JNCC Report
No: 592

Semi-automated mapping of rock in the North Sea

Downie, A.L.[1], Dove, D.[2], Westhead, R.K.[2], Diesing, M.[1], Green, S.L.[2] & Cooper, R.[2]

July 2016

ISSN 0963-8901

1 Centre for Environment, Fisheries and Aquaculture Science
2 British Geological Society

For further information please contact:

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY
jncc.defra.gov.uk

This report should be cited as:

Downie, A.L., Dove, D., Westhead, R.K., Diesing, M., Green, S. & Cooper, R. 2016. Semi-automated mapping of rock in the North Sea. *JNCC Report No. 592*. JNCC, Peterborough.

Summary

This report describes the results from a semi-automated approach to the mapping of bedrock outcropping at the seabed, applied to Charting Progress 2 (CP2) regions 1 (Northern North Sea) and 2 (Southern North Sea). It represents the second phase of this work, following development and initial application of the approach in the Western Channel (CP2 region 3) and Celtic Sea (CP2 region 4) (Diesing *et al* 2015).

As described previously (Diesing *et al* 2015), the method consists of two elements, namely: 1) the automated spatial prediction of the presence and absence of rock at the seabed using a random forest ensemble model, and; 2) manual editing of the model outputs based on ancillary geological data and expert knowledge.

The random forest prediction yielded satisfying results with an overall accuracy of 86 % and a kappa of 0.53 based on a test set of samples not involved in model building. However, visual inspection did reveal that mis-classifications occurred in places, and the model outputs were adjusted accordingly. In particular, polygons were removed where it was clear they were associated with sediment features such as sand waves. Conversely, the automated method under-predicted the distribution of rock where rock outcrop is not associated with discrete morphological features (e.g. rock pavements).

The confidence in the developed rock layer was also assessed according to the method used by Diesing *et al* (2015), based on the type (quality) of bathymetric data, probability of rock presence based on the random forest ensemble and agreement between predictions and observations in a spatially explicit way.

The final output gives a significantly improved representation of the presence of bedrock at the seabed in the North Sea.

Contents

1 Introduction

1.1 Background

In order to prepare a set of data layers to be used in the first stage of the Marine Conservation Zones (MCZ) project, the Department for Environment, Food and Rural Affairs (Defra) let contract MB0103 in 2009-10 to produce a UK-wide data layer showing areas of rock and hard substrate at or near the seabed surface (Gafeira et al., 2010). The British Geological Survey (BGS) carried out this work as a subcontractor of ABPmer.

The outputs were:

1. Rock and hard substrate polygon layer.
2. Rock and cobbles point layer.
3. Confidence layer.
4. Layer showing areas in which multi-beam bathymetry data has been collected.

In 2011, BGS updated the polygon dataset and named it DigHardSubstrate250, which is provided alongside version 3 of DigSBS250.

The Joint Nature Conservation Committee (JNCC) has a responsibility for reporting on the status of the UK's reefs, which is a habitat defined under Annex I of the Habitats Directive[1]. Reefs are made up of three sub-types: bedrock, stony and biogenic. DigHardSubstrate is therefore a useful product for JNCC in that it indicates the potential location and extent of bedrock reef and would benefit from it being updated as new data becomes available and methods are developed.

In 2015 JNCC partially funded BGS and Cefas to carry out a semi-automated mapping of rock at the seabed surface in Charting Progress 2 (CP2) regions 3 and 4 (adjusted for the new EEZ boundary). The aim was to demonstrate a method that maximises the benefits of both automated mapping approaches and in-depth geological knowledge, and that would subsequently be applicable to other sea areas around the UK. The methods and results can be found in Diesing *et al* (2015) and are shown in Figure 1.

[1] Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Communities No L 206/7.

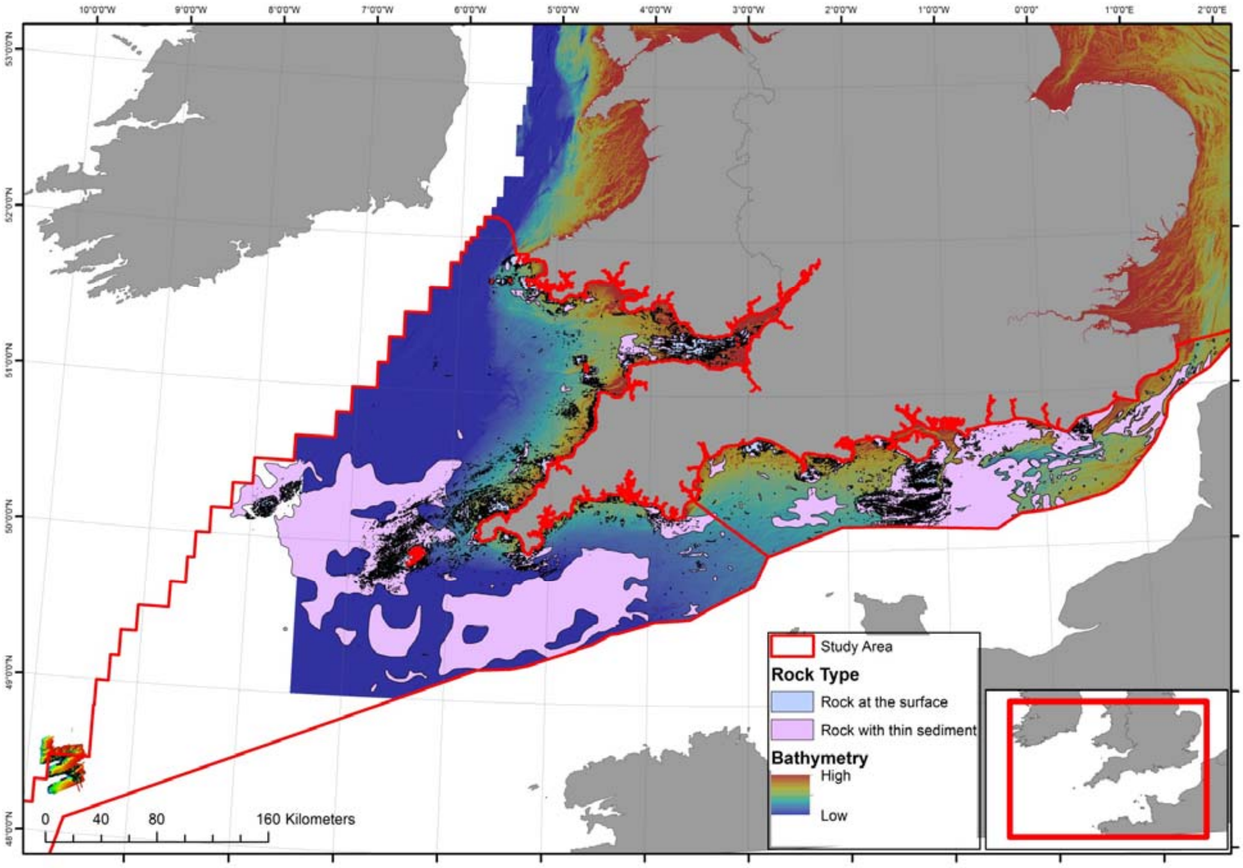


Figure 1: Results from the preceding study (Diesing *et al* 2015). Distribution of rock at the seabed surface and rock covered with thin sediment (0.5m) within Charting Progress 2 (CP2) regions 3 (English Channel) and 4 (Celtic Sea)

1.2 Aims and objectives

1.2.1 Project Aims

This report describes the second phase of this work, where the aim was to extend the coverage of the mapping started by Diesing *et al* (2015) to produce a standard interpretation for rock distribution for CP2 regions 1 and 2 (northern North Sea and southern North Sea) and part of CP2 region 7 adjacent to region 1.

1.2.2 Objectives

1. Develop a vector-based geospatial data product showing the potential extent of rock at, or near, the sea floor for subtidal areas of the regions shown in Figure 2, at a spatial scale equivalent to 1:250,000.

2. Identify sub-types as follows: rock at the surface, rock with thin sediment (up to 0.5m), according to the following definitions:

    a. **Rock at the surface**: Rock present at outcrop. This suggests a habitat dominated by exposed bedrock. Whilst it is unlikely that large areas of exposed rock will exist with zero sediment present, this classification should capture areas of negligible or highly mobile, patchy sediments where the veneer is minimal.

  b. **Rock with thin sediment**: These are essentially subcrops of bedrock, i.e. areas where bedrock rises to the seabed surface, but remains largely covered by a thin veneer of sediment. This will be derived by subtracting areas predicted as ‘rock at the surface’ from previously mapped rock areas (DigHardSubstrate).

3. Keep a record of manual edits made to allow for efficient updates in future.

4. Carry out a three-step confidence assessment for each polygon and include scores in the output data product.

5. Make a plan to incorporate the output into the EMODnet Geology seabed substrate map – this may involve identifying seabed sediment types in areas previously identified as rock.

2 Materials and Methods

2.1 Study Site

The study site comprises CP2 regions 1 (Northern North Sea; extended to include areas of potential bedrock around Orkney and Shetland identified in DigHardSubstrate) and 2 (Southern North Sea) (Figure 2).

