ref. 28 that there is a change in PE with depth. However, it appears that this change becomes less discernible below 75 cm. More importantly, MOC formation grows substantially within the top 10 cm and continues to increase below this depth (Fig. 4d) to the point where MOC exceeds the POC content below 50 cm (further discussed in Supplementary Section 4).

In conclusion, our results reveal that, aside from conventionally known processes (hydrolysis, mixing and remineralization), kinetic sorption and geopolymerization probably play a major role in OC preservation. Kinetic sorption creates a mineral shuttle that effectively removes DOC from the active surface layer and releases it at depth, while geopolymerization as an age-dependent process renders OC less reactive. Kinetic sorption leads to the formation of a mineral-associated fraction that becomes larger than the POC pool below ~50 cm, even though the globally averaged POC profile (Fig. 4d, green curve) and PE of POC remain similar to those predicted by the conventional paradigm. Thus, the preservation of OC in marine sediments, hitherto conceived to be due to the burial of POC, is a result of several DOC-related processes, including hydrolysis, geopolymerization and net sorption of DOC to minerals. Consideration of these processes in a single model reconciles the otherwise mismatch of modelled PE with field data and sheds light on the concurrent and complex roles of different processes. We suggest that our conceptualization should be considered in models of carbon cycling and may lead to better quantification and understanding of global carbon budgets, present and past climate dynamics and long-term implications for potential ocean-based carbon dioxide removal technologies<sup>7,8,11</sup>. We acknowledge that, although we have followed the best modelling practice and our model presents advancement in several facets, other models may still be developed in the future that better describe the carbon cycle in marine sediments.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41561-024-01606-y>.

