AGRI-FOOD
& BIOSCIENCES
INSTITUTE

Mapping of potential sandeel habitat within the Northern Irish inshore region

January 2022

Document version control

Version	Issue date	Modifier	Note	Issued to
1.0	05/01/2022	EO'K	First draft	AB, AC
1.1	21/01/2022	EO'K	Second draft	MS, EH
1.2	31/01/2022	EO'K	Final version	DAERA
1.3	20/05/2022	EO'K	Final version incorporating DAERA comments	DAERA

Further information

Eugene O'Kane
Marine Ecology
Fisheries & Aquatic Ecosystems Branch
Newforge Lane
Belfast
BT9 5PX

Tel: 028 9025 5653
email: Eugene.O'Kane@afbini.gov.uk

The GIS project "Sandeel habitat mapping 2022" should be available for use in conjunction with this report

Recommended citation:
AFBI, 2022. Mapping of potential sandeel habitat within the Northern Irish inshore region. Report to the Department of Agriculture, Environment and Rural Affairs (DAERA).

Acknowledgements
The author wishes to thank Christian Wilson and Louise Brown (both AFBI) for their assistance in the creation and review of the Random Forest modelling produced here. I would also like to thank the masters and crews of FPV *Queen of Ulster*, R.V. *Corystes* and F.V. *Causeway Lass*.

Table of Contents

Table of Figures

Executive summary

Sandeels, *Ammodytes* spp., are a vital component in marine food webs, providing an important food source for a range of protected birds and mammals, as well as commercial fish species. Under the EU Birds Directive (2009/147/EC) Special Protection Areas (SPAs) are designated to protect bird species, usually at nesting/overwintering sites. Recently however, emphasis has shifted to foraging grounds with policy makers recognising the necessity to protect birds' food species. Following the UK exit from the EU, the EU Birds Directive has been transposed into NI law through the Conservation (Natural Habitats, etc.) (Amendment) (Northern Ireland) (EU Exit) Regulations 2019.

Sandeels are known to completely bury themselves in sediment at night and for most of the winter, with research from the North Sea consistently showing a preference for sediment of a very specific composition: > 15% particles 0.25 – 2 mm (medium and coarse sand), as well as avoidance of sediment with high proportions of < 0.25 mm (fine sand and silt). This study reviewed AFBI's historic ground-truthing data and multibeam echosounder holdings within the Northern Irish inshore zone (< 12 nM), ascertaining spatial coverage and highlighting where data gaps existed. A complementary survey plan, focussed on grab sampling, was then designed in order to enhance the seabed ground-truthing data holdings and enable adequate, informative habitat suitability modelling to take place.

A total of 273 stations were sampled during the project. At each of these stations a sediment sample was taken using a 0.1 m^2 Day grab. Particle size analysis (PSA) was carried out on these and results were then compiled along with 721 historic AFBI stations where PSA data or appropriate underwater video data were available. Stations were then assigned a sandeel habitat suitability (Suitable/Unsuitable) based on studies which have investigated sandeel habitat preferences. This allowed for the creation of an interpolated suitability raster in ArcGIS which was then used along with available environmental variables as inputs to produce habitat prediction models. Due to constraints in geographical extent of the input parameters three models were produced;

- **Model 1** used sediment suitability, MBES bathymetry and its derivatives, slope angle (degrees), aspect (northness and eastness), Bathymetric Position Index (both 3 and 50 radii resolution) for the entire Northern Irish coast.
- **Model 2** used all Model 1 inputs plus current speed (m s^{-1}), derived from a Delft3D hydrodynamic model providing average, maximum 25th and 90th percentiles, for the eastern coast from Garron Point to Carlingford Lough.
- **Model 3** used all Model 1 inputs plus MBES backscatter for the north coast area from Magilligan Point to Torr Head.

Random Forest modelling was then implemented using R Studio to create sandeel habitat suitability prediction maps. Input variable analysis showed that Model 1 was 80% reliant on the sediment suitability layer and as such, was not used for any subsequent analysis. Model 2 showed high probability of suitable sandeel habitat ($\geq$ 60%) at the mouth of Belfast Lough, around the Copeland Islands, east and south-east of the Feathers and a thin strip close to the shore from The Feathers north to Skullmartin. Patches with a slightly lower probability (approx. 50%) were identified at the mouth of Strangford Narrows, the southern extreme of Murlough Bay and the approaches to Carlingford Lough. Model 3 shows the majority of the seabed from Portstewart east to Whitepark

Bay and extending out to the extent of the inshore zone as being $\geq$ 60% likely to be suitable habitat for sandeels.

Following identification of potential sandeel habitat an evaluation was then completed to determine the overlap between these areas and SPAs and other Marine Protected Areas. It was shown that on the east coast of Northern Ireland, probable sandeel habitat overlaps the East Coast Marine pSPA and also North Channel SAC. This is not the case however on the north coast, where only a tiny percentage of probable sandeel habitat is found within the SPA network. Roughly 1/3 in this area overlaps the Skerries and Causeway SAC, leaving approximately 19,280 Ha of $\geq$ 60% probability suitable sandeel habitat outside of current MPA boundaries.

1. Introduction

1.1 Introduction to the study

Ammodytes marinus and *A. tobianus* are planktivorous pelagic fish that are vital constituents of marine food webs and provide an important trophic link between primary producers and top predators (Macer, 1966; Frederiksen *et al.*, 2006; van der Kooij *et al.*, 2008; Eliasen *et al.*, 2011). They are a primary component in the diet of a number of protected seabirds, for example, terns (*Sterna* spp.), razorbill (*Alca torda*), Manx shearwater (*Puffinus puffinus*), common guillemot (*Uria aalge*), cormorants and shags (*Phalacrocorax* spp.) (Thompson, 1987; Barrett *et al.*, 1990; Harris & Wanless, 1991; Engelhard *et al.*, 2014). Grey seals (*Halichoerus grypus*), harbour porpoise (*Phocoena phocoena*) and minke whale (*Balaeonoptera acutorostrata*), all of which are protected under the habitats directive (92/43/EEC), also feed intensively on *Ammodytes* spp., and sandeels provide a significant component in the diet of commercially important fish species, for example, mackerel (*Scomber scombrus*), haddock (*Melanogrammus aeglefinus*), cod (*Gadus morhua*) and whiting (*Merlangius merlangus*) (Santos *et al.*, 2004; Greenstreet *et al.*, 2006; Engelhard *et al.*, 2014).

The importance of *Ammodytes* spp. as a food source for protected birds cannot be understated and a number of sites in Northern Irish inshore waters have been designated as Special Protection Areas (SPAs) under the Birds Directive (79/409/EEC) as a result of the presence of breeding populations of these birds in Carlingford Lough, Strangford Lough, Outer Ards, Copeland Islands, Belfast Lough, Larne Lough, Rathlin Island and Sheep Island. It should be noted that following the UK exit from the EU, the EU Birds Directive has been transposed into NI law through the Conservation (Natural Habitats, etc.) (Amendment) (Northern Ireland) (EU Exit) Regulations 2019. Frederiksen *et al.* (2006) linked seabird breeding success in south-east Scotland to sandeel recruitment in the preceding year, and likewise, Furness & Tasker (2000) illustrated the sensitivity of certain North Sea seabird breeding colonies to declines in sandeel abundance, notably five species of tern, whose small size and limited foraging range identified them as particularly vulnerable. Four species of tern, common (*Sterna hirundo*), roseate (*S. dougallii*), Arctic (*S. paradisaea*) and sandwich (*Thalasseus sandvicensis*) are qualifying species in the SPA network listed above, present in both nationally and internationally important populations. Studies on the fragility of seabird breeding colonies highlight the failings in protecting bird nesting sites only whilst neglecting consideration of the availability of prey items required to ensure breeding success (Thaxter *et al.*, 2012). This oversight has the potential to negate the protective intentions of the existing SPA network (Chivers *et al.,* 2013).

Extensive work has been carried out on sandeel behaviour and their habitat preferences due to the historic importance of the fish as a commercial species in the North Sea. Targeted for industrial purposes, fish meal and fish oil, rather than for human consumption (Macer, 1966; Holland *et al.*, 2005; Jensen *et al.*, 2011) the sandeel fishery began in the 1950's, expanding rapidly in the 1970s and by the 1990s it had become the largest single species fishery in the North Sea, with over 1 million tonnes being landed in some years (Pedersen *et al.*, 1999; Furness, 2002; Engelhard *et al.*, 2014). Since 1997 there has been a sharp decline in landings but the 2009-2018 average still exceeded 300,000 tonnes per year (ICES, 2019). The enormity of the fishery and the dependence on sandeels by top predators led to the need for greater understanding of sandeel distribution (Wright *et al.*, 2000) and so a number of studies were carried out to investigate the determining physical factors. The importance of sediment particle size was recognized (Wright *et al.*, 2000) and both laboratory and field based observations (Holland *et al.*, 2005) identified a strong habitat preference for a very particular sediment grain size range of 0.25 mm – 2 mm.

