


Predicting the standing stock of organic carbon in surface sediments of the North–West European continental shelf

Markus Diesing  · Silke Kröger · Ruth Parker · Chris Jenkins · Claire Mason · Keith Weston

Received: 11 November 2016 / Accepted: 6 February 2017 / Published online: 15 February 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Shelf seas and their associated benthic habitats represent key systems in the global carbon cycle. However, the quantification of the related stocks and flows of carbon are often poorly constrained. To address benthic carbon storage in the North–West European continental shelf, we have spatially predicted the mass of particulate organic carbon (POC) stored in the top 10 cm of shelf sediments in parts of the North Sea, English Channel and Celtic Sea using a Random Forest model, POC measurements on surface sediments from those seas and relevant predictor variables. The presented model explains 78% of the variance in the data and we estimate that approximately 250 Mt of POC are stored

in surficial sediments of the study area (633,000 km²). Upscaling to the North–West European continental shelf area (1,111,812 km²) yielded a range of 230–882 Mt of POC with the most likely estimate being on the order of 476 Mt. We demonstrate that the largest POC stocks are associated with coarse-grained sediments due to their wide-spread occurrence and high dry bulk densities. Our results also highlight the importance of coastal sediments for carbon storage and sequestration. Important predictors for POC include mud content in surficial sediments, annual average bottom temperature and distance to shoreline, with the latter possibly a proxy for terrestrial inputs. Now that key variables in determining the spatial distribution of POC have been identified, it is possible to predict future changes to the POC stock, with the presented maps providing an accurate baseline against which to assess predicted changes.

Data availability The presented model results on POC concentrations are available at <https://doi.pangaea.de/10.1594/PANGAEA.871584>.

Responsible Editor: Martin Solan.

Electronic supplementary material The online version of this article (doi:10.1007/s10533-017-0310-4) contains supplementary material, which is available to authorized users.

M. Diesing (✉) · S. Kröger · R. Parker ·
C. Mason · K. Weston
Centre for Environment, Fisheries and Aquaculture
Science, Lowestoft, Suffolk NR33 0HT, UK
e-mail: markus.diesing@cefas.co.uk

C. Jenkins
Institute of Arctic and Alpine Research, University of
Colorado, Boulder, CO 80309-0450, USA

Keywords Organic carbon · Continental shelf ·
Sediment · Spatial prediction · Europe

Introduction

Carbon dioxide (CO₂) from the atmosphere is taken up by seawater, where it is fixed by primary producers such as phytoplankton with a proportion of this particulate organic carbon (POC) supporting the food web within the water column, while another part sinks

to the seabed (Miller 2004). The latter may be incorporated into the surface sediments or support the benthic food web and be respired or buried, along with POC from terrestrial sources. The sedimentation of POC is therefore a key process in transferring CO₂ from the atmosphere to the seabed where it may be stored long term (decades–centuries) mitigating increases in atmospheric CO₂ associated with climate change (Pachauri and Meyer 2014). There are, however, many natural mechanisms that affect the incorporation of POC into sediments, from physical processes such as particle movement or bedform migration due to storms and tides (Jenness and Duineveld 1985) or water column temperature (Bernier 1980; Middelburg 1989 and references there-in), to biological processes such as infaunal activity and mode of feeding (Aller 1982). POC supply, incorporation and storage may also be perturbed by human activities such as bottom trawling (Duplisea et al. 2001; Trimmer et al. 2005) through direct mixing or indirect impact on the infaunal community. These processes can also be altered through changes in supply through localised eutrophication or redox effects associated with differing bottom oxygen regimes and anoxia (Diaz and Rosenberg 2008; Middelburg and Levin 2009). Ultimately, it is the balance between the supply from the water column and remineralisation rate of benthic POC, which controls the POC stock in the surficial sediments and which will dictate the incorporation rate of POC. An increasing POC stock, either over space or time, therefore implies an increasing input or decreasing remineralisation. This may be due to either changing/contrasting lability of POC (i.e. terrestrial vs marine source), natural conditions or human pressures (Burdige 2005, 2007).

To understand how this range of processes affects POC storage and hence how they will be affected under future change or human pressure conditions, it is first necessary to quantify the stock. Generally, these natural and anthropogenic processes affect the upper sediment layer and this study therefore estimates the amount of POC in the surface sediments (0–10 cm) of the North-West (NW) European continental shelf using a data set of directly sampled stations to better constrain their POC stocks and hence their role in the carbon cycle. Understanding the functioning and distribution of these coastal and shelf sea POC stores is also critical since there are considerable pressures on these systems both from local and far field effects

(Bauer et al. 2013) and by improving our quantification and the underpinning conditions associated with changes in the levels of such stores, we can better understand the effects of future changes to them and in turn the global carbon cycle.

Data on POC measurements have been reported in the literature. These datasets are largely a result of detailed, small scale studies on sediment biogeochemical processing of carbon and associated pathways or linked to faunal community analysis (Basford et al. 1993; Willems et al. 2007; Stockdale et al. 2009 and references there-in). As a result, comparatively few attempts have been made so far to spatially predict POC in surficial sediments across defined areas of seabed: Mollenhauer et al. (2004) presented a map of organic carbon in surface sediments of the South Atlantic Ocean based on 1118 samples; Seiter et al. (2004) spatially predicted organic carbon content in the top 5 cm of global deep sea sediments at a 1° × 1° grid resolution by the application of a combined qualitative and quantitative-geostatistical method; Acharya and Panigrahi (2016) mapped the distribution of organic carbon on the Eastern Arabian shelf with Empirical Bayesian Kriging; Serpetti et al. (2012) mapped the organic content of coastal sediments from hydro-acoustic reflectance data in an area of seabed off the east coast of Scotland; and Neto et al. (2016) assessed the suitability of seismic peak amplitude as a predictor of total organic carbon content in shallow marine sediments, based on data collected in the Cabo Frio mud belt in an upwelling zone off south-eastern Brazil. This indicates that maps of POC concentration in surficial sediments have either been derived by means of acoustic methods (Serpetti et al. 2012; Neto et al. 2016) with limited spatial coverage or some type of kriging (Mollenhauer et al. 2004; Seiter et al. 2004; Acharya and Panigrahi 2016).

More recently, machine learning algorithms have made inroads into spatial prediction and have been used to spatially predict categorical (e.g. sediment types; Stephens and Diesing (2014)) and continuous (e.g. sediment composition; Stephens and Diesing (2015)) data. Machine learning algorithms are data-driven flexible statistical prediction techniques that ‘learn’ patterns in data to predict an associated value. Machine learning is defined as “programming computers to optimise a performance criterion using example data or past experience” (Alpaydin 2010). Such predictive mapping methods entail a two-step approach: Initially, the relationship between a set of

predictor variables and a response variable is modelled from observations (samples). The established model is then employed to predict the response variable at unsampled locations for which values of the predictor variables are known.

The aim of this study is to map and quantify POC in surficial sediments over a large area. Hence, acoustic methods are not applicable due to the lack of suitable data. Geostatistical (kriging) methods would be applicable in this case. However, there are advantages in using a machine learning approach for the task in hand, as such methods do not need to satisfy strict statistical assumptions as is the case for kriging. Additionally, such methods allow investigation of predictor-response variable relationships, which might shed light on factors and processes controlling POC in surficial sediments at a regional scale, in contrast to site specific studies undertaken previously.

The objectives of this paper are to (i) develop a machine learning methodology that allows to spatially predict POC concentrations in surficial sediments of the NW European continental shelf in an accurate and validated way, (ii) estimate the mass of POC stored in surface sediments and (iii) elucidate relationships between POC concentrations and relevant environmental variables.

Regional setting

The study area is part of the NW European continental shelf adjacent to the North–East Atlantic Ocean. It includes parts of the North Sea, English Channel and Celtic Seas (Fig. 1) and measures approximately 633,000 km². Water depths range between 0 and 1235 m below sea level, with a mean depth of 67 m. Most of the area (98.6%) is continental shelf with depths shallower than 200 m. Within the study area, greater water depths are only found in the Norwegian Trench and on the continental slope off the Scottish West coast.

Data

Response variable

A total of 1111 measurements of the concentration of POC in the sediment fraction <2 mm collected

between 1996 and 2015 were collated from the Centre for Environment, Fisheries and Aquaculture Science (Cefas) in-house data holdings (Mason et al. 2017). Sediment samples were freeze-dried and any material >2 mm was removed. The sediment was subsequently ground and inorganic carbon removed using a sulphurous acid digest. POC concentrations were measured using an elemental analyser. For some sites, more than one measurement of POC concentration existed in the database. These were either repeat measurements (replicates) or measurements for different depth horizons (0–5 cm and 5–10 cm). Average values of POC concentration were calculated in such instances, which reduced the number of records to 1004. Of those, 849 records co-occurred with all available predictor variables (see below) and these were used for further analysis. The statistics of POC concentrations for varying depth intervals are shown in Table 1.

The POC concentrations are reported as a proportion (weight-%). In such a case, an arcsine transformation is advisable (Sokal and Rohlf 1981):

$$Y = \arcsin \sqrt{X}, \quad (1)$$

with X being measured POC as a fraction (ranging from 0 to 1) and Y being the transformed POC concentration. Back-transformation of the predicted values is achieved via:

$$X = (\sin Y)^2 \quad (2)$$

The data set was randomly split into training and test data with a ratio of 2:1 respectively, yielding 566 samples for model training and 283 samples for model testing.

