Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

A synthetic map of the northwest European Shelf sedimentary environment for applications in marine science

Robert J. Wilson [1], Douglas C. Speirs [1], Alessandro Sabatino [1], and Michael R. Heath [1]

[1]812 Livingstone Tower, Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH

Correspondence to: Robert Wilson (robert.wilson@strath.ac.uk)

Abstract.

Seabed sediment mapping is important for a wide range of marine policy, planning and scientific issues, and there has been considerable national and international investment around the world in the collation and synthesis of sediment data sets. However, in Europe at least, much of this effort has been directed towards seabed classification and mapping of discrete habitats. Scientific users often have to resort to reverse-engineering these classifications to recover continuous variables such as mud content and median grain size that are required for many ecological and biophysical studies. Here we present a new set of 0.125 by 0.125° resolution synthetic maps of continuous properties of the northwest European sedimentary environment, extending from the Bay of Biscay to the northern limits of the North Sea and the Faroe Islands. The maps are a blend of gridded survey data, and statistically modelled values based on distributions of bed shear stress due to tidal currents and waves, and bathymetric properties. Recent work has shown that statistical models can predict sediment composition in British waters and the North Sea with high accuracy, and here we extend this to the entire shelf and to the mapping of other key seabed parameters. The maps include percentage compositions of mud, sand and gravel; porosity and permeability; median grain size of the whole-sediment and of the sand and the gravel fractions; carbon and nitrogen content of sediments; percentage of seabed area covered by rock; mean and maximum depth-averaged tidal velocity and wave-orbital velocity at the seabed; and mean monthly natural disturbance rates. A number of applications for these maps exist, including species distribution modelling and the more accurate representation of seafloor biogeochemistry in ecosystem models. The data products are available from http://dx.doi.org/10.15129/07bc686e-a354-40de-8c08-372ced7aad64.

1 Introduction

Knowledge of the geographic variation of the sedimentary environment of the seabed is required for a wide variety of marine planning and science tasks. Benthic species have differing sediment requirements and seabed mapping can therefore help identify ecologically distinct habitats (Robinson et al., 2011). Sediment type and wave and tidal regime are important determinants of the rate of natural disturbance of the seabed (Aldridge et al., 2015; Bricheno et al., 2015). The composition of sediments also has a large influence on the consequences of anthropogenic disturbance on the seabed, particularly those due to trawling

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

(Diesing et al., 2013). Mapping the sediment composition and physical environment of the seabed is therefore an integral part of understanding and managing benthic environments.

The northwest European Shelf is one of the world's sea regions most impacted by human activities (Halpern et al., 2014). These impacts are dominated by fishing, and it has been estimated that over 99% of human impact on the seabed is from trawling (Foden et al., 2011). Existing maps of seabed sediments for this region have almost exclusively focused on the territorial waters of individual states (e.g. the British Geological Survey's DigSBS250 product) or subregions (e.g. the North Sea (Basford et al., 1993)). Currently the EU Mesh project is mapping benthic habitat classes across the northwest European Shelf (Vasquez et al., 2015). However, no existing research has mapped the continuous properties of sediments across this region. Here we map key parameters related to the sediment composition and the physical environment of the seabed in an area extending from the Bay of Biscay to the northern limits of the North Sea.

This study was motivated by the need for openly available data sets of the sedimentary environment for parameterizing shelf sea ecosystem models (e.g. Baretta et al. (1995); Blackford (1997); Heath (2012); Ruardij and Van Raaphorst (1995)) and for habitat mapping. Hence, we set out to map mud, sand and gravel percentage compositions and a set of parameters which are of particular relevance for ecosystem modelling and habitat mapping.

A key challenge to mapping seabed sediments across the northwest European Shelf is that sediment data is unavailable across the entire region. In areas with high quality spatial sediment data, it is relatively easy to provide credible maps of sediment composition using statistical interpolation techniques. However, an alternative method is needed where there is poor or no data coverage. Recently, Stephens and Diesing (2015) demonstrated that the mud, sand and gravel percentages of the seabed in British territorial waters and a large part of the North Sea can be predicted using random forest models (Liaw and Wiener, 2002) which have environmental conditions at the seabed as predictors. Further, other work has shown clear relationships between the sediment composition of the seafloor and the energetic regime at the seafloor Porter-Smith et al. (2004); Ward et al. (2015); Heath et al. (2016)

We extend this method by predicting sediment composition of the seabed across the entire northwest European Shelf. However, our approach to mapping differs from that taken by Stephens and Diesing (2015), who only mapped predictions of sediment composition. Since these predictions will be less reliable than interpolated values in regions with good data coverage, we interpolate sediment composition where data is available, and predict it where it is not, and thus create a synthetic picture of the seabed over the northwest European Shelf. Further, we expand the approach of Stephens and Diesing (2015) and map a number of other key parameters including seabed rock, median grain sizes of the whole-sediment, sand and gravel fractions, and porosity and permeability; and the outputs of these models are combined with time series of tidal and wave orbital velocities, and a model of natural disturbance to provide a map of natural disturbance rate on the shelf.

The motivation for the choice of seabed parameters is as follows. Mud, sand and gravel percentages, and rock cover are key determinants of the suitability of habitat for benthic species (Gray, 2002; Thrush et al., 2003) and they strongly influence the median grain size of sediments, which plays a key role in determining natural disturbance of the seabed (Aldridge et al., 2015). Similarly, the median grain size of the sand and gravel fraction play key roles in the properties of sandy and gravelly sediments. The median grain size of the mud fraction is critical for cohesion in muddy regions, but there is insufficient data for this to

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

be mapped credibly. A complete representation of seabed biochemistry in ecosystem models requires knowledge of porosity and permeability, which are related to whole-sediment median grain size (Ruardij and Van Raaphorst, 1995; Lohse et al., 1993). Carbon and nitrogen content of seabed sediments were mapped because of their importance to benthic communities, sediment re-suspension and the potential importance of sediment carbon stores in national carbon inventories (Avelar et al., 2017). Quantitative information about the physical environment on the seabed is necessary for the production of benthic habitat maps (Vasquez et al., 2015) and as a means to compare rates of natural and physical disturbance (Diesing et al., 2013).

2 Methods

2.1 Overview

Our goal was to produce synthesized maps of the sedimentary environment of the northwest European Shelf, which we define to be areas shallower than 500m within the longitude and latitude range 17° W to 9° E and 44° N to 63° N (Fig. 1). There are minimal sediment data for deeper areas within this region and almost all of of the observations are dominated by mud (George and Hill, 2008), and so it is reasonable to assume that these regions are comprised largely of mud and will have negligible natural disturbance rates. Data products were created with a spatial resolution of 0.125° longitude by 0.125° latitude, and are listed in Table 1.

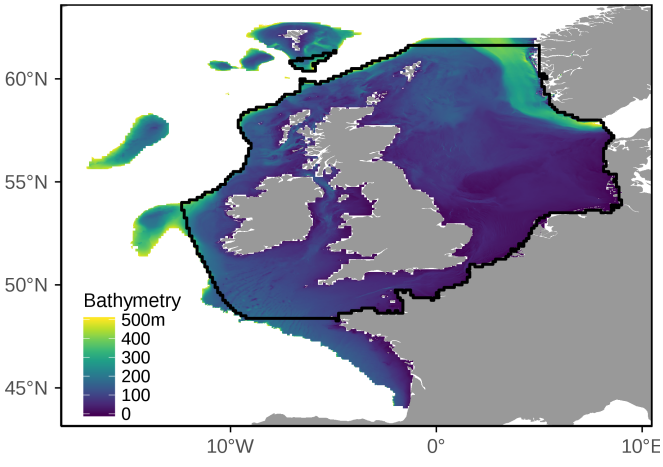

Figure 1. Region where the sedimentary environment was mapped. We defined the northwest European Shelf as the region between 17° W and 9° E and 44° N and 63° N , where bathymetry was less than 500 m. The solid black line demarcates the region where tidal velocities were taken from the Scottish Shelf Model, as described in section 2.2.4.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

Table 1. Data products created at a spatial resolution of 0.125° by 0.125°.

Variable	Description	Unit
Mud, sand & gravel %	Percentage of surface sediment on seabed composed of mud, sand and gravel	%
Whole-sediment D50	Median grain size of the whole sediment	mm
Sand D50	Median grain size of the sand fraction of sediment	mm
Gravel D50	Median grain size of the gravel fraction of sediment	mm
Porosity	Porosity of sediment	
Permeability	Permeability of sediment	m^{-2}
Rock %	Percentage of area made of up of surface rock or rock in top 50cm	%
POC	Carbon content of organic sediment	%
TN	Nitrogen content of organic sediments	%
Orbital velocity	Maximum and mean seabed wave orbital velocity	ms^{-1}
Tidal velocity	Maximum and mean depth-averaged tidal velocity	ms^{-1}
Natural disturbance %	Monthly natural disturbance rate of seabed sediments	

Seabed sample coverage of this shelf region is highly heterogeneous with large expanses of the domain lacking accessible data. Hence, our strategy was to in-fill these voids in the sample coverage with statistically modelled values. The steps involved in mapping the sedimentary environment were therefore as follows:

1. Sediment data from a number of sources (Table 2 were compiled to create a composite data set of mud, sand and gravel percentages, rock cover, carbon and nitrogen content of sediments, and median grain sizes.
2. In areas where we have data, we spatially interpolate the relevant statistic on to the study grid.
3. Using observations we developed random forest (RF) models to predict sediment composition using wave and tidal velocities, bathymetric properties of the seabed and distance from the coast.
4. We then used RF predicted values to in-fill regions of the mapping domain where the observed data density was insufficient for direct gridding.
5. Sediment porosity and permeability at each map grid point were derived from the whole-sediment median grain size using empirically based relationships assembled from literature data.
6. Natural disturbance rates of sediments at each gridded location were then calculated from wave and current bed shear stress and grain size estimates using sediment dislocation theory.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

2.2 Data sources

2.2.1 Raw sediment data and processing

We compiled data on the sediment composition of the seabed from a large number of sources. Our analysis uses the following data: mud, sand, and gravel percentages, rock cover, and the median grain size of the whole-sediment, sand fraction and gravel fraction. The data sources are summarized in Table 2 and the geographic locations where sediment data were available are shown in Fig. 2.

Table 2. Summary of data sources used in sediment analysis. Data sets 1 (http://www.bgs.ac.uk/geoindex/wms.htm) and 8 (http://www.vliz.be/vmdcdata/nsbs/) were open access. Data sets 3, 4, 5, 9 and 10 were available from the Trans-national database of North Sea sediment data (Valerius et al., 2015), which is a collation of data compiled by the EMODnet-Geology (http://www.emodnet.eu/geology), TOLES (http://www.belspo.be/belspo/brain-be/projects/TILES_en.PDF), and AufMod (http://www.kfki.de/de/projekte/aufmod) projects. Data sets 2, 6 and 7 were available from institutional contacts. Data set 11 was downloaded from https://jetstream.gsi.ie/iwdds/index.html.

No.	Source	Sediment percentages	Whole-sediment D50	Sand D50	Gravel D50	Rock presence
1	British Geological Survey	20,857	-	13,289	-	20,744
2	Cefas	3,813	1,879	-	1,865	-
3	Federal Maritime and Hydrographic Agency (Germany)	20,629	-	-	-	-
4	Geological Survey of Denmark and Greenland	475	-	-	-	1,594
5	Geological Survey of the Netherlands	6,346	-	-	-	5,774
6	Geopotenzial Deutsche Nordsee	-	-	-	-	862
7	INFOMAR	1,392	-	-	-	-
8	Marine Scotland	1,214	1,214	-	-	-
9	Marine Scotland(2)	-	898	-	-	-
10	North Sea Benthos Survey	-	219	-	-	-
11	Rikswaterstaat (Ministerie van Infrastructuur en Milieu (Netherlands))	6,114	-	-	-	-
12	Royal Belgian Institute of Natural Sciences	3,433	-	-	-	-
	All data	**64,273**	**4,210**	**13,289**	**1,865**	**28,974**

The British Geological Survey (BGS, 2013) provides mud, sand and gravel percentages, and the median grain size of the sand fraction for locations in most of the United Kingdom's territorial waters. Data were downloaded from the BGS website using the offshore GeoIndex tool (http://www.bgs.ac.uk/geoindex/wms.htm) . The raw BGS data set included 26,259 records of sediment composition. However, to provide a consistent measure of mud, sand and gravel content we only used grab samples. This reduced the total number of records of sediment percentages and sand D50 records to 20,857 and 13,289 respectively.