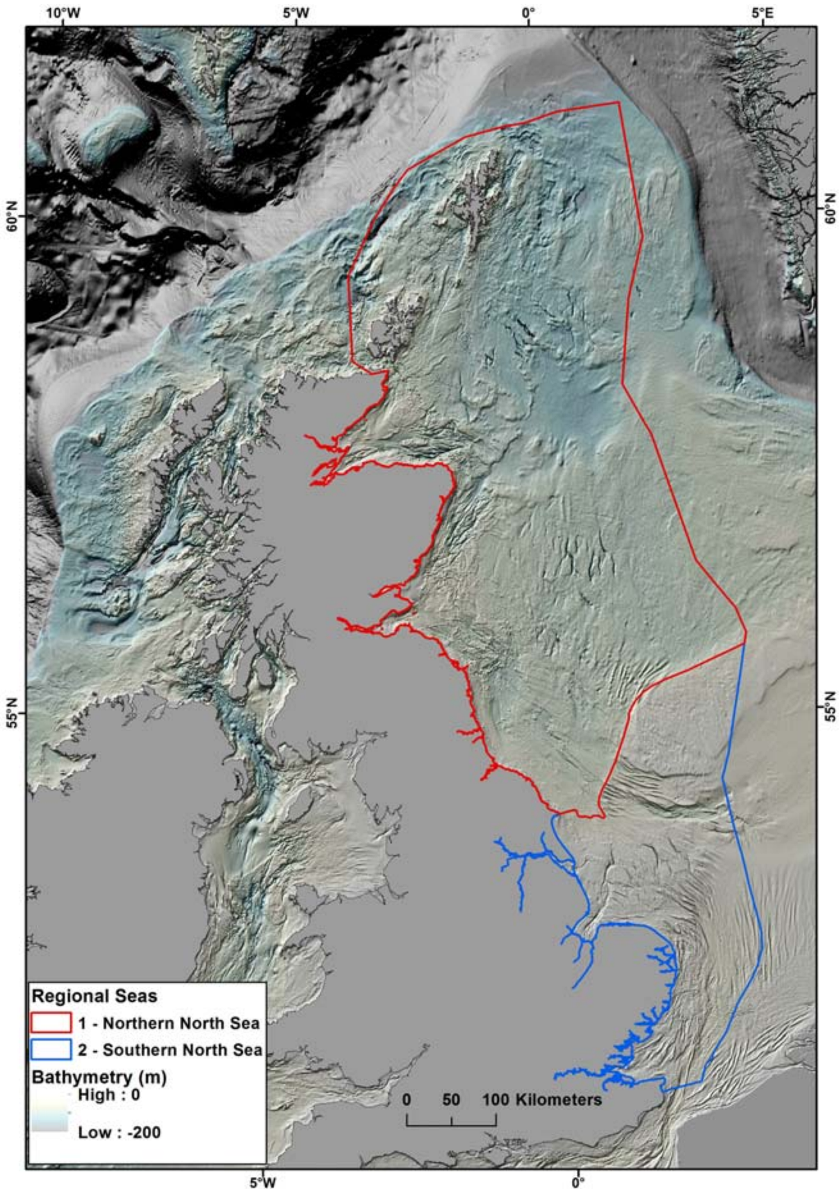


Figure 2: Bathymetry and Charting Progress 2 regional sea boundaries. This study addresses regions 1(extended) and 2.

2.2 Data

2.2.1 Substrate observations

The input dataset contained 14,228 substrate observations within the study area (Figure 3). These were obtained from the Defra marine vector dataset (JNCC 2011). These data have been successfully used in previous studies involving the mapping of rocky substrates (Diesing *et al* 2015; Stephens *et al* 2014). Of these data, 11,317 (79.5%) were recorded as indicting the unambiguous absence of rock. 2,236 (15.7%) were recorded as indicating the unambiguous presence of rock, i.e. rock and no other substrate type was recorded. In 464 (3.2%) cases rock occurred together with other substrate types. No information on substrate type was recorded in 211 (1.5%) cases. For further analysis, ‘no data’ and ambiguous records were removed. This decision was based on previous experience (Diesing *et al* 2015), which indicated that excluding ambiguous samples would give the most accurate predictions. This meant that 13,553 observations were retained for further analysis (Figure 3).

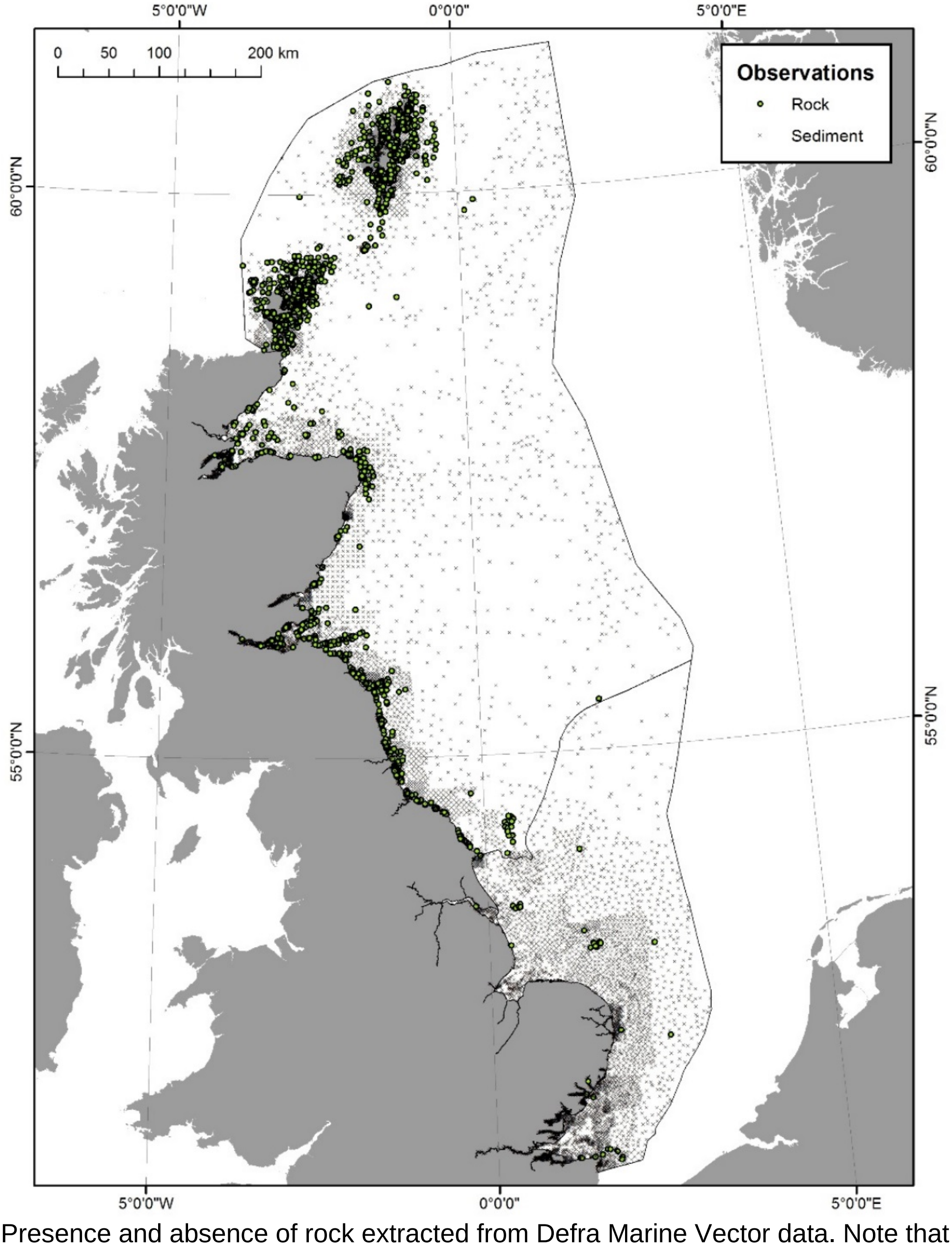


Figure 3: Presence and absence of rock extracted from Defra Marine Vector data. Note that inconclusive cases (i.e. those observations that included both rock and sediment) were ignored.

2.2.2 Predictor features

In order to predict rock presence at unobserved locations, the substrate observations had to be related to auxiliary variables (referred to as features) that have continuous coverage across the study area. These predictor features are comprised of a bathymetry digital elevation model (DEM), topographic characteristics derived from the bathymetry (such as slope and roughness), outputs from hydrodynamic modelling and polygon layers indicating properties of the seabed. Detailed descriptions of all features are given in Table 1.

Table 1: Predictor features utilised in this study.