Unlike the North Sea, there is no sandeel fishery in the Irish Sea, North Channel area (ICES Area VIIa and south VIa) and as such, the considerable body of work that has been carried out on sandeel habitat preference in the North Sea has not been replicated in the latter areas. These areas also lack the documented and well-defined sandeel inhabited areas as seen in Macer (1966), Pedersen *et al.* (1999), Christensen *et al.* (2008), Jensen *at al.* (2011) and Engelhard *et al.* (2014), among others, that come with decades of commercial fishing, leaving a knowledge gap that limits decision making in both conservation and sustainable fishery management in the Northern Irish inshore region. Through the use of acoustic surveys followed by targeted grab surveys for ground-truthing, this study aims to determine locations throughout the Northern Irish inshore zone that are suitable for *Ammodytes* spp. habitation. Furthermore, an assessment of overlap between identified sandeel habitat and the boundaries of existing SPAs and other MPAs will be completed and analysis of seabird feeding habits and foraging ranges will be undertaken to ascertain potential feeding areas.

Random Forest modelling will be undertaken on all data used within the project. This technique is a decision tree based classifier used for a supervised classification and was developed by Breiman (2001). It has been successfully applied in ecological fields over recent years (Bargiel, 2013; Torres & Qiu, 2014; Turner *et al.*, 2018) and was deemed the most applicable method to achieve the project aims.

1.2 Sandeel ecology and behaviour

Sandeel is the common name applied to fishes of the family Ammodytidae, a group of phylogenetically related fish species (van Deurs *et al.*, 2012), five of which occur in the north-east Atlantic. Identification of these species can be difficult and as a result, they are frequently grouped (Camphuysen & Henderson, 2017; Jørgensen *et al.*, 2017). Two of the smaller of these species belong to the genus *Ammodytes*, *A. marinus* and *A. tobianus*, and they both receive the name lesser sandeel. The two species are difficult to differentiate as they vary only in the number of dorsal fin rays and vertebrae and with *A. marinus* lacking scales on the base of the tail fin lobes (Wheeler, 1978). The key difference between the species is their spatial distribution where, typically, *A. tobianus* would be considered an almost exclusively inshore species, found in depths of up to 30 m, including intertidally during high tide (Gibson *et al.*, 1996; Jovanovich *et al.*, 2007; Kellnreitner *et al.*, 2011), while *A. marinus* is predominantly found in offshore waters, 30 – 150 m deep (Reay, 1970; Wheeler, 1978; Wright *et al.*, 2000; Tien *et al.*, 2017). The two species will be referred to hereafter as sandeels.

Sometimes described as forage fish, sandeels are a small, highly abundant, pelagic, schooling species that occupy a mid-trophic position feeding on zooplankton (Frederiksen *et al.*, 2006; Engelhard *et al.*, 2014). Although a range of prey items are taken, Macer (1966) showed that copepods such as *Temora*, *Calanus* and *Pseudocalanus* contribute most to their diet. Sandeels are specialist burrowers, spending the majority of their life buried, either completely or partially, in sandy substrates, emerging only to feed in the water column in late spring and summer and to spawn in the winter (Winslade, 1974; van der Kooij *et al.*, 2008). The particle size composition of these sediments is the primary determining factor for habitation, with a very specific range of particle size being preferred. Medium and coarse sands, in the range of 0.25 – 2 mm, have been identified in laboratory choice experiments (Wright *et al.*, 2000) and field observations (Wright *et al.*, 2000; Holland *et al.,* 2005) as being actively selected, with high proportions of fine sands and silt, < 0.25 mm, being avoided. More specifically, sediments containing over 4% silt (silt fraction < 63 µm) were significantly less likely to be chosen, and those containing over 10% were avoided completely (Wright *et al.*, 2000; Holland *et al.,* 2005). Tagging experiments suggest high levels of sand bank fidelity, and in a study of otolith microchemistry Wright *et al.* (2018) found no extensive movement of post-settled juveniles. Even as adults, sandeels remain relatively close to their night-time habitats when foraging, staying within approximately 20 km, utilising tidal currents to travel due to limited swimming abilities (Engelhard *et al.*, 2008). This information indicates that determining the preferred substrate habitat choice for sandeels will also provide approximate locations when they utilise the water column, important for generating policy regarding foraging seabirds.

For more detail on *Ammodytes* spp. ecology, life history, and the reasons behind this project see accompanying literature review “Habitat preferences of sandeels, *Ammodytes* spp.” (AFBI, 2020.)

2. Methodology

2.1 Ground-truthing data collation

2.1.1 Historic data collation

AFBI historic data holdings were reviewed to locate geo-referenced sediment analysis samples and underwater video data. Only sediment samples with full PSA results and GRADISTAT output were considered. A variety of data sources proved valuable and were made available to the project, as listed in Table 1.

Table 1. Historic data sources made available for this project

Survey and Area	Data type	Number of sample points used
INIS Hydro, Dundrum Bay & Mourne Coast, 2011	Sediment PSA	47
AFBI Strangford Lough Habitat Map, 2015	Sediment PSA	60
AFBI Habitat Data for Marine Conservation Zones (MCZs); Rathlin Island, Ballycastle Bay, 2015	Sediment PSA	7
AFBI Outer Ards *Modiolus* Assessment, 2016	Sediment PSA	25
Ulster University RV *Celtic Explorer* students cruise, Skerries, 2009	Sediment PSA	13
AFBI Belfast Lough Dredge & Grab Survey, 2012	Sediment PSA	50
AFBI Annual NMP Survey, 2016	Sediment PSA	6
AFBI Habitats Regulations Assessment, Killowen, 2016	Sediment PSA	9
British Geological Survey/JNCC	Sediment PSA	140
AFBI, Nearshore Subtidal Habitat Mapping, Rathlin 2009	Underwater Video	32
MESH project, Laconia Bank, 2006	Underwater Video	31
AFBI Marine substratum and biotope maps, Maidens, 2009	Underwater Video	73
AFBI Fairhead Tidal Energy Benthic Assessment, 2014	Underwater Video	104
MESH project, Shamrock Pinnacle, 2006	Underwater Video	19
AFBI Cruise CO1519, North Channel & Skerries Causeway, 2019	Underwater Video	63
AFBI Annual Seed mussel surveys, 2017-2019	Underwater Video	42

These data were examined within ArcGIS v10.6 to ascertain spatial coverage and to highlight where data gaps existed. Once completed, a complementary survey plan, focussed on grab sampling, was designed in order to enhance the seabed ground-truthing data holdings and enable adequate, informative habitat suitability modelling to take place.

Underwater video data points were only included where certainty of substrate suitability for sandeels could be derived, i.e. when footage showed sandeels clearly exiting the substrate or where bedrock and boulders dominated making sandeel habitation impossible. Although a potential form of bias, these strict rules were adhered to in order to utilise an otherwise unusable data source, and allowed for the inclusion of 364 appropriate sample points in the study.

2.1.2 Benthic grab sample surveys

Sampling surveys were carried out from the Department of Agriculture, Environment and Rural Affair's Fisheries Protection Vessel *Queen of Ulster*, along with one survey from Portrush charter vessel *Causeway Lass*, during daylight hours between October 2019 and April 2021. A 0.1 m^2 Day grab was deployed from *Queen of Ulster* to obtain quantifiable seabed sediment samples. Sample depth retained by the grab was recorded (cm). The sample was photographed, a description of the sediment surface was documented and a subsample of sediment was taken for particle size analysis (PSA). The sample was then examined for presence of *Ammodytes* spp. and where present, these were counted and Total Length of each specimen was measured in mm (tip of the snout to tip of the longer lobe of the caudal fin).

Day grab sampling was not possible on *Causeway Lass* so a pipe dredge with a closed end was deployed instead. The collected sample was photographed, a description of the sediment surface was documented and a sediment subsample was taken for PSA and the dredge was examined for *Ammodytes* spp.

A total of 273 stations were sampled by these methods (251 Day grab, 22 pipe dredge), which were then compiled along with datasets from the 217 historic sediment samples identified earlier in the process for use in model development.

2.2 Multibeam echosounder data

AFBI data holdings were also reviewed for relevant multibeam echosounder (MBES) bathymetry data from within the Northern Irish inshore zone (< 12 nautical miles). MBES data came from a variety of surveys and different vessels, as can be seen in Table 2, all of which were acquired to IHO Order 1.

Table 2. AFBI Multibeam echo sounder data sources holdings used in modelling.

Survey Project	Survey Area
Joint Irish Bathymetric Survey (JIBS) 2008	Tuns Bank to Runabay Head
United Kingdom Hydrographic Office, Civil Hydrography Programme	Runabay Head to Belfast Lough, 0-80 m
British Geological Survey	Outer Belfast Lough & North Channel
United Kingdom Hydrographic Office, Civil Hydrography Programme	Belfast Lough to St. John's Point
INIS Hydro 2011	Dundrum Bay, Mourne Coast & Carlingford Lough

These available data were post-processed to yield Bathymetric Position Index (both 3 and 50 radii resolution), aspect (northness and eastness) and slope angle (degrees) in ArcGIS 10.6 using Benthic Terrain Modeler 3.0 (Walbridge et al., 2018).