An extensive data set of surface mud, sand and gravel percentages was compiled for the Trans-national database of North Sea sediments (Valerius et al., 2015). This provides 36,997 records of sediment composition, with data coming from historical

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

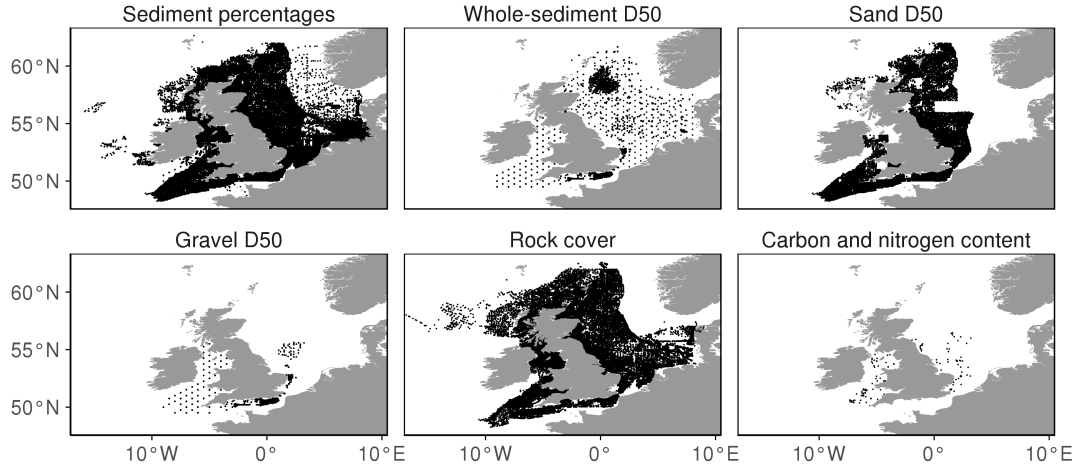

Figure 2. Locations with field estimates of each seabed sediment parameter. Data sources are listed in Table 2.

records of the Federal Maritime and Hydrographic Agency (Germany), the Geological Survey of Denmark and Greenland, the Geological Survey of the Netherlands, Rikswaterstatt (Netherlands), and the Royal Belgian Institute of Natural Sciences.

Records of whole-sediment median grain size were available from the North Sea Benthos Survey (NSBS) (Basford and Eleftheriou, 1988; Basford et al., 1993). NSBS data was downloaded from the http://www.vliz.be/vmdcdata/nsbs/ website. In total there were 219 records of the whole-sediment median grain size. This data is available as separate web pages for each location and we used the *rvest* package in R to convert the html code into columned csv format.

The Centre for Environment, Fisheries and Aquaculture Science (Cefas) provided sediment data which included the mud, sand and gravel percentages and the distribution of sediments by grain size. Data provided by Cefas covered a large part of English and Welsh waters. In total, Cefas provided 3,814 records of mud, sand and gravel percentages and sediment distribution. However, to provide a consistent estimate of sediment type we restricted our analysis to sediments analysed using laser methodology and from the top 10cm of the seabed. This resulted in a total of 1,879 sediment records being used. Cefas did not provide estimates of median grain size. We therefore calculated the median grain size as follows. For the sediment record at each location, a cumulative curve of sediment weight percentage was calculated. We then calculated the median point of this curve and classified this as the median grain size. This was carried out for the entire sediment and also for the gravel fraction.

Two data sets were provided by Marine Scotland. The first included mud, sand and gravel percentages and whole-sediment median grain sizes for a large part of the North Sea. In total, this data set had 1,214 sediment records. The second data set included estimates of the median grain size of the combined mud and sand fraction. These grain size data were not directly usable, so we filed out samples where the percentage of gravel was small enough that the median grain size of the mud/sand fraction was close to that of the whole sediment. To do this we analysed Cefas's data and established that when the whole-sediment D50 is calculated with and without the gravel fraction for sediments with less than 10% gravel there is negligible

difference between the estimates of D50. We therefore used the BGS data set to identify regions where the gravel fraction was below 10%. This was carried out by first calculating the number of BGS observations in each 0.5° N by 1° W cell. We excluded cells with fewer than 10 observations. We then further excluded all cells where more than 10% of the observations had 10% of higher gravel content. In these regions we accepted the Marine Scotland data as a reasonable estimate of the whole-sediment D50.

The Infomar project http://www.infomar.ie is mapping the seabed in Ireland's territorial waters. It has compiled a historical dataset of grab samples which show the surface mud, sand and gravel percentages in many locations in Irish waters. Data was downloaded in shape file format from the https://jetstream.gsi.ie/iwdds/map.jsp website. In total, there were 1,392 records of surface sediment composition.

2.2.2 Rock data

Our aim was to classify locations as non-rock, rock at surface (i.e. approximately top 10cm of sediment, and rock in the approximately top 50cm of sediment and to map the percentage of surface area in each rock classification. Historically, areas have only been mapped in a discrete fashion (e.g. the British Geological Survey's Digirock map (Gafeira et al., 2010)), with relatively broad areas placed in one rock category or another. Further, there are no published large-scale datasets explicitly identifying whether locations have rock at or near the seafloor. We therefore created a composite dataset using historical survey logs for British territorial waters and bore hole records for the territorial waters of the Denmark, Germany and the Netherlands.

The British Geological Survey provides a database of downloadable historical logs of sediment sampling surveys (available from http://www.bgs.ac.uk/geoindex/wms.htm), which provide good spatial coverage for British territorial waters. These logs come in the form of scanned PDFs, and they provide written summaries of each sampling event. The analysis was restricted to corer records, which typically provided sufficient information to determine if there was strong evidence of rock at or near the seabed. Grab sample survey data was initially analyzed, however the use of grab sample records will underestimate rock levels as the grab can return sediment despite there being rock at or close to the surface. We therefore ignored grab samples.

Before analyzing the PDFs we created the following categories for the records: 1) evidence shows there is no rock at the location; 2) written logs are consistent with rock at the surface or rock covered by a thin skin of sediment (approximately 10cm); 3) written logs show that there is probably a significant layer of sediment covering rock; 4) ambiguous or an unreadable record. A Python script was written that will move through each PDF and allow an analyst to classify it. This process was randomized to ensure there was no spatial bias in classification error. In total there were 20,709 initial PDFs. Of these 149 could not be classified as rock or non-rock and were discarded. There were 18,871 records with no evidence of rock, 747 with evidence of rock at or near the surface and 942 records showing rock in the top 50cm.

Borehole records provide reliable records of the rock composition of the seabed and the layers below it. German borehole data is available from the Geopotenzial Deutsche Nordsee project. The http://www.gpdn.de/ website provides visual records of borehole logs at a large number of locations in German territorial waters in the North Sea. A total of 862 records were visually inspected and we found no evidence of rock at or near the surface in any record. The Geological Survey of the Netherlands provides extensive borehole data. These were downloaded as individual text files from the https://www.dinoloket.nl/en/subsurface-

data website. Each text file provided a record of the sediment type in each layer of the borehole in a consistent format. We first identified whether there was rock in the top 50 cm of any of the core records, and found there was none. We therefore found no evidence of surface rock in Dutch waters. The Geological Survey of Denmark and Greenland provide borehole data for Danish waters in the North Sea. Data was available from the http://www.geus.dk website. Each borehole record is available as a separate webpage, and we therefore used the R package *rvest* to save the relevant html code and convert the depth profile of sediment type to csv format. We were then able to identify the sediment type in the top 10 and 50 cm at each location.

2.2.3 Carbon and nitrogen content

Diesing et al. (2017) showed that carbon content of sediment could be credibly predicted based on a series of environmental predictors. Here we take a similar approach to the predictive mapping of carbon and nitrogen content of sediments. Particulate organic carbon (POC) and total nitrogen content (TN) were downloaded from the Cefas Data Hub (http://dx.doi.org/10.14466/CefasDataHub.32) and taken from Serpetti et al. (2012). It is clear that carbon and nitrogen levels in sediment are strongly determined by mud content (Serpetti et al., 2012), and each record of carbon and nitrogen content is associated with a field estimate of mud content. We therefore used mud as a predictor. However, the mud measurements in the Cefas dataset alternate between using laser and sieve methodology and therefore do not provide consistent and comparable estimates of mud content. The Cefas dataset contained 182 sediment samples where the mud content was estimated using laser and sieve methodology, which showed a strong statistical relationship between each measure. We therefore converted each sieve estimate of mud content to a laser-equivalent using a statistical relationship modelled using the lm function in R (Laser-mud = $3.157 \times (\text{Sieve-mud})^{0.7225}$, p-value: < 2.2e-16, $r^2 = 0.93$).

2.2.4 The physical environment

Depth-averaged tidal velocities were calculated as follows. For most of the study region tidal velocities were taken from output from the Scottish Shelf Model, which is an implementation of the unstructured, finite-volume 3D hydrodynamic model FVCOM. The spatial domain of this model covers approximately 80% of our study domain (Fig 1). A full description of the model is provided by De Dominicis et al. (2017), and here we use the same model run described therein. A one year climatology 1990-2014) of atmospheric forcings were used to run the model.

For the rest of the model domain we derived tidal velocities as follows. The Oregon State University Tidal Prediction Software (OTPS) is a well-known open source barotropic tidal model based on the Oregon State University tidal inversion of TOPEX/POSEIDON altimeter data and tide gauge data (Egbert et al., 2010). This model was used to derive the relevant tidal components. The model can be obtained from http://volkov.oce.orst.edu/tides/otps.htmlhttp://volkov.oce.orst.edu/tides/otps.html. We obtained a regional tidal solution using the Oregon State University Tidal Inversion Software (OTIS), with the spatial reso-

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

lution of 1/30 degree. The model satisfies the depth-integrated two dimensional shallow water equations describing momentum balance as follows:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{f} \times \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + \mathbf{F} + \mathrm{A_H} \nabla^2 \mathbf{u} = -g \nabla (\eta - \eta_{\mathrm{EQ}}), \tag{1}$$

and volume conservation

$$-\frac{\partial \eta}{\partial t} = \nabla \cdot (H + \eta) \mathrm{u}, \tag{2}$$

where η is sea surface elevation, $\mathbf{u}$ is the horizontal velocity vector, $\mathbf{f}$ is the Coriolis parameter, F is the fractional damping, $\mathrm{A_H}$ is an eddy coefficient, which is assumed to be constant, H is bathymetry, and η_{EQ} is the equilibrium tide, allowing for the body tide, and tidal self-attraction and loading.

Wave conditions were acquired from the ERA-Interim reanalysis (Dee et al., 2011). Significant wave height, mean wave period and mean wave direction were downloaded from the ECMWF website, http://www.ecmwf.int/en/research/climate-reanalysis/. The ERA-Interim re-analysis has a spatial resolution of approximately 79 km and a temporal resolution of 6 hours. Orbital-velocities at the seabed were calculated using the equations of Soulsby (2006), and the relevant equations are given in this paper's appendix. Bathymetry for the wave and tidal model runs was attained from the high resolution (30 arc-second) General Bathymetric Chart of the Oceans (GEBCO). With the exception of the Scottish Shelf model output we used 2012 as the year for analysis of wave and tidal conditions.

To calculate the bed shear stress we used the equations of Soulsby and Clarke (2005) under combined wave and currents conditions on smooth and rough beds. This set of equations are reproduced in the appendix to this paper. Root mean square shear stress for waves plus currents were used. Calculation of bed shear stress requires the bathymetry, depth-averaged current speed, current direction, significant wave height, wave period and wave direction.

For the statistical modelling of sediment composition we used EMODnet bathymetry data, This has a spatial resolution of 1/8 arc minutes by 1/8 arc minutes and was downloaded from the EMODnet website (http://www.emodnet.eu/bathymetry/). Data processing and calculations were carried out in R using the packages *dplyr* (Wickham and Francois, 2016) and *Rcpp* (Eddelbuettel et al., 2011).