Predictor variables

Predictor environmental variables were initially selected based on their expected relevance to the spatial distribution in POC and their availability. The predictor variables are comprised of bathymetry, Euclidean distance to the nearest shoreline, geographic position (eastings, northings), sediment composition (mud, sand and gravel fraction), earth observation data (chlorophyll-a, depth of the euphotic zone and suspended particulate matter (SPM) concentrations) from the moderate resolution imaging spectroradiometer (MODIS), hydrodynamic model data

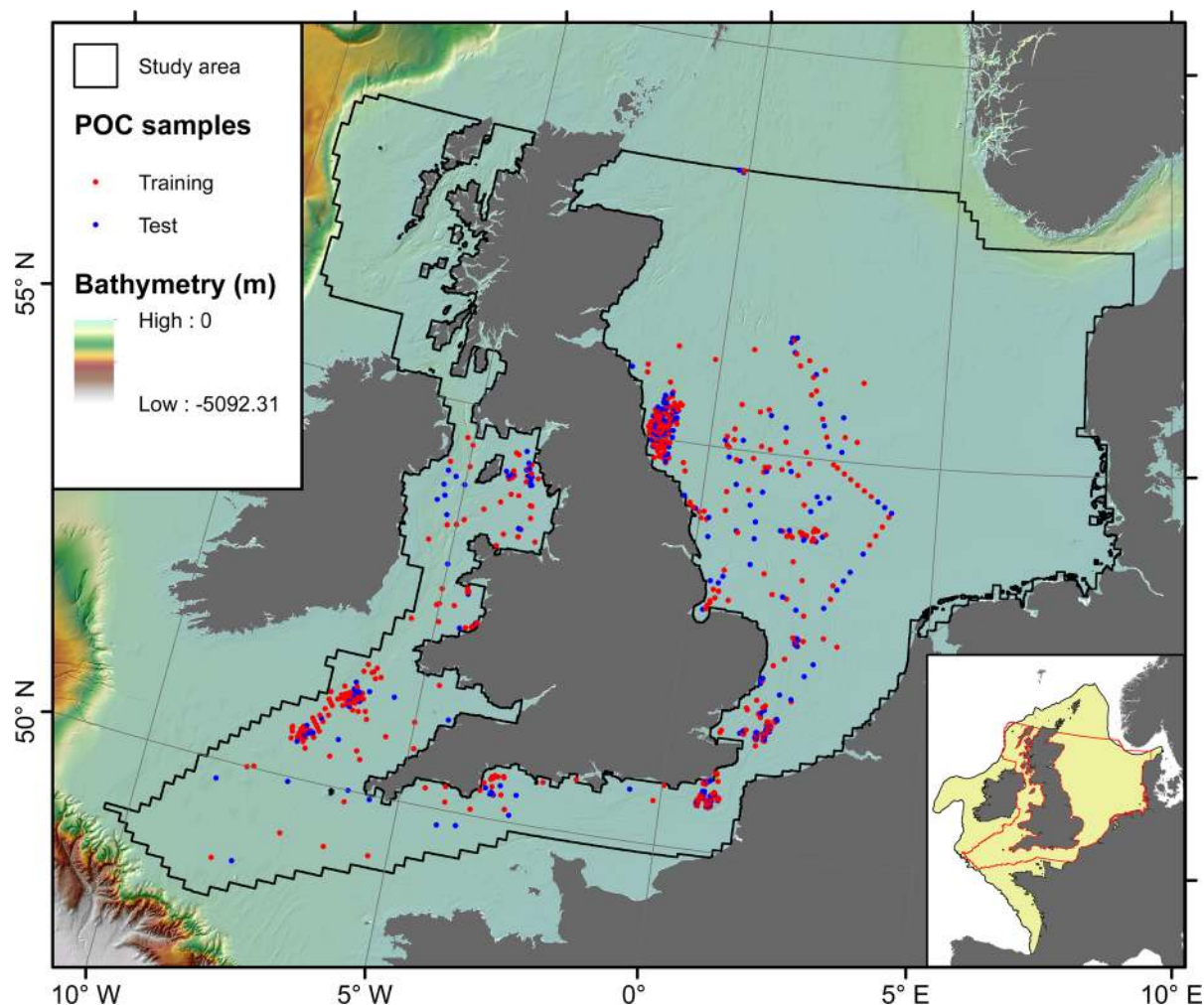


Fig. 1 Location of the study area on the NW European continental shelf (*inset*). Also shown are the locations of POC samples, split into training and test datasets

Table 1 Statistics of POC concentration by depth interval

Depth	N	Mean (%)	SD (%)	Min (%)	Max (%)
Surface (nominally 0–2 cm)	711	0.46	0.51	0.02	4.49
Surface layer; variable depths; max depth = 5 cm	33	0.21	0.17	0.03	0.70
0–5 cm	33	0.22	0.20	0.03	1.00
0–10 cm	72	0.52	0.36	0.07	1.64

(depth averaged mean and peak current speed, peak wave orbital velocity and peak wave-current shear stress), water-column bottom salinity (annual average and range), water-column bottom temperature (annual average and range) and stratification (thermal and salinity).

Methods

Random forest regression

The random forest (RF) prediction algorithm (Breiman 2001) was chosen as the modelling tool for the

analysis because it has shown high predictive accuracy in a number of domains (Prasad et al. 2006; Huang et al. 2012; Mutanga et al. 2012; Oliveira et al. 2012; Huang et al. 2014). RF can be used without extensive parameter tuning, it can handle a large number of predictor variables, is insensitive to the inclusion of some noisy/irrelevant features, makes no assumptions regarding the shape of distributions of the response or predictor variables (Cutler et al. 2007) and is therefore suitable for this analysis. The RF is an ensemble technique that ‘grows’ many regression trees. It is called a random forest because two elements of randomness are introduced. Firstly, each tree is constructed from a bootstrapped sample of the training data. Secondly, only a random subset of the predictor variables is used at each split in the tree building process. This has the effect of making every tree in the forest unique. The underlying principle of the technique is that although each tree in the forest may individually be a poor predictor and that any two trees could give different answers, by aggregating the predictions over a large number of uncorrelated trees, prediction variance is reduced and accuracy improved (James et al. 2013: p. 316). Observations not included in each tree construction, the ‘out-of-bag’ (OOB) samples, are then used to create a form of cross-validated prediction error. RF also provides a relative estimate of predictor variable importance. This is measured as the relative increase in mean squared error associated with each variable when it is assigned random but realistic values and the rest of the variables are left unchanged. The randomForest package (Liaw and Wiener 2002), executed via the Marine Geospatial Ecology Tools v08a.58 (Roberts et al. 2010), was used for the implementation of the model. The forest had 500 trees, and the number of variables tested at each split equalled the number of predictor variables divided by three (rounded down). These are the default settings, which were selected as an increase in the number of trees or variables tested at each split did not lead to improved results (measured as variance explained, see below).

Model validation

The RF implicitly carries out a form of cross-validation (CV) using the OOB observations. This usually gives a reliable measure for real model performance assuming enough trees are grown (Liaw

and Wiener 2002). In addition to this performance indicator, the model constructed here is tested against the test set of observations. For both the CV and the test set, the performance is assessed by calculating the mean of the squared prediction error:

$$MSE_{\hat{y}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

where y are observed and \hat{y} are predicted values. The ‘variance explained’ (VE) by the model is then calculated by taking the ratio of the MSE to the variance (σ^2) of the observed values:

$$VE = 1 - \frac{MSE_{\hat{y}}}{\sigma_y^2} \quad (4)$$

The predictions of the transformed and back-transformed response variable were compared with the observed values from the test set and Pearson product-moment correlation coefficients calculated.

Variable selection

Variable selection reduces the number of predictor variables to a subset that is relevant to the problem. The aims are to reduce redundancy without losing information content and to increase the interpretability of the model. Predictor variables were selected in a two-step approach: Initially, the Boruta variable selection wrapper algorithm (Kursa and Rudnicki 2010) was employed to identify important predictor variables. Wrapper algorithms identify relevant features by performing multiple runs of predictive models, testing the performance of different subsets (Guyon and Elisseeff 2003). The Boruta algorithm creates copies of all variables and randomises them. These so-called shadow variables are added to the predictor variable data set and the RF algorithm is run to compute variable importance scores for predictor and shadow variables. The maximum importance score among the shadow variables (MZSA) is determined. For every predictor variable, a two-sided test of equality is performed with the MZSA. Predictor variables that have a variable importance score significantly higher than the MZSA are deemed important. Likewise, predictor variables that have a variable importance score significantly lower than the MZSA are deemed unimportant. Tentative variables have a variable importance score that is not significantly different from the MZSA. Second, a RF model was run

with the remaining predictor variables to establish the variable importance. Beginning with the most important variable, correlated variables ($|r| > 0.5$) with lower importance were subsequently removed.

Partial dependence plots

Partial dependence plots (Hastie et al. 2009: pp. 369–370) give a graphical depiction of the marginal effect of a predictor variable on the response. They allow to visualise the effect of a predictor variable on the response variable, while averaging out the effects of all other predictors. Partial dependence plots are a useful tool for data exploration. We used the Marine Geospatial Ecology Tools v08a.58 (Roberts et al. 2010) to create partial dependence plots for selected predictor variables.

Estimation of dry bulk density

Porosity (ϕ) of the surficial sediment layer was derived from predicted mud content (Stephens and Diesing 2015) employing an equation from Jenkins (2005):

$$\phi = 0.3805 \cdot C_{mud} + 0.42071, \quad (5)$$

with ϕ and C_{mud} (mud content) both given as dimensionless fractions. The equation is based on data from the Mississippi–Alabama–Florida shelf. By applying this equation to our study area, we assumed that the relationship is not site specific. To test the validity of this assumption, we compared estimates made this way with porosity measurements carried out at 55 stations in the Celtic Sea (Silburn et al. 2017). We also apply an alternative mud-porosity relationship based on the same porosity measurements and predicted mud content (Stephens and Diesing 2015) to assess the differences in the estimation of porosity.