Feature	Description	Unit	Reference
Bathymetry	Bathymetry (water depth) projected to UTM 30 North at a resolution of 25m.	m	(Astrium Oceanwise 2011)
Roughness	Derived from bathymetry; the difference between minimum and maximum of cell and its eight neighbours.	m	(Wilson *et al* 2007)
Slope	Derived from bathymetry, the maximum slope gradient.	degree	(Wilson *et al* 2007)
Aspect	Derived from bathymetry, direction of steepest slope, expressed as Eastness (sine of aspect) and Northness (cosine of aspect).		(Wilson *et al* 2007)
Curvature	Derived from bathymetry, rate of change of slope. Profile curvature is measured parallel to maximum slope; plan curvature is measured perpendicular to slope.		(Wilson *et al* 2007)
Bathymetric Position Index (BPI)	Derived from bathymetry, vertical position of cell relative to neighbourhood (identifies topographic peaks and troughs). Radii of 5, 10, 15, 20, 25, 30, 40 and 50 pixels were used.	m	(Lundblad *et al* 2006)
BGS Hard Substrate	DigHardSubstrate250 data product. Delineates areas of rock at outcrop, or overlain by thin (<0.5m) sediment based on bathymetric data, the BGS legacy sample database and expert interpretation.		(Gafeira *et al* 2010)
Indicators of Mobile Sediments	Seabed morphologies characteristic of mobile sediments were delineated using hillshade, slope and rugosity data.		(Westhead *et al* 2014)
Quaternary Thickness	Data layer detailing thickness of Quaternary cover on the UK Continental Shelf categorised into three classes: 0-5m; 5-50m; >50m.		(Westhead *et al* 2014)
Relative Resistance	Representation of the relative resistivity of bedrock on the UK Continental Shelf based on age and lithology. Derived utilising BGS DigRock250 (Westhead *et al (*2013) following the method described by Clayton and Shamoon (1998).		(Clayton & Shamoon 1998)
Distance to Coast	Euclidean distance to nearest coastline.	m	
Current Velocity	Mean M2 tidal current velocity averaged across water column calculated using a Telemac model with an unstructured grid of variable resolution.	m/s	Cefas in-house product. Not yet published.
Peak Wave Orbital Velocity	Peak wave orbital velocity at seabed. Surface wave parameters (wave height and period) were output from a POLCOM model for the years 2000 to 2008. The original resolution of the data is 12km. Bottom orbital velocities were calculated from these and the 1 arcsec Defra DEM. Maximum, mean and standard deviation of peak orbital velocity were calculated.	m/s	(Aldridge *et al* 2015; Bricheno *et al* 2015; Holt & James, 2001)

2.3 Methods

2.3.1 Pre-processing of observations

The first step was to extract the values of each predictor feature at the location of each substrate observation. Not all predictor features had the same spatial extent or resolution. This meant that for some predictor features there were gaps, mainly around the coast or far offshore, resulting in some observation locations having no data values (NA) for some or all features. Any observations that contained NA values for at least one predictor feature were discarded. Of the 13,553 observations, 11,803 intersected all predictor layers and were retained in the dataset (see Section 2.2.1). The breakdown of the number of NA values by feature indicates the highest number was for the wave orbital velocity layer. Of the 11,803 observations, 2,145 (18%) observations indicated 'presence' (P) of bedrock and 9,658 (82%) were 'absence' (A).

The quality and reliability of the bathymetry data is not constant across the study area. The dataset is a collation of available data, mostly collected since the 1980s and differing techniques were used to collect and process the data. This means that, although the grid resolution is constant at 25m across the study area, the underlying data are of varying quality and this will also affect the topographic variables derived from the bathymetry. Table 2 shows the number of observations in each category of bathymetry quality. The effect of bathymetric quality on model performance and prediction was investigated by building two models, one using all 11,803 observations and another using only the 8,561 observations with acoustic source data (singlebeam and multibeam echosounder). Only the model using all observations is described here, as there was no clear difference in the performance or spatial predictions of the two models.

Table 2: Observations by bathymetry quality class. Ordered by reliability from left to right.

Type	*Chart*	*Singlebeam echosounder*	*Multibeam echosounder*	*Total*
Number of observations	3,242	5,662	2,899	11,803
Percent of observations	27 %	48 %	25 %	

In order to test the model predictions, the data were split randomly into training (70% of observations) and test datasets (30%). The ratio of presence to absence records were kept equal in both the training and test datasets (Table 3).

Table 3: Training and test datasets.

	Training	Test	Total
P	1,502	643	2,145 (18 %)
A	6,761	2,897	9,658 (82 %)
Total	**8,263**	**3,540**	**11,803 (100 %)**

2.3.2 Model training

Probability of rock was modelled using a random forest (RF) classification model (Breiman 2001). RF has become one of the most widely used and successful statistical learning models for classification and regression, showing good performance in a large number of domains (Chan & Paelinckx 2008; Chapman *et al* 2010; Che Hasan *et al* 2012, 2014; Cutler *et al* 2007; Diesing *et al* 2014; Huang *et al* 2012, 2014; Lucieer *et al* 2013; Oliveira *et al* 2012; Pal 2005; Prasad *et al* 2006; Stephens & Diesing 2014, 2015). RF is an ensemble technique, which aggregates the results of a large number of classification trees, each built using a random subset of the training data and a trialling random subset of predictor variables at each node split. The result is a 'forest' of tree models, with the number of splits determined either by achieving pure end nodes (default) or a number of end nodes determined by the user. RF is a non-parametric technique, i.e. no assumptions regarding the shape of distributions of the response or predictor variables are made (Cutler *et al* 2007). It can handle complex, non-linear relationships between predictor and response variables. As well as using the test set to validate the model, RF implicitly generates a cross-validated (CV) measure of model accuracy. RF also provides a relative estimate of predictor feature importance. This is a measure of the variability explained by each feature, averaged across every tree in the RF.

The model was built in the free statistical computing software R (R Development Core Team 2011) using the 'randomForest' package (Liaw & Wiener 2002). In this case, the 'forest' included 2,500 classification trees. The number of observations in each random subset was set at 300 presences and absences each. Other settings were kept at their defaults. Predictions were made for the probability of encountering rock at each raster cell. Probability predictions by RF are derived based on the fraction of votes given for a specified class by the ensemble of trees.

Prior to training the model, a feature selection step was implemented to test the statistical significance of the predictor features for the presence/absence prediction of rock. The Boruta algorithm (Kursa & Rudnicki 2010) is a feature selection wrapper (Guyon & Elisseeff 2003) based on the RF model. The algorithm uses the feature importance score generated by RF to test each of the predictor features against the effect of random noise. The importance score of each variable is compared to the range of importance scores achieved by random permutations of the predictor variables. Only features with importance scores significantly higher than those of the random variables were retained for use in the final model. A further feature selection step removed correlated variables from the set of predictors used in the final model. Out of any two variables with a correlation coefficient above 0.7, only the variable with the higher Boruta score was retained in the final set of predictors.

To evaluate the predictive performance of the model, four validation statistics were calculated, including Area Under the Curve (AUC), Sensitivity[2], Specificity[3] and Kappa statistics (Cohen 1960). The AUC statistic is a threshold-independent goodness measure between 0 and 1, where a value of 0.5 signifies a model that is no better than random, whilst a value of 1 corresponds to perfect discrimination. As a general guideline, AUC values over 0.9 indicate excellent, 0.8 - 0.9 very good, 0.7 - 0.8 satisfactory and below 0.7 poor discriminative ability (Hosmer & Lemeshow 2000). The AUC statistic is, however, sensitive to the prevalence of rock presence observations in the test data (Lobo *et al* 2008) and it is good practice to combine it with other validation measures. Sensitivity, Specificity and Kappa are all calculated from dichotomised, presence/absence predictions. The threshold used to convert probabilities to presence and absence, and consequently to calculate Sensitivity, Specificity and Kappa, affects the overall accuracy and the likelihood of false negatives and false positives. The objectives of mapping will define which threshold is appropriate. Where the cost associated with false negatives (i.e. not predicting rock where it occurs) is high, a threshold which places more weight on sensitivity would be more appropriate. Conversely, where false positives (i.e. predicting rock in places where it does not occur) have a high cost, the threshold should place more weight on specificity. Using equal sensitivity and specificity, i.e. the threshold where positive observations are just as likely to be wrong as negative observations, gives the most unbiased prediction, whilst optimising the sum of specificity and sensitivity gives the highest balanced accuracy (Manel *et al* 2001). In unbalanced datasets overall accuracy can be high, although the class with fewer numerous samples is poorly predicted. A high balanced accuracy indicates good prediction success in both classes. In large datasets with comprehensive spatial cover and a sampling design unbiased by class, the prevalence of a class reflects its expected likelihood of occurrence. Consequently,

[2] Sensitivity: the amount of true presence predictions as a proportion of the total number of presence observations.
[3] Specificity: the amount of true absence predictions as a proportion of the total number of absence observations.

selecting a threshold that produce predicted presences with the same prevalence is likely to predict an accurate spatial extent of the class.

The output probability prediction was initially split into five classes, to encompass the range of probabilities. Class boundaries were set as follows:

1) High confidence in predicted absence of rock: threshold was set to yield 0.99 sensitivity; almost all rock presence observations in the test dataset occur above this cut-off.
2) Absences and presences with lower confidence: threshold was set to maximise the sum of sensitivity and specificity, with absences below and presences above threshold.
3) Maintaining the prevalence of rock: threshold was set to maintain the fraction of rock presences observed in training data in the predictions across raster cells.
4) High confidence in predicted presence of rock: threshold was set to yield 0.99 specificity; almost all rock absences are below this cut-off.

2.3.3 Knowledge-based enhancements

The output of the RF predictions was then reviewed manually by a mapping geologist, in order to assess its validity in terms of the established geology of the area. The first stage of the process was to convert the modelled output into a readily editable ESRI Shape file, in accordance with the project requirement.

The steps detailed below were followed:

- Conversion of the RF output to 20m raster in order to perform generalisation.
- Each cell was replaced with a majority of eight neighbouring cells. This essentially reduces smaller areas and increases a large (majority) area.
- Boundary cleaning; smoothing of the boundaries between zones by buffering and debuffering. This results in smaller areas being engulfed into larger ones, according to boundary length. Large areas have a higher priority to expand into smaller ones.
- Conversion of raster back to polygons.
- Elimination of polygons smaller than 0.015625km^2 (based on a minimum mappable unit feature with a diameter of 125m for 1:250K mapping).
- Aggregation of polygons with less than 125m distance between features, then removal of holes.

Following the generalisation process the modelled output was reviewed against available multibeam bathymetry data, published mapping and sample data, and polygons were deleted or re-attributed in accordance with the geological understanding of the region.

A number of small, irregular polygons were removed on the basis that they represented artefacts from the bathymetric data.