2.3 Spatial gridding and predictive modelling

The synthetic maps of mud, sand and gravel percentages, and rock cover were created as follows. First we identified regions where a statistical interpolation of the relevant parameter would give a reasonable estimate across that region. In other regions we used statistical models to predict the parameter. We assume that the environmental drivers of sediment composition are consistent across space.

Sampling coverage of sediment composition covered almost all of the North Sea, the United Kingdom's territorial waters and parts of Ireland's territorial waters (Fig. 2). Observations almost universally come from sampling programmes that aimed

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

to provide consistent spatial coverage of a specific region (e.g. the North Sea Benthos Survey (Basford et al., 1993)), and parameters can be interpolated in those regions. These regions were selected by creating an alphahull around each unique set of coordinates using the R package *alphahull* (Pateiro-l and Rodr, 2010). An alphahull is a convex envelope around the data points which will exclude areas outside the sampled regions and exclude large holes in the data coverage. Data was first interpolated on to a 1/16° by 1/16° grid and then means were calculated for each 0.125° by 0.125° cell. Parameters were spatially interpolated using bilinear spline interpolation using the interpp function from the R package *akima* (Akima and Gebhardt, 2015).

For areas outside the alphahulls we used random forest (Liaw and Wiener, 2002) models to predict each parameter. This class of model has been used to predict seabed sediments (Diesing et al., 2014; Huang et al., 2012; Li et al., 2011) and carbon and nitrogen content (Diesing et al., 2017). Random forest was developed by Breiman (Breiman, 2001). It is an ensemble based modelling approach that makes no assumptions about the form of the relationships between predictor and response variables, that does not require extensive parameterization, and one that performs internal cross-validation and avoids over-fitting. Random forest takes an ensemble-based approach to regression. This is carried out by first growing a number of regression trees (Loh, 2011). Each tree is composed of a bootstrapped sample from, and of the same size, as the fitting data. Bootstrapped samples are drawn with replacement. Each split in the tree-building process only uses a subset of the predictor variables. Splitting the trees in this way reduces the dominance of individual variables, and thus decorrelates the trees, making the trees less variable and more reliable (James et al., 2013). The average across all trees is then used for predictions. This ensemble averaging makes random forest robust to over-fitting (Breiman, 2001).

The observed mud, sand and gravel percentages summed to 100. However, there is no guarantee separately predicted mud, sand and gravel percentages will sum to 100. We therefore predicted the mud, sand and gravel percentages separately and then a multiplier was applied to each prediction, so that the predictions were adjusted to total to 100. Random forests were created in R using the *ranger* package (Wright and Ziegler, 2017), which is a computationally efficient implementation of random forest for high-dimensional data. The number of trees was set to 2000, with mtry set to 3.

A similar process was carried out for median grain sizes and carbon and nitrogen content. Grain size data was available for large parts of the United Kingdom's territorial waters and some parts of the North Sea, while carbon and nitrogen content was exclusively available in parts of the United Kingdom's territorial waters (Fig. 2). First we used the alphahull approach to identify regions where we can interpolate the parameter. We then used statistical models (discussed in section 2.3.1) to predict each parameter. In each case the sediment percentage maps discussed above were used as predictors in the mapping exercise. Maps and figures were produced using the R package *ggplot2* (Wickham, 2016) and ternary diagrams were produced using the R package *ggtern* (Hamilton, 2016).

2.3.1 Environmental predictors for random forest models and model validation

The environmental predictors used for the random forest models that predicted mud, sand and gravel percentage, rock cover, and carbon and nitrogen content are listed in Table 3. Predictors were chosen based on a review of evidence on the environmental influences on the seabed and the requirement that data was available at the necessary spatial resolution.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Tidal and wave energy levels at the seabed should strongly influence mud, sand and gravel percentages. Large grain sizes require more energy to dislodge from the seabed, and therefore high bed shear stress is associated with increases in average grain size and reductions in mud content (Ward et al., 2015; Heath et al., 2016). There is scarce evidence to determined if seabed composition is influenced by year-round bed shear stress or individual high-energy events. We therefore used mean and maximum annual tidal and wave orbital velocities as predictors in the models of sediment composition, and carbon and nitrogen content. The supply of sediment from river discharges and coastal erosion influences seabed sediment composition and carbon and nitrogen. We therefore included distance from the coast as a model predictor. Distance from the coast was calculated as follows. Shape files of coasts were attained from the Global Self-consistent, Hierarchical, High-resolution Geography (GSHHG) Database (Wessel and Smith, 1996). Distance of each data point from the coast was then calculated using the R package *geosphere* (Hijmans et al., 2012).

Table 3. Predictors used for statistical models for predicting sediment parameters. When mud, sand and gravel percentages and whole-sediment median grain sizes were used as predictors, raw field data were used in the creation of the statistical models, whereas the synthetic maps created in this study were used for model predictions.

		Predictor used?			
Predictor	**Unit**	**Mud/Sand/Gravel/Rock**	**TOC/TN**	**D50s**	**Porosity/Permeability**
Maximum and mean wave orbital velocity	ms^{-1}	Y	Y	-	-
Max. and mean depth-average tidal velocity	ms^{-1}	Y	Y	-	-
Bathymetry	m	Y	Y	-	-
Standard deviation of residual topography	m	Y	Y	-	-
Standard deviation of slope	°	Y	Y	-	-
Distance from coast	km	Y	Y	-	-
Mean annual salinity	-	Y	Y	-	-
Mean annual chlorophyll	$mg\ m^{-3}$	Y	Y	-	-
Mud percentage	%	-	Y	Y	-
Sand percentage	%	-	-	Y	-
Gravel percentage	%	-	-	Y	-
Total D50	%	-	-	-	Y

Smoothness of the seabed will influence seabed disturbance and sediment accumulation, and is likely an indicator of the existence of rocky outcrops. We therefore included measures of seabed roughness as predictors in each random forest. A number of methods exist to quantify the roughness of the seabed (Wilson et al., 2007). However, many of them are not independent of the slope of the seafloor, and are arguably not purely measures of roughness. For example, the standard deviation of bathymetry would classify a steeply sloping, but smooth, part of a continental shelf as being very rough. We therefore used the standard deviation of slope and the standard deviation of the residual topography as predictors in the random forests. Residual topography is the difference between the bathymetry at a specific point and the mean bathymetry within a specified

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

spatial window. The residual topography was calculated using a 25 cell moving window. First the mean bathymetry was calculated within each window. The standard deviation of residual topography (σ) was then calculated using the formula of Cavalli et al. (2008): $\sigma = 1/25\sqrt{\sum_{i=1}^{25}(x_i - x_m)^2}$, where x_i is the bathymetry in a specific cell in the moving window and the respective moving window mean bathymetry. Slope was calculated using the slope function from the R package *SDMTools* (VanDerWal et al., 2014). We then calculated the standard deviation of slope in a similar 25 cell moving window.

The above predictors were used for the mud, sand and gravel percentage and rock cover models. For the models of carbon and nitrogen content we also included chlorophyll, salinity and seabed temperature. Carbon and nitrogen content are influenced by biological activity, and thus should be influenced by primary production levels and temperature at the seabed. The MetO-NWS-REAN-PHYS-bed-daily re-analysis was used for seabed temperature. This data was downloaded from the Copernicus Marine Environmental Monitoring Service website (http://marine.copernicus.eu/services-portfolio/access-to-products/). Daily seabed temperatures from 1995-2014 were interpolated onto each location and an annual climatology was calculated for each model grid point. Climatological (1997-2015) annual mean chlorophyll (mg m^{-3}) data was derived from the level 4 North Atlantic chlorophyll concentration from satellite observations reprocessed data product, available from the Copernicus website. Proximity to river outflows likely influence levels of carbon and nitrogen, and salinity levels act as a proxy for this. We therefore calculated an annual climatological mean (1985-2014) of salinity from the MetO-NWS-REAN-PHYS-monthly-SAL reanalysis product available from the Copernicus website.

Our methodology involves predicting the sedimentary environment in geographically distinct regions. We therefore tested the ability of random forest models to do this credibly by using a cross validation technique involving spatially disaggregated training and test data sets. Spatial disaggregation has been shown to be a reasonable method to avoid the excessive overconfidence that can possibly result from other training and testing methodologies of spatial models (Bahn and McGill, 2013; Roberts et al., 2017). The cross validation method was as follows. We chose to use a spatial blocking method Roberts et al. (2017). This places data into consistently sized and spatially separate blocks or bins. We chose to bin data at a resolution of 1° longitude by 1° latitude. We then used 100 iterations in which each bin was randomly assigned to training and test datasets. In each iteration the random forest was trained using the training data set and this model was then used to predict the relevant parameter using the test data. We therefore evaluated the predictive ability of the model by calculating the mean value of each statistic in the test data for each 0.125 by 0.125 ° cell. The number of observations, and thus observation reliability, in each cell varies significantly. We therefore calculate the weighted r^2 between predicted and observed values in each cell, with the number of observations used as the weighting value. Weighted correlation coefficients were calculated using the function corr from the R package *boot* (Canty, 2002). For the full predictive models over the entire European shelf we retrained the random forests using all available data.

2.3.2 Median grain sizes

Sufficient median grain size data were available to provide a spatial interpolation of whole-sediment D50 in most of the North Sea and large parts of the English Channel and Irish Sea. We therefore interpolated whole-sediment D50 in these regions. This was carried out in the same way as for the distribution of sediment percentages, using bilinear spline interpolation and

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

interpolating solely within the alphahull which surrounds the relevant data points. Outside the alphahulls we predict the relevant D50 using the mud, sand and gravel percentages in the synthetic maps created in this study.

In contrast to the mud, sand and gravel percentages, we chose not to predict median grain sizes using environmental variables. Predicting both the sediment percentages and median grain sizes separately is likely to result in contradictory predictions. For example, a model might predict a much higher median grain size than is possible given the predicted sediment percentages. We therefore chose to create a statistical model which predicts the median grain size using mud, sand and gravel percentages.

The median grain size of the gravel fraction has previously been shown to relate strongly to the mud to sand ratio (Aldridge et al., 2015). We therefore chose to model the whole-sediment D50, the sand D50 and the gravel D50 in relation to mud, sand and gravel percentages. In all cases we used general additive models (GAMs) (Wood, 2006), which marginally outperformed random forests in terms of predictive ability.

Median grain size of the whole-sediment varied by 4 orders of magnitude. Consequently, a GAM model which uses the D50 unaltered was incapable of credibly predicting the D50 for the small grained muddy sediments. We therefore used the following log transformation for the general additive model of the total sediment median grain size, $\log_{10}(\text{D50}) \sim \text{te(Mud, Sand, Gravel)}$, where the interactions between mud, sand and gravel percentages are accounted for using a tensor product smooth (te). This provides realistic predictions of whole-sediment median grain sizes across all sediment types, while also ensuring that predictions for the median grain size were never negative. As with the sediment percentages, data was split into training and testing data. We randomly selected 70% of the data and used it as the training data, and then used the remaining 30% as the test data. Likewise, the final predictive model was created using all of the data.

For the sand and gravel fractions we used a GAM model of the form $\text{D50} \sim \text{te(Mud, Sand, Gravel)}$, with a log-link function to ensure predictions were never negative. Finally, a small number of predictions for the sand and gravel D50 were outside the grain size boundaries for gravel or sand respectively. In these cases we forced the modelled D50 to be the largest or smallest possible grain size where appropriate. General additive models were created using the R package *mgcv* (Wood, 2001).

2.3.3 Porosity and permeability

The porosity and permeability of sediments is quantitatively related to grain size distribution, with coarser grained sediments having lower porosity and higher permeability. We evaluated the relationship between porosity and whole-sediment median grain size by compiling published data (Table 4). Porosity is conventionally expressed as the percentage volume of sediment occupied by void spaces of water. However, some data (Wiesner et al., 1990) expressed water percentage by weight. In this case of we converted the water content data (by weight) to porosity assuming a solid material density of 2.65 g cm^{-3} and a fluid density of 1.025 g cm^{-3}. There was a sigmoidal relationshp between log-transformed porosity and $\log_{10}$-grain size (mm). We therefore fitted a logistic relationship between them using Nelder Mead optimization in the *optim* package in R. This equation is shown below and the parameters are given in Table 5.

$$\log_{10}\text{porosity} = p_1 + p_2 \frac{1}{1 + e^{\frac{-(\log_{10}\text{D50} - p_3)}{p_4}}} \quad (3)$$

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

Table 4. Published literature with porosity estimates. This data was used to statistically model porosity in terms of whole-sediment median grain size.