Dry bulk density (ρ_d) of the sediment was then derived from sediment porosity and grain density ($\rho_s = 2650 \text{ kg m}^{-3}$) according to:

$$\rho_d = (1 - \phi)\rho_s \quad (6)$$

All calculations were carried out in ArcGIS v10.1 using the Raster Calculator tool.

Estimation of the total mass of POC stored in surficial sediments

The mass of POC (m_{POC}) per grid cell was calculated by multiplying POC concentration (as a dimensionless

fraction) with dry bulk density (in kg m^{-3}), the sediment depth ($d = 0.1 \text{ m}$) and the area of the grid cell ($A = 250,000 \text{ m}^2$):

$$m_{POC} = POC \cdot \rho_d \cdot d \cdot A \quad (7)$$

A summation of the values of all grid cells yielded the total standing stock of POC.

Calculation of statistics for POC and dry bulk density per sediment class

Statistics of POC concentrations and dry bulk density were calculated for Folk sediment classes as mapped by Stephens and Diesing (2015). For each sediment class, 1000 random samples were generated and the predicted values of POC concentration and dry bulk density extracted. Due to low spatial extent, the number of samples was lower for some classes (gravelly mud, muddy gravel and slightly gravelly sandy mud). The mean, standard deviation, 5th and 95th percentile of POC concentrations and dry bulk density were calculated.

Scaling up to the NW European continental shelf

Results from the study area were scaled up to the NW European continental shelf (Fig. 1, inset) in two ways: First, we calculated m_{POC} by multiplying the area of the NW European continental shelf ($A = 1,111,812 \text{ km}^2$) with the average mass of POC per m^2 and to a depth of 0.1 m as derived for the study area. Second, we employed statistics on POC and dry bulk density and estimates of spatial extent for the different Folk sediment classes (Supplement 1). We used mean values of POC and dry bulk density to derive an estimate of m_{POC} . To account for uncertainty in scaling up our results, we calculated estimates from the extreme low and high end of the distributions of POC and dry bulk density. We used the 5th and 95th percentiles to avoid influence of outliers in either direction.

Results

Variable selection and importance

The Boruta variable selection process indicated that nine variables were deemed important (Table 2). Subsequent removal of correlated variables reduced

the number of variables to six. The final selected variables were mud content, annual average water column bottom temperature, eastings, distance to shoreline, gravel content and peak wave orbital velocity. Figure 2 shows the relative importance of the six variables to prediction accuracy. Mud content in surface sediments is the most important variable in predicting POC, followed by the annual average water column bottom temperature (Table 3).

Model validation

Approximately 75% of the variability of the transformed POC values is explained by the RF model (Table 3). The good agreement between cross-validated and test set statistics indicates that the model is not over-fitted to the training data and it is generalising real patterns in the data. Figure 3 shows the predicted versus the observed values for the transformed POC (left) and POC concentration (right). From this it is apparent that the model tends to slightly under-predict

POC concentration (predicted POC = 0.7896 * observed POC when forced through origin). Pearson product-moment correlation coefficients for the transformed POC and POC concentration are $r = 0.880$ ($n = 275$, $p < 2.2e-16$) and 0.842 ($n = 275$, $p < 2.2e-16$), respectively.

Data exploration

Partial dependence plots (Fig. 4) allow graphical exploration of the relationships between the response variable (transformed POC) and selected predictor variables. Also shown is the relationship between

Table 3 Cross-validation and test set performance

Statistic	Value
MSE (cross-validation)	0.000273
MSE (test set)	0.000273
Variance explained (cross-validation)	74.9%
Variance explained (test set)	77.5%

Table 2 List of predictor variables, results of the Boruta variable selection process and final selection of variables after removal of correlated variables

Predictor variable	Boruta	Final	Source
Bathymetry	Tentative		EMODnet-Bathymetry (http://www.emodnet-bathymetry.eu/), Atrium Oceanwise (2011)
Distance to shoreline	Important	Selected	Calculated
Eastings	Important	Selected	Calculated
Northings	Tentative		Calculated
Mud	Important	Selected	Stephens and Diesing (2015)
Sand	Important		Stephens and Diesing (2015)
Gravel	Important	Selected	Stephens and Diesing (2015)
Chlorophyll-a	Tentative		Gohin et al. (2005)
Depth of euphotic zone	Unimportant		Gohin et al. (2005)
SPM (Winter)	Unimportant		Gohin et al. (2005)
SPM (Summer)	Unimportant		Gohin et al. (2005)
Average current speed	Tentative		Aldridge et al. (2015), Bricheno et al. (2015)
Peak current speed	Important		Aldridge et al. (2015), Bricheno et al. (2015)
Peak wave orbital velocity	Important	Selected	Aldridge et al. (2015), Bricheno et al. (2015)
Peak wave-current stress	Tentative		Aldridge et al. (2015), Bricheno et al. (2015)
Annual average bottom salinity	Tentative		Berx and Hughes (2009)
Annual amplitude bottom salinity	Important		Berx and Hughes (2009)
Annual average bottom temperature	Important	Selected	Berx and Hughes (2009)
Annual amplitude bottom temperature	Tentative		Berx and Hughes (2009)
Stratification salinity	Unimportant		Calculated from Berx and Hughes (2009)
Stratification temperature	Tentative		Calculated from Berx and Hughes (2009)

Fig. 2 Variable importance scores. The importance of predictor variables as indicated by the random forest algorithm. The *x-axis* indicates the relative increase in mean squared error when the variable is assigned random but realistic values, the *y-axis* indicates the variables of the final model

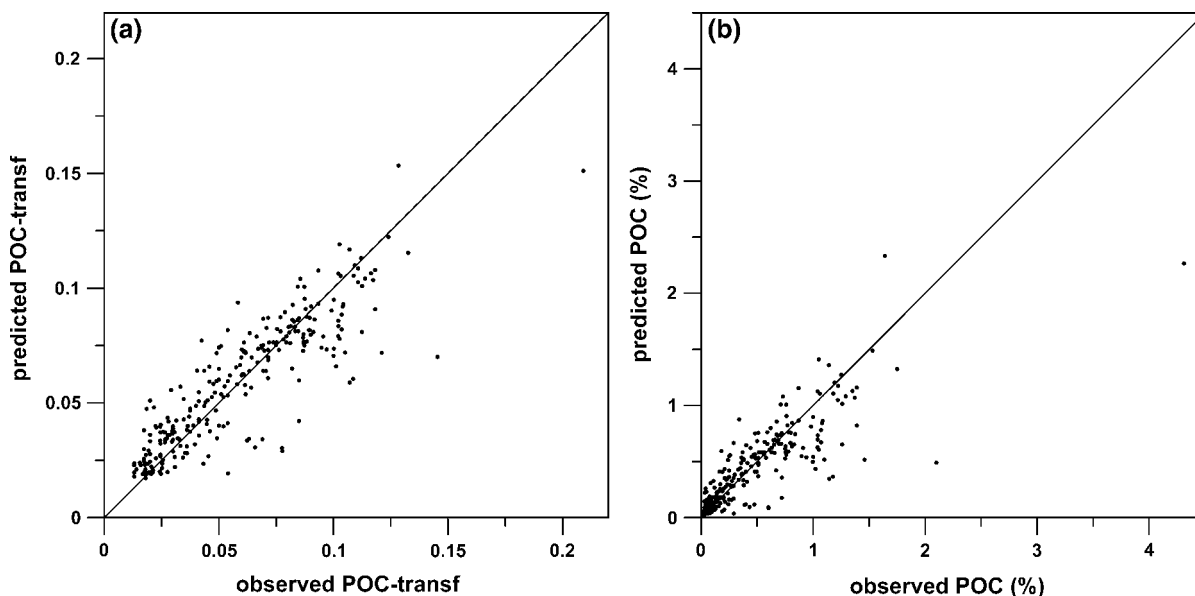
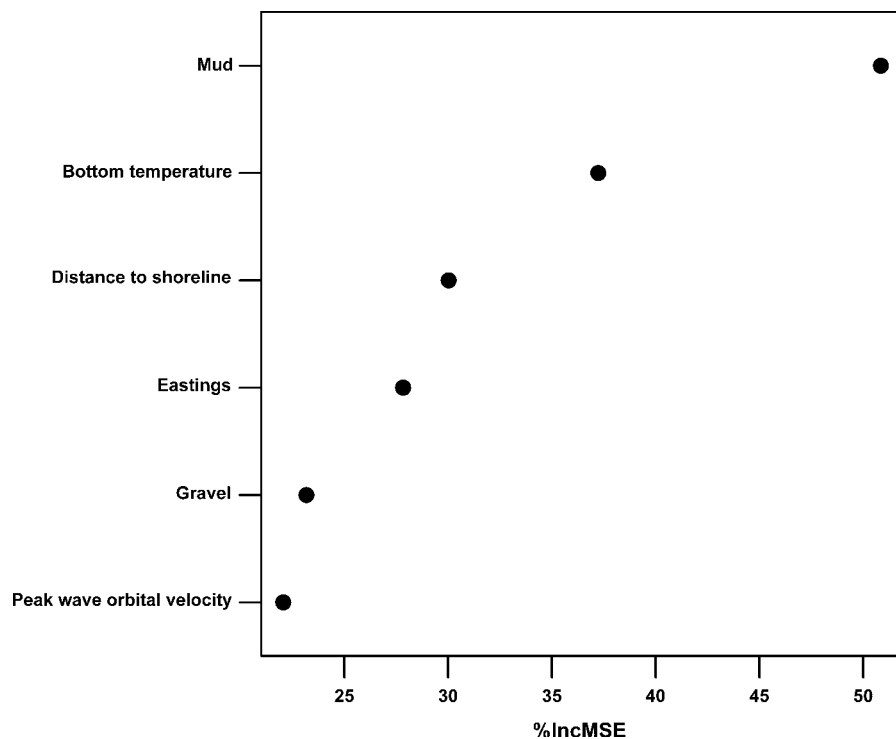