2.3.4 Confidence assessment

The confidence assessment method follows that used in Diesing *et al* (2015), which is based on the three-step confidence assessment framework of Lillis (2016). The assessment was performed on a per-polygon basis due to the possible heterogeneity of inputs into the model across the output area. The method requires the assessor to follow the flow diagram shown in Figure 4 and score the polygon appropriately at each stage.

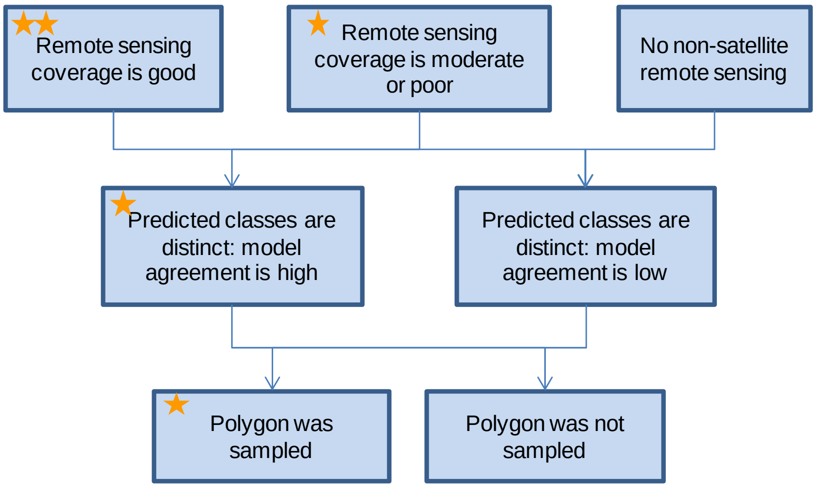


Figure 4: Three-step confidence decision tree; the assessor starts at the top and follows the arrows. Stars/points are awarded according to the answers given and the final score is the sum of the stars/points.

From this method, a maximum qualitative score of 4 can be achieved by a polygon (Table 4). The final score should not be taken as a quantitative probability of the habitat's likelihood in extent or presence, the measurement is a qualitative score based on the data inputs and level of agreement between the predictive models.

Application for polygons identified as rock at outcrop

The remote-sensing coverage confidence was assessed based on the type of acoustic data that were available: A score of two was given where multibeam echosounder data were present, a score of one for singlebeam echosounder data and a score of zero for all other data types. Beyond the multibeam data that were built into the compiled bathymetry grid used in the automated process, additional data from the Maritime and Coastguard Agency's Civil Hydrography Programme and Defra's Marine Conservation Zone mapping programme were also included as part of the knowledge-based review. As such, a value of two was scored for these areas.

The distinctness of class boundaries criterion was scored in two stages:

1. Initially the agreement of the RF ensemble outputs were used: A score of one was attained where the value was above an agreement threshold set to yield a sensitivity of 0.99 - indicating high probability of presence of rock - or where the value was below an agreement threshold set to yield a specificity of 0.99 - indicating high probability of absence of rock. Intermediate values were given a score of zero.
2. Following the knowledge-based enhancements, where expert judgement led to modification or addition of a polygon the initial score was overwritten with a score of one. This therefore indicates higher confidence associated with validation of the presence of an area of rock outcrop by more detailed study or assessment by a geologist.

In the case of the amount of sampling criterion, a score of one was given if a polygon was sampled and the majority of samples agreed with the prediction. Both the sample database used in the automated process, as well as the BGS core database that was used as part of the knowledge-based review were included. A score of zero was attained if a polygon was not sampled or the majority of samples within the polygon disagree with the prediction.

Application for polygons identified as rock with thin sediment

The BGS DigHardSustrate250 dataset includes an assessment of confidence based on data density. However, for production of the shapefile in this project a standard value of zero was applied for Remote Sensing Coverage as limited bathymetry data was available for the production of this shapefile. A value of one for the Agreement confidence criteria, to reflect the influence of human judgement was also applied. As the same legacy database was used for the samples, the same Sample criteria were applied as for the updated 'Rock at outcrop' polygons.

Table 4: All possible combinations of scores under the three-step scheme. Polygons with equal scores are therefore assumed to have roughly similar levels of confidence, regardless of the route through the decision tree.

Score	Remote sensing coverage	Distinctness of class boundaries	Amount of sampling
4	★★	★	★
3	★★		★
	★★	★	
	★	★	★
2	★★		
	★		★
	★	★	
		★	★
1	★		
		★	
			★
0			

3 Results

3.1 Random forest predictions

The feature selection process indicated that all potential predictor variables contributed significantly to the presence/absence predictions. Table 5 gives the permutation importance scores for all tried variables and indicates the variables that were selected for the model based on removal of correlated features.

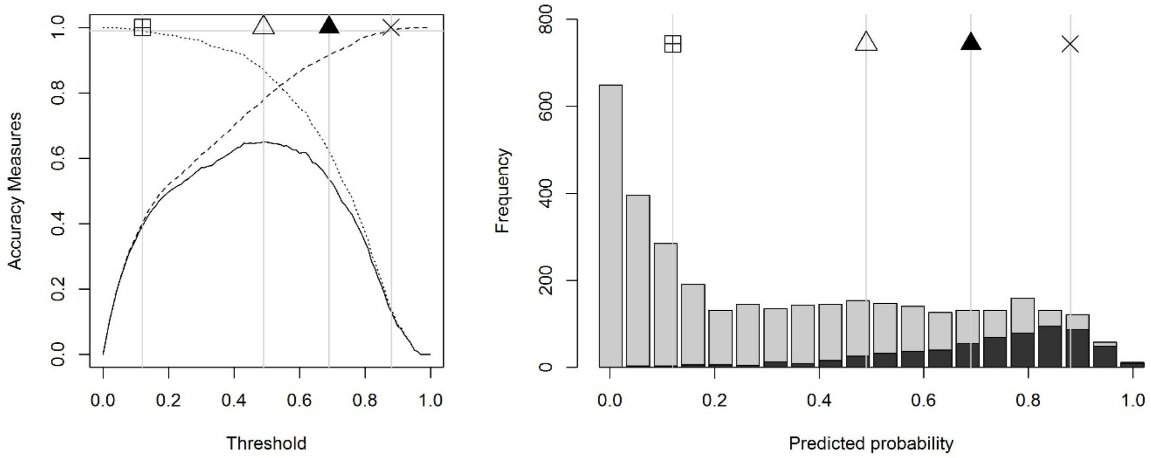
Table 5: Permutation importance scores for potential predictor variables derived from the Boruta algorithm.

Predictor variable	Boruta permutation importance	Selected Predictor variables
BGS Hard Substrate	48.3	Selected
Current Velocity	38.0	Selected
Relative Resistance	33.0	Selected
Bathymetry	28.0	Selected
Peak Wave Orbital Velocity - Standard Deviation	27.3	Selected
BPI30	22.5	Selected
BPI25	22.3	
BPI20	21.7	
Roughness	21.1	Selected
Peak Wave Orbital Velocity - Maximum	21.1	
BPI10	20.8	
Curvature - Combined	20.5	Selected
Peak Wave Orbital Velocity - Mean	20.5	
BPI15	20.4	
BPI05	20.2	
Curvature - Planar	19.7	
BPI50	19.1	
Curvature - Profile	18.7	
Slope	18.6	
BPI40	16.8	
Quaternary Thickness	12.0	Selected
Aspect - Northness	6.6	
Aspect - Eastness	6.4	

The threshold independent AUC score for the model was 0.9, which indicates a well performing model. Accuracy statistics calculated using the four selected thresholds are given in Table 6. In the final output map, high sensitivity and high specificity thresholds of 0.99 were used to indicate high confidence in absences and presences of rock, respectively. Setting a high sensitivity of 0.99 yielded a probability threshold of 0.12 (Table 6 and Figure 5a). Almost all rock presences in the test dataset occur above this cut-off (Figure 5b). Likewise, setting specificity at 0.99 yielded a probability threshold of 0.88 (Table 6 and Figure 5a) and almost all rock absences are found below this cut-off (Figure 5b). The overall cut-off between rock presence/absence was set at the threshold which ensured that prevalence of rock in the predicted output remained the same as in the input data. The selected threshold avoids over-prediction, assuming the model input data is an unbiased estimate of the prevalence of rock in the area. It was also found to give the highest overall accuracy, at 86% of all observations in the test dataset correctly classified, as well as the highest Kappa value at 0.53. The resulting spatial predictions of rock presence and absence with high and low confidence is shown in Figure 6.

Table 6: Accuracy statistics for each threshold used to convert probability of presence to presence/absence classes. The threshold used in the final map is highlighted. PCC = Percent Correctly Classified.

Threshold Method	*Threshold*	*PCC*	*Sensitivity*	*Specificity*	*Kappa*
Maximising confidence in predicted absence of rock, i.e. High Sensitivity	0.12	51 %	0.99	0.40	0.19
Maximising balanced accuracy, i.e. Maximum of Sensitivity + Specificity	0.49	80 %	0.87	0.78	0.49
Maintaining the prevalence of rock, i.e. Predicted Prevalence= Observed Prevalence	0.69	86 %	0.62	0.92	0.53
Maximising confidence in predicted presence of rock, i.e. High Specificity	0.88	84 %	0.14	0.99	0.19



a) b)

— Sensitivity — Specificity
— Sum of sensitivity and specificity
Presence
Absence

W High sensitivity, △ Maximise the sum of sensitivity and specificity,
▲ Predicted prevalence matches observed prevalence, × High specificity

Figure 5: Selected thresholds for converting probability of presence to presence and absence plotted against sensitivity and specificity (a) and the frequency of observed presences and absences (b).