Although MBES backscatter (dB) had been acquired during these surveys only those backscatter data from the JIBS were of a usable quality. However, the project had been conducted from three different survey vessels using different sensors and settings. Therefore, harmonisation of the three datasets was required in order to utilise these backscatter data in the creation of sandeel habitat maps. This was possible due to the overlap between the surveys and a statistical calibration approach was undertaken to implement this. This work was subcontracted to Ulster University and the full report is attached in Appendix I.

2.3 Sediment analysis

2.3.1 PSA and habitat suitability criteria

The collected sediment samples were analysed with a combination of dry sieve and laser diffraction following NMBAQC guidelines. Results were via GRADISTAT output providing proportion of each grade of particle size, from clay (grain size < 0.0019 mm) to very coarse gravel (grain size = 64.0 mm).

A sandeel habitat suitability criterion was then allocated to each sample station, largely based on two methods used from comprehensive North Sea studies into *Ammodytes* spp. habitat preferences:

a) After Greenstreet *et al.* (2010) where particles 0.0001 mm – 0.25 mm were assigned "silt & fine sand" and 0.25 mm – 2.00 mm were assigned "coarse sand". The proportion of these classes in each sediment sample would determine suitability for *Ammodytes* spp. habitat on the scale "Unsuitable", "Suitable", "Sub-prime" or "Prime".

b) After Holland *et al.* (2005) where sediments were characterised as either "Suitable" or "Unsuitable". A sample having one of the following determinants defined it as suitable: ≤ 1% 0.0039 mm – 0.016 mm OR > 55% 0.25 mm – 0.71 mm OR ≤ 2% 0.016 mm – 0.063 mm OR > 15% 0.71mm – 2.00 mm.

   A sample was unsuitable when all of the following was determined:

   > 1% 0.0039 mm – 0.016 mm AND ≤ 55% 0.25 mm to 0.71 mm AND >2% 0.016 mm – 0.063 mm AND ≤ 15% 0.71 mm – 2.00 mm

Classifications resulting from these methods agreed for the majority of samples. However, 135 of 490 samples were given conflicting classifications, with the Greenstreet *et al.* (2010) method showing unsuitability while Holland's concluded suitability. In these cases consideration was therefore given to the silt content as a whole (< 0.0625 mm) and also the Fine Gravel content (2 – 8 mm) as Holland *et al.* (2005) showed sediments with a composition between 5 – 20% Fine Gravel were favourable to sandeel habitation and sediments containing > 30% Fine Gravel were avoided. This allowed for an appropriate suitability criteria to be applied to these conflicting classifications.

2.3.2 Inclusion of British Geological Survey data

Data from BGS provided ternary classification only (percentage gravel, sand and mud) which was not sufficiently comprehensive in the gravel and sand categories to be utilised for sandeel habitat criteria allocation. However, used in isolation, the mud category could be informative since it is known that *Ammodytes* spp. actively avoid sediment containing >10% silt. In order to strengthen this assumption, descriptive statistics were gathered from all samples that had been assigned a suitability according to both Greenstreet and Holland. Mean and median silt content (0.0001 mm – 0.063 mm) of samples grouped by suitability, according to both Greenstreet and Holland's methods, were determined and boxplots produced. This then allowed for an acceptable mud percentage threshold to be established, over which samples from the BGS dataset could be deemed "Unsuitable". A value of 14% was chosen based upon:

Sediment sample "Unsuitable" category	- Mean 14.30%	Median 12.50%	
Sediment sample "Suitable" category	- Mean 4.66%	Median 3.84%	75th percentile 6.73%

This method could only be used to define unsuitable areas, not suitable areas, due to the significant contribution towards suitability by a combination of the other parameters not described in the dataset (Fine-, Medium- and Coarse sand, and Fine gravel). Therefore, only 140 of these sample points were included in the habitat suitability modelling.

2.4 Model

An interpolated sediment suitability raster was created using the Kriging tool in ArcGIS on a point dataset based on the suitability criterion applied to the ground-truthing data. All bathymetry and environmental rasters were standardised to a 5 m by 5 m resolution grid and projected in UTM Zone 29N co-ordinate reference system using the Spatial Analyst and Data Management toolsets in ArcGIS v10.6. Random Forest was used to produce the habitat suitability model, using R Studio 1.4.1717.

Due to limitations in the geographical extent of variables, three models were produced using different combinations of available parameters. Bathymetry and its derivatives were available for the majority of the Northern Irish inshore zone and so were used along with the interpolated sediment suitability raster as the inputs for Model 1 (Figure 1). Current speed (m s^{-1}) derived from a Delft3D hydrodynamic model was available for the area from Garron Point to Carlingford Lough and so was included along with the parameters from Model 1 to make Model 2 (Figure 2). These current data were interpolated into raster format utilising the Natural Neighbour tool from the Spatial Analyst toolset in ArcGIS v10.6 producing average, maximum, 90th and 25th percentile as parameters. The harmonised MBES backscatter for the JIBS area was included along with the parameters for Model 1 to produce Model 3 (Figure 3). A detailed report of the harmonisation process can be viewed in Appendix I, and for Model 3 the H1 mosaic was judged most appropriate for use, due to its slightly better performance with smaller variance from the original.

The rasters were stacked for each model and values were extracted from all the input rasters at each previously classified sample point, providing the input dataset for the Random Forest model. Splitting the dataset into two, a training set to fit the model and a test set for evaluation, the machine learning model then creates predictions for the areas with no ground-truthing data, providing the values to create the habitat prediction rasters.

Independent data from other AFBI Fisheries and Aquatic Branch (FAEB) surveys from the past 5 years were collated in a point shapefile in ArcGIS to provide a visual comparison of predicted habitat suitability and where *Ammodytes* spp. have been caught historically.

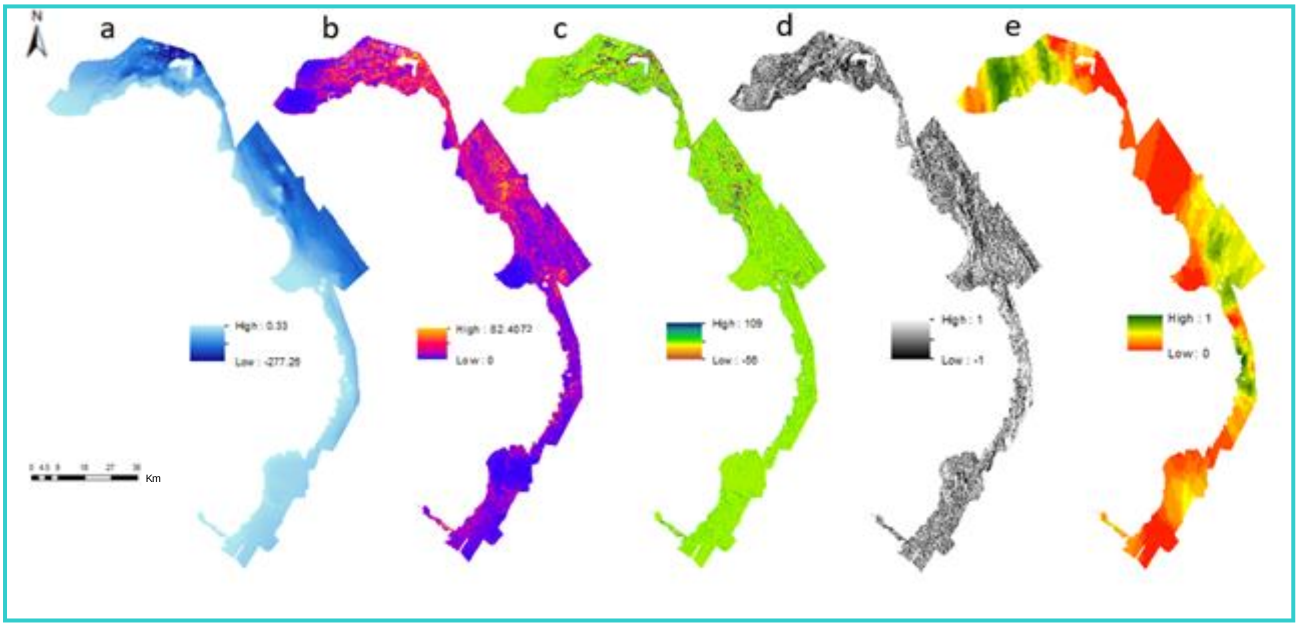


Figure 1. Input rasters for Model 1, (a) bathymetry, (b) slope angle, (c) BPI, (d) aspect and (e) interpolated seabed suitability.