Fig. 3 Observed versus predicted values for transformed POC (a) and POC concentrations (b). The *diagonal line* indicates $y = x$

transformed POC and POC (Eq. 1) to aid the interpretation. Transformed POC, and likewise POC, increases with an increasing mud content; however, this increase is not uniform and levels off towards

higher mud content of >0.5 (i.e. 50 weight-%). For bottom water-column temperatures below approximately $8\text{ }^{\circ}\text{C}$ transformed POC stays constant at around 0.072 (0.52% POC), then drops steeply to 0.063 (0.4%

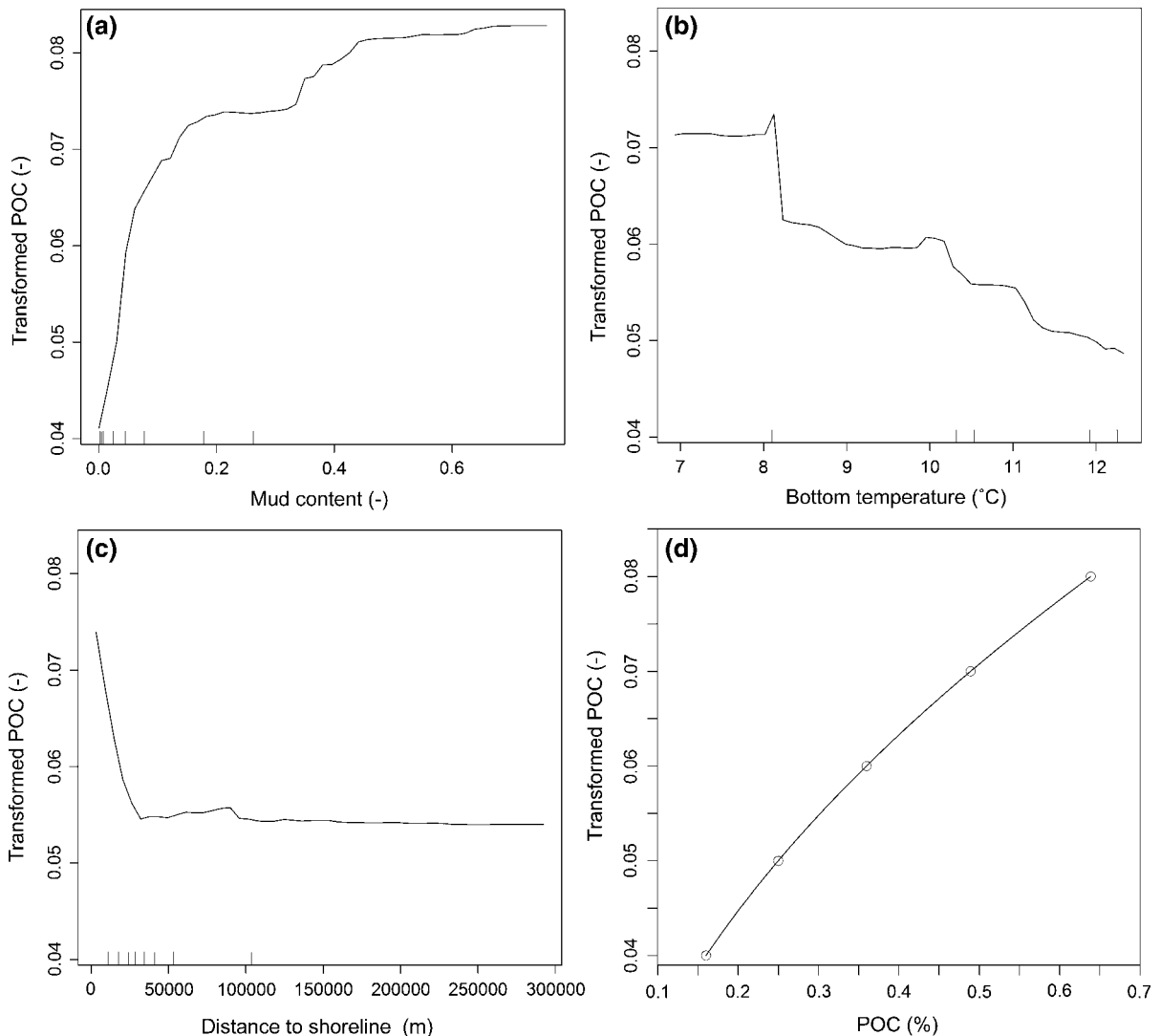


Fig. 4 Partial dependence plots showing the relationships between mud content (a), annual average water column bottom temperature (b), distance to shoreline (c) and transformed POC. Also shown is the relationship between transformed POC and POC (d)

POC). Beyond that, increasing bottom temperatures relate to a broadly linear decrease in transformed POC. Transformed POC values drop sharply with increasing distance from the shoreline up to about 30 km, then stay broadly constant at about 0.055 (0.3% POC).

Estimates of porosity and dry bulk density

Figure 5a shows values of measured porosity plotted against estimates of porosity derived from predicted mud content (Stephens and Diesing 2015) using Eq. 5. Both are strongly (Pearson product-moment

correlation coefficient, $r = 0.803$) and significantly ($p = 1.65e-13$) correlated and points plot along the diagonal line that indicates perfect agreement. Plotting measured porosity against predicted mud content (Fig. 5b) yields a relationship that closely resembles Eq. 5:

$$\phi = 0.4013 \cdot \text{mud} + 0.4265 \quad (8)$$

However, Eq. 5 systematically yields slightly lower estimates for sediment porosity and consequently slightly higher estimates of dry bulk density (Fig. 6). Differences are highest in muddy basins.

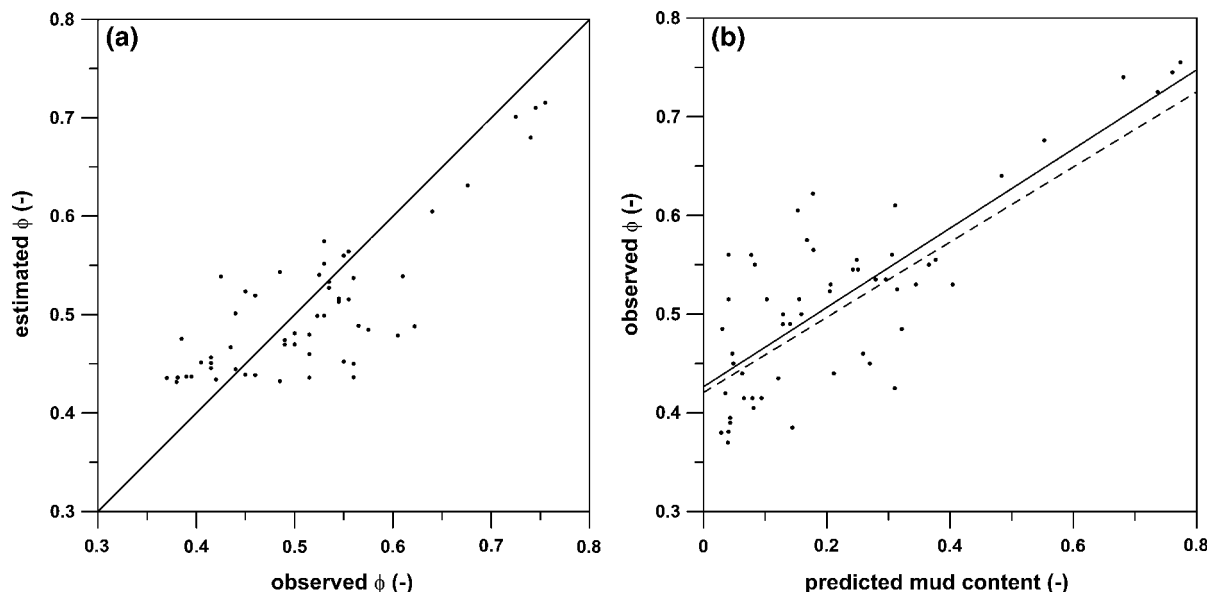


Fig. 5 **a** Observed versus estimated values of sediment porosity (ϕ). The *diagonal line* indicates $y = x$. **b** Predicted mud content versus observed porosity. The *solid line* indicates the best fit linear regression (Eq. 8), the *dashed line* indicates Eq. 5