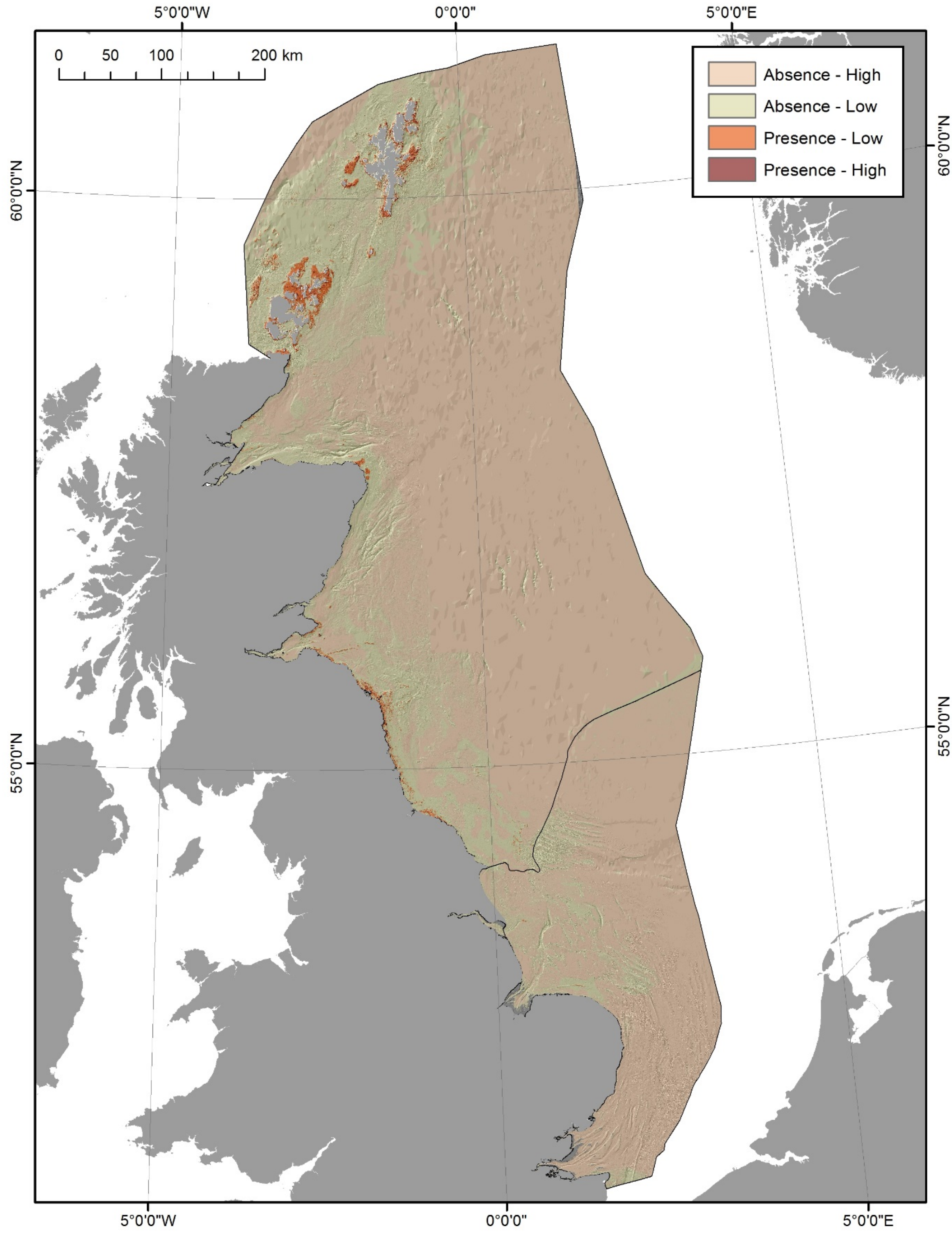


Figure 6: Resulting predictions of rock presence/absence and associated confidence. The boundary between ‘Absence – High’ and ‘Absence – Low’ confidence occurs at an agreement value of 0.12 (high sensitivity). The boundary between ‘Absence – Low’ and ‘Presence – Low’ occurs at an agreement value of 0.69 (maintaining prevalence). The boundary between ‘Presence – High’ and ‘Presence – Low’ confidence occurs at an agreement value of 0.88 (high specificity).

Feature importance scores (Table 5) indicated the BGS rock layer as being the most important variable, followed by tidal velocity, relative resistance, bathymetry and variability in wave orbital velocity. Rock was indicated to be more likely to be outcropping at higher tidal velocities and variable wave regime, low resistance index (hard-wearing rock type), high local BPI (elevated from surrounding seabed), high bathymetric roughness and areas with thin quaternary sediment deposits (Figure 7).

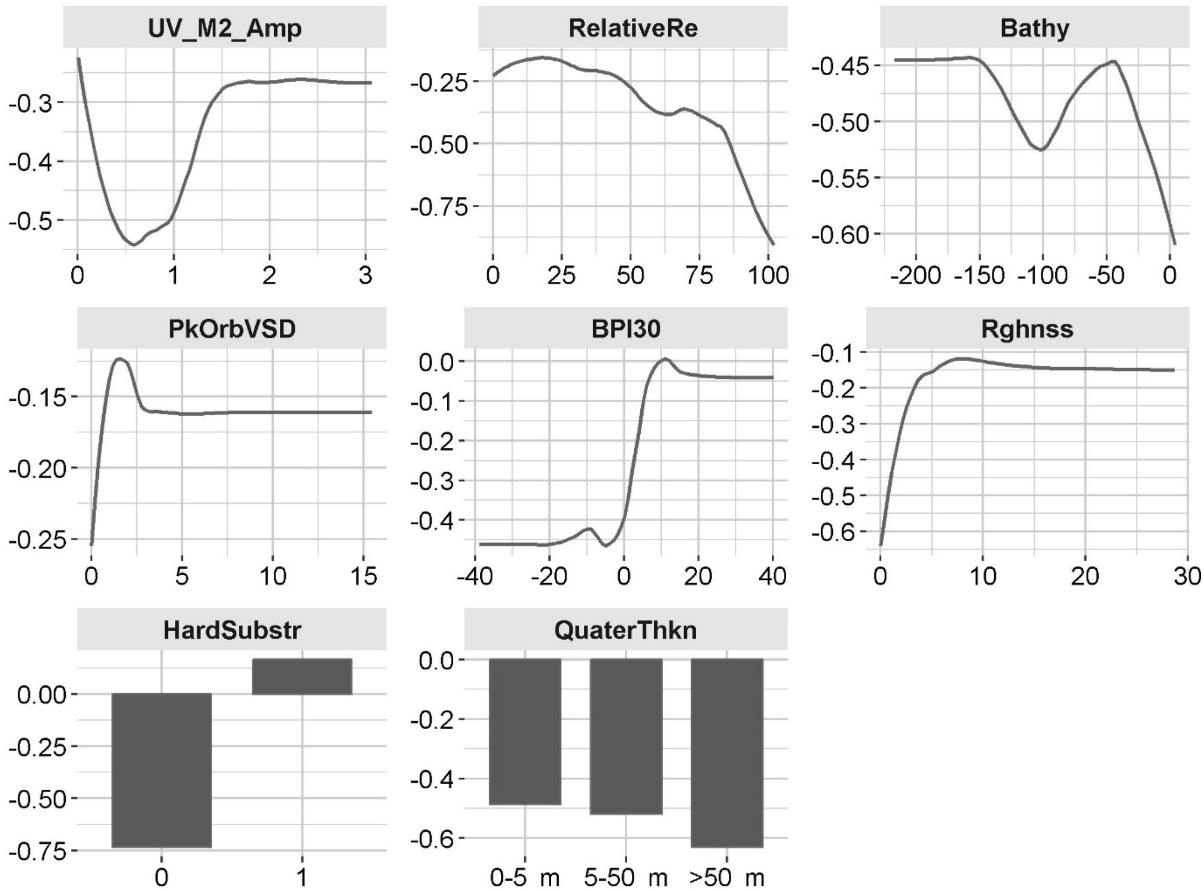


Figure 7: Partial dependence plots showing the response to chosen predictor variables: UV_M2_Amp – M2 tidal current speed; RelativeRe – Relative resistance to erosion; Bathy – Bathymetry; PkOrbVSD – Standard deviation of peak orbital velocity; BPI30 – Bathymetric Position Index with a radius of 30 cells; Rghnss – Roughness; HardSubstr – Hard substrate; QuaterThkn – Quaternary thickness.

3.2 Knowledge-based enhancements

The rock prediction dataset was reviewed against available multibeam bathymetry and sample data, as well as previously published maps (Table 7). This validation process was conducted in accordance with the geological understanding of each region. Where available, the high-resolution multibeam bathymetry data (2 -12m resolution) were particularly useful for this exercise. The multibeam data employed (as they account for the largest datasets) were acquired either by the MCA's Civil Hydrography Programme (CHP) or for the purposes of mapping England's Marine Conservation Zones (MCZ).

Table 7: Published data and mapping used to support expert interpretation.