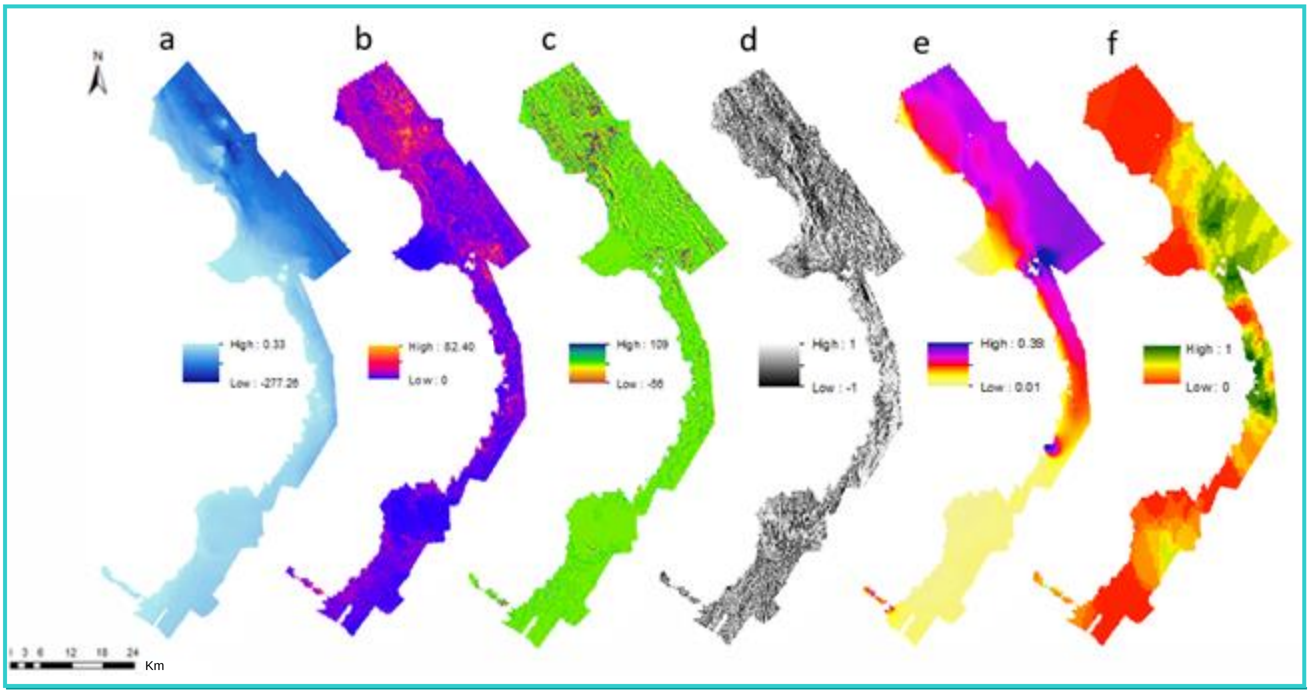


Figure 2. Input rasters for Model 2, (a) bathymetry, (b) slope angle, (c) BPI, (d) aspect, (e) current speed and (f) interpolated seabed suitability.

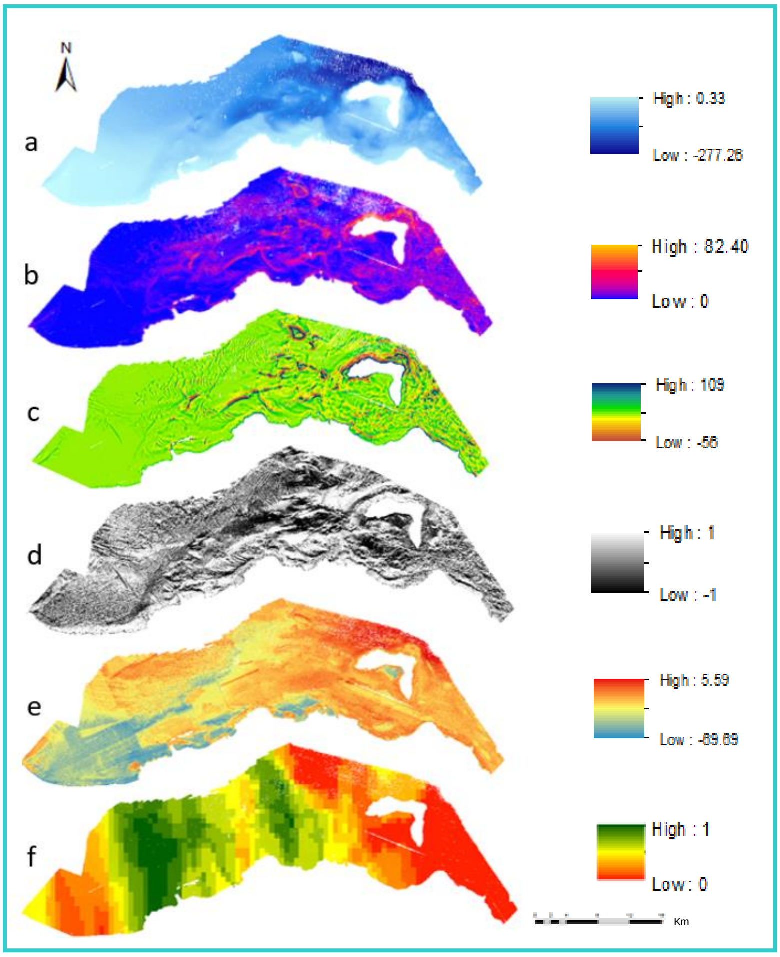


Figure 3. Input rasters for Model 3, (a) bathymetry, (b) slope angle, (c) BPI, (d) aspect, (e) MBES backscatter and (f) interpolated seabed suitability.

2.5 Sandeel habitat assessment

The areas predicted with a high probability (> 60% and > 70%) of being potential sandeel habitat were then assessed to locate and calculate areas falling within current SPAs and other Marine Protected Areas (MPAs), such as Special Areas of Conservation (SACs) and Marine Conservation Zones (MCZs), using ArcGIS. A minimum threshold value of 60% was chosen to highlight the greatest area of potentially suitable sandeel habitat to decision makers that confidence of sandeel presence could be applied to. The threshold of 70% was also presented in order to illustrate a reduction in area of suitable sandeel habitat resulting from higher confidence in species presence. Areas with probabilities higher than 70% were very small and would not benefit wider decision making when trying to balance the needs of protecting the seabed versus sustainable seabed use or development. Seabird count data was obtained through JNCC Seabird Monitoring Programme database online (https://app.bto.org/seabirds/public/data.jsp, only data from 2019-2021 used) in order to ascertain which identified high probability sandeel habitat could be potential feeding areas for vulnerable breeding seabird species.

3. Results

3.1 Sandeel habitat suitability predictions

Spatial coverage and a visualisation of the outputs from Model 1 which was developed using bathymetry and its derivatives alongside the sediment suitability raster are shown in Figure 4. Figure 5 presents the spatial coverage and a visualisation of the outputs from Model 2 (developed utilising the same inputs as Model 1 with the addition of current speed) and Figure 6 shows the same for Model 3 (developed utilising the same inputs as Model 1 with the addition of harmonised backscatter data). For the predicted seabed suitability (Figures 4, 5 & 6), areas illustrated red to yellow on the maps show a ≥ 50% probability of being suitable sandeel habitat whilst those areas which range from green to blue represent a ≤ 50% probability of being suitable sandeel habitat.

Performance statistics for the models are presented in Table 3 and Table 4 and variable importance plots for each model are attached in Appendix II.

Table 3. Performance statistics for all models carried out.

Model	Overall accuracy			Class accuracy		% variance explained
	Accuracy	Sensitivity	Specificity	Sensitivity	Specificity	
1	0.86	0.86	0.86	0.66	0.93	52.21
2	0.86	0.86	0.86	0.67	0.93	48.83
3	0.86	0.86	0.86	0.74	0.90	50.72

Table 4. Performance stats continued

Model	Root Mean Square Error	R^2	Kappa	Mean Absolute Error
1	0.285	0.610	0.619	0.184
2	0.286	0.616	0.631	0.201
3	0.295	0.546	0.624	0.176

The sea immediately off the Ards Peninsula shows the greatest potential for sandeel habitat in the east coast model area, specifically identifying a large patch to the east and south east of the Copeland Islands, a patch to the south and east of The Feathers and a thin strip running close to the shore (Figure 5). These areas show between 85% – 95% chance of being suitable sandeel habitat. There is another sizeable area just off the mouth of Outer Belfast Lough which has a 75% – 80% probability of being sandeel habitat.

In the Model 3 area, the north coast region of Northern Ireland, a much larger and less patchy area of potential sandeel habitat is found (Figure 6). Quite a number of large areas of sandbanks can be seen with very pronounced ripples, some very near shore and others out to 6 nM from the coast, and it is these banks which show the highest probability of being suitable sandeel habitat (approximately 90%).