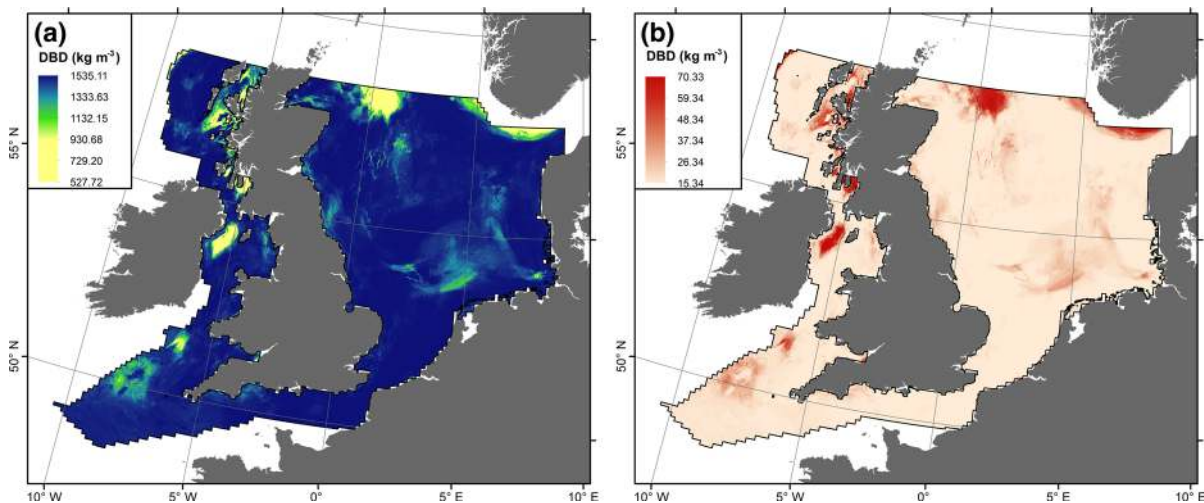


Fig. 6 **a** Dry bulk density based on Eq. 5 for porosity estimation. **b** Difference between dry bulk density based on Eq. 5 for porosity estimation and dry bulk density based on Eq. 8 for porosity estimation

POC in surficial sediments

Spatial patterns of POC are shown in Fig. 7. Highest POC concentrations are associated with the Norwegian Trench, shelf basins and coastal areas around Scotland and north-east England. Lowest concentrations are found in the southern North Sea, on Dogger Bank, the English Channel and in the deeper parts of the Irish Sea. Dry bulk density is negatively correlated

with mud content. Areas of high mud content, mainly the shelf basins, consequently have a low dry bulk density and vice versa (Fig. 6). This means that the highest mass of POC per unit area (m^2) is associated with coastal areas around Scotland and north-east England and the rims of shelf basins. Using Eq. 5 for the estimation of porosity yields a total standing stock of POC in the top 10 cm of shelf sediments of $m_{\text{POC}} = 247.1 \text{ Mt}$ over an area of $632,881 \text{ km}^2$, equal

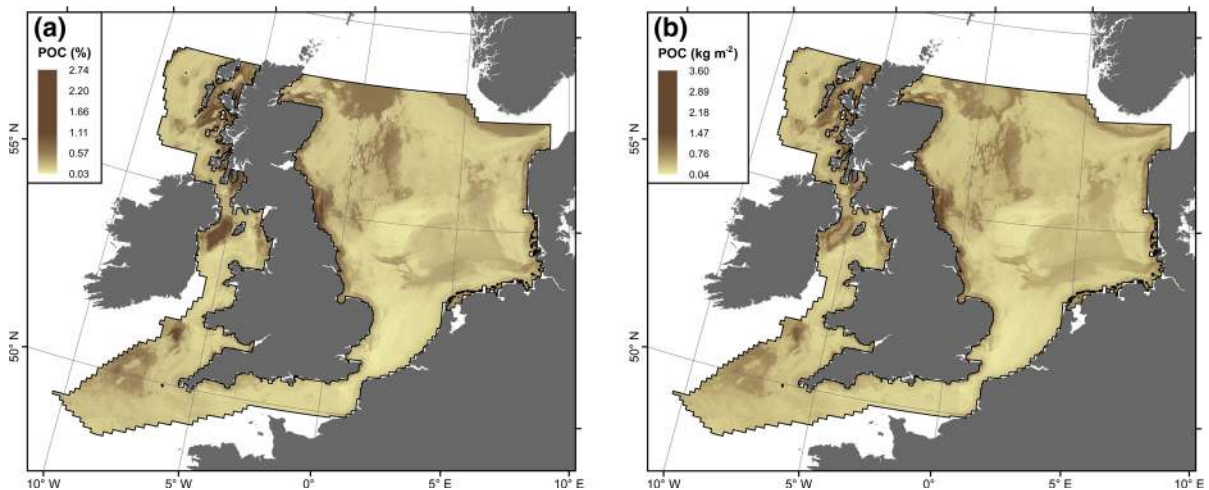


Fig. 7 **a** Predicted concentrations of POC across the study site; **b** Predicted mass of POC per unit area seabed to a depth of 10 cm

to 390.4 t of POC per 1 km². Alternatively, using Eq. 8 for the estimation of porosity yields $m_{\text{POC}} = 243.3$ Mt, which is 3.8 Mt (1.5%) lower than the previous estimate. Due to these limited differences, we will subsequently use estimates of POC derived by utilising Eq. 5.

Statistical values for POC concentrations and dry bulk density by Folk sediment class are summarised in Table 4. Note that no estimates could be made for slightly gravelly mud. Statistics reported for gravelly mud and muddy gravel are based on a very limited number of data points as these sediment types rarely occur within the study site. The highest POC concentrations are associated with gravelly mud, mud and sandy mud. Conversely, gravel and sandy gravel exhibit the lowest POC concentrations. The highest standing stock of POC is associated with sand due to the large area and high dry bulk density. Sediments with the highest POC concentrations provide a minor contribution to the overall POC stock. The total standing stock of POC estimated from mean values per sediment class as reported in Table 4 is $m_{\text{POC}} = 252.2$ Mt, which is slightly higher (2%) than estimated by summing up predicted values as reported above.

Scaling up to the NW European continental shelf

A simple scaling up of m_{POC} to the area of the NW European continental shelf based on the average mass of POC per unit seabed and to a depth of 10 cm yields

$m_{\text{POC}} = 434$ Mt. Based on mean values of POC and dry bulk density (Table 4), we derive $m_{\text{POC}} = 476$ Mt. The lower and upper bounds of our estimates (based on the 5th and 95th percentiles) are calculated to 230 Mt and 882 Mt, respectively. Note that the contribution of slightly gravelly mud was based on statistics for the mud class, as no statistical values could be derived.

Discussion

We have described a quantitative spatial model of POC concentrations in surficial sediments of parts of the NW European continental shelf. The results were produced with a repeatable method and validated with independent, i.e. not used for model building, sample data. The derived statistics indicate that the model is not over-fitted to the training data and more than three quarters of the variance in the response data are explained by the model. Predicted POC concentrations are highly and significantly correlated with measured POC concentrations, although the model appears to slightly under-predict POC concentrations.

Model appraisal and limitations

Every model has limitations as it is a simplification of reality and this is also true in this case. In the following, we will discuss the major limitations of our RF model:

Table 4 Statistical values for POC concentrations and dry bulk density by Folk sediment class

Folk class	Area (km ²)	POC (%)				Dry bulk density (kg m ⁻³)				POC stock (Mt)
		P5	P95	Mean	SD	P5	P95	Mean	SD	
Mud	3080	0.59	1.11	0.88	0.20	536	624	580	29	1.56
Sandy mud	13,656	0.54	1.11	0.78	0.21	646	1011	828	120	8.81
Muddy sand	64,043	0.27	0.92	0.54	0.22	1111	1429	1323	99	45.49
Sand	323,200	0.10	0.50	0.24	0.12	1454	1535	1511	25	116.24
Slightly gravelly sandy mud	122	0.55	0.93	0.67	0.16	789	1030	945	73	0.08
Slightly gravelly muddy sand	5772	0.32	0.82	0.54	0.22	1192	1433	1357	80	4.20
Slightly gravelly sand	92,414	0.07	0.43	0.22	0.11	1467	1534	1512	21	31.13
Gravelly mud	2	0.70	1.69	0.91	0.51	845	1080	1011	102	0.00
Gravelly muddy sand	1638	0.30	0.77	0.49	0.23	1287	1447	1397	51	1.12
Gravelly sand	90,987	0.12	0.44	0.23	0.10	1486	1534	1515	16	32.35
Muddy gravel	1	0.62	0.62	0.62	0.01	1234	1394	1314	125	0.00
Muddy sandy gravel	802	0.16	0.45	0.29	0.10	1438	1510	1482	25	0.34
Sandy gravel	35,222	0.12	0.35	0.19	0.09	1492	1534	1521	13	10.33
Gravel	1942	0.13	0.25	0.18	0.05	1511	1535	1529	8	0.55
Sum	632,881									252.21

Firstly, care was taken to include as many potentially relevant predictor variables as possible. Initially, a conceptual model was developed and potentially relevant variables were identified. Subsequently, those were selected for which full coverage information of the study site was available. These were then subjected to a formal variable selection process. Important variables may be not included in the final model due to two reasons: (i) important additional variables may have been omitted when creating the conceptual model or (ii) it was not possible to obtain data on a variable with full coverage. The latter is especially likely to have occurred as at the time of model development we were not able to source an even more comprehensive set of physico-chemical, chemical and biological variables including water column and porewater nutrients concentrations, oxygen saturation in bottom waters and sediment pH distributions. Furthermore, the inclusion of biological variables such as sediment microbial or faunal communities was beyond the scope of this stage of model development. However, such variables might be accounted for to a certain extent indirectly via correlated variables included in the model.

We believe that further improvements in model performance are generally possible through a better

understanding of causal relationships (e.g. the processes by which temperature or distance to shoreline might affect carbon storage and mineralisation) leading to better conceptual models underpinning the applied statistical approaches, through additional data on relevant predictor variables becoming available as a consequence of advances in parameter measurement and increases in spatial coverage of relevant observations. However, we believe that under the given circumstances our model is a significant achievement as indicated by the fact that 78% of the variance in the response variable is explained by the model.