## References

- Alcott, L. J., Mills, B. J. W. & Poulton, S. W. Stepwise Earth oxygenation is an inherent property of global biogeochemical cycling. *Science* **366**, 1333–1337 (2019).
- Berner, R. A. Biogeochemical cycles of carbon and sulfur and their effect on atmospheric oxygen over Phanerozoic time. *Glob. Planet. Change* **1**, 97–122 (1989).
- Amon, R. M. W. & Benner, R. Rapid cycling of high-molecular-weight dissolved organic matter in the ocean. *Nature* **369**, 549–552 (1994).
- Li, Z., Zhang, Y. G., Torres, M. & Mills, B. J. W. Neogene burial of organic carbon in the global ocean. *Nature* **613**, 90–95 (2023).
- Abelson, P. H. Organic matter in the Earth's crust. *Annu. Rev. Earth Planet. Sci.* **6**, 325–351 (1978).
- Hatcher, P. G., Spiker, E. C., Szeverenyi, N. M. & Maciel, G. E. Selective preservation and origin of petroleum-forming aquatic kerogen. *Nature* **305**, 498–501 (1983).
- Smetacek, V. et al. Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. *Nature* **487**, 313–319 (2012).
- Babakhani, P. et al. Potential use of engineered nanoparticles in ocean fertilization for large-scale atmospheric carbon dioxide removal. *Nat. Nanotechnol.* **17**, 1342–1351 (2022).
- Arndt, S. et al. Quantifying the degradation of organic matter in marine sediments: a review and synthesis. *Earth Sci. Rev.* **123**, 53–86 (2013).
- Kleber, M. et al. Dynamic interactions at the mineral–organic matter interface. *Nat. Rev. Earth Environ.* **2**, 402–421 (2021).
- Burd, A. B. et al. 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 Sea Res. Part II* **57**, 1557–1571 (2010).
- Burdige, D. J. Preservation of organic matter in marine sediments: controls, mechanisms, and an imbalance in sediment organic carbon budgets? *Chem. Rev.* **107**, 467–485 (2007).
- Burdige, D. J. & Komada, T. in *Biogeochemistry of Marine Dissolved Organic Matter* (eds Hansell, D. A. & Carlson, C. A.) 535–577 (Elsevier, 2015).
- Burdige, D. J., Komada, T., Magen, C. & Chanton, J. P. Modeling studies of dissolved organic matter cycling in Santa Barbara Basin (CA, USA) sediments. *Geochim. Cosmochim. Acta* **195**, 100–119 (2016).
- Burdige, D. J., Berelson, W. M., Coale, K. H., McManus, J. & Johnson, K. S. Fluxes of dissolved organic carbon from California continental margin sediments. *Geochim. Cosmochim. Acta* **63**, 1507–1515 (1999).
- Keil, R. G., Montluçon, D. B., Prahl, F. G. & Hedges, J. I. Sorptive preservation of labile organic matter in marine sediments. *Nature* **370**, 549–552 (1994).
- Lalonde, K., Mucci, A., Ouellet, A. & Gélinas, Y. Preservation of organic matter in sediments promoted by iron. *Nature* **483**, 198–200 (2012).
- Hemingway, J. D. et al. Mineral protection regulates long-term global preservation of natural organic carbon. *Nature* **570**, 228–231 (2019).
- Hedges, J. I. & Keil, R. G. Sedimentary organic matter preservation: an assessment and speculative synthesis. *Mar. Chem.* **49**, 81–115 (1995).
- Dittmar, T. et al. Enigmatic persistence of dissolved organic matter in the ocean. *Nat. Rev. Earth Environ.* **2**, 570–583 (2021).
- Moore, O. W. et al. Long-term organic carbon preservation enhanced by iron and manganese. *Nature* **621**, 312–317 (2023).
- Arrieta, J. M. et al. Dilution limits dissolved organic carbon utilization in the deep ocean. *Science* **348**, 331–333 (2015).
- Middelburg, J. J. Escape by dilution. *Science* **348**, 290–290 (2015).
- Hansell, D. A. Recalcitrant dissolved organic carbon fractions. *Annu. Rev. Mar. Sci.* **5**, 421–445 (2013).
- Sexton, P. F. et al. Eocene global warming events driven by ventilation of oceanic dissolved organic carbon. *Nature* **471**, 349–352 (2011).
- Lenton, T. M., Daines, S. J. & Mills, B. J. W. COPSE reloaded: an improved model of biogeochemical cycling over Phanerozoic time. *Earth Sci. Rev.* **178**, 1–28 (2018).
- Katsev, S. & Crowe, S. A. Organic carbon burial efficiencies in sediments: the power law of mineralization revisited. *Geology* **43**, 607–610 (2015).
- Bradley, J. A., Hülse, D., LaRowe, D. E. & Arndt, S. Transfer efficiency of organic carbon in marine sediments. *Nat. Commun.* **13**, 7297 (2022).
- Hedges, J. I. The formation and clay mineral reactions of melanoidins. *Geochim. Cosmochim. Acta* **42**, 69–76 (1978).
- Schrumpf, M. et al. Storage and stability of organic carbon in soils as related to depth, occlusion within aggregates, and attachment to minerals. *Biogeosciences* **10**, 1675–1691 (2013).
- Mayer, L. M. The inertness of being organic. *Mar. Chem.* **92**, 135–140 (2004).
- Henrichs, S. M. & Sugai, S. F. Adsorption of amino acids and glucose by sediments of Resurrection Bay, Alaska, USA: functional group effects. *Geochim. Cosmochim. Acta* **57**, 823–835 (1993).
- Henrichs, S. M. Sedimentary organic matter preservation: an assessment and speculative synthesis—a comment. *Mar. Chem.* **49**, 127–136 (1995).
- Dale, A. W. et al. A revised global estimate of dissolved iron fluxes from marine sediments. *Glob. Biogeochem. Cycles* **29**, 691–707 (2015).