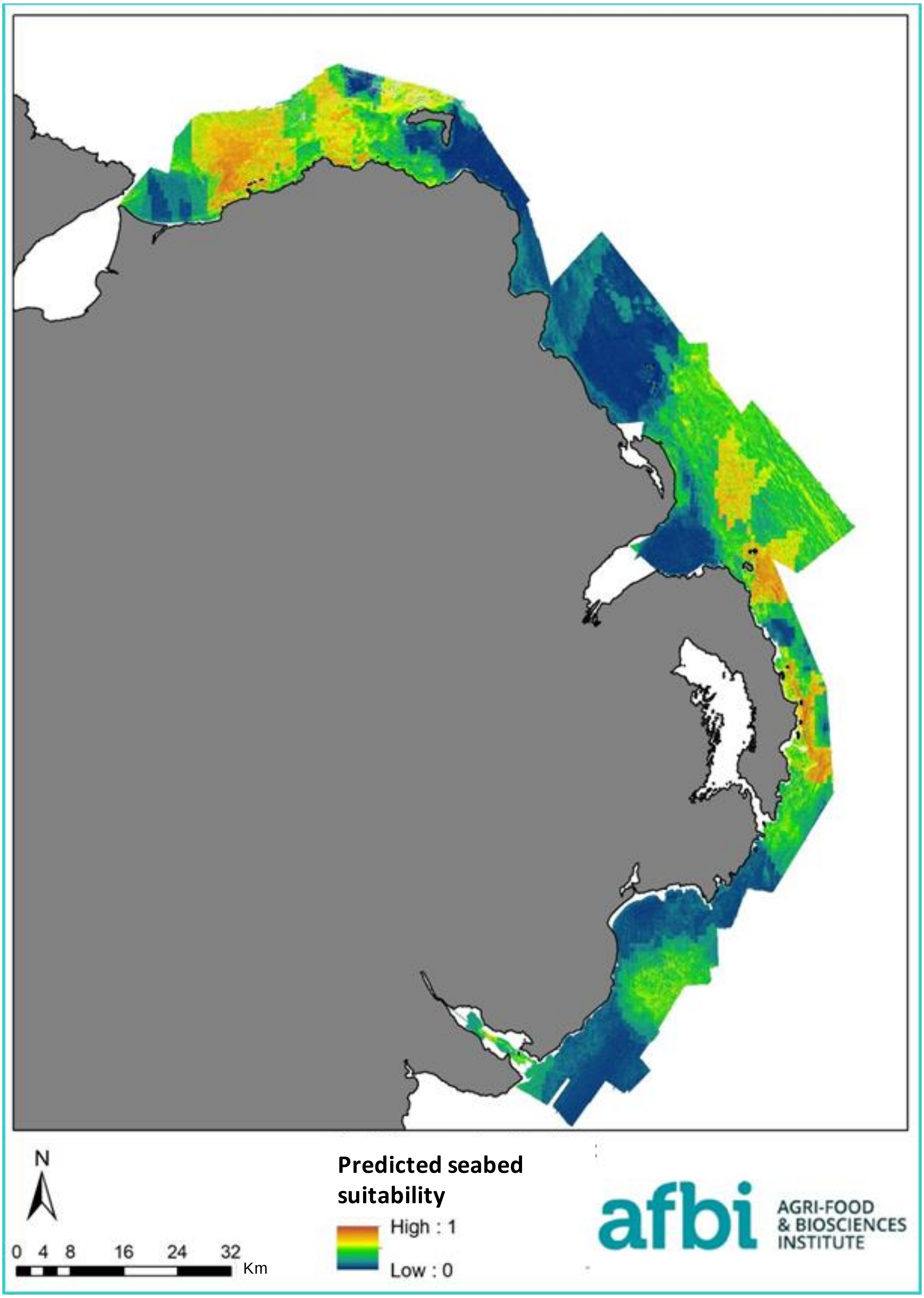


Figure 4. Predicted seabed suitability for sandeel using Model 1.

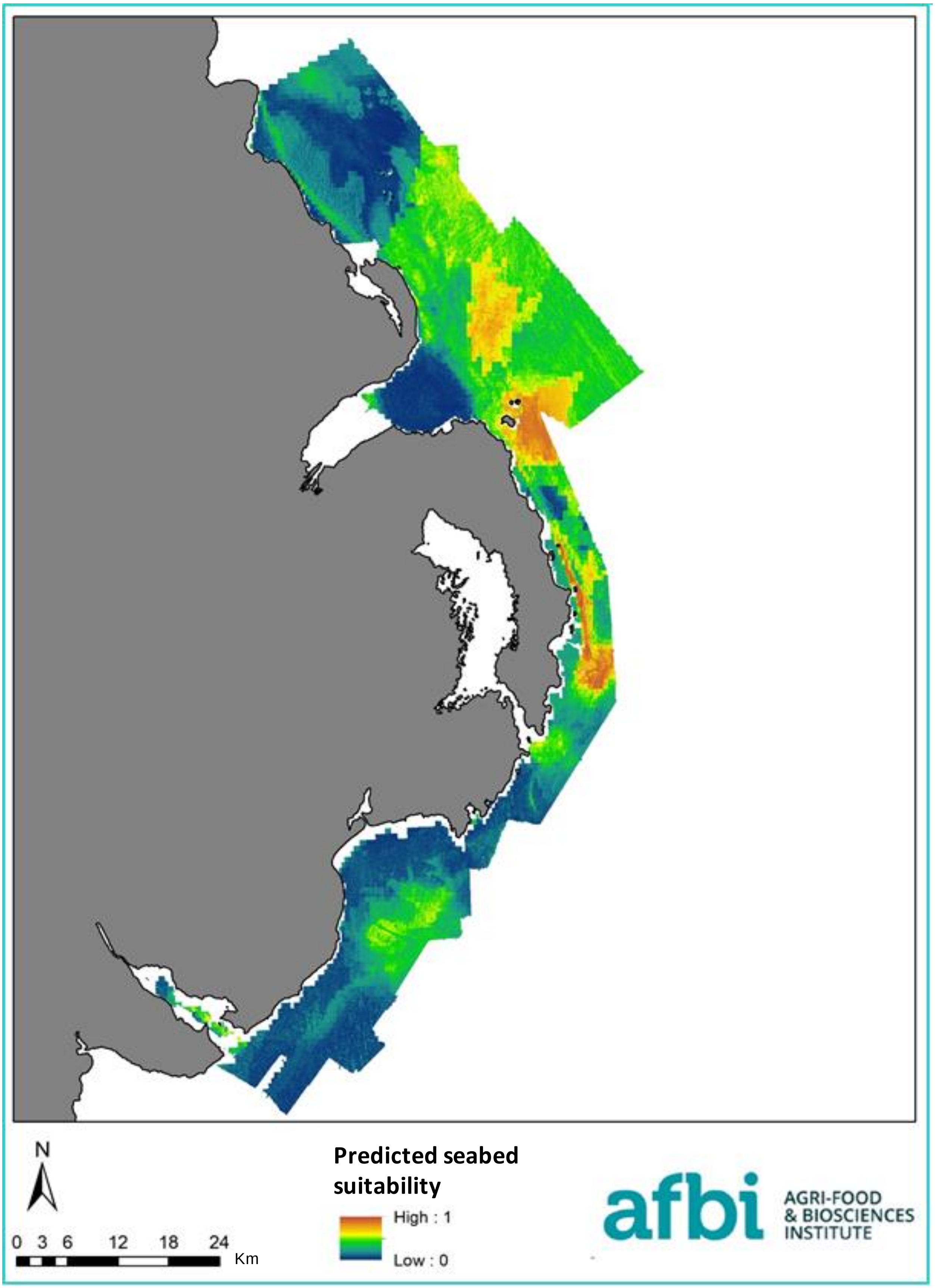


Figure 5. Predicted seabed suitability for sandeel using Model 2.