The availability of predictor and response variables is an external constraint that will influence model performance. However, even if data on a variable exist in principle, they might relate to a certain time interval when the data were collected or for which they were modelled (e.g. hydrodynamic models). Additionally, predictor variables are gridded to a certain spatial resolution. In our model, mismatches between variables do exist both temporally and spatially: e.g. samples were taken from 1996 to 2015, but peak wave orbital velocity was modelled for the period 1999–2008 (Bricheno et al. 2015) and annual average bottom temperature refers to a climatology for the years 1971–2000 (Berox and Hughes 2009). The

implicit assumption of our model therefore is that the response and predictor variables are constant through time. Such an assumption is unlikely to hold. However, we note that the mentioned time intervals do at least partially overlap and it might be assumed that changes in predictor variables are likely to act on longer time scales before they become significant for our model.

With regard to the spatial resolution of the gridded predictor variables we note that four out of six (distance to shoreline, eastings, mud and gravel) existed at the same resolution as the POC model. Annual average bottom temperature was provided at a lower resolution of approximately 10 km and was interpolated to match the 500 m grid of the other predictor variables. We believe that this approach is defensible as bottom temperature is likely to exhibit relatively gradual changes over distances of 10 km. Regarding peak wave orbital velocity, we obtained modelled surface wave properties (significant wave height and zero crossing period) at a spatial resolution of approximately 12 km and calculated orbital velocities at the seabed utilising a high-resolution bathymetry (c. 200 m) in an attempt to better account for small-scale variability in the predictor variable. Again, the assumption was that surface wave properties change gradually over distances of 12 km (at least in water depths beyond the wave base) and that fine-scale variability in orbital velocity is mainly driven by changes in bathymetry, for which high-resolution data existed. We are thus confident that we have accounted for the differing spatial resolutions of the predictor variables in an adequate way.

Factors controlling POC

Mud content, annual average water column bottom temperature, eastings, distance to shoreline, gravel content and peak wave orbital velocity were identified as important predictor variables. Our results demonstrate that mud content in surficial seabed sediments is the most important variable in predicting POC concentrations, which increase with an increase in mud content. Such general relationships have been observed before in North Sea sediments (Cadée 1984; Lohse et al. 1996; de Haas et al. 1997; Trimmer et al. 2005); however, these were often based on a limited number of samples in a spatially restricted area and such studies focused on other sediment mechanisms

rather than regional scale POC-sediment composition relationships directly. Various mechanisms have been proposed, including sorption of organic matter to mineral surfaces and its subsequent concentration in fine-grained sediments with large surface areas (Keil and Hedges 1993), preservation under anoxic conditions in a static situation (e.g. Black Sea and Baltic Sea) and high primary productivity in a dynamic system (coastal upwelling primarily on the western continental margins). Studies of sediment biogeochemistry and carbon cycling have often noted this relationship between sediment type and POC concentration, both in terms of POC driving remineralisation processes and towards an improved understanding of conditions which control POC stocks. Again, these studies have been geographically constrained and focused on carbon or nutrient cycling directly so have not described the regional scale relationships across shelf sea areas.

When critically evaluating the observed link between mud and gravel content and POC concentrations, it is useful to consider the two end members of such sediments and the spectrum of sediments between. A relationship of increasing POC concentrations with mud content is expected given previous findings. Fine sediments are often associated with high natural organic matter loading due to proximity to terrestrial inputs, sedimentary hydrographic environments of low natural disturbance or create an environment where POC that is deposited naturally accumulates due to the diffusional environment that the higher mud percentage creates. This favours a more reducing environment within the sediments and anoxic bacterial processes of remineralisation, which may lead to increased carbon storage in locations where anaerobic degradation processes based on alternative electron acceptors such as nitrogen, iron, manganese or sulphur are dominant. Conversely, substrates with lower mud percentages have more open structures where advective flow in the upper layers deepens the oxic layers and so POC drawn into the sediments is more rapidly respired (Huettel et al. 1996; Huettel and Rusch 2000; Ehrenhauss et al. 2004). The transitions between these two end members which control remineralisation rates through sediment and oxygen are rarely described at a regional scale and so this study bridges the gaps in understanding of the interplay between these end-point mechanisms and its relevance to POC stocks.

Similarly, the observed trend of an increase in POC concentrations with decreases in average bottom water temperature in the range of 7–12 °C covered by the observational data and model agrees with many studies on the remineralisation of POC in marine sediments, both seasonal and geographical (Berner 1980; Middelburg 1989; Burdige 1991). The decrease may be forced by the temperature dependency on bacterial processes (both oxic and anoxic), as remineralisation occurs faster at higher temperatures or alternatively by a negative relationship between primary production (supply level and POC lability) and temperature in the study area. A noticeable step change at ca. 8 °C might potentially correlate to the onset of additional metabolic pathways or non-linear temperature controls involved in POC remineralisation or additionally conditions of these low-temperature shelf areas with disproportionately slow POC remineralisation, for example refractory terrestrial organic matter (TOM) inputs and low oxygen regimes in sea-lochs or deep stratified regions (Glud 2008; Burrows et al. 2014).

The observed influence of the distance to a shoreline on POC concentrations might indicate a change in POC sources: Close to the shore, terrestrial inputs (such as drainage of peats or river catchments) and benthic primary production (e.g. Duarte et al. 2005), where benthic systems lie within the euphotic zone, might dominate measured nearshore POC concentrations. Further offshore, water column primary and secondary productivity or detrital dominated POC sources are more likely to prevail. The input of TOM within these areas can also illustrate the significance of this pool of POC closer to shore. It is a more refractory component of POC and therefore will accumulate in the stock to a greater extent than more labile marine derived POC (Burdige 2005). Chlorophyll-a concentration, which is a proxy for pelagic primary productivity, was not found to be an important predictor. This might be explained by the overriding dominance of other physical or physico-chemical determinants, variabilities in transfer of POC production within the water column and transfer to the bed (sinking, recycling) or significant lateral transport by currents during the sinking process. The derived chlorophyll-a concentrations might also be confounded by POC sources not quantified well by satellite systems (for example deep chlorophyll maxima within stratified regions) or by coloured dissolved organic matter,

which is frequently found in coastal waters but was not explicitly accounted for when deriving products from the MODIS data (e.g. Gohin et al. 2005).

POC stocks in surface sediments of the NW European continental shelf

The estimation of the POC stock (Eq. 7) also depends on robust estimates of dry bulk density. We used Eq. 5 (Jenkins 2005) to estimate sediment porosity and subsequently dry bulk density (Eq. 6). Equation 5 is based on a large data set; however, the samples were collected on the Mississippi–Alabama–Florida shelf. We have demonstrated that this relationship generally holds for our study area and that estimates of the POC stock are hardly affected. We are thus confident that our estimates of the mass of POC stored in surficial sediments of the NW European continental shelf are realistic.

Our results (Figs. 4a, 7a) support the concept that the highest concentrations of POC are associated with muddy sediments. However, these do not always translate into the highest values in terms of mass per unit area, as dry bulk densities of muddy sediments are usually low (Fig. 6). Counterintuitively, muddy sediments (mud, slightly gravelly mud, slightly gravelly sandy mud, sandy mud, gravelly mud) contribute little to the total stock due to their spatially restricted areas and low dry bulk densities. Conversely, sand, slightly gravelly sand and gravelly sand contribute 71% of the POC stock due to high dry bulk densities and widespread occurrence in the study area (Table 4). Hence, our results challenge the view that POC is mainly stored in soft, fine-grained sediments. These results indicate that future research needs to consider previously under-studied coarse-grained sediments with low mud contents.

So far, estimates of POC in the NW European shelf area only exist for localised areas (Serpetti et al. 2012; Burrows et al. 2014). Previous shelf wide approaches to carbon budgets have been carried out but have focussed on annual sedimentation rates in order to generate annual carbon budgets (de Haas et al. 1997, 2002). This latter approach, although key to our understanding of the autotrophy versus heterotrophy of shelf seas and overall carbon cycling, does not address the fundamental question of the overall stock. The results presented here are therefore the first large scale estimate of POC in this area from field samples. Our estimate of 250 Mt carbon stored in surficial

sediments of the study area ($A = 633,000 \text{ km}^2$) contrasts with an estimate of 18.1 Mt stored in the top 10 cm of marine sediments off Scotland covering an area of approximately $470,000 \text{ km}^2$ (Burrows et al. 2014), an order of magnitude lower than our estimate. Given that there is significant spatial overlap between the two study areas, it is likely that the observed discrepancy relates to differences in the approaches taken to derive the carbon stocks. Burrows et al. (2014) derived their estimate by combining average POC concentrations for different sediment types taken from published studies with the estimates of the spatial extent of these sediment types. POC concentrations were assumed to be zero for coarse-grained sediments. This might to some extent explain the discrepancies between the two estimates.

Our figure of 476 Mt (230–882 Mt) for the NW European continental shelf shows the stock to be important for example in comparison to estimates of 9000 and 25,000 Mt of carbon for the whole of Europe's forest vegetation and soils respectively (Kauppi et al. 1992; Dixon et al. 1994). It is important to note that the area of the NW European continental shelf is $1,111,812 \text{ km}^2$ in comparison to $2,830,000 \text{ km}^2$ for forest cover, i.e. $\sim 39\%$ of the surface area for forest. Shelf sea sediments therefore have an aerial storage of ca 3.6% that of terrestrial soils in Europe showing that this store albeit smaller is nonetheless important.