Feature	**Description**	**Reference**
Bathymetry	EMODnet – medium resolution bathymetry compilation (~150m resolution). High resolution (~2-12m resolution) swath bathymetry acquired by MCA's Civil Hydrography Programme (CHP), and Marine Conservation Zones (MCZ).	**EMODnet**: http://www.emodnet-hydrography.eu/ MCA's CHP: https://www.gov.uk/guidance/the-civil-hydrography-programme **MCZ:** http://jncc.defra.gov.uk/page-4525
BGS Seabed Sediments	DigSBS250 V3; 1:250 000 scale seabed sediments mapping for the UKCS.	Cooper *et al* 2010a
BGS Hard Substrate	DigHardSubstrate250 data product. 1:250 000 scale. Delineates areas of rock at outcrop, or overlain by thin (<0.5m) sediment based on bathymetric data, the BGS legacy sample database and expert interpretation.	Gafeira *et al* 2010; Cooper *et al* 2010b
BGS Indicators of Mobile Sediments	Seabed morphologies characteristic of mobile sediments were delineated using hillshade, slope and rugosity data. 1:250 000 scale.	Westhead *et al* 2014
BGS Quaternary Thickness	Data layer detailing thickness of Quaternary cover on the UK Continental Shelf categorised into three classes: 0-5m; 5-50m; >50m.	Westhead *et al* 2014
BGS Quaternary Deposits	1:1M scale mapping of the UKCS, compiled digitally from analysis of information displayed on BGS 1:250,000 scale paper maps across the UKCS, supplemented by expert interpretation.	Holmes *et al* 1993
BGS Bedrock Geology	DiGROCK250 1:250k scale bedrock mapping of the UKCS.	Westhead *et al* 2013
OSEA3 Hard substrates and non-rock hard substrates	Hard substrates and indicators of non-rock hard substrates – Multiple maps and supporting datasets.	Dove *et al* (*in press*)
OSEA 3 Geology and Surficial Processes	Characterization of coastal and seabed geology – Multiple maps and supporting datasets.	Dove *et al* (*in press*)
ABPmer - Seabed Geological and Geomorphologial features	Compilation of previously mapped geological and geomorphological features.	ABPmer 2009

In this North Sea study area, the predicted rock polygons were either validated or deleted (i.e. no manual digitisation or polygon editing was conducted) following the integrated review. An exception relates to some polygons that were originally predicted as 'rock at surface', that on review were clearly observed to comprise a thin veneer of sediment. In these circumstances the relevant polygons were re-classified as 'rock with thin sediment'.

The primary reasons for deleting polygons were as follows:

- A number of small, irregular polygons were removed on the basis that they represented artefacts from the bathymetric data.
- Numerous elevated features incorrectly predicted as rock are in fact sedimentary bedforms. These features include mobile sediment bedforms (e.g. sediment waves and banks) and glacial bedforms (e.g. moraines).

A general trend is that predictions made from the automated approach are typically more accurate in the coastal zone than farther offshore (i.e. a greater percentage of polygons were deleted farther from the coast). Conversely, the coastal zone is also where greatest under-predictions were observed, where the predicted rock polygons don't sufficiently account for the more extensively observed rock outcrops and platforms. Multibeam bathymetry was typically required to make this assessment. In particular, the automated approach is not effective at discriminating rock pavements, relatively flat areas of bedrock present at seabed. Figure 8 demonstrates this phenomenon (under-predicting rock at surface) as well as a situation in which a sedimentary bedform has been incorrectly predicted as rock and was subsequently deleted during the knowledge-based review. The areas shown in Figure 8 was chosen as it clearly demonstrates these separate causes of error (i.e. where the automated model fails to accurately predict the observed data), however we note that this represents an extreme example of model inaccuracy. These features are quite clear upon visual inspection as fractures, folds, and bedding planes distinguish these features from sedimentary forms.

While time restrictions in the current project did not allow for the manual digitisation of new or revised polygons, a number of observations can be made on how future iterations could improve the rock mapping:

- For most of CP2 region 2 (Southern North Sea), the automated identification exercise appeared to have captured potential rock areas fairly comprehensively, although as a follow-up exercise, it would be possible to improve the assessments particularly in nearshore rock platform areas e.g. off the SE and NE England coasts (e.g. Chalk and Carboniferous bedrock) which are under-represented in places.

- For most of the southern part of CP2 region 1 (Northern North Sea), again the automated process appears to have captured most potential rock areas. However, for the northern part, surrounding Orkney and Shetland in particular, there were areas where seabed rock platforms were significantly under-represented (e.g. Figure 8). In some instances this relates to the rock pavements described above, whereas in other places the automated method has simply not captured the full extent of more rugged bedrock outcrop. Review of multibeam bathymetry was typically required to detect these discrepancies, underlying the importance of utilising these data for such efforts. Another common source of error occurred where predicted rock polygons were clipped/constrained by DigHardSubstrate-250k layer (which was not based on the review of multibeam bathymetry), where in fact rock at surface persisted beyond these boundaries.

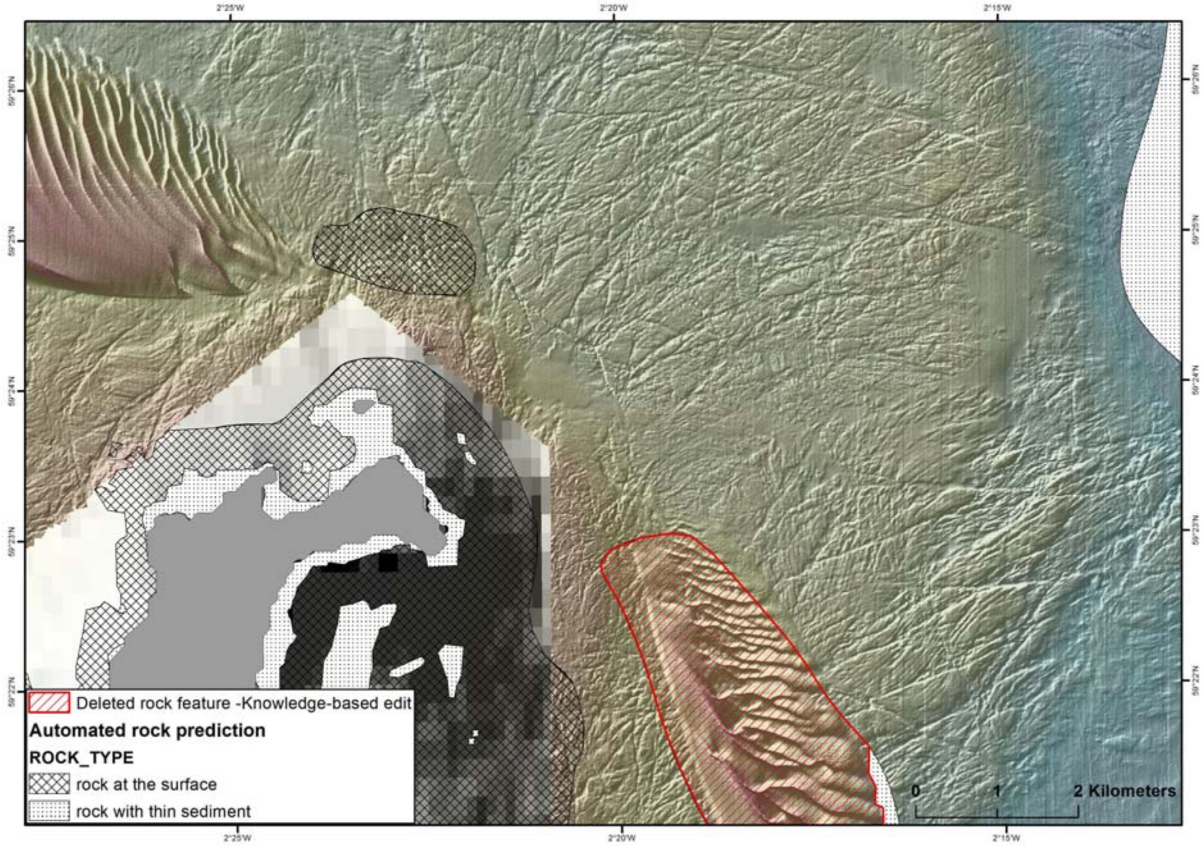


Figure 8: Example off NE coast of Orkney Islands demonstrating where sedimentary bedform was incorrectly predicted as bedrock, and automated approach underestimated rock at surface due to relatively flat rock pavement.

As with the previous Channel/Irish Sea assessment (Diesing *et al* 2015), the second category, 'Rock with thin sediment' was derived by subtracting the 'Rock at outcrop' polygons (following expert assessment) from the BGS DigHardSubstrate250 mapping, and this was combined with the re-assessed rock polygons above to derived the final output layer (Figure 9).

In addition to the rock outputs, a 'changes' shapefile was also generated in order to document the modifications to the modelled output. This will allow the same expert-driven modifications to be quickly applied to future reruns of the model.

3.3 Confidence assessment

The confidence assessment recorded results between zero and four (Figure 10). Polygons based on model predictions scored between two and four depending on sample and multibeam bathymetry coverage. The highest concentration of polygons with a score of four is around the Orkney Isles where a significant portion of rock was predicted which overlaps with extensive multibeam data. Polygons derived from the BGS DigHardSubstrate layer and classified as 'rock with thin sediment' received lower confidence scores as a value of zero was assigned to the remote sensing data that underpinned the original analysis (Gafeira *et al* 2010). Total confidence scores of zero however only apply to relatively small areas which have also not been directly sampled.