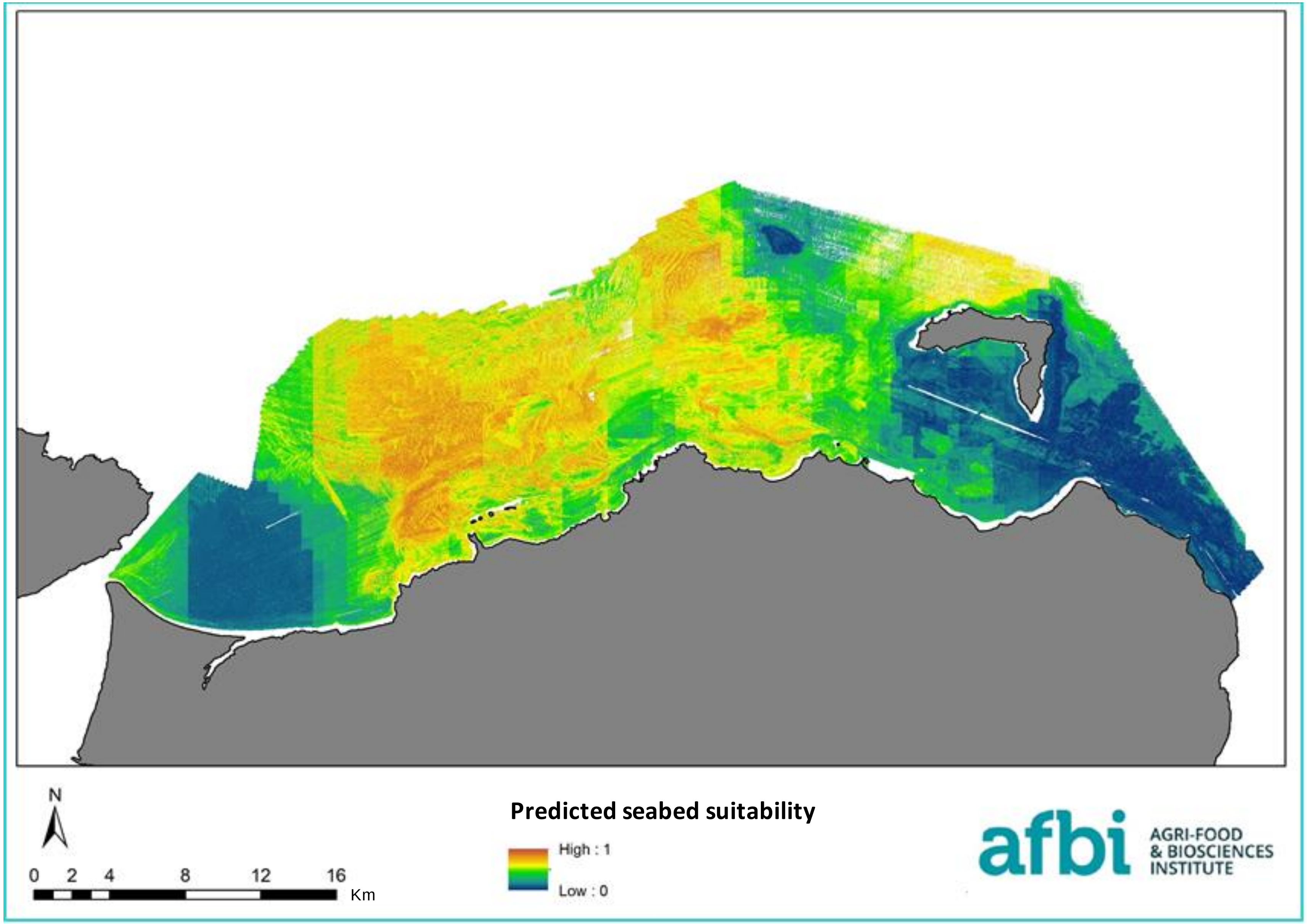


Figure 6. Predicted seabed suitability for sandeel using Model 3.

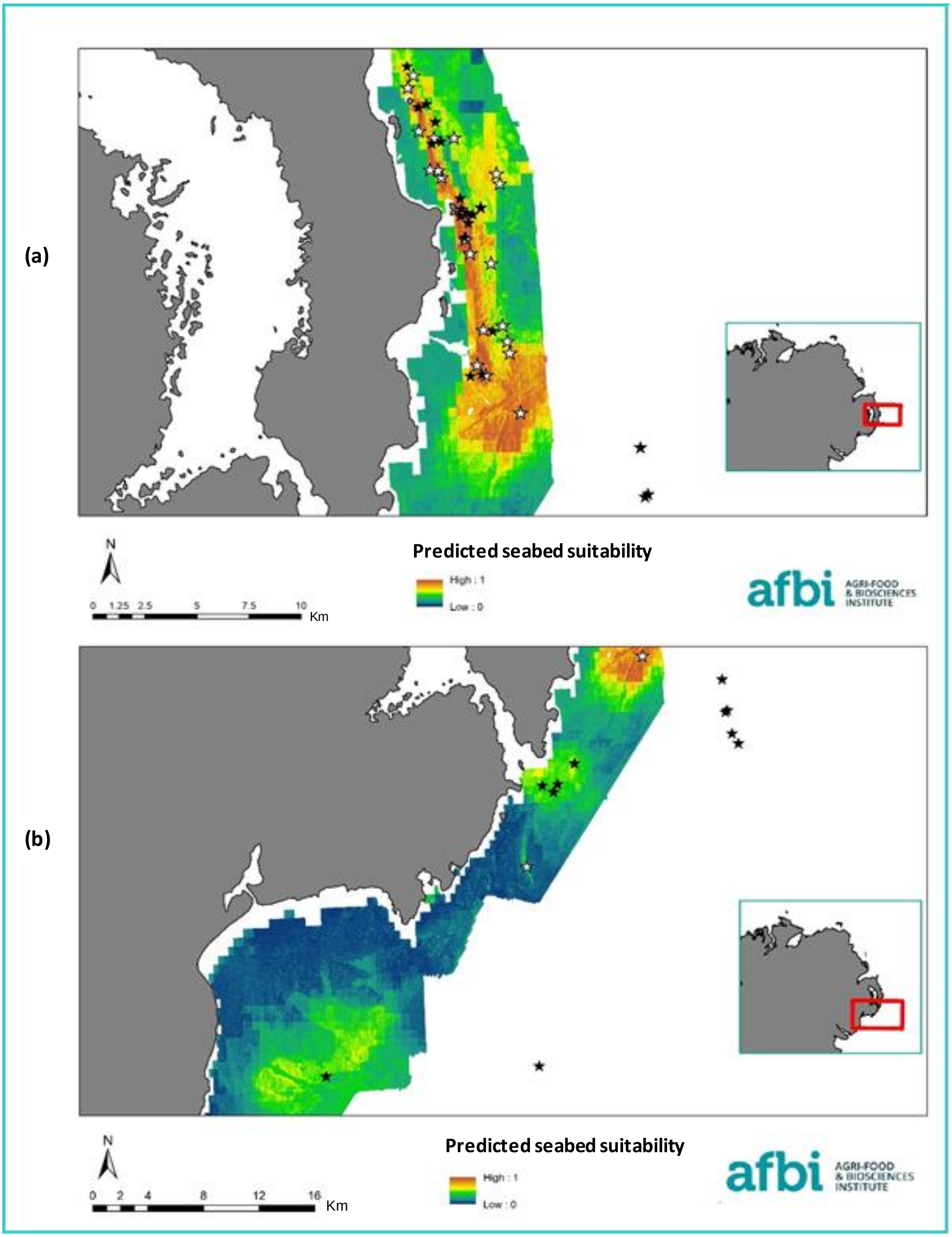


Figure 7. Predicted seabed suitability for sandeel plotted with grab samples collected in this study containing sandeels (white stars) and independent FAEB survey data recording captured sandeels (black stars) in two selected areas of the Co. Down coast – (a) Ards Peninsula from Skullmartin to south of The Feathers; (b) Lecale coast and Murlough Bay.

3.2 Assessment of potential sandeel habitat in relation to Marine Protected Areas (MPAs)

To assess potential sandeel habitat covered by designated areas, the boundaries of MPAs (SPAs, marine SACs and MCZs) within the model areas were compared with the predicted distributions of the sandeel habitat suitability models.

Model 2 overlapped with multiple SPAs including: East Coast Marine pSPA, Strangford Lough SPA, Larne Lough SPA, Belfast Lough SPA, Belfast Lough Open Water SPA, Outer Ards SPA and Carlingford Lough extended SPA, and was adjacent to the Copeland Islands SPA (Figure 8). Model 2 also had substantial spatial overlap with Carlingford Lough MCZ, Murlough SAC, Strangford Lough MCZ, North Channel SAC, Outer Belfast MCZ, The Maidens SAC, Red Bay SAC and Waterfoot MCZ (Figure 9).

The domain of Model 3 encompassed Rathlin Island SPA, was adjacent to the boundary of Sheep Island SPA (Figure 10) and overlapped with Skerries and Causeway SAC, Rathlin Island SAC, Red Bay SAC and Rathlin MCZ (Figure 11).

The total area of currently designated MPAs which contained areas of seabed with high probabilities of being suitable sandeel habitat was then calculated for Model 2 (Table 5) and Model 3 (Table 6).

Table 5. Area of potential sandeel habitat predicted using Model 2, falling under existing Marine Protected Areas.

Sandeel habitat probability	Area Protected under SPA (Ha)	Percentage of Total Area Identified	Area protected under SAC and MCZ (Ha)	Percentage of Total Area Identified
0.6 – 1.0	13,763	91.3%	9,265	61.4%
0.7 – 1.0	8,558	98.2%	6,375	73.1%

Table 6. Area of potential sandeel habitat predicted using Model 3, falling under existing Marine Protected Areas.