We have identified key variables influencing the distribution of POC but now that we have a baseline of POC we can in the future address variables that influence change in this stock. Long term changes in water column variables such as light penetration (Capuzzo et al. 2015) and changes in anthropogenic physical pressures such as beam trawling intensity (Callaway et al. 2007) may be important in terms of the overall primary production reaching the bed and sediment storage. Additionally, potential changes in terrestrial land-use and catchment management policies are likely to alter inputs of terrestrially derived POC. The maps (Fig. 7) we have presented may therefore be used in the future to show how changes of these and other variables alter the sediment stock. It is likely that the factors identified here in controlling regional POC stock (temperature, terrestrial inputs, sediment composition) will all act in the future via climate change or through human activities to alter these stocks. The implications for spatially restricted muddy substrates with high POC concentrations and for the spatially

extensive sand/gravel substrates with low POC concentrations are likely to be contrasting and different. These changes are important since management decisions on the distribution of fishing effort (Duineveld et al. 2007) or marine aggregate extraction (Desprez 2000) might be able to increase the storage capacity of key areas such as in coastal zones where POC stocks are at their highest. The potential and importance of coastal sediments for carbon sequestration has so far rarely been investigated. Current research has focussed on coastal vegetated habitats such as salt marsh (Chmura et al. 2003), sea grass (Fourqurean et al. 2012) and macroalgae (Krause-Jensen and Duarte 2016). Our results suggest that coastal sediments and their role in carbon storage and sequestration deserve more attention in the future.

Human activities may result in net release of CO_2 due to sediment resuspension, or conversely an increase in POC stocks by shifting sediments which normally process carbon quickly towards regimes which favour POC storage, with consequences for the overall autotrophic/heterotrophic balance of shelf seas. The link between spatial significance of stock levels and factors that interact geographically to produce this stock and the magnitude and direction of change under future conditions or pressures will be key to future management of carbon stocks. The maps of POC stored in surface sediments presented here with the variables (e.g. mud content and bottom temperature) identified in the model provide a starting point to identify controls and vulnerabilities of present day stocks.

The magnitude of the stock on the NW European continental shelf and by extension in other temperate shelf seas shows how important the management of this stock may be globally. Information on POC stocks, their spatial distribution and the factors controlling them will allow improved management of shelf sea areas into the future or prediction of changes in POC storage as a response to climate change. We now need to apply these methods more widely to estimate the stock of all temperate shelf seas so that their importance is quantified and recognised.

Conclusions

We have presented a method that allows to spatially predict and quantify POC stored in surficial sediments

of shelf seas. We show that the surface sediments of the NE European continental shelf are a regionally important POC store. Conversely to expectations, the largest POC stocks are associated with coarse-grained sediments. We highlight the previously overlooked importance of coastal sediments as a store for and site of carbon sequestration. Key variables influencing POC concentrations in shelf sediments of the NW European continental shelf are mud content, annual average bottom temperature and distance to coastline. The resulting model outputs might be useful for a variety of purposes including assessing surface sediment carbon stores by benthic habitat and location, typifying different habitats in terms of their expected surface POC concentration and mass and evaluating their corresponding vulnerability to disturbance. Such potential future uses of the presented results would not only improve our scientific understanding of likely sediment surface POC distributions but could also form valuable information for marine policy development and management decisions.

Acknowledgements This work was conducted under work package 5 ‘Data Synthesis and Management of Marine and Coastal Carbon’ (C6090, NE/K001914/1) of the Shelf Sea Biogeochemistry programme jointly funded by the National Environment Research Council (NERC) and the Department for Environment, Food and Rural Affairs (Defra) of the United Kingdom. The views expressed are those of the authors and do not necessarily represent those of NERC or Defra.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Acharya SS, Panigrahi MK (2016) Evaluation of factors controlling the distribution of organic matter and phosphorus in the Eastern Arabian Shelf: a geostatistical reappraisal. *Cont Shelf Res* 126:79–88. doi:[10.1016/j.csr.2016.08.001](https://doi.org/10.1016/j.csr.2016.08.001)
- Aldridge JN, Parker ER, Bricheno L et al (2015) Assessment of the physical disturbance of the northern European Continental shelf seabed by waves and currents. *Cont Shelf Res* 108:121–140. doi:[10.1016/j.csr.2015.03.004](https://doi.org/10.1016/j.csr.2015.03.004)
- Aller RC (1982) The effects of macrobenthos on chemical properties of marine sediment and overlying water. In: McCall PL, Tevesz MJS (eds) *Animal-sediment relations: the biogenic alteration of sediments*. Springer, Boston, pp 53–102
- Alpaydin E (2010) *Introduction to machine learning*, 2nd edition. MIT Press, Cambridge
- Astrium Oceanwise (2011) Creation of a high resolution digital elevation model (DEM) of the British Isles continental shelf
- Basford DJ, Eleftheriou A, Davies IM et al (1993) The ICES North Sea benthos survey: the sedimentary environment. *ICES J Mar Sci* 50:71–80. doi:[10.1006/jmsc.1993.1008](https://doi.org/10.1006/jmsc.1993.1008)
- Bauer JE, Cai W-J, Raymond PA et al (2013) The changing carbon cycle of the coastal ocean. *Nature* 504:61–70
- Berner RA (1980) *Early diagenesis: a theoretical approach*. Princeton University Press, New Jersey
- Berx B, Hughes SL (2009) Climatology of surface and near-bed temperature and salinity on the north–west European continental shelf for 1971–2000. *Cont Shelf Res* 29:2286–2292. doi:[10.1016/j.csr.2009.09.006](https://doi.org/10.1016/j.csr.2009.09.006)
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Bricheno LM, Wolf J, Aldridge J (2015) Distribution of natural disturbance due to wave and tidal bed currents around the UK. *Cont Shelf Res* 109:67–77. doi:[10.1016/j.csr.2015.09.013](https://doi.org/10.1016/j.csr.2015.09.013)
- Burdige DJ (1991) The kinetics of organic matter mineralization in anoxic marine sediments. *J Mar Res* 49:727–761. doi:[10.1357/002224091784995710](https://doi.org/10.1357/002224091784995710)
- Burdige DJ (2005) Burial of terrestrial organic matter in marine sediments: A re-assessment. *Global Biogeochem Cycles*. 19:GB4011. doi:[10.1029/2004GB002368](https://doi.org/10.1029/2004GB002368)
- Burdige DJ (2007) Preservation of organic matter in marine sediments: controls, mechanisms, and an imbalance in sediment organic carbon budgets? *Chem Rev* 107:467–485. doi:[10.1021/cr050347q](https://doi.org/10.1021/cr050347q)
- Burrows MT, Kamenos NA, Hughes DJ, Stahl H, Howe JA, Tett P (2014) Assessment of carbon budgets and potential blue carbon stores in Scotland’s coastal and marine environment. Scottish Natural Heritage Commissioned Report No. 761, pp 83. http://www.snh.org.uk/pdfs/publications/commissioned_reports/761.pdf
- Cadée G (1984) Macrobenthos and macrobenthic remains on the oyster ground, North Sea. *Netherlands J Sea Res* 18:160–178. doi:[10.1016/0077-7579\(84\)90030-9](https://doi.org/10.1016/0077-7579(84)90030-9)
- Callaway R, Engelhard GH, Dann J et al (2007) One century of North Sea epibenthos and fishing: comparison between 1902–1912, 1982–1985 and 2000. *Mar Ecol Prog Ser* 346:27–43
- Capuzzo E, Stephens D, Silva T et al (2015) Decrease in water clarity of the southern and central North Sea during the 20th century. *Glob Chang Biol* 21:2206–2214. doi:[10.1111/gcb.12854](https://doi.org/10.1111/gcb.12854)
- Chmura GL, Anisfeld SC, Cahoon DR, Lynch JC (2003) Global carbon sequestration in tidal, saline wetland soils. *Global Biogeochem Cycles* 17:12. doi:[10.1029/2002gb001917](https://doi.org/10.1029/2002gb001917)
- Cutler D, Edwards T, Beards K et al (2007) Random Forests for classification in ecology. *Ecology* 88:2783–2792
- de Haas H, Boer W, van Weering TCE (1997) Recent sedimentation and organic carbon burial in a shelf sea: the