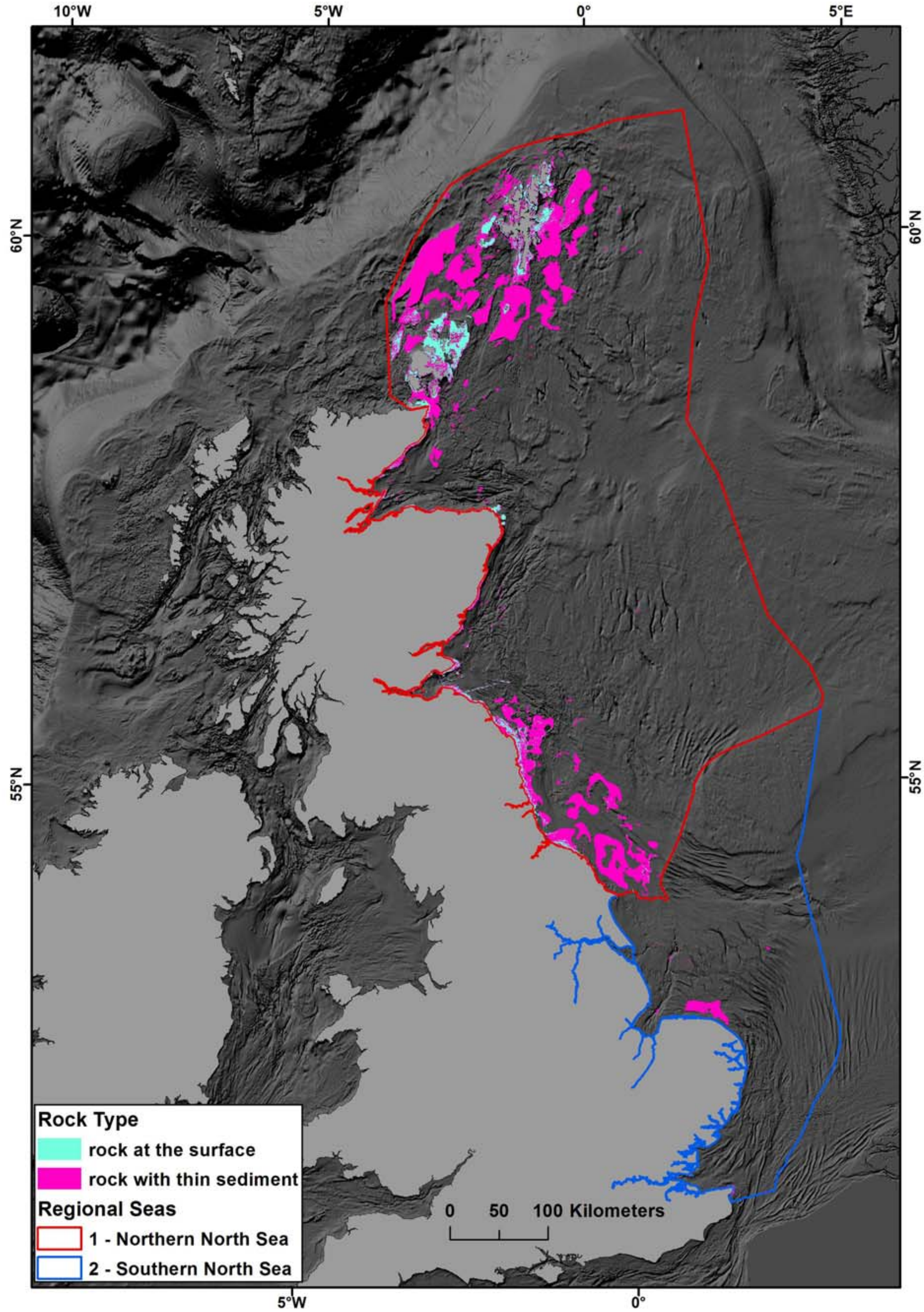


Figure 9: Distribution of rock at the seabed surface and rock covered with thin sediment (<0.5m) within the study area.

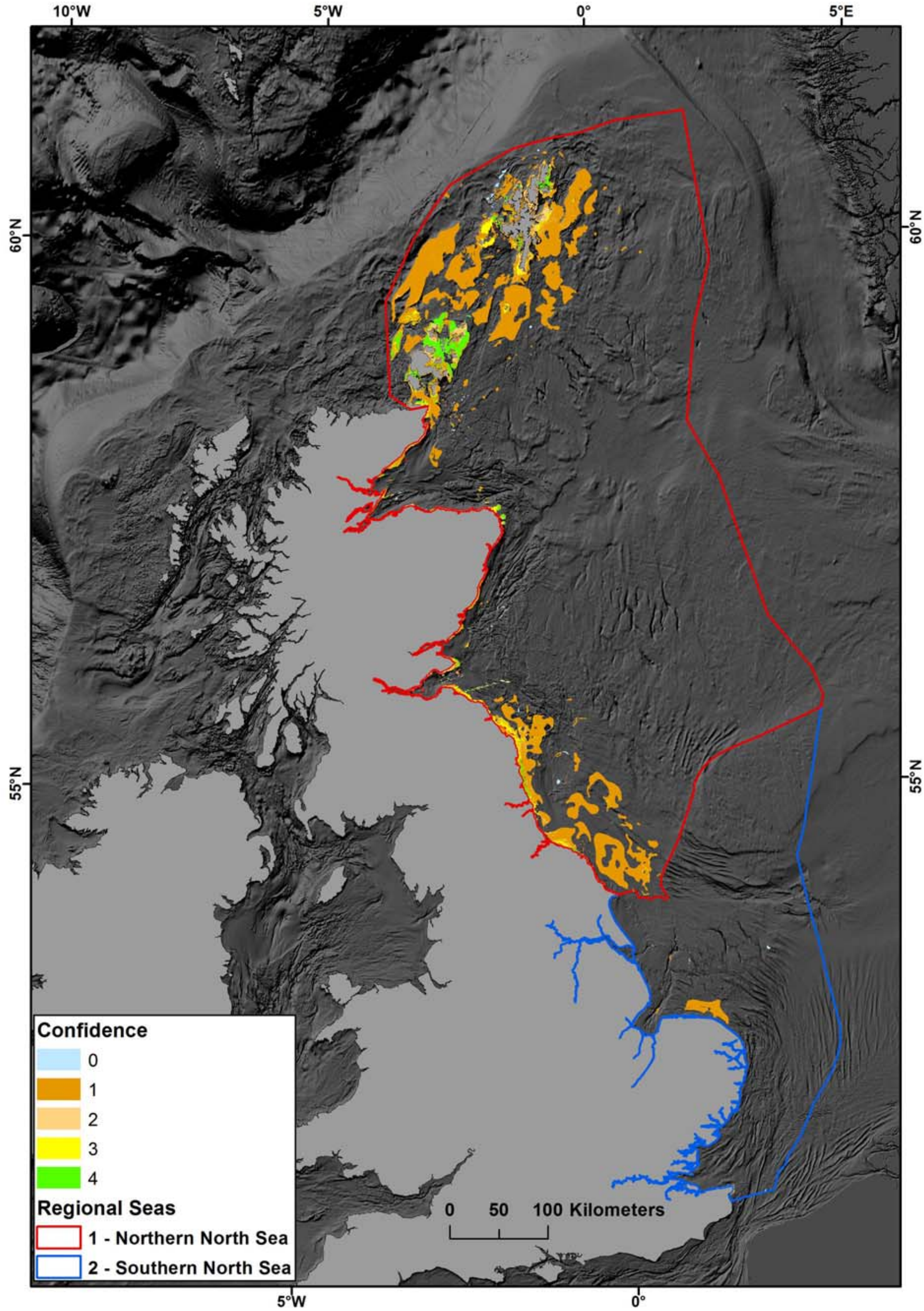


Figure 10: Confidence assessment of the updated map output. Values between zero and four where four indicates maximum confidence.

4 Discussion

We have derived a new data layer of rock in CP2 regions 1 (Northern North Sea) and 2 (Southern North Sea) and a small part of region 7 around Orkney and Shetland, representing the second phase of roll-out of the method developed by Diesing *et al* (2015). The derived data layer has a nominal scale of 1:250,000 and as such gives a sufficiently detailed indication of the distribution of rock at or near the seabed in the North Sea at a regional scale. Whilst the data layer was derived by using the best available data sources and methods, it should be noted that the derived results are unlikely to be sufficient for detailed monitoring of change in reef extent, due to the inherent and unavoidable inaccuracies in data and methods.

We have demonstrated how automated approaches to seabed mapping and in-depth geological knowledge can be combined to derive an improved representation of bedrock at and near the seabed in CP2 regions 1 (Northern North Sea) and 2 (Southern North Sea). In essence, this means that the applied method could be described as semi-automated, as it contains aspects of automated mapping as well as expert intervention. It would certainly be desirable to develop a fully automated method with the aim to reduce subjectivity and increase reproducibility. However, this would require:

i) a complete understanding of the underlying processes that lead to exposure of bedrock at the seabed; and
ii) exhaustive datasets that fully describe the predictor-response relationships.

With incomplete knowledge and data, the best option to derive meaningful predictions is a combined approach as demonstrated in this report. It is noteworthy that we have made an effort to include as much knowledge as possible at the automated prediction stage by including predictor variables that are known or expected to influence the presence of rock at the seabed. Likewise, it should be noted that tools like variable importance plots are useful in understanding which variables are suitable predictors. Unlike in the English Channel and the Celtic Sea, BGS hard substrate and relative resistance were among the three most important features. Conversely, terrain features (e.g. slope and BPI) were of lower importance, while hydrodynamics were important in both cases. These results seem to indicate that the importance of features varies between sites.