Sandeel habitat probability	Area Protected under SPA (Ha)	Percentage of Total Area Identified	Area protected under SAC and MCZ (Ha)	Percentage of Total Area Identified
0.6 – 1.0	241	1.2%	6,206	32.2%
0.7 – 1.0	10	0.1%	2,778	30.6%

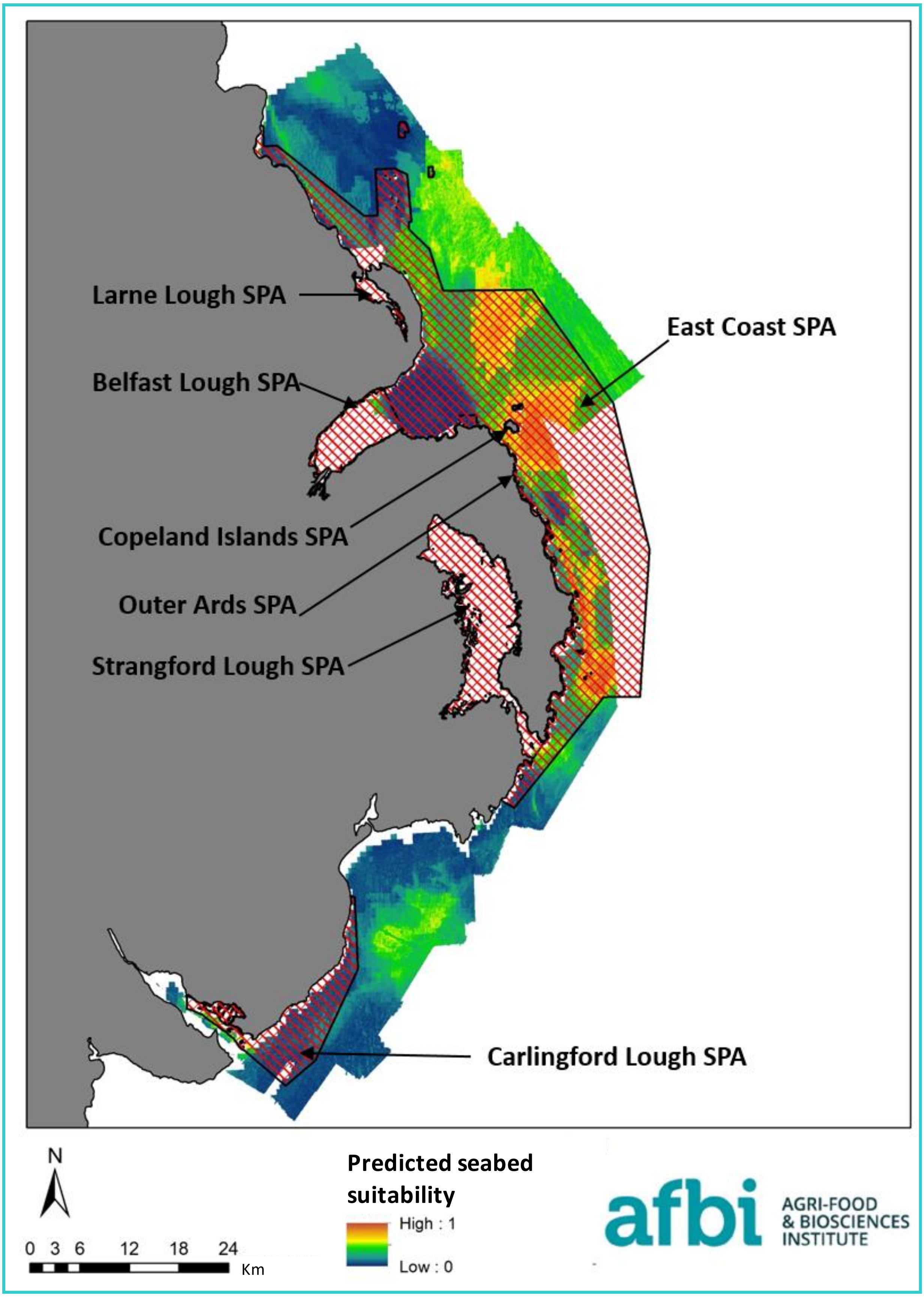


Figure 8. Predicted seabed suitability for sandeel from Model 2 plotted with Special Protected Areas, shown in red crosshatch and labelled.

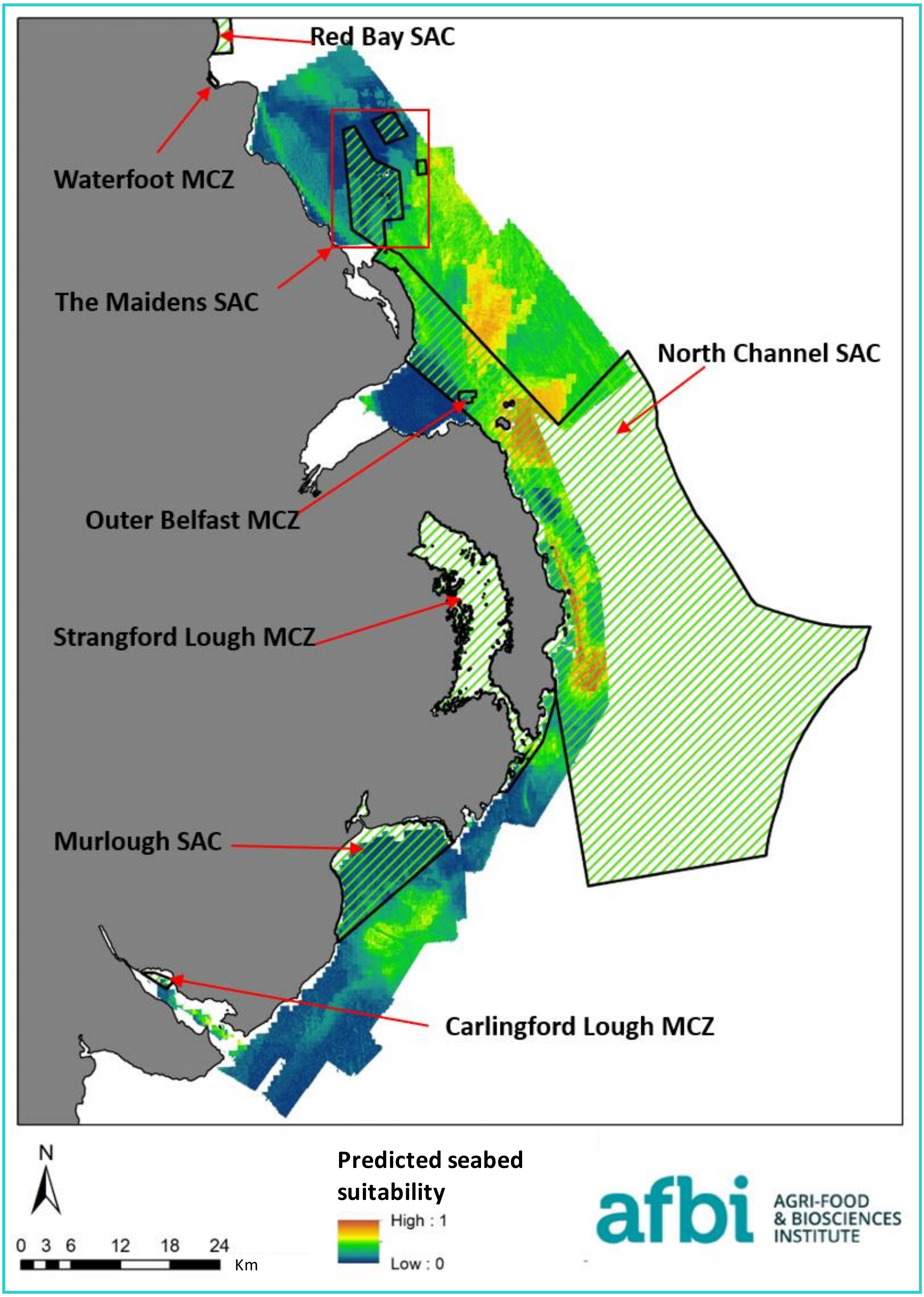


Figure 9. Predicted seabed suitability for sandeel from Model 2 plotted with Special Areas of Conservation and Marine Conservation Zones, shown in green hatch and labelled.