- North Sea. *Mar Geol* 144:131–146. doi:[10.1016/S0025-3227\(97\)00082-0](https://doi.org/10.1016/S0025-3227(97)00082-0)
- de Haas H, van Weering TC, de Stigter H (2002) Organic carbon in shelf seas: sinks or sources, processes and products. *Cont Shelf Res* 22:691–717. doi:[10.1016/S0278-4343\(01\)00093-0](https://doi.org/10.1016/S0278-4343(01)00093-0)
- Desprez M (2000) Physical and biological impact of marine aggregate extraction along the French coast of the Eastern English Channel: short- and long-term post-dredging restoration. *ICES J Mar Sci* 57:1428–1438
- Diaz RJ, Rosenberg R (2008) Spreading dead zones and consequences for marine ecosystems. *Science* 80(321):926–929
- Dixon R, Brown S, Houghton R et al (1994) Carbon pools and flux of global forest ecosystems. *Science* 80(263):185–191
- Duarte CM, Middelburg JJ, Caraco N (2005) Major role of marine vegetation on the oceanic carbon cycle. *Biogeochemistry* 2:1–8. doi:[10.5194/bg-2-1-2005](https://doi.org/10.5194/bg-2-1-2005)
- Duineveld GCA, Bergman MJN, Lavaleye MSS (2007) Effects of an area closed to fisheries on the composition of the benthic fauna in the southern North Sea. *ICES J Mar Sci* 64:1–10
- Duplisea DE, Jennings S, Malcolm SJ et al (2001) Modelling potential impacts of bottom trawl fisheries on soft sediment biogeochemistry in the North Sea. *Geochem Trans* 2:112. doi:[10.1186/1467-4866-2-112](https://doi.org/10.1186/1467-4866-2-112)
- Ehrenhauss S, Witte U, Janssen F, Huettel M (2004) Decomposition of diatoms and nutrient dynamics in permeable North Sea sediments. *Cont Shelf Res* 24:721–737. doi:[10.1016/j.csr.2004.01.002](https://doi.org/10.1016/j.csr.2004.01.002)
- Fourqurean JW, Duarte CM, Kennedy H et al (2012) Seagrass ecosystems as a globally significant carbon stock. *Nat Geosci* 5:505–509. doi:[10.1038/ngeo1477](https://doi.org/10.1038/ngeo1477)
- Glud RN (2008) Oxygen dynamics of marine sediments. *Mar Biol Res* 4:243–289. doi:[10.1080/17451000801888726](https://doi.org/10.1080/17451000801888726)
- Gohin F, Loyer S, Lunven M et al (2005) Satellite-derived parameters for biological modelling in coastal waters: illustration over the eastern continental shelf of the Bay of Biscay. *Remote Sens Environ* 95:29–46. doi:[10.1016/j.rse.2004.11.007](https://doi.org/10.1016/j.rse.2004.11.007)
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. Springer, New York
- Huang Z, Nichol SL, Siwabessy JPW et al (2012) Predictive modelling of seabed sediment parameters using multibeam acoustic data: a case study on the Carnarvon Shelf, Western Australia. *Int J Geogr Inf Sci* 26:283–307. doi:[10.1080/13658816.2011.590139](https://doi.org/10.1080/13658816.2011.590139)
- Huang Z, Siwabessy J, Nichol SL, Brooke BP (2014) Predictive mapping of seabed substrata using high-resolution multi-beam sonar data: a case study from a shelf with complex geomorphology. *Mar Geol* 357:37–52
- Huettel M, Rusch A (2000) Transport and degradation of phytoplankton in permeable sediment. *Limnol Oceanogr* 45:534–549. doi:[10.4319/lo.2000.45.3.0534](https://doi.org/10.4319/lo.2000.45.3.0534)
- Huettel M, Ziebis W, Forster S (1996) Flow-induced uptake of particulate matter in permeable sediments. *Limnol Oceanogr* 41:309–322. doi:[10.4319/lo.1996.41.2.0309](https://doi.org/10.4319/lo.1996.41.2.0309)
- James G, Witten D, Hastie T, Tibshirani R (2013) *An introduction to statistical learning*. Springer, New York
- Jenkins C (2005) Summary of the onCALCULATION methods used in dbSEABED. <http://pubs.usgs.gov/ds/2006/146/docs/onCALCULATION.pdf>. Accessed 2 Sep 2016
- Jenness MI, Duineveld GCA (1985) Effects of tidal currents on chlorophyll a content of sandy sediments in the southern North Sea. *Mar Ecol Prog Ser* 21:283–287
- Kauppi PE, Mielikäinen K, Kuusela K (1992) Biomass and carbon budget of European forests, 1971 to 1990. *Science* 256(80):70
- Keil RG, Hedges JI (1993) Sorption of organic matter to mineral surfaces and the preservation of organic matter in coastal marine sediments. *Chem Geol* 107:385–388. doi:[10.1016/0009-2541\(93\)90215-5](https://doi.org/10.1016/0009-2541(93)90215-5)
- Krause-Jensen D, Duarte CM (2016) Substantial role of macroalgae in marine carbon sequestration. *Nat Geosci* 9:737–742
- Kursa M, Rudnicki W (2010) Feature selection with the Boruta Package. *J Stat Softw* 36:1–11
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2:18–22. doi:[10.1159/000323281](https://doi.org/10.1159/000323281)
- Lohse L, Epping EHG, Helder W, Van Raaphorst W (1996) Oxygen pore water profiles in continental shelf sediments of the North Sea: turbulent versus molecular diffusion. *Mar Ecol Prog Ser* 145:63–75. doi:[10.3354/meps145063](https://doi.org/10.3354/meps145063)
- Mason C, Elson J, Lyons B et al (2017) Sediment organic carbon in marine sediments around England and Wales 1996–2015. *Centre Environ Fish Aquac Sci*. doi:[10.14466/CefasDataHub.32](https://doi.org/10.14466/CefasDataHub.32)
- Middelburg JJ (1989) A simple rate model for organic matter decomposition in marine sediments. *Geochim Cosmochim Acta* 53:1577–1581. doi:[10.1016/0016-7037\(89\)90239-1](https://doi.org/10.1016/0016-7037(89)90239-1)
- Middelburg JJ, Levin LA (2009) Coastal hypoxia and sediment biogeochemistry. *Biogeochemistry* 6:1273–1293. doi:[10.5194/bg-6-1273-2009](https://doi.org/10.5194/bg-6-1273-2009)
- Miller C (2004) *Biological Oceanography*, 2nd edition. Oxford University Press, UK
- Mollenhauer G, Schneider RR, Jennerjahn T et al (2004) Organic carbon accumulation in the South Atlantic Ocean: its modern, mid-Holocene and last glacial distribution. *Glob Planet Change* 40:249–266. doi:[10.1016/j.gloplacha.2003.08.002](https://doi.org/10.1016/j.gloplacha.2003.08.002)
- Mutanga O, Adam E, Cho MA (2012) High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int J Appl Earth Obs Geoinf* 18:399–406. doi:[10.1016/j.jag.2012.03.012](https://doi.org/10.1016/j.jag.2012.03.012)
- Neto AA, Mota BB, Belem AL et al (2016) Seismic peak amplitude as a predictor of TOC content in shallow marine sediments. *Geo-Marine Lett* 36:395–403. doi:[10.1007/s00367-016-0449-3](https://doi.org/10.1007/s00367-016-0449-3)
- Oliveira S, Oehler F, San-Miguel-Ayanz J et al (2012) Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For Ecol Manage* 275:117–129. doi:[10.1016/j.foreco.2012.03.003](https://doi.org/10.1016/j.foreco.2012.03.003)
- Pachauri RK, Allen MR, Barros VR, Broome J, Cramer W, Christ R, Church JA, Clarke L, Dahe Q, Dasgupta P, Dubash NK, Edenhofer O, Elgizouli I, Field CB, Forster P, Friedlingstein P, Fuglestvedt J, Gomez-Echeverri L, Hallegatte S, Hegerl G, Howden M, Jiang K, Jimenez Cisneros B, Kattsov V, Lee H, Mach KJ, Marotzke J, Mastrandrea MD, Meyer L, Minx J, Mulugetta Y, O'Brien K,

- Oppenheimer M, Pereira JJ, Pichs-Madruga R, Plattner GK, Pörtner HO, Power SB, Preston B, Ravindranath NH, Reisinger A, Riahi K, Rusticucci M, Scholes R, Seyboth K, Sokona Y, Stavins R, Stocker TF, Tschakert P, van Vuuren, D, van Ypserle JP (2014) Climate Change 2014: synthesis report. Contribution of working groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Pachauri R, Meyer L (eds) IPCC. Geneva, Switzerland, pp 151. http://epic.awi.de/37530/1/IPCC_AR5_SYR_Final.pdf
- Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9:181–199. doi:[10.1007/s10021-005-0054-1](https://doi.org/10.1007/s10021-005-0054-1)
- Roberts JJ, Best BD, Dunn DC et al (2010) Marine geospatial ecology tools: an integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environ Model Softw* 25:1197–1207. doi:[10.1016/j.envsoft.2010.03.029](https://doi.org/10.1016/j.envsoft.2010.03.029)
- Seiter K, Hensen C, Schröter J, Zabel M (2004) Organic carbon content in surface sediments—defining regional provinces. *Deep Sea Res Part I Oceanogr Res Pap* 51:2001–2026. doi:[10.1016/j.dsr.2004.06.014](https://doi.org/10.1016/j.dsr.2004.06.014)
- Serpetti N, Heath M, Rose M, Witte U (2012) High resolution mapping of sediment organic matter from acoustic reflectance data. *Hydrobiologia* 680:265–284
- Silburn B, Sivyer D, Kroeger S et al (2017) Shelf sea biogeochemistry sediment characterisation. *Br Oceanogr Data Cent*. doi:[10.5285/47110529-757c-40b5-e053-6c86abc0eddc](https://doi.org/10.5285/47110529-757c-40b5-e053-6c86abc0eddc)
- Sokal RR, Rohlf FJ (1981) *Biometry*. San Francisco, USA
- Stephens D, Diesing M (2014) A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. *PLoS ONE* 9:e93950. doi:[10.1371/journal.pone.0093950](https://doi.org/10.1371/journal.pone.0093950)
- Stephens D, Diesing M (2015) Towards quantitative spatial models of seabed sediment composition. *PLoS ONE* 10:e0142502. doi:[10.1371/journal.pone.0142502](https://doi.org/10.1371/journal.pone.0142502)
- Stockdale A, Davison W, Zhang H (2009) Micro-scale biogeochemical heterogeneity in sediments: a review of available technology and observed evidence. *Earth-Science Rev* 92:81–97. doi:[10.1016/j.earscirev.2008.11.003](https://doi.org/10.1016/j.earscirev.2008.11.003)
- Trimmer M, Petersen J, Sivyer DB et al (2005) Impact of long-term benthic trawl disturbance on sediment sorting and biogeochemistry in the southern North Sea. *Mar Ecol Prog Ser* 298:79–94
- Willems W, Rees HL, Vincx M et al (2007) Relations and interactions between environmental factors and biotic properties. *ICES Coop Res Rep* 288:69–90