Reference	Region
Wiesner et al. (1990)	North Sea
Lohse et al. (1993)	North Sea
Ruardij and Van Raaphorst (1995)	North Sea
Serpetti et al. (2012)	North Sea

Table 5. Fitted values and standard errors of the four parameters required for the function relating sediment porosity to median grain size.

Parameter	Fitted value	Standard error
p_1	-0.436	0.023
p_2	0.366	0.050
p_3	-1.227	0.063
p_4	-0.270	0.046

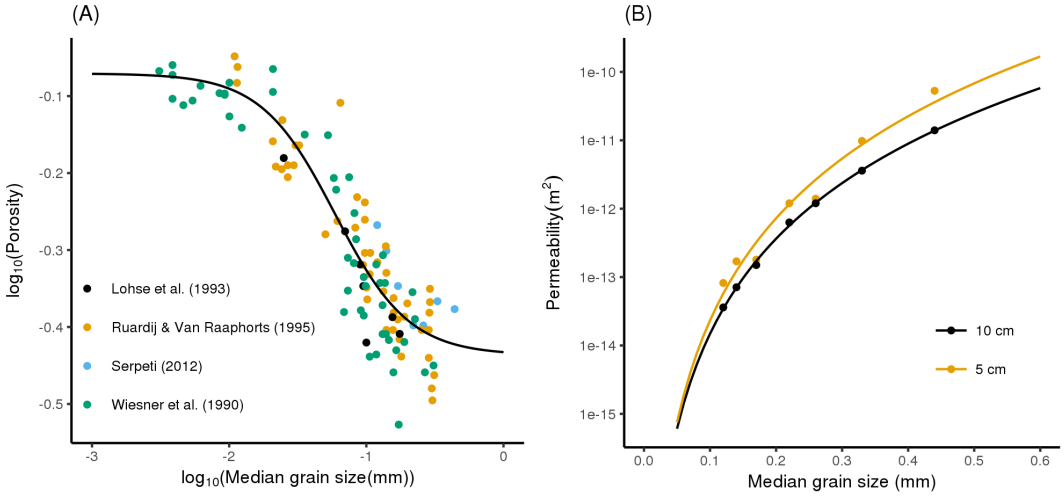

Figure 3. A: Assembled data on sediment porosity and median grain size (filled circles), and the fitted relationship (solid line). B: Annual average permeability m^{-2} of sediments from 7 sites off the north east coast of Scotland; Data from Serpetti (2015).

To our knowledge the best data set available on the relationship between whole-sediment permeability and median grain size is that of Serpetti (2015). This data set covered muddy sand, sand and mixed sediments sampled at approximately monthly intervals over a year at seven sites off the east coast of Scotland. Permeability and median grain size were measured on cores from the upper 5cm and upper 10cm of the seabed at each site. Most sediments are sampled at a depth of 10cm and we therefore

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

chose to only map permeability at this depth. The differences in annual average permeability (m^{-2}) can be explained using a power function of median grain size (D50, mm) (r^2 = 0.999 for 10cm cores). The equation was as follows:

$$\text{Permeability} = 10^{-9.213}\text{D50}^{4.615}(\text{10cm cores})$$

Porosity and permeability were mapped across the study region using the above equations and the synthetic map median grain size.

We then used then porosity estimates and the maps of TOC and TN to derive additional maps of the density of carbon (kg C m^{-2}) and nitrogen (kg N m^2) stored in the surface sediment layer across the shelf. This was derived from the carbon and nitrogen percentages of sediment, and porosity values using the following equation,

$$\begin{aligned}\text{Carbon density}(\text{kgm}^{-2}) &= \text{TOC} \times \text{Sediment depth (m)} \times \text{Dry sediment density (kgm}^{-3}) \times (1 - \text{Porosity}) \\ &= \text{TOC} \times 0.1 \times 2650 \times (1 - \text{Porosity}) \\ \text{Nitrogen density}(\text{kgm}^{-2}) &= \text{TN} \times 0.1 \times 2650 \times (1 - \text{Porosity})\end{aligned}$$

2.3.4 Natural disturbance

We modelled the extent to which the surface layers of the sediment were disturbed by waves and tides during the year. Disturbance was defined as an event which results in physical movement of the surface sediments due to the effects of bed shear stress. We then estimated the average percentage area disturbed per month in each 0.125° by 0.125° cell over our model region. We assumed that sediments are mobilized when the bed shear stress exceeds a critical Shields threshold, and that this threshold is given by the equation provided by Wilcock et al. (2009).

Disturbance could be heterogeneous in space and time within each of our 0.125° by 0.125° cells due to variations in grain size and shear stress. We accounted for this heterogeneity as follows. The bed shear stress on the seabed is determined by the wave and tide conditions and the whole-sediment D50. However, for the mud, sand and gravel fraction the critical threshold is determined by the D50 of the relevant fraction.

We therefore estimate natural disturbance using the following procedure. For each day of the year do the following,

1. Calculate the bed shear stress at each 15 minute time interval using equations shown in this paper's appendix and the whole-sediment D50.
2. Determine the critical threshold at each time step for mud, sand and gravel using the respective D50 and equation A50.
3. Percentage of area disturbed = $(\text{Mud}_{\%} \times \text{Mud}_{\text{dist}}) + (\text{Sand}_{\%} \times \text{Sand}_{\text{dist}}) + (\text{Gravel}_{\%} \times \text{Gravel}_{\text{dist}})$, where Mud_{dist}, $\text{Sand}_{\text{dist}}$, $\text{Gravel}_{\text{dist}}$ denote whether the Shields stress exceeded the critical threshold for mud, sand and gravel respectively

We follow Aldridge et al. (2015) and use a one day time window to classify disturbance events. Monthly disturbance rates are then calculated by aggregating the areas disturbed in each day of the month. It is important to note that modelled disturbance rate ignores the existence of rock at the surface. We are therefore only modelling the disturbance rate in regions with sediment cover.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

3 Results

3.1 Sediment percentages and median grain sizes

Fig. 4 shows the derivation of the synthetic map of sediment percentages. The interpolated map shows that mud (regions with greater than 50% mud) is largely concentrated in the deep Norwegian Trench, an area in the northwestern North Sea, part of the western Irish Sea and in patches on the Scottish west coast. Sandy sediments (greater than 50% sand) dominate in the North Sea, except for those areas with high mud, and a small region on the southeastern English coast with high gravel levels. High gravel levels are seen exclusively in shallow coastal regions, with most of the English Channel having more than 50% gravel.

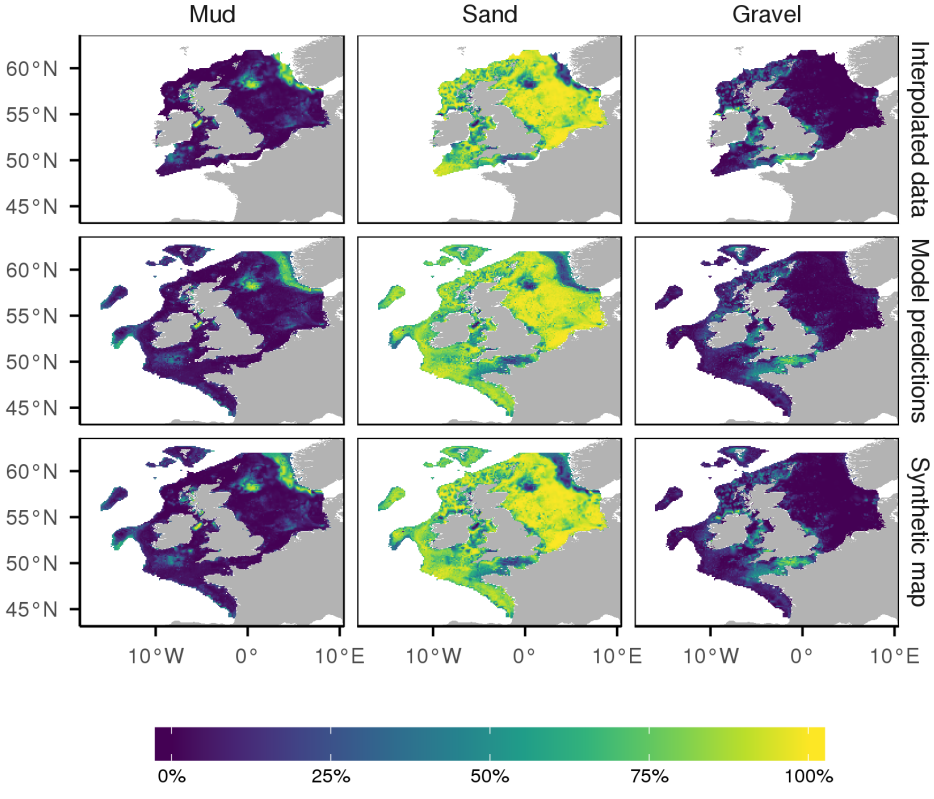

Figure 4. The derivation of the synthetic map of sediment percentages. The interpolated map uses bilinear spline interpolation using sediment data over the region. The random forest map predicts the sediment percentages using a random forest model which relates the percentage to the bed shear stress and the distance to the coast. The synthesized map merges the two, using spatial interpolations where we have data, and the random forest predictions where we do not.

The predictions of the random forest models reproduce the large-scale geographic patterns of sediment composition. The R^2 of the predictions of mean sediment percentage in each 0.125 by 0.125 grid cell on the test data were 0.444, 0.412 and 0.476 for mud, sand and gravel percentages respectively. The models pick up most of the key geographic features revealed by

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

the spatially interpolated map. The high levels of mud in the Norwegian Trench, west of the Isle of Man, and region of the northern North Sea are reproduced. Regions of the western North Sea with relatively high mud levels are also well represented. Similarly, the model predicts the existence of relatively high levels of mud south of Ireland.

The GAM model of whole-sediment D50 created using the training data set performed well against the test data. R^2 was 0.85 on the D50 values and 0.95 on the log10(D50) values. This model had an R^2 of 0.98. Fig. 5 shows the modelled relationship between percentages mud, sand and gravel and the median grain size of the whole-sediment. The GAM model relating the sand D50 to the mud, sand and gravel percentages which was trained on the training data set, had an R^2 0.42 when compared with the test data. The R^2 for the GAM model relating the gravel D50 to the mud, sand and gravel percentages was 0.38.

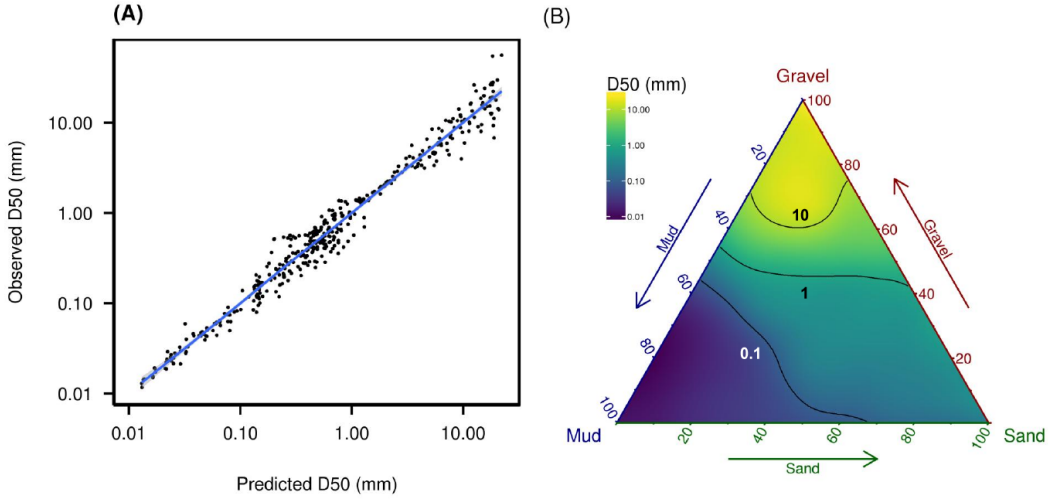

Figure 5. A: Predictions of a GAM model that relates whole-sediment D50 to the mud, sand and gravel percentages. B: Relationship between total sediment median grain size and percentage mud, sand and gravel. The relationship was derived using a general additive model which relates the D50 to the mud, sand and gravel percentage.