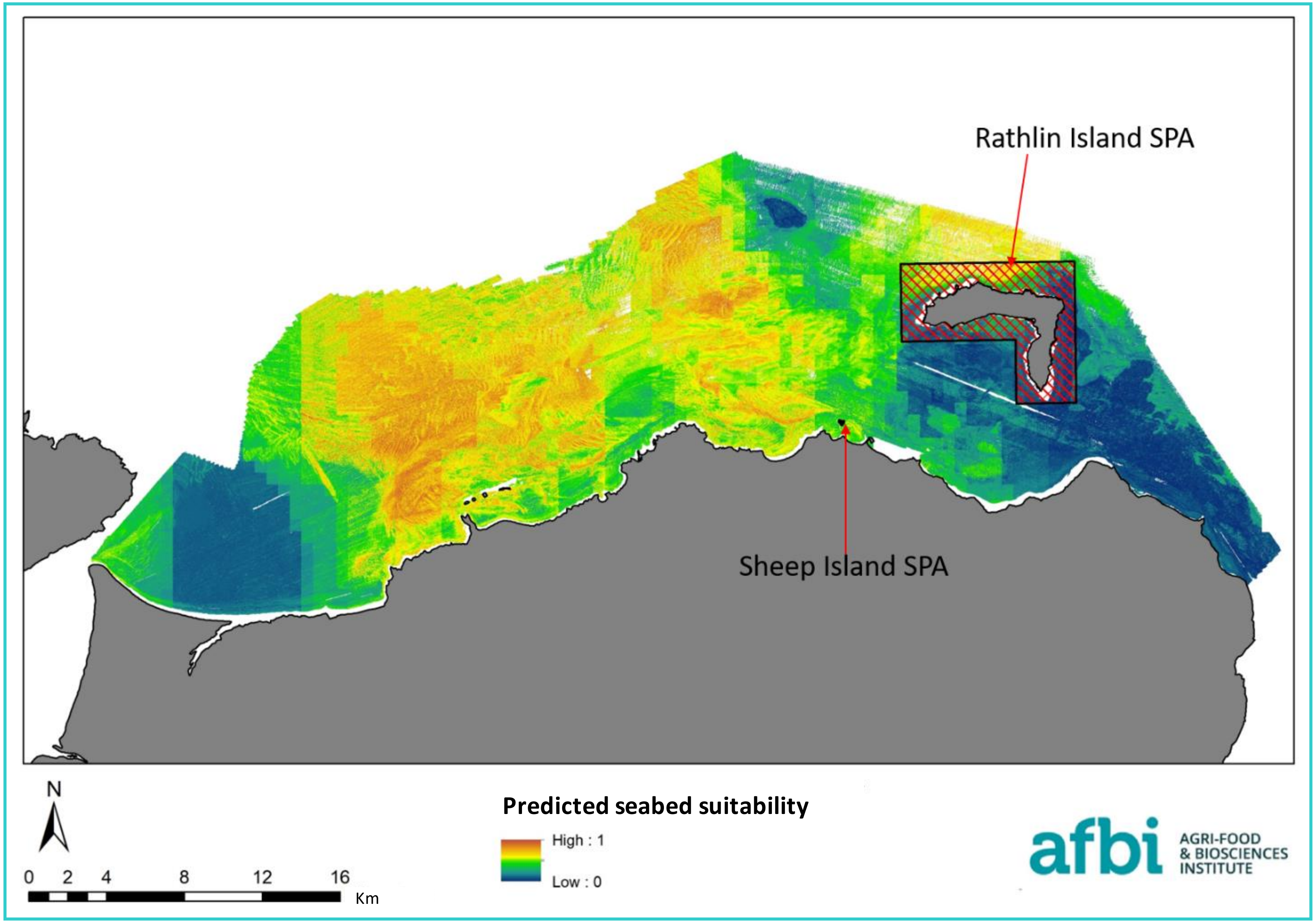


Figure 10. Predicted seabed suitability for sandeel from Model 3 plotted with Special Protected Areas, shown in red crosshatch and labelled.

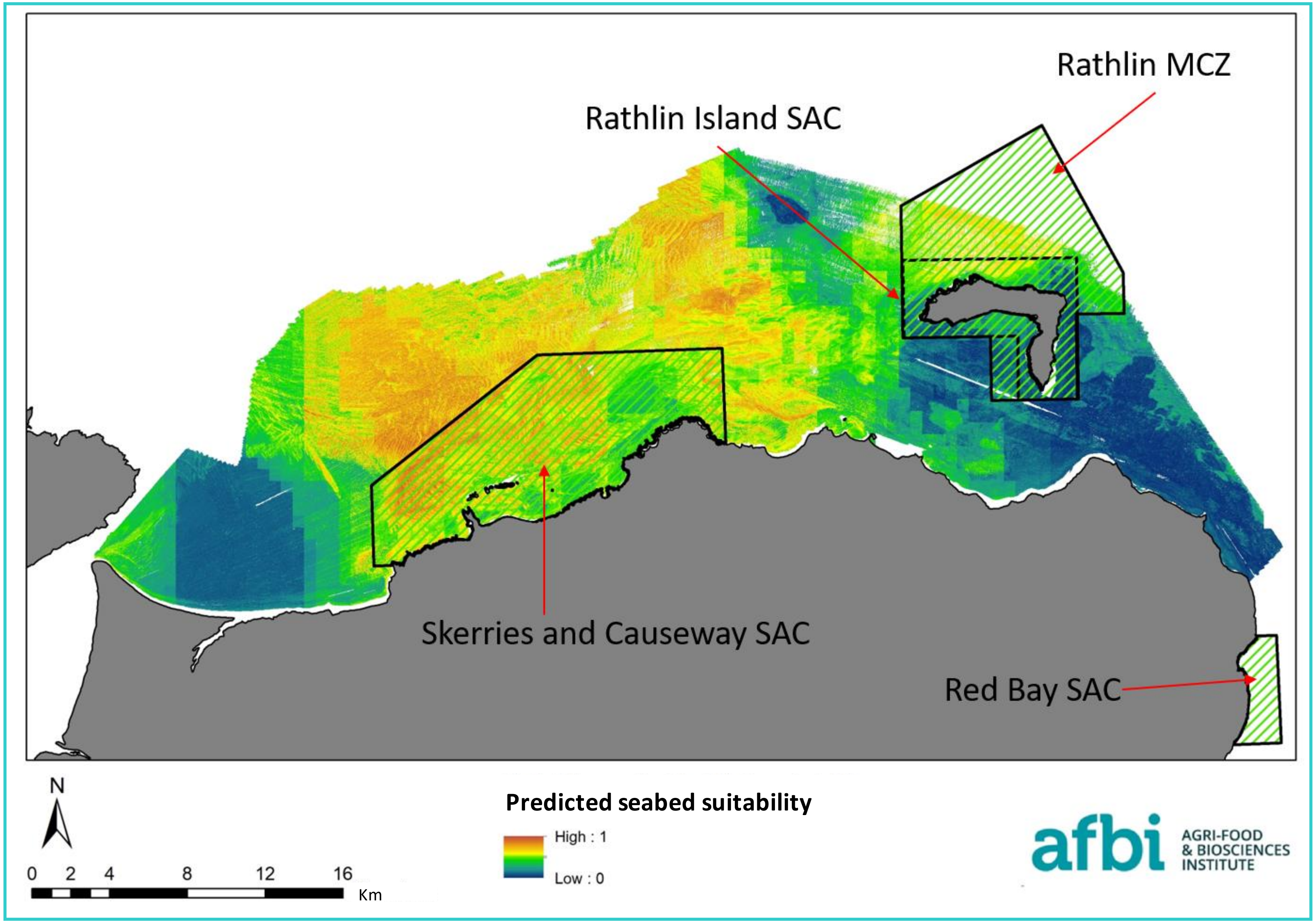


Figure 11. Predicted seabed suitability for sandeel from Model 3 plotted with Special Areas of Conservation and Marine Conservation Zones, shown in green hatch and labelled.

4. Discussion

4.1 Modelling

Random Forest proved to be an effective method of predicting sediment suitability for sandeel habitation over large areas when only relatively sparse or limited ground-truthing datasets are available. Both Model 2 and Model 3 show clear differentiation from Model 1, most notable on the Ards coast and west of Skerries, due to their additional parameters providing strength to the predictions. Performance statistics for each model are shown in Table 3 and although there is no difference in overall accuracies, slight differences can be seen in the class accuracies, where Model 1 showed the lowest sensitivity (ability to detect true positives). This figure was highest for Model 3. Model 1 is 80% reliant on the interpolated sediment suitability layer, in contrast to the other two models where this reliance drops to approximately 40% (see variable importance plots, Appendix II). This knowledge reduces the application of Model 1 to a reference model and its predictions were not used for any of the subsequent analysis. Its use is limited to a visual representation of where the additional variables used in Model 2 and Model 3 strengthened the machine learning (Figure 12).

The various formats of current speed were shown to be the most important variables to Model 2 after sediment suitability. Similarly, backscatter in Model 3 was the next most important after sediment. Since current speed is one of the drivers behind seabed characteristics it follows that it is helpful for locating sandeel habitat, i.e. areas with lower current speeds tend to be dominated by silty sediments not suitable for sandeels, and coarse sands tend to be found in areas of higher current speeds. Coupled with the knowledge that sandeels are thought to require higher current speeds in order to maintain sufficient oxygen concentrations in their burrows (Wright *et al.*, 2000) this variable would likely prove important for future refinement of the models.

Previous studies have shown the value in combining MBES backscatter with bathymetric derivatives and current speed projections in order to deliver a robust set of variables with strong predictive capabilities (Pearman *et al.*, 2020). Applied to the work already completed in this study this combination would enable determination of suitable sandeel habitat with higher accuracy. AFBI data holdings contain substantial areas of MBES backscatter and through the use of the harmonisation methods applied in this project more areas could be investigated for sandeel habitat suitability. The project has also highlighted where gaps in knowledge still exist, in particular, the North Channel area from Fair Head southward to Larne. Ground-truthing can prove difficult in this area due to the combined effects of severe tidal currents, hard ground and rough weather conditions. AFBI also currently holds quite limited MBES data for the northern half of this area. Similarly, the Ards coast has only a relatively thin strip (1-3 nM) of MBES data. If MBES data acquisition were to be extended to the unsurveyed areas further offshore, a clearer picture would be presented of areas adjacent to those that this study has demonstrated to be of high suitability for *Ammodytes* habitat.

Temperature and salinity were omitted from all the models produced through this study because the only datasets made available were of too coarse a resolution and as such would undermine the quality generated by the other variables. Finer scale resolution data may be available at a later date and can be included in future refinement. However, various studies