The insights gained from the variable importance plot and the manual reclassification of seemingly misclassified objects could be fed back to the automated classification stage and it could be expected that such an iterative process will improve automated prediction results and reduce the amount of expert intervention required. For example, it has become clear that the BGS hard substrate layer, which was the most important feature, could be improved based on new data. Likewise, it is obvious that relatively flat rock pavements are insufficiently captured with the available features. Future work should therefore focus on improving existing features and finding new one that lead to improved predictions. Such an iterative approach could be repeated until no further improvements in classification accuracy are achieved. Additionally, new or improved data become available over time (e.g. improvements to the Defra DEM reflecting new hydrographic survey data). It would therefore be desirable to regularly update the predictions in order to reflect improvements in data, methods and knowledge. The general methodology that was set up as part of the project lends itself to such a task as processes of automated prediction and knowledge-based enhancements have been formalised.

The shapefiles produced by this approach represent a significant update to our previous understanding of the distribution of rock at, or near the seabed CP2 regions 1 (Northern North Sea) and 2 (Southern North Sea). This may be used to inform future updates to seabed substrate maps such as those included in EMODnet Geology and the BGS map series. The outputs were not completed in time for inclusion in the final 2016 update to the EMODnet Geology substrate data products; however, they will be considered for inclusion in further updates.

5 References

ABPmer. 2009. Assessing and developing the required biophysical datasets and datalayers for Marine Protected Areas network planning and wider marine spatial planning purposes. Report No.8 task 2A. Mapping of geological and geomorphological features. Report to Defra by ABPmer for Project MB102.

Aldridge, J.N., Parker, E.R., Bricheno, L., Green, S.L. & van der Molen, J. 2015. Assessment of the physical disturbance of the northern European Continental shelf seabed by waves and currents. Continental Shelf Research, 108: 121–140. http://www.sciencedirect.com/science/article/pii/S0278434315000576 (Accessed 18 March 2015).

Astrium Oceanwise. 2011. Creation of a high resolution digital elevation model (DEM) of the British Isles continental shelf. 21pp.

Bricheno, L.M., Wolf, J. & Aldridge, J. 2015. Distribution of natural disturbance due to wave and tidal bed currents around the UK. Continental Shelf Research, 109: 67–77. http://www.sciencedirect.com/science/article/pii/S0278434315300583 (Accessed 19 October 2015).

Callaway, A., Smyth, J., Brown, C.J., Quinn, R., Service, M. & Long, D. 2009. The impact of scour processes on a smothered reef system in the Irish Sea. *Estuarine, Coastal and Shelf Science*, 84(3), pp.409–418.

Chan, J.C.W. & Paelinckx, D. 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 112: 2999–3011.

Chapman, D.S., Bonn, A., Kunin, W.E. & Cornell, S.J. 2010. Random Forest characterization of upland vegetation and management burning from aerial imagery. Journal of Biogeography, 37: 37–46.

Che Hasan, R., Ierodiaconou, D., Laurenson, L. & Schimel, A. 2014. Integrating Multibeam Backscatter Angular Response, Mosaic and Bathymetry Data for Benthic Habitat Mapping. PLOS ONE, 9: e97339. Public Library of Science. http://dx.doi.org/10.1371%2Fjournal.pone.0097339.

Che Hasan, R., Ierodiaconou, D. & Monk, J. 2012. Evaluation of Four Supervised Learning Methods for Benthic Habitat Mapping Using Backscatter from Multi-Beam Sonar. Remote Sensing, 4: 3427–3443. http://www.mdpi.com/2072-4292/4/11/3427.

Clayton, K. & Shamoon, N. 1998. A new approach to the relief of Great Britain II. A classification of rocks based on relative resistance to denudation. Geomorphology, 25: 155–171.

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20: 37–46.

Cooper, R., Green, S. & Long, D. 2010a. User Guide for the British Geological Survey DiGSBS250K Dataset. *British Geological Survey Internal Report*, IR/11/026.

Cooper, R., Green, S. & Long D. 2010b. User Guide for the British Geological Survey DiGHardSubstrate250k Dataset. *British Geological Survey Internal Report*, IR/11/027

Cutler, D., Edwards, T., Beards, K., Cutler, A., Hess, K., Gibson, J. & Lawler, J. 2007. Random Forests for classification in Ecology. Ecology, 88: 2783–2792.

Diesing, M., Green, S.L., Stephens, D., Cooper, R. & Mellett, C.L. 2015. Semi-automated mapping of rock in the English Channel and Celtic Sea. Peterborough, UK. 19pp.

Diesing, M., Green, S.L., Stephens, D., Lark, R.M., Stewart, H.A. & Dove, D. 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Continental Shelf Research, 84: 107–119. http://www.sciencedirect.com/science/article/pii/S0278434314001629.

Dove, D., Mellett, C., Green, S., Akhurst, M., Carter, G., Evans, D., Gatliff, R., Long, D., Tappin, D., Westhead, K., Stewart, H. & Stevenson, A. (*in press*). Geology and Surficial Processes in DECC's Offshore Energy Strategic Environmental Assessment (SEA) 3.

Dove, D., Green, S. & Stevenson, A. (*in press*). Hard substrates and indicators of non-rock hard substrates in DECC's Offshore Energy Strategic Environmental Assessment (SEA) 3.

Gafeira, J., Green, S.L., Dove, D., Morando, A., Cooper, R., Long, D. & Gatliff, R.W. 2010. Developing the necessary data layers for Marine Conservation Zone selection - Distribution of rock/hard substrate on the UK Continental Shelf. British Geological Survey, Edinburgh. 72pp.

Guyon, I. & Elisseeff, A. 2003. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 3: 1157–1182.

Holmes, R., Jeffrey, D.H., Ruckley, N.A. & Wingfield, R.T.R. 1993. Quaternary geology around the United Kingdom (north and south sheets). Edinburgh, British Geological Survey

Holt, J.T. & James, I.D. 2001. An s coordinate density evolving model of the northwest European continental shelf: 1. Model description and density structure. Journal of Geophysical Research: Oceans, 106: 14015–14034. http://dx.doi.org/10.1029/2000JC000304.

Hosmer, D. & Lemeshow, S. 2000. Applied Logistic Regression. John Wiley & Sons Inc, New York, USA.

Huang, Z., Nichol, S.L., Siwabessy, J.P.W., Daniell, J. & Brooke, B.P. 2012. 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. <Go to ISI>://WOS:000300611200006.

Huang, Z., Siwabessy, J., Nichol, S.L. & Brooke, B.P. 2014. Predictive mapping of seabed substrata using high-resolution multibeam sonar data: A case study from a shelf with complex geomorphology. Marine Geology, 357: 37–52. http://www.sciencedirect.com/science/article/pii/S0025322714002205.

JNCC. 2011. DEFRA Marine Reference Data.

Kursa, M. & Rudnicki, W. 2010. Feature selection with the Boruta Package. Journal of Statistical Software, 36: 1–11. http://www.jstatsoft.org/v36/i11/paper/.

Liaw, A. & Wiener, M. 2002. Classification and regression by randomForest. R News, 2: 18–22.

Lillis, H. (2016). Three-step confidence assessment framework for classified seabed maps. *JNCC Report No. 591*. JNCC, Peterborough.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17: 145–151. Blackwell Publishing Ltd. http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x.

Lucieer, V., Hill, N.A., Barrett, N.S. & Nichol, S. 2013. Do marine substrates 'look' and 'sound' the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images. Estuarine, Coastal and Shelf Science, 117: 94–106. http://www.sciencedirect.com/science/article/pii/S0272771412004246.

Lundblad, E.R., Wright, D.J., Miller, J., Larkin, E.M., Rinehart, R., Naar, D.F., Donahue, B.T., Anderson, M. & Battista, T. 2006. A Benthic Terrain Classification Scheme for American Samoa. Marine Geodesy, 29: 89–111. http://www.informaworld.com/10.1080/01490410600738021.

Manel, S., Williams, H.C. & Ormerod, S.J. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38: 921–931. Blackwell Science Ltd. http://dx.doi.org/10.1046/j.1365-2664.2001.00647.x.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A. & Pereira, J.M.C. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. Forest Ecology and Management, 275: 117–129.

Pal, M. 2005. Random forest classifier for remote sensing classification.

Prasad, A.M., Iverson, L.R. & Liaw, A. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems, 9: 181–199.

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 1706pp.

Stephens, D., Capuzzo, E., Aldridge, J. & Forster, R.M. 2014. Potential interactions of seaweed farms with natural nutrient sinks in kelp beds. Marine Estate Research Report. 1-34 pp.

Stephens, D. & Diesing, M. 2014. A Comparison of Supervised Classification Methods for the Prediction of Substrate Type Using Multibeam Acoustic and Legacy Grain-Size Data. PLOS ONE, 9: e93950. Public Library of Science. http://dx.doi.org/10.1371%2Fjournal.pone.0093950.

Stephens, D. & Diesing, M. 2015. Towards quantitative spatial models of seabed sediment composition. PLOS ONE, 10: e0142502. Public Library of Science.

Westhead, R.K., Diaz Doce, D., Cooper, R. & Stevenson, A. 2013. User guide for the British Geological Survey DiGRock250k dataset. *British Geological Survey Open Report, OR/13/038.*

Westhead, R.K., Campbell, E., Carter, G.D.O., Diaz Doce, D., Gafeira, J.D.L., Gales, J.A., Hobbs, P.R.N., Long, D. & Mellett, C.L. 2014. Geological Constraints on Development across the UK Continental Shelf: a study for The Crown Estate. British Geological Survey Commissioned Report, CR/14/050.

Wilson, M.F.J., O'Connell, B., Brown, C., Guinan, J.C. & Grehan, A.J. 2007. Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. Marine Geodesy, 30: 3–35. Taylor & Francis. http://dx.doi.org/10.1080/01490410701295962.