Fig. 6 shows the derivation of the synthetic maps of median grain sizes. Whole-sediment median grain size can be interpolated for most of the North Sea, English Channel, and the Irish and Celtic Seas. It varies by approximately three orders of magnitude, with median grain sizes above 10 mm in the gravelly regions in the English Channel and other coastal regions, and median grain sizes close to 0.01 mm in muddy regions such as that in the northwestern North Sea. The median grain size of the sand fraction can be interpolated for most British territorial waters and is highest in regions which are predominantly gravelly. The median grain size of the gravel fraction can only be interpolated for parts of southern British territorial waters, and it is highest in regions of high gravel content.

Fig. 7 shows the derived maps of porosity and permeability. Porosity is similar across most regions, with the exception of the muddy areas in the Norwegian Trench, northwestern North Sea, and the Irish Sea. Permeability varies by 18 orders of

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

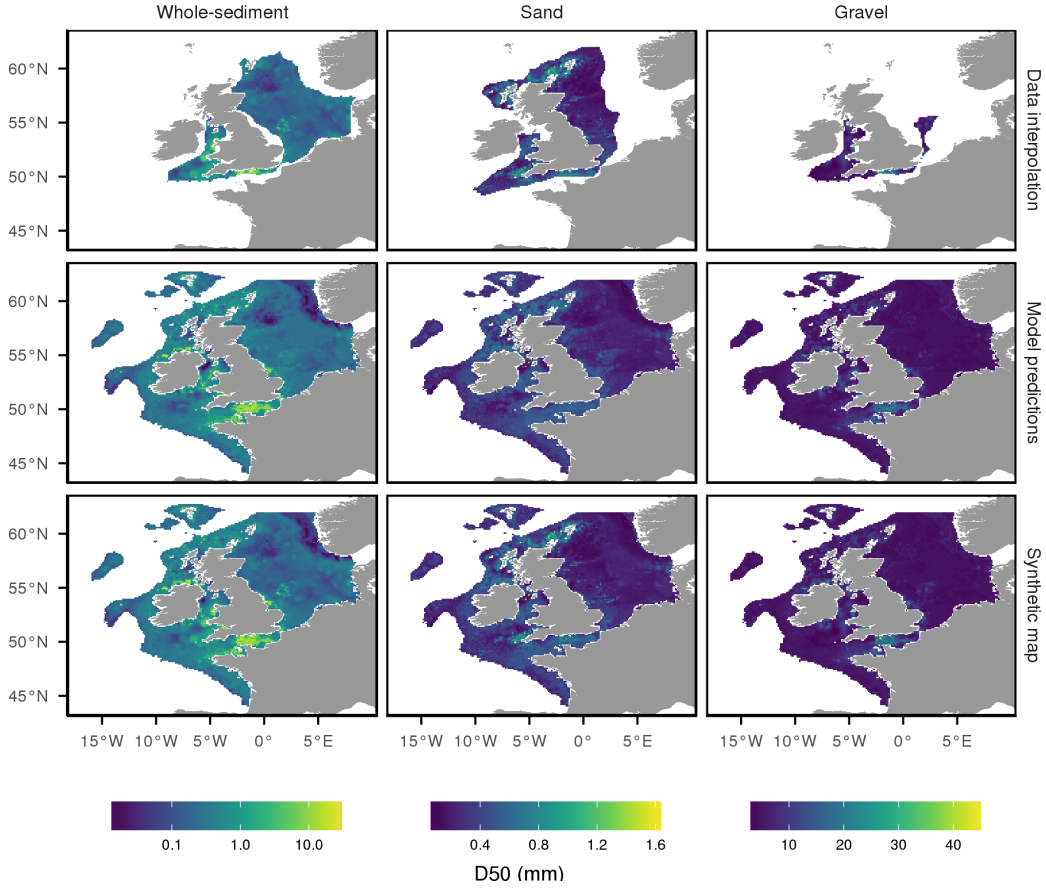

Figure 6. Summary of derivation of the synthetic median grain size maps. Where we have sufficient median grain size data we spatially interpolated a map of D50. In other locations we used the synthetic map of mud, sand and gravel percentages and a GAM model which relates the D50 to the mud, sand and gravel percentages to predict the D50.

magnitude. It is highest in the gravelly regions in the English Channel and some coastal regions, and it is lowest in muddy regions.

The synthetic maps of rock cover are shown in Fig. 8. Observed data indicates that the eastern North Sea is almost entirely free of surface rock. There are large concentrations of surface rock in the English Channel, south west of England, the Bristol Channel and west of the Hebrides Islands on the west coast of Scotland. The predictions of the random forest model of rock provide credible large-scale reproductions of the geographic patterns of sediment composition. Predictions of surface rock and rock in the top 50cm have r^2 of 0.104 and 0.1991 when compared with mean values in each 0.125 by 0.125 grid cell. The random forest predictions (Fig. 8 reproduces the key rock areas shown by the spatially interpolated map. Regions where we rely on predictions are largely rock-free, with the notable exceptions of the high energy English Channel, northwest of France and west of the Faroe Islands.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

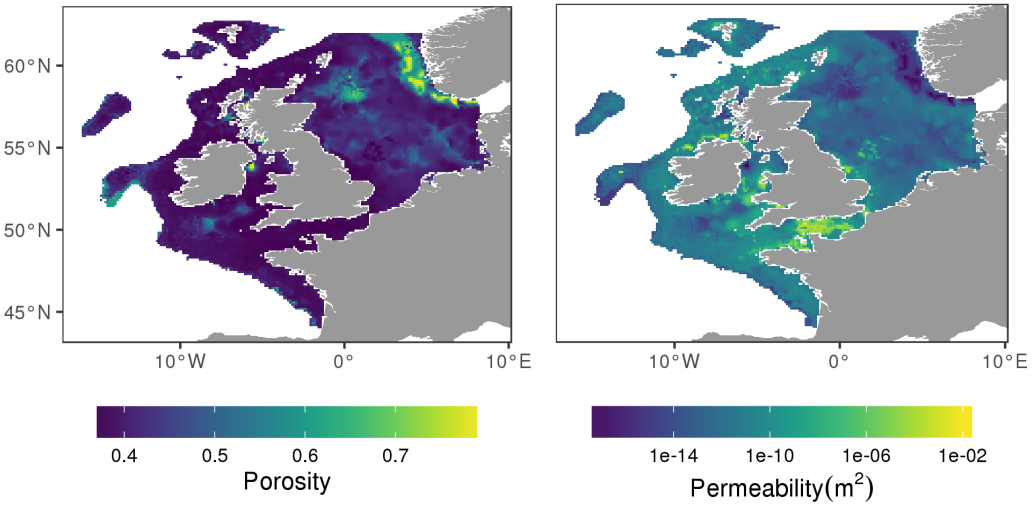

Figure 7. Maps of porosity and permeability. The relationship between porosity and permeability and median grain size was estimated using published field data. We then predicted porosity and permeability using the synthetic map of median grain size.

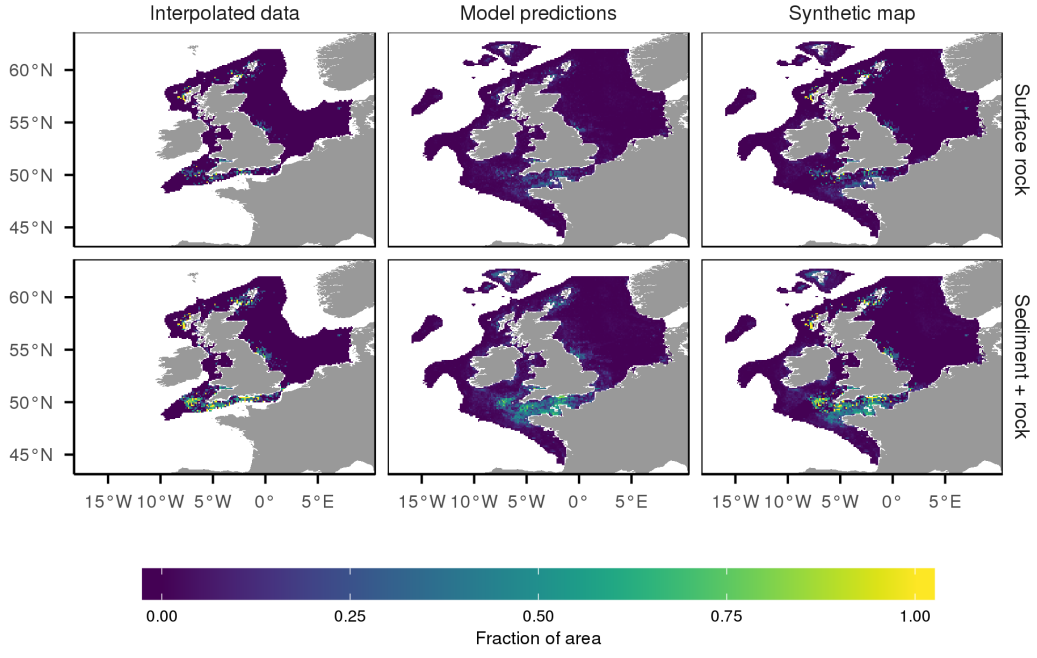

Figure 8. Proportion of area in each rock classification. Areas were classified by whether there was rock at the surface or a surface sediment layer plus rock in the top 50 cm. Historical survey logs and borehole records were first interpolated to provide a map of rock cover where we have sufficient data. Random forests were used to predict rock cover elsewhere using wave and tidal velocities, bathymetry, measures of bathymetry variation and distance from the coast as predictors.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

Mapped carbon and nitrogen content of sediment are shown in Fig. 9. The random forest predictions show close agreement with observations. Across 100 iterations where training and test data was spatially disaggregated, 70% of data in the training data, there was a mean r^2 of 0.59 and 0.70 between predicted and observed POC and PON respectively. Carbon and nitrogen content is largely determined by mud content. Therefore the regions of high carbon and nitrogen content reflect those of large mud content.

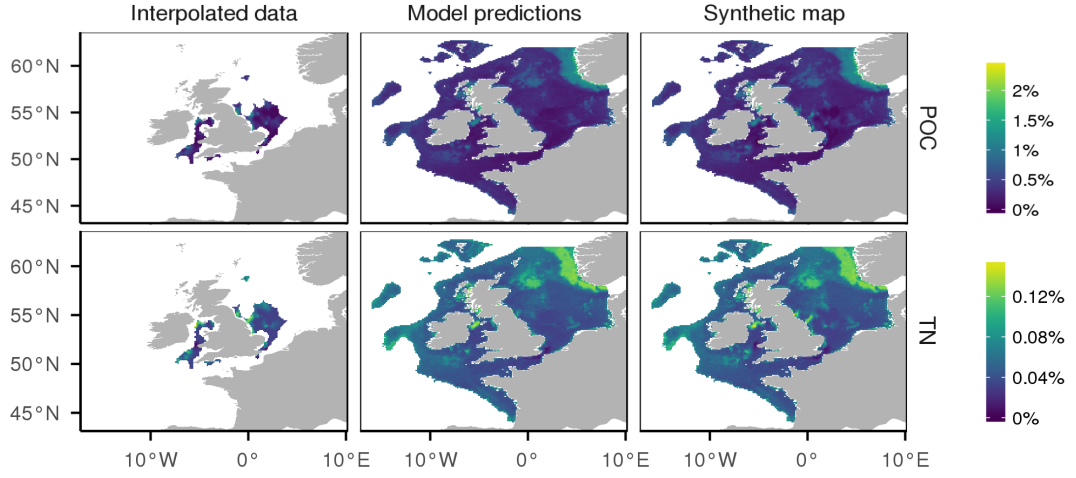

Figure 9. Derivation of the synthetic maps of particulate organic carbon (POC) and total nitrogen (TN) of sediments. Data was interpolated based on field observations in areas with good spatial coverage. In other regions, parameters were predicted using a random forest which had mud content and physical environmental variables as predictors.