35. Babakhani, P., Bridge, J., Doong, R. A. & Phenrat, T. Parameterization and prediction of nanoparticle transport in porous media: a reanalysis using artificial neural network. *Water Resour. Res.* **53**, 4564–4585 (2017).
36. Canfield, D. E. Factors influencing organic carbon preservation in marine sediments. *Chem. Geol.* **114**, 315–329 (1994).
37. Boudreau, B. P. *Diagenetic Models and Their Implementation* Vol. 410 (Springer, 1997).
38. Berner, R. A. Inclusion of adsorption in the modelling of early diagenesis. *Earth Planet. Sci. Lett.* **29**, 333–340 (1976).
39. Sohma, A., Sekiguchi, Y. & Nakata, K. Modeling and evaluating the ecosystem of sea-grass beds, shallow waters without sea-grass, and an oxygen-depleted offshore area. *J. Mar. Syst.* **45**, 105–142 (2004).
40. Mayer, L. M. Surface area control of organic carbon accumulation in continental shelf sediments. *Geochim. Cosmochim. Acta* **58**, 1271–1284 (1994).
41. Lee, C. Kitty litter for carbon control. *Nature* **370**, 503–504 (1994).
42. Mayer, L. M. Relationships between mineral surfaces and organic carbon concentrations in soils and sediments. *Chem. Geol.* **114**, 347–363 (1993).
43. Keil, R. G. & Mayer, L. M. in *Treatise on Geochemistry* 2nd edn (eds Holland H. D. & Turekian K. K.) 337–359 (Elsevier, 2014).
44. Rothman, D. H. & Forney, D. C. Physical model for the decay and preservation of marine organic carbon. *Science* **316**, 1325–1328 (2007).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

## Methods

### Methodology

**Overview of the modelling procedure.** A RTM is developed to consider the cycling and preservation of DOC in marine sediment. This RTM is then emulated with artificial intelligence, which is used as a tool for conducting a robust process importance analysis. The flowchart of the model deployment is shown in Supplementary Fig. 3. The artificial intelligence technique we use in this study is an ANN, and the RTM is emulated with ANN through a Monte Carlo approach (Supplementary Fig. 3, stage 1), whereby the RTM is executed multiple times (for example, 1,000–2,000) using randomly varied input parameters to obtain the model outputs (for example, the preservation rates for different MOC fractions). These input–output datasets are then used to train an ANN (Supplementary Fig. 3, stage 2)<sup>35</sup>, which allows assessment of the sensitivities of each output to different input parameters grouped into a number of processes.

Among the 68 unknown input parameters of the model, we obtain the ranges and statistical distributions for 6 of the input parameters, including water depth<sup>45</sup>, sediment accumulation rate<sup>46</sup>, sediment surface porosity<sup>47</sup> and sediment–water interface concentrations of POC, which are assumed to be the same as total OC<sup>47</sup>, NO<sub>3</sub> (refs. 48,49) and O<sub>2</sub> (refs. 48,49) (Supplementary Tables 1 and 8 and Supplementary Figs. 1 and 2) from globally gridded data. For other typical unknown parameters, we compiled data from ten previous studies that had conducted reactive transport modelling on field data (summarized in Supplementary Tables 1 and 2). The ranges for new parameters were also taken either from the literature<sup>24</sup> or from fitting our model to the field-modelling data from the literature (Burdige et al.<sup>14,50</sup>). Although we have tested the RTM against field datasets and have selected the ranges for some of the new model parameters by considering the calibrated parameter values (Supplementary Section 2.2), such model fittings to field datasets are useful but not necessary. This is because the RTM, which is based on established or hypothetical concepts, already embodies process-based knowledge in the approach, and the additional application of ANN and Monte Carlo allows the use of parameter ranges instead of set values. A more detailed description of the overall modelling procedure is provided in Supplementary Section 1.1.

**Formulation of the RTM.** The RTM considers all common early diagenesis reactions for different compounds including dissolved species (O<sub>2</sub>, SO<sub>4</sub>, NH<sub>4</sub>, NO<sub>3</sub>, dissolved inorganic carbon (DIC), H<sub>2</sub>S, CH<sub>4</sub>, Fe<sup>2+</sup>, Mn<sup>2+</sup>, DOC<sub>1</sub> to DOC<sub>*m*</sub>, GPS<sub>1</sub> to GPS<sub>*p*</sub>, and IrDOC, where subscript *m* represents the maximum number of DOC pools and subscript *p* represents the maximum number of GPS pools) and particulate species (highly reactive iron oxide, Fe(OH)<sub>3</sub><sup>HR</sup>, moderately reactive iron oxide, Fe(OH)<sub>3</sub><sup>MR</sup>, non-reactive iron oxide, Fe(OH)<sub>3</sub><sup>UR</sup>, MnO<sub>2</sub>, FeS, FeS<sub>2</sub>, S<sup>0</sup>, and POC<sub>1</sub> to POC<sub>*n*</sub>, where subscript *n* represents the maximum number of POC pools) as listed in Supplementary Table 10. Here, we considered seven species for POC, four species for DOC and two species for GPS in addition to IrDOC. The number of different carbon species are selected mainly on the basis of consistency with previous literature<sup>24</sup> and with consideration of other aspects of modelling, including alignment with the conceptual model, smooth transition of rates across different species for numerical solution considerations and minimizing the number of unknown model parameters. It should be noted that, as demonstrated in the Supplementary Section 1.2.2, the selection of the number of these species, for example, DOC pools, per se, does not affect the model outputs noticeably.