3.2 Natural disturbance

Fig. 10 shows modelled natural disturbance in each month. The deep Norwegian Trench is notable for lacking any disturbance year round. Disturbance is highest in the southern North Sea, where sandy regions on the French, Belgian and Dutch coasts see disturbance events almost on a daily basis. There is a notable seasonal pattern in disturbance rates, with summer months seeing lower disturbance rates, which reflects the lower wind and wave regime in this time period.

4 Discussion

The underlying goal of this study was to synthesize large-scale information about the physical environment of the seabed, both in terms of the characteristics of sediment and the wave and tidal regimes which cause disturbance. Using field estimates of the sediment composition of the seabed we were able to map, with high confidence, the sediment composition of the North Sea and British territorial waters, and we were able to make credible statistical predictions of the sediment composition in other regions. The compiled data sets of sediment composition and disturbance regime are, as far as we know, the most extensive

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

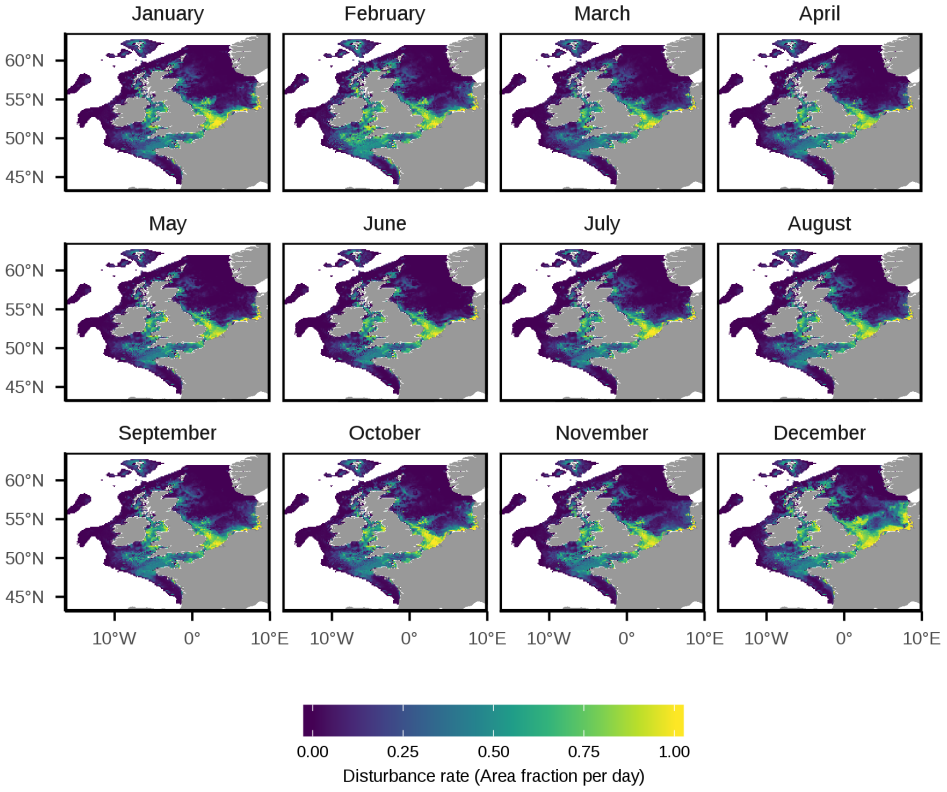

Figure 10. Modelled monthly disturbance rate. A disturbance event was defined as a time when the bed shear stress exceeded the threshold required to move either the mud, sand or gravel portion of the sediment. Monthly disturbance rate was defined as the mean fraction of the total mud, sand and gravel area disturbed per day.

that exist over such a large spatial scale. A number of applications exist for these data sets, including habitat mapping and quantification of anthropogenic disturbance on the seabed.

Habitat mapping requires knowledge of the composition of seabed sediments (Galparsoro et al., 2012), and the maps we produced can be seen as complementary to previous work (e.g. the EU Mesh project (Vasquez et al., 2015)). Existing habitat maps typically use discontinuous categories, and the continuous nature of the maps we have produced may be advantageous for some researchers.

4.1 Limitations and assumptions

A simplifying assumption of our study was that sedimentary environments are in a state of equilibrium or near-equilibrium throughout the European Shelf. However, this is unlikely to be true everywhere. Ward et al. (2015) have argued that the coarser sediments found south east of Ireland were inherited from prior stress regimes. Furthermore, the Irish Sea has linear tidal sand ridges, which are likely relics from an earlier more energetic stress regime (Uehara et al., 2006; Scourse et al., 2009).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Reconstructions of historical tidal conditions on the European Shelf (e.g. Uehara et al. (2006); Neill et al. (2010, 2009)) could potentially be included as model predictors in future modelling studies.

Our maps of rock area are broadly comparable with the existing hard substrate map for British territorial waters produced by the British Geological Survey (Gafeira et al., 2010). Both maps largely draw on historical British Geological Survey logs from seafloor surveys, however the philosophy and motivation of our study differed from that of the British Geological Survey's. The British Geological Survey's was motivated by mapping rocky reef areas for marine conservation planning purposes. Regions were classified as rock or non-rock, which inevitably leads to an overestimation of rock cover if analysts assume that all mapped rock regions are made up exclusively of rock. This is illustrated in the region west of the Hebridean Islands on the west coast of Scotland, where the British Geological Survey historical records show that the seabed is a complex mixture of rock-free seabed and rocky outcrops. However, the British Geological Survey substrate map classifies almost this entire area as rock. This classification was justifiable, given the aim of identifying broad regions that may have rocky reefs. However, in applications such as species distribution modelling this approach is problematic. Classification of mixed habitats as rock could result in a-priori ruling out a large-amount of biological activity, such as fish spawning (Ellis et al., 2012), that are known to take place in these areas. In this case the continuous mapping approach taken by our study is likely more informative.

The confidence in our rock data products is significantly lower than those for mud, sand and gravel percentages. However, this was an expected result and was consistent with existing work (Diesing et al., 2015; Stephens and Diesing, 2015; Downie et al., 2016). The survey data we rely on was explicitly designed to estimate mud, sand and gravel percentages. In contrast, the rock data was based on interpretations of historical survey logs, which creates an additional level of uncertainty. Predictive modelling is also complex due to the array of conditions that appear to result in a rocky seabed. The English Channel and Bristol Channel are rocky due to the strong tidal energy regime, whereas the region west of the Hebridean Islands on the Scottish west coast is relatively rocky due to the existence of rocky outcrops. It is also possible that underlying geology plays a key role in determining rock levels. A previous study that took a similar predictive modelling approach in British waters used information about rock formations as predictors (Diesing et al., 2015; Downie et al., 2016), however we were unable to find any comparable data sets that covered the entire northwest European Shelf.

We excluded the influence of rivers from predictive models because of a lack of large-scale data. However, it is likely that this is a key influence near large estuaries. This can be seen in the high energy Bristol Channel, where there is both a high level of rock and a relatively high level of mud due to the contradictory influences of strong tidal currents and the sediment deposits from the river Severn (McLaren et al., 1993). The influence of river outflows is implicitly captured by the inclusion of distance from the coast as a predictor. For example, there is a large increase in carbon content of sediments close to coasts, which is likely influenced by sediment deposits from rivers. We therefore cannot rule out that certain parameters were over- or under-predicted in coastal regions due to the influence of estuaries.

Previously, Aldridge et al. (2015) mapped the natural disturbance rates of the seabed in English territorial waters and a large part of the North Sea. Despite using different methodology and assumptions our modelled disturbance rates were broadly similar for sandy and muddy regions. However, they deviated drastically for gravelly sediments, in particular in the English Channel. Our model typically predicted disturbance events to occur at least 10 times more often in gravelly sediments compared

with Aldridge et al. (2015). This difference likely results from the assumption for median grain size. A key difference in assumptions between our work and Aldridge et al. (2015) is that we used the whole-sediment D50 as the basis for the bed roughness term in the shear-stress and disturbance rate calculations, whereas Aldridge et al. (2015) used the D50 of the gravel fraction only. Where the seabed sediments are composed of mixtures of mud, sand and gravel fractions this leads to large differences in estimated disturbance rates. It is not clear which approach is more correct. The critical Shields stress calculations are parameterized from empirical studies of sorted sediments. The extent to which these calculations apply to poorly sorted sediments is uncertain. In fact, there is a lack of a theoretical and empirical basis for estimating the suspension and transport dynamics of sediments comprising mixtures of mud, sand and gravel. Sensitivity analysis (unshown) indicated that only using the gravel D50 to determine disturbance in our model resulted in comparable disturbance levels to that in Aldridge et al. (2015). Further research is therefore necessary to reduce the level of uncertainty in our knowledge of disturbance of mixed coarse sediments.

Data availability

The data products listed in Table 1 can been be downloaded in csv, netcdf and ESRI grid format from http://dx.doi.org/10.15129/07bc686e-a354-40de-8c08-372ced7aad64.

Acknowledgements

We thank John Aldridge (Cefas), Simon Greenstreet (Marine Scotland), Mike Robertson (Marine Scotland) and Jennifer Valerius (Federal Maritime and Hydrographic Agency, Germany) for providing access to sediment data. We are grateful to Michela De Dominicis and Judith Wolf (National Oceanography Centre, Liverpool) for providing outputs from the Scottish Shelf Model. British Geological Survey data was provided under Open Government Licence Contains British Geological Survey materials ©NERC. Valuable technical support was provided by Ian Thurlbeck. This paper received funding under the NERC Marine Ecosystem Programme (NE/L003120/1), and from the EPSRC TeraWatt and EcoWatt projects (EP/J010170/1 and EP/K012851/1).

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

Appendix A: Bed shear stress and natural disturbance

The bed shear stress and sediment dislocation rates were calculated by combining the equations of Soulsby and Clarke (2005), Soulsby (2006) and Wilcock et al. (2009).

We calculated the bed shear stress using the equations of Soulsby (2006), who developed equations to estimate the bed shear-stresses under combined waves and current on smooth and rough beds. Bed shear stress was then calculated using the equations of Soulsby and Clarke (2005) who developed equations which calculate combined bed shear stress under waves and currents. Finally, we use the equations of Wilcock et al. (2009) to estimate the critical threshold required for the bed shear stress to cause dislocation of sediment from the seafloor. Model inputs and outputs are given in Tables A1 and A2 respectively.

Table A1. Calculation inputs

Parameter	Symbol	Unit
Water depth	h	m
Water density	ρ	kgm^{-3}
Kinematic viscosity	υ	m^2s^{-1}
Median grain diameter of bed	D_{50}	m
Wave orbital velocity amplitude	U_w	ms^{-1}
Wave period	P_x	s
Current direction	ϕ_c	
Wave direction	ϕ_w	
Sediment density	se_d	kgm^{-3}
Significant wave height	H_s	m
Depth-averaged current speed	$\overline{U}$	ms^{-1}
Angle between wave and current direction	ϕ_d	degrees
Acceleration due to gravity	g	ms^{-2}

Table A2. Calculation outputs

Parameter	Symbol	Unit
Current-alone bed shear stress	τ_c	Nm^{-2}
Wave-only bed shear-stress amplitude	τ_w	Nm^{-2}
Mean wave + current bed shear-stress	τ_m	Nm^{-2}
Maximum wave + current bed shear-stress	τ_{max}	Nm^{-2}
Root-mean-square wave + current bed shear-stress	τ_{rms}	Nm^{-2}

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

A1 Calculation of wave orbital velocities

We calculate the wave orbital velocities using the equations of Soulsby (2006). All waves are assumed to be irregular (spectral)

The the zero-crossing period, T_z (s), and the peak period, T_p (s), the mean wave crossing period T_m (s), are calculated depending on whether there is a JONSWAP spectrum

$$T_z = \begin{cases} P_x/1.28 & \text{if JONSWAP} = 1, \\ P_x & \text{if JONSWAP} = 0. \end{cases} \tag{A1}$$

$$T_p = \begin{cases} P_x & \text{if JONSWAP} = 1, \\ 1.28 \times P_x & \text{if JONSWAP} = 0. \end{cases} \tag{A2}$$

We calculate the natural scaling period, T_n, as follows:

$$T_n = \left(\frac{h}{g}\right)^{0.5} \tag{A3}$$

Soulsby and Smallnan (1986) formulated equations which approximate the wave orbital velocity at the seabed, U_w (1.48×10^{-3} ms^{-1}), as follows.

$$U_w = 0.25 \left(\frac{H_s}{T_n \left(1 + A_{pw}^2\right)^3}\right) \tag{A4}$$

where

$$t_{pw} = \frac{T_n}{T_z} \tag{A5}$$

and

$$A_{pw} = \left[6500 + (0.56 + 15.54 \times t_{pw})^6\right]^{1/6} \tag{A6}$$

A2 Intermediate calculations

We must then calculate a number of intermediate terms for the shear stress calculation.

We then relate the kinematic viscosity υ (m^{2}s^{-1}) to sea water density (1026.96 kgm^{-3}) and kinematic velocity μ (kg (ms)$^{-1}$),

$$\upsilon = \frac{\mu}{\rho} \tag{A7}$$

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

If we define ϕ_c as the current direction and ϕ_w as the wave direction, then we can calculate ϕ_c as the angle between the current and wave direction, as follows

$$\phi = \begin{cases} (\phi_c - \phi_w) * \pi/180 & \text{if } \phi_c > \phi_w, \\ (\phi_w - \phi_c) * \pi/180 & \text{if } \phi_w > \phi_c, \\ 0 & \text{if } \phi_w = \phi_c. \end{cases} \tag{A8}$$

The bed roughness length, z_0 as follows,

$$z_0 = \frac{D_{50}}{12} \tag{A9}$$

The Reynolds number for currents is calculated as

$$Re_c = \frac{\overline{U}h}{\upsilon} \tag{A10}$$

The current drag coefficient for smooth turbulent flow is calculated as follows:

$$CD_s = 0.0001615 \times \exp\left(6 \times Re_c^{-0.08}\right) \tag{A11}$$

The current drag coefficient for rough flow is calculated using the equation:

$$C_{Dr} = \left(\frac{0.4}{\log(h/z_0) - 1}\right)^2 \tag{A12}$$

The wave semi-orbital excursion is calculated using the equation:

$$A = \frac{U_w T_p}{2\pi} \tag{A13}$$

The Reynolds number for waves is calculated as follows:

$$Re_w = \frac{U_w A}{\upsilon} \tag{A14}$$

The wave friction factor for smooth flow is calculated using the equation:

$$f_{ws} = 0.0521 \times Re_w^{-0.187} \tag{A15}$$

The wave friction factor for rough flow is calculated using the equation:

$$f_{wr} = 1.39 \left(\frac{A}{z_0}\right)^{-0.52} \tag{A16}$$

The bed shear stress depending on whether there are currents or waves only or whether there is a combination of waves and currents.

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

Case 1: $\overline{U} > 0$ and $U_w = 0$; current only and no waves

The current-only shear stress is calculated as follows. When $Re_c <= 2000$, we calculate the current bed shear stress under laminar flow equations are.

$$\tau_m = \tau_{max} = \frac{3\rho\upsilon\overline{U}}{h} \tag{A17}$$

When $Re_c > 2000$, we calculate the current bed shear stress under turbulent equations are.

$$\tau_{mr} = \rho C_{Dr}\overline{U}^2 \tag{A18}$$

$$\tau_{ms} = \rho CD_s\overline{U}^2 \tag{A19}$$

$$\tau_m = \tau_{max} = \max(tc_r, tc_s) \tag{A20}$$

Case 4: $\overline{U} = 0$ and $U_w > 0$; waves only, not currents

We calculate wave-only stress as follows.

4.1 Laminar flow: $Re_w <= Rec_{wr}$

$$\tau_m = 0 \tag{A21}$$

$$\tau_m = \tau_{max} = \rho Re_w^{-0.5} U_w^2 \tag{A22}$$

4.2 Turbulent flow: $Re_w > Re_{w,cr}$

$$\tau_{wr} = \frac{1}{2}\rho f_{wr} U_w^2 \tag{A23}$$

with f_{wr} calculated using equation A15.

$$\tau_{ws} = \frac{1}{2}\rho f_{ws} U_w^2 \tag{A24}$$

with f_{ws} calculated using equation A16.

$$\tau_m = 0 \tag{A25}$$

$$\tau_{max} = \max(tw_r, tw_s) \tag{A26}$$

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

Case 5: $\overline{U} > 0$ and $U_w > 0$; combined wave and current flow

First we calculate the critical current Reynolds for transition from laminar to turbulent flow,

$$Re_{c,cr} = 2000 + \left(5.92 \times 10^5 \times Re_w\right)^{0.35} \quad \text{(A27)}$$

The critical wave Reynolds for transition from laminar to turbulent flow,

$$Re_{w,cr} = 1.5 \times 10^5 \quad \text{(A28)}$$

5.1 $RE_c \leq Re_{c,cr}$ and $RE_w \leq Re_{w,cr}$

$$\tau_m = \tau_c \quad \text{(A29)}$$

$$\tau_{max} = \left((\tau_m + \tau_w|\cos(\phi)|)^2 + (\tau_w|\sin(\phi)|)^2\right)^{0.5} \quad \text{(A30)}$$

5.2 Turbulent flow: $Re_c > Re_w$ or $Re_w > Re_{w,cr}$

We must recalculate τ_c and τ_w, since this is an or statement it is possible that one or other of these could appear as laminar flow when estimated independently. Current only component of stress under combined wave and current turbulent conditions

$$tc_r = \rho C_{Dr} \overline{U}^2 \quad \text{(A31)}$$

$$tc_s = \rho C D_s \overline{U}^2 \quad \text{(A32)}$$

$$tc = \max(tc_r, tc_s) \quad \text{(A33)}$$

The wave-only component of stress under combined wave and current turbulent conditions:

$$tw_r = \frac{1}{2}\rho f_{wr} U_w^2 \quad \text{(A34)}$$

$$tw_s = \frac{1}{2}\rho f_{ws} U_w^2 \quad \text{(A35)}$$

$$tw = \max(tw_r, tw_s) \quad \text{(A36)}$$

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

5.2.1 Rough turbulent flow: $\tau_{max,r} > \tau_{max,s}$

Calculate tm_r and $tmax_r$

$$a_r = 0.24 \tag{A37}$$

$$T_1 = \max\left(a_r\left(\frac{f_{wr}}{2}\right)^{0.5}\left(\frac{A}{z_0}\right), 12\right) \tag{A38}$$

$$T_2 = \frac{h}{T_1 z_0} \tag{A39}$$

$$T_3 = \left(C_{Dr}^2 + \left(\frac{f_{wr}}{2}\right)^2\left(\frac{U_w}{\overline{U}}\right)^4\right)^{0.25} \tag{A40}$$

$$A_1 = T_3\frac{\log(T_2) - 1}{2\log(T_1}} \tag{A41}$$

$$A_2 = 0.4\frac{T_3}{\log(T_1)} \tag{A42}$$

$$CD_m = \left(\left(A_1^2 + A_1\right)^{0.5} - A_1\right)^2 \tag{A43}$$

$$CD_{max} = \left(\left(CD_m + T_3\frac{U_w}{\overline{U}}\left(\frac{f_{wr}}{2}\right)^{0.5}|\cos(\phi)|\right)^2 + \left(T_3\frac{U_w}{\overline{U}}\left(\frac{f_{wr}}{2}\right)^{0.5}|\sin(\phi)|\right)^2\right)^{0.5} \tag{A44}$$

$$\tau_r = \rho CD_m\overline{U}^2 \tag{A45}$$

$$\tau_{max} = \rho CD_{max}\overline{U}^2 \tag{A46}$$

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

Shields number calculation

We calculate the bed shear velocity as

$$u_{star} = \left(\frac{\tau_{max}}{\rho}\right)^{0.5} \tag{A47}$$

The Reynolds particle number is calculated as follows:

$$Ren = u_{star} \times \frac{D_{50}}{\upsilon} \tag{A48}$$

The Shields stress

$$Tc_S = \frac{\tau_{max}}{(se_d - \rho) \times g \times D_{50}} \tag{A49}$$

The critical shields stress is calculated as follows from the empirical relationship shown in Wilcock et al. (2009):

$$TcS_{cr} = (0.105 \times Ren^{-0.3})) + 0.045 \times \exp(-35 \times Ren^{-0.59})) \tag{A50}$$

RMS shear stress for waves and currents

The root-mean-square shear stress is calculated as follows:

$$\tau_{rms} = \left(\tau_m^2 + \tau_w^2\right)^{0.5} \tag{A51}$$

Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-88
Manuscript under review for journal Earth Syst. Sci. Data
Discussion started: 18 August 2017

Open Access Earth System Science Data Discussions

References

Akima, H. and Gebhardt, A.: akima: Interpolation of Irregularly and Regularly Spaced Data. R package version 0.5-12, 2015.

Aldridge, J. N., Parker, E. R., Bricheno, L. M., Green, S. L., and van der Molen, J.: Assessment of the physical disturbance of the northern European Continental shelf seabed by waves and currents, Continental Shelf Research, 108, 121–140, doi:10.1016/j.csr.2015.03.004, 2015.

Avelar, S., van der Voort, T. S., and Eglinton, T. I.: Relevance of carbon stocks of marine sediments for national greenhouse gas inventories of maritime nations, Carbon Balance and Management, 12, 10, doi:10.1186/s13021-017-0077-x, 2017.

Bahn, V. and McGill, B. J.: Testing the predictive performance of distribution models, Oikos, 122, 321–331, doi:10.1111/j.1600-0706.2012.00299.x, 2013.

Baretta, J. W., Ebenhöh, W., and Ruardij, P.: The European regional seas ecosystem model, a complex marine ecosystem model, Netherlands Journal of Sea Research, 33, 233–246, doi:10.1016/0077-7579(95)90047-0, 1995.

Basford, D. and Eleftheriou, A.: The benthic environment of the North Sea (56° to 61°N), Journal of Marine Biology Association UK, 68, 125–141, doi:10.1017/S0025315400050141, 1988.

Basford, D. J., Eleftheriou, A., Davies, I. M., Irion, G., and Soltwedel, T.: The ICES North Sea benthos survey: the sedimentary environment, ICES Journal of Marine Science, 50, 71–80, 1993.

BGS: BGS Legacy Particle Size Analysis uncontrolled data export, Tech. rep., British Geological Survey, 2013.

Blackford, J. C.: An analysis of benthic biological dynamics in a North Sea ecosystem model, Journal of Sea Research, 38, 213–230, doi:10.1016/S1385-1101(97)00044-0, 1997.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, doi:10.1023/A:1010933404324, 2001.

Bricheno, L. M., Wolf, J., and Aldridge, J.: Distribution of natural disturbance due to wave and tidal bed currents around the UK, Continental Shelf Research, 109, 67–77, doi:10.1016/j.csr.2015.09.013, 2015.

Canty, A. J.: Resampling methods in R: the boot package, R News, 2, 2–7, 2002.

Cavalli, M., Tarolli, P., Marchi, L., and Dalla Fontana, G.: The effectiveness of airborne LiDAR data in the recognition of channel-bed morphology, Catena, 73, 249–260, doi:10.1016/j.catena.2007.11.001, 2008.

De Dominicis, M., O'Hara Murray, R., and Wolf, J.: Multi-scale ocean response to a large tidal stream turbine array, Renewable Energy, doi:10.1016/j.renene.2017.07.058, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

Diesing, M., Stephens, D., and Aldridge, J.: A proposed method for assessing the extent of the seabed significantly affected by demersal fishing in the Greater North Sea, ICES Journal of Marine Science, 70, 1085–1096, doi:10.1093/icesjms/fst048, 2013.

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D.: Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches, Continental Shelf Research, 84, 107–119, doi:10.1016/j.csr.2014.05.004, 2014.