The three governing equations of RTM for dissolved species, particulate species and sorbed species, respectively, are as follows, while the full details of the model development and validation are provided

in Supplementary Sections 1 and 2:

(1) The governing equation for dissolved species:

$$\begin{aligned} (\varphi + \rho_s \varepsilon K_{d_i}) \frac{\partial C_{d_i}}{\partial t} = & \frac{\partial}{\partial z} \left( (\varphi D + \rho_s \varepsilon K_{d_i} D_b) \frac{\partial C_{d_i}}{\partial z} \right) \\ & - \frac{\partial}{\partial z} \left( (\varphi v_d + \rho_s \varepsilon K_{d_i} v_p) C_{d_i} \right) + \varphi \alpha (C_{d_i}(0) - C_{d_i}(z)) \quad (1) \\ & + \varphi \sum_{j=1} R_{d_{i,j}} - k_{\text{sorp}_i} C_{d_i} + \frac{k_{\text{sorp}_i}}{K_{d_{\text{sorp}_i}}} S_{d_i} \end{aligned}$$

(2) The governing equation for particulate species:

$$\rho_s \varepsilon \frac{\partial C_{p_i}}{\partial t} = \rho_s \frac{\partial}{\partial z} \left( \varepsilon D_b \frac{\partial C_{p_i}}{\partial z} \right) - \rho_s \frac{\partial}{\partial z} (\varepsilon v_p C_{p_i}) + \rho_s \varepsilon R_{p_i} \quad (2)$$

(3) The governing equation for mineral phase, MOC, resulting from the kinetically sorbed fraction of dissolved species:

$$\begin{aligned} \rho_s \varepsilon \frac{\partial S_{d_i}}{\partial t} = & \rho_s \frac{\partial}{\partial z} \left( \varepsilon D_b \frac{\partial S_{d_i}}{\partial z} \right) - \rho_s \frac{\partial}{\partial z} (\varepsilon v_p S_{d_i}) + \rho_s \varepsilon \sum_{j=1} R_{s_{d_{i,j}}} \\ & + k_{\text{sorp}_i} C_{d_i} - \frac{k_{\text{sorp}_i}}{K_{d_{\text{sorp}_i}}} S_{d_i}, \quad (3) \end{aligned}$$

where  $C_{d_i}$  is the concentration of dissolved species *i* (mM or  $\mu\text{mol cm}^{-3}$  of pore water),  $C_{p_i}$  is the concentration of particulate species *i* ( $\text{g g}^{-1}$ ),  $S_{d_i}$  is the concentration of dissolved species *i* kinetically sorbed to sediment minerals ( $\mu\text{mol g}^{-1}$  of solid sediments),  $\varphi$  is porosity,  $\varepsilon$  is the solid fraction of sediments, which is equal to  $1 - \varphi$ ,  $v_d$  and  $v_p$  are the burial velocities of pore water and particulate species ( $\text{cm yr}^{-1}$ ),  $\rho_s$  is the dry density of sediments ( $\text{g cm}^{-3}$ ),  $D_i$  is the apparent diffusion coefficient of dissolved species *i* ( $\text{cm}^2 \text{yr}^{-1}$ ),  $\alpha$  is the bio-irrigation coefficient ( $\text{cm}^2 \text{yr}^{-1}$ ),  $D_b$  is the bioturbation coefficient ( $\text{cm}^2 \text{yr}^{-1}$ ),  $z$  is the sediment depth with respect to the coordinate system located at the sediment–water interface (cm),  $R_p$ ,  $R_d$  and  $R_{s_{d_i}}$  stand for reaction rates of particulate, dissolved and kinetically sorbed species ( $\text{yr}^{-1}$ ,  $\mu\text{mol cm}^{-3} \text{yr}^{-1}$  and  $\mu\text{mol g}^{-1} \text{yr}^{-1}$ ), respectively, which are temporally and spatially variable,  $k_{\text{sorp}}$  is the mass transfer rate between the dissolved and kinetically sorbed phases to minerals (MOC pools) ( $\text{yr}^{-1}$ ), and  $K_{d_{\text{sorp}}}$  is the so-called distribution coefficient in the kinetic mass transfer expression ( $\text{cm}^3 \text{g}^{-1}$ ).

The first stage of the hydrolysis is considered similar to the conventional first-order multi-POC degradation model known as the multi-G model<sup>51</sup> with a series of POC pools converting to a single DOC pool, DOC<sub>*i*</sub>, in parallel:

$$\left\{ \frac{\partial C_{p_i}}{\partial t} \right\}_{\text{Hydrolysis}} = R_{p_i} = k_i C_{p_i}, \quad (4)$$

where  $k_i$  is the hydrolysis rate constant, which was considered in a similar way to the degradation rate constants of POC in the continuum model following previous studies<sup>9,34,52</sup>.

The sequential stage of the hydrolysis has been described using a consecutive first-order reaction expression<sup>20,53</sup>:

$$\left\{ \varphi \frac{\partial C_{d_i}}{\partial t} \right\}_{\text{Hydrolysis}} = \left\{ \varphi \sum_{j=1} R_{d_{i,j}} \right\}_{\text{Hydrolysis}} = \lambda_{\text{DOC}_{i-1}} C_{d_{i-1}} - \lambda_{\text{DOC}_i} C_{d_i}, \quad (5)$$

where  $\lambda_{\text{DOC}_i}$  is the conversion rate of DOC<sub>*i*</sub> to DOC<sub>*i+1*</sub>, and  $\lambda_{\text{DOC}_{i-1}}$  is the conversion rate of DOC<sub>*i-1*</sub> to DOC<sub>*i*</sub> in  $\text{yr}^{-1}$ .

The same mathematical formula is used to describe geopolymerization<sup>54,55</sup> as provided in Supplementary Section 1 along with the other details.

**Calculation of PE.** PE, elsewhere known as BE, conventionally considered for POC<sup>28,36</sup> is defined as follows:

$$PE = \frac{\text{POC flux at depth } L}{\text{Total POC flux at sediment surface}} \times 100, \quad (6)$$

where  $L$  is a given depth herein considered as 1 m. In the present study, owing to the full consideration of the fate of DOC in our model, we are able to present a more accurate consideration of PE that includes the fraction of solid phase OC that has undergone hydrolysis and sorption to minerals:

$$PE = \frac{\text{POC flux at depth } L + \text{Sorption rate integrated over depth } L}{\text{Total POC flux at sediment surface}} \times 100. \quad (7)$$

The sorption rate is the net DOC kinetic sorption rate (sorption rate minus desorption rate or the net MOC formation rate). Further explanations about the rates are provided in Supplementary Sections 2.4 and 3. Although the consideration of depth-versus-age horizons in early diagenesis modelling can be important for global predictions, as was recently highlighted<sup>28</sup>, in the present study, we considered only a constant depth horizon as the aim is to obtain insight into the processes that control OC preservation rather than making global predictions.

**ANN for process importance analysis.** The ANN is a versatile and universal tool for function approximation problems and is notable for its application to complex, nonlinear systems<sup>35,56</sup>. The commonly used ANN structure is a three-layer configuration comprising input, hidden and output layers<sup>35,57–59</sup>. Each of these layers is composed of a series of nodes (neurons) with their numbers in the input and output layers corresponding to the number of input and output variables, respectively. The number of neurons on the hidden or middle layers should be optimized when finding the best fit during the training process<sup>57,60</sup>. The main equation used for processing the information (or signal) in the structure of the ANN is a simple algebraic equation in the form of  $y = w \times x + b$ , which applies to each neuron in the hidden layer. The information is then summed up for all nodes; additional functions called transfer (or activation) functions that are exerted on the input and output information that are detailed elsewhere<sup>35,57</sup>. Here,  $x$  stands for the input information (or signal),  $y$  stands for output information,  $w$  is weights and  $b$  is biases. Weights and biases are the hyperparameters of the ANN, which are determined after fitting ANN to data, and once they are determined, they form an empirical network that can be used for new predictions. In the scope of the present study, we use ANN only for process importance analysis, not for prediction. Here, we use the partial derivative method<sup>35,57,58,61</sup> for process importance analysis. In brief, in this method, the derivatives of the equations used in the structure of ANN are used to represent the role of each ANN input in controlling the ANN output; for example, for the main equation  $y = w \times x + b$ , the derivative is equal to  $w$ . In this way, the  $w$  value for each neuron represents the strength of the signals passing through that neuron<sup>35,57–59</sup>. This implies that in the process importance analysis, the input parameter values do not play a marked role; rather, it is their variations that are important and reflected in the structure of the ANN. Details of the ANN model used here have been selected following ref. 35 and are described in Supplementary Section 1.5.

**Model validation.** We validate our model in a number of ways. We validate our developed governing equations of the RTM on the basis of an analytical approach. In this approach, for the condition where equilibrium adsorption and kinetic sorption are expected to behave similarly, that is, at high exchange rates, we first run the model after turning off equilibrium adsorption and then run the model again turning off the kinetic sorption expression. Then, the model outputs for these two types of simulation are compared. We used existing field-modelling data (principally from Santa Barbara Basin, given the comprehensive

dataset available)<sup>14,62</sup> as shown in Supplementary Figs. 5–11 and Supplementary Table 6. We also validated the model on the basis of mass budgets. The use of the ANN is validated on the basis of its ability to describe the data (measured using goodness-of-fit criteria described in Supplementary Section 2.2) and using the uncertainties it yields in the process importance analysis. Finally, validation of the overall RTM modelling process was carried out using mass budgets of the averaged model outputs over the multiple runs of the Monte Carlo. This was done using the concept of mass flow in the model illustrated in Supplementary Fig. 14.

The results of the model output comparison between the cases when kinetic sorption is operative and when the equilibrium adsorption is operative at a high mass transfer rate show an excellent match ( $R^2 = 1.000$ ; Supplementary Fig. 4) verifying our approach towards development and implementation of sorption formulation in the governing equations.

The results of the model fit to several field or modelling datasets, including Meysman et al.<sup>62</sup> (Supplementary Fig. 5), Kraal et al.<sup>63</sup> (Supplementary Figs. 6 and 8 and Supplementary Tables 3–5) and Burdige et al.<sup>14,50</sup> (Supplementary Figs. 9–11 and Supplementary Tables 6) show excellent matches between our model and existing field-model datasets for most of the concentration-versus-depth profiles. The exceptions are generally  $\text{FeS}_2$  and  $\text{Mn}^{2+}$  profiles, which show poorer model fit due to the lack of carbonate species in our model. The added complexity of our model (more unknown parameters) is validated in eight steps against the Santa Barbara Basin dataset<sup>14,50</sup> (Supplementary Table 6), showing that each step is justifiable in terms of improvements in model fits to the data for the cost of complexity, according to model selection criteria<sup>64</sup> increasing from 0.626 in step 2 to 0.843 in step 7. Burdige et al.<sup>14,50</sup> further considered  $\delta^{13}\text{C}$ ,  $\Delta^{14}\text{C}$  and carbon-to-nitrogen ratios in their model and matched them with field measurements that are not conducted here.

The ANN model could fit the data in all cases with the best predictive fit Nash–Sutcliffe model efficiency criterion<sup>65</sup> ranging from 0.923 to 0.944 (Supplementary Table 7 and Supplementary Fig. 12). The uncertainties in the ANN process importance analysis determined as a 95% confidence interval were relatively minor (see error bars in Fig. 3). These uncertainties range from 4.6% to 29.9% (12.6% on average) of the mean values for the cases investigated and shown in Fig. 3 and Supplementary Fig. 13.

The mass budgets for different cross-sections of the simplified conceptual model shown in Supplementary Fig. 14 were calculated on the basis of averaged results of all Monte Carlo model runs (1,450) at stage 1. According to these results, the mass budget in cross-section A–A is  $\text{MB}_{\text{A-A}} = 57.136 \mu\text{mol cm}^{-2} \text{yr}^{-1}$ , in B–B is  $\text{MB}_{\text{B-B}} = 57.197 \mu\text{mol cm}^{-2} \text{yr}^{-1}$  and in C–C is  $\text{MB}_{\text{C-C}} = 57.189 \mu\text{mol cm}^{-2} \text{yr}^{-1}$ , demonstrating an overall mass balance error of  $-0.1\%$ , which is less than the acceptable mass balance error of 1% considered in our general modelling process. It should be noted that, despite our extensive model validation process, in the present study, we use the model only for process importance analysis and finding insight into the underlying processes responsible for OC preservation, not for making global predictions, which is the subject of future study. Finally, based on acceptable uncertainties (95% confidence interval) related to the sum of all >1,000 RTM runs, shown as the shaded area around the curves in Fig. 4 and Supplementary Fig. 16, and the uncertainties of process importance analysis obtained from the ANN stage shown in Fig. 3, our general approach of random variation of input parameters is appropriate.

It should be noted that limitation in the capacity of sorption sites, for example, monolayer sorption<sup>16,40</sup>, typically does not apply to kinetic sorption because the kinetic sorption model in the present study mainly represents the processes that internalize DOC into the mineral matrix, such as occlusion, co-precipitation and aggregation, and thus, limited-capacity sorption considered in the literature mostly through the monolayer surface adsorption hypothesis is not



applicable to our MOC production. Furthermore, kinetic sorption is slower than equilibrium adsorption, which is known to be instantaneous. Thus, kinetic sorption, which is limited by pore water concentrations that are also controlled by hydrolysis, degradation and so on, is less likely to face a second limitation by the capacity of sorption sites compared with instantaneous equilibrium adsorption for which different types of isotherm, such as linear, Langmuir and Freundlich, have been defined<sup>66,67</sup>. Adding an additional parameter to force a proportion of the OC to be taken up by the kinetic and equilibrium sorption sites would add more unknown parameters and is deemed unnecessary in this case.

### Data availability

Data associated with Figs. 2–4 are available via figshare at <https://doi.org/10.6084/m9.figshare.27273006> (ref. 68). Data related to Supplementary Figs. 13, 15 and 16 and the main datasets (such as those generated by RTM through the Monte Carlo process (stage 1) and used in the ANN process importance analysis (stage 2)) are available via figshare at <https://doi.org/10.6084/m9.figshare.27273030> (ref. 69).

### Code availability

Computer codes are available from the authors upon request.

### References

45. Burwicz, E. B., Rüpke, L. H. & Wallmann, K. Estimation of the global amount of submarine gas hydrates formed via microbial methane formation based on numerical reaction-transport modeling and a novel parameterization of Holocene sedimentation. *Geochim. Cosmochim. Acta* **75**, 4562–4576 (2011).
46. Restrepo, 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-Mar. Lett.* **40**, 755–763 (2020).
47. Lee, T. R., Wood, W. T. & Phrampus, B. J. A machine learning (kNN) approach to predicting global seafloor total organic carbon. *Glob. Biogeochem. Cycles* **33**, 37–46 (2019).
48. WOA18 Data Access (National Centers for Environmental Information, 2018); <https://www.nodc.noaa.gov/OC5/woa18/woa18data.html>
49. Bohlen, L., Dale, A. W. & Wallmann, K. Simple transfer functions for calculating benthic fixed nitrogen losses and C:N:P regeneration ratios in global biogeochemical models. *Global Biogeochem. Cycles* **26**, GB3029 (2012).
50. Burdige, D. J., Komada, T., Magen, C. & Chanton, J. P. Methane dynamics in Santa Barbara Basin (USA) sediments as examined with a reaction-transport model. *J. Mar. Res.* **74**, 277–313 (2016).
51. Jørgensen, B. B. A comparison of methods for the quantification of bacterial sulfate reduction in coastal marine sediments: I. Measurement with radiotracer techniques. *Geomicrobiol. J.* **1**, 11–27 (1978).
52. Bradley, J. A. et al. Widespread energy limitation to life in global subseafloor sediments. *Sci. Adv.* **6**, eaba0697 (2020).
53. Rogalinski, T., Liu, K., Albrecht, T. & Brunner, G. Hydrolysis kinetics of biopolymers in subcritical water. *J. Supercrit. Fluids* **46**, 335–341 (2008).
54. Leong, L. P. & Wedzicha, B. L. A critical appraisal of the kinetic model for the Maillard browning of glucose with glycine. *Food Chem.* **68**, 21–28 (2000).
55. Babakhani, P. et al. Comparison of a new mass-concentration, chain-reaction model with the population-balance model for early- and late-stage aggregation of shattered graphene oxide nanoparticles. *Colloids Surf. A* **582**, 123862 (2019).
56. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
57. Yu, H. et al. Prediction of the particle size distribution parameters in a high shear granulation process using a key parameter definition combined artificial neural network model. *Ind. Eng. Chem. Res.* **54**, 10825–10834 (2015).
58. Nourani, V. & Sayyah-Fard, M. Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Adv. Eng. Softw.* **47**, 127–146 (2012).
59. Gevrey, M., Dimopoulos, I. & Lek, S. Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecol. Modell.* **195**, 43–50 (2006).
60. Yerramareddy, S., Lu, S. C. Y. & Arnold, K. F. Developing empirical models from observational data using artificial neural networks. *J. Intell. Manuf.* **4**, 33–41 (1993).
61. Gevrey, M., Dimopoulos, I. & Lek, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modell.* **160**, 249–264 (2003).
62. Meysman, F. J. R., Middelburg, J. J., Herman, P. M. J. & Heip, C. H. R. Reactive transport in surface sediments. II. Media: an object-oriented problem-solving environment for early diagenesis. *Comput. Geosci.* **29**, 301–318 (2003).
63. Kraal, P., Slomp, C. P., Reed, D. C., Reichart, G. J. & Poulton, S. W. Sedimentary phosphorus and iron cycling in and below the oxygen minimum zone of the northern Arabian Sea. *Biogeosciences* **9**, 2603–2624 (2012).
64. Akaike, H. A new look at the statistical model identification. *Autom. Control IEEE Trans.* **19**, 716–723 (1974).
65. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models part I—a discussion of principles. *J. Hydrol.* **10**, 282–290 (1970).
66. Clark, M. M. *Transport Modeling for Environmental Engineers and Scientists* (Wiley, 2011).
67. Langmuir, D. *Aqueous Environmental Geochemistry* (Prentice Hall, 1997).
68. Babakhani, P. et al. Babakhani\_et al\_main\_text\_figure\_source\_data. *figshare* <https://doi.org/10.6084/m9.figshare.27273006> (2024).
69. Babakhani, P. et al. Babakhani\_et al\_supplementary\_data. *figshare* <https://doi.org/10.6084/m9.figshare.27273030> (2024).

### Acknowledgements

We acknowledge D. Burdige for sending field data and B. Mills for helpful discussion and review of an earlier version of our paper. The use of High-Performance Computing (ARC) facilities at the University of Leeds is acknowledged. We acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 725613 MinOrg). C.L.P. gratefully acknowledges the Royal Society Wolfson Research Merit Award (WRM/FT/170005) and Leverhulme Research Fellowship (RF-2023-395\4).

### Author contributions

Conceptualization: P.B., C.L.P. and A.W.D. Data curation: P.B., C.L.P. and A.W.D. Formal analysis: P.B. Investigation: P.B. Methodology: P.B., A.W.D. and C.L.P. Visualization: P.B. Funding acquisition: C.L.P. Project administration: C.L.P. Resources: C.L.P. Software: P.B. and A.W.D. Supervision: C.L.P., A.W.D. and C.W. Validation: P.B., A.W.D., C.L.P. and K.-Q.X. Writing—original draft: P.B. Writing—review and editing: P.B., C.L.P., A.W.D., C.W., O.W.M., K.-Q.X. and L.C.

### Competing interests

The authors declare no competing interests.

**Peer review information** *Nature Geoscience* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Xujia Jiang, in collaboration with the *Nature Geoscience* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41561-024-01606-y>.

**Correspondence and requests for materials** should be addressed to Peyman Babakhani.