Diesing, M., Green, S., Stephens, D., Cooper, R., and Mellett, C.: Semi-automated mapping of rock in the English Channel and Celtic Sea, p. 19, doi:569, 2015.

Diesing, M., Kroger, S., Parker, R., Jenkins, C., Mason, C., and Weston, K.: Predicting the standing stock of organic carbon in surface sediments of the North-West European continental shelf, Biogeochemistry, pp. 1–18, doi:10.1007/s10533-017-0310-4, 2017.

Downie, A. L., Diesing, R. K., and Cooper, S. L. &.: Semi-automated mapping of rock in the North Sea, doi:592, 2016.

Eddelbuettel, D., François, R., Allaire, J., Chambers, J., Bates, D., and Ushey, K.: Rcpp: Seamless R and C++ integration, Journal of Statistical Software, 40, 1–18, 2011.

Egbert, G. D., Erofeeva, S. Y., and Ray, R. D.: Assimilation of altimetry data for nonlinear shallow-water tides: Quarter-diurnal tides of the Northwest European Shelf, Continental Shelf Research, 30, 668–679, doi:10.1016/j.csr.2009.10.011, 2010.

Ellis, J. R., Milligan, S. P., Readdy, L., Taylor, N., and Brown, M. J.: Spawning and nursery grounds of selected fish species in UK waters, Science Series Technical Report, 147, 56, 2012.

Foden, J., Rogers, S. I., and Jones, A. P.: Human pressures on UK seabed habitats: a cumulative impact assessment, Marine Ecology Progress Series, 428, 33–47, 2011.

Gafeira, J., Green, S., Dove, D., Morando, A., Cooper, R., Long, D., and Gatliff, R. W.: Developing the necessary data layers for Marine Conservation Zone selection - Distribution of rock / hard substrate on the UK Continental Shelf, Tech. rep., British Geological Survey, 2010.

Galparsoro, I., Connor, D. W., Borja, Á., Aish, A., Amorim, P., Bajjouk, T., Chambers, C., Coggan, R., Dirberg, G., Ellwood, H., Evans, D., Goodin, K. L., Grehan, A., Haldin, J., Howell, K., Jenkins, C., Michez, N., Mo, G., Buhl-Mortensen, P., Pearce, B., Populus, J., Salomidi, M., Sánchez, F., Serrano, A., Shumchenia, E., Tempera, F., and Vasquez, M.: Using EUNIS habitat classification for benthic mapping in European seas: Present concerns and future needs, Marine Pollution Bulletin, 64, 2630–2638, doi:10.1016/j.marpolbul.2012.10.010, 2012.

George, D. A. and Hill, P. S.: Wave climate, sediment supply and the depth of the sand-mud transition: A global survey, Marine Geology, 254, 121–128, doi:10.1016/j.margeo.2008.05.005, 2008.

Gray, J. S.: Species richness of marine soft sediments, Marine Ecology Progress Series, 244, 285–297, doi:10.3354/meps244285, 2002.

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., Casey, K. S., Ebert, C., Fox, H. E., Fukita, R., Heinemann, D., Lenihan, H. S., Madin, E. M. P., Perry, M. T., Selig, E. R., Spalding, M., Steneck, R., and Watson, R.: A Global Map of Human Impact on Marine Ecosystems, Science, 319, 948–953, doi:10.1126/science.1149345, 2014.

Hamilton, N.: ggtern: An Extension to 'ggplot2', for the Creation of Ternary Diagrams, R package version, 2, 2016.

Heath, M., Sabatino, A., Serpetti, N., McCaig, C., and O'Hara Murray, R.: Modelling the sensitivity of suspended sediment profiles to tidal current and wave conditions, Ocean & Coastal Management, doi:10.1016/j.ocecoaman.2016.10.018, 2016.

Heath, M. R.: Ecosystem limits to food web fluxes and fisheries yields in the North Sea simulated with an end-to-end food web model, Progress in Oceanography, 102, 42–66, doi:10.1016/j.pocean.2012.03.004, 2012.

Hijmans, R. J., Williams, E., and Vennes, C.: geosphere: Spherical Trigonometry. R package version 1.2–28, CRAN. R-project. org/package= geosphere, 2012.

Huang, Z., Nichol, S. L., Siwabessy, J. P., Daniell, J., and Brooke, B. P.: Predictive modelling of seabed sediment parameters using multibeam acoustic data: a case study on the Carnarvon Shelf, Western Australia, International Journal of Geographical Information Science, 26, 283–307, doi:10.1080/13658816.2011.590139, 2012.

James, G., Witten, D., Hastie, T., and Tibshirani, R.: An introduction to Statistical Learning, Springer, doi:10.1007/978-1-4614-7138-7, 2013.

Li, J., Heap, A. D., Potter, A., Huang, Z., and Daniell, J. J.: Can we improve the spatial predictions of seabed sediments? A case study of spatial interpolation of mud content across the southwest Australian margin, Continental Shelf Research, 31, 1365–1376, doi:10.1016/j.csr.2011.05.015, 2011.

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R news, 2, 18–22, doi:10.1177/154405910408300516, 2002.

Loh, W.-Y.: Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1, 14–23, doi:10.1002/widm.8, 2011.

Lohse, L., Malschaert, J. F. P., Slomp, C. P., Helder, W., and Vanraaphorst, W.: Nitrogen cycling in North Sea sediments: interaction of denitrification and nitrification in offshore and coastal areas, Marine Ecology Progress Series, 101, 283–296, doi:10.3354/meps101283, 1993.

McLaren, P., Collins, M. B., Gao, S., and Powys, R. I. L.: Sediment dynamics of the Severn Estuary and Inner Bristol Channel, Journal of the Geological Society, 150, 589–603, doi:10.1144/gsjgs.150.3.0589, 1993.

Neill, S. P., Scourse, J. D., Bigg, G. R., and Uehara, K.: Changes in wave climate over the northwest European shelf seas during the last 12,000 years, Journal of Geophysical Research, 114, doi:10.1029/2009JC005288, 2009.

Neill, S. P., Scourse, J. D., and Uehara, K.: Evolution of bed shear stress distribution over the northwest European shelf seas during the last 12,000 years, Ocean Dynamics, 60, 1139–1156, doi:10.1007/s10236-010-0313-3, 2010.

Pateiro-l, B. and Rodr, A.: Generalizing the Convex Hull of a Sample : The R Package Alphahull, Journal of Statistical software, 34, 1–28, doi:http://dx.doi.org/10.18637/jss.v034.i05, 2010.

Porter-Smith, R., Harris, P. T., Andersen, O. B., Coleman, R., Greenslade, D., and Jenkins, C. J.: Classification of the Australian continental shelf based on predicted sediment threshold exceedance from tidal currents and swell waves, Marine Geology, 211, 1–20, doi:10.1016/j.margeo.2004.05.031, 2004.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schr??der, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, pp. 1–17, doi:10.1111/ecog.02881, 2017.

Robinson, K. A., Ramsay, K., Lindenbaum, C., Frost, N., Moore, J., Wright, A. P., and Petrey, D.: Predicting the distribution of seabed biotopes in the southern Irish Sea, Continental Shelf Research, 31, S120–S131, doi:10.1016/j.csr.2010.01.010, 2011.

Ruardij, P. and Van Raaphorst, W.: Benthic nutrient regeneration in the ERSEM ecosystem model of the North Sea, Netherlands Journal of Sea Research, 33, 453–483, doi:10.1016/0077-7579(95)90057-8, 1995.

Scourse, J., Uehara, K., and Wainwright, A.: Celtic Sea linear tidal sand ridges, the Irish Sea Ice Stream and the Fleuve Manche: Palaeotidal modelling of a transitional passive margin depositional system, Marine Geology, 259, 102–111, doi:10.1016/j.margeo.2008.12.010, 2009.

Serpetti, N.: Modelling and mapping the physical and biogeochemical properties of sediments on the North Sea coastal waters, Ph.D. thesis, University of Aberdeen, 2015.

Serpetti, N., Heath, M., Rose, M., and Witte, U.: High resolution mapping of sediment organic matter from acoustic reflectance data, Hydrobiologia, 680, 265–284, doi:10.1007/s10750-011-0937-4, 2012.

Soulsby, R.: Simplified calculation of wave orbital velocities, Tech. rep., HR Wallingford Ltd., Wallingford, 2006.

Soulsby, R. L. and Clarke, S.: Bed shear-stresses under combined waves and currents on smooth and rough beds, HR Wallingford, Report TR137, 2005.

Soulsby, R. L. and Smallnan, J. V.: A direct method of calculating bottom orbital velocity under waves, Tech. rep., Report SR76, Hydraulics Research Wallingford, 1986.

Stephens, D. and Diesing, M.: Towards quantitative spatial models of seabed sediment composition, PLoS ONE, p. e0142502, doi:10.1371/journal.pone.0142502, 2015.

Thrush, S. F., Hewitt, J. E., Norkko, A., Nicholls, P. E., Funnell, G. A., and Ellis, J. I.: Habitat change in estuaries: Predicting broad-scale responses of intertidal macrofauna to sediment mud content, Marine Ecology Progress Series, 263, 101–112, doi:10.3354/meps263101, 2003.

Uehara, K., Scourse, J. D., Horsburgh, K. J., Lambeck, K., and Purcell, A. P.: Tidal evolution of the northwest European shelf seas from the Last Glacial Maximum to the present, Journal of Geophysical Research: Oceans, 111, 1–15, doi:10.1029/2006JC003531, 2006.

Valerius, J., V. L., S., V. H., Let, J. O., and Zeiler, M.: Trans-national database of North Sea sediment data. Data compilation by Federal Maritime and Hydrographic Agency (Germany); Royal Belgian Institute of Natural Sciences (Belgium); TNO (Netherlands) and Geological Survey of Denmark and Greenland (Denmark)., Tech. rep., 2015.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., and Storlie, C.: SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises, R package version, p. 1, 2014.

Vasquez, M., Mata Chacón, D., Tempera, F., O'Keeffe, E., Galparsoro, I., Sanz Alonso, J. L., Gonçalves, J. M. S., Bentes, L., Amorim, P., Henriques, V., McGrath, F., Monteiro, P., Mendes, B., Freitas, R., Martins, R., and Populus, J.: Broad-scale mapping of seafloor habitats in the north-east Atlantic using existing environmental data, Journal of Sea Research, 100, 120–132, doi:10.1016/j.seares.2014.09.011, 2015.

Ward, S. L., Neill, S. P., Van Landeghem, K. J. J., and Scourse, J. D.: Classifying seabed sediment type using simulated tidal-induced bed shear stress, Marine Geology, 367, 94–104, doi:10.1016/j.margeo.2015.05.010, 2015.

Wessel, P. and Smith, W. H. F.: A global, self-consistent, hierarchical, high-resolution shoreline, Journal of Geophysical Research, 101, 8741–8743, 1996.

Wickham, H.: ggplot2: elegant graphics for data analysis, Springer, 2016.

Wickham, H. and Francois, R.: dplyr: A grammar of data manipulation, URL http://CRAN. R-project. org/package= dplyr. R package version 0.43, 2016.

Wiesner, M. G., Haake, B., and Wirth, H.: Organic facies of surface sediments in the North Sea, Organic Geochemistry, 15, 419–432, doi:10.1016/0146-6380(90)90169-Z, 1990.

Wilcock, P. R., Pitlick, J., and Cui, Y.: Sediment transport primer, estimating bed-material transport in gravel-bed rivers. Gen Tech Rep RMRS-GTR-226, Tech. rep., US Department of Agriculture, Forest service, Rocky Mountain Research Station: Fort Collins, CO; 78., 2009.

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., and Grehan, A. J.: Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope, vol. 30, doi:10.1080/01490410701295962, 2007.

Wood, S. N.: mgcv: GAMs and generalized ridge regression for R, R News, 1, 20–25, doi:10.1159/000323281, 2001.

Wood, S. N.: Generalized Additive Models: an introduction with R, Chapman and Hall, London, doi:10.1111/j.1541-0420.2007.00905_3.x, 2006.

Wright, M. N. and Ziegler, A.: ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, Journal of Statistical Software, 77, doi:10.18637/jss.v077.i01, 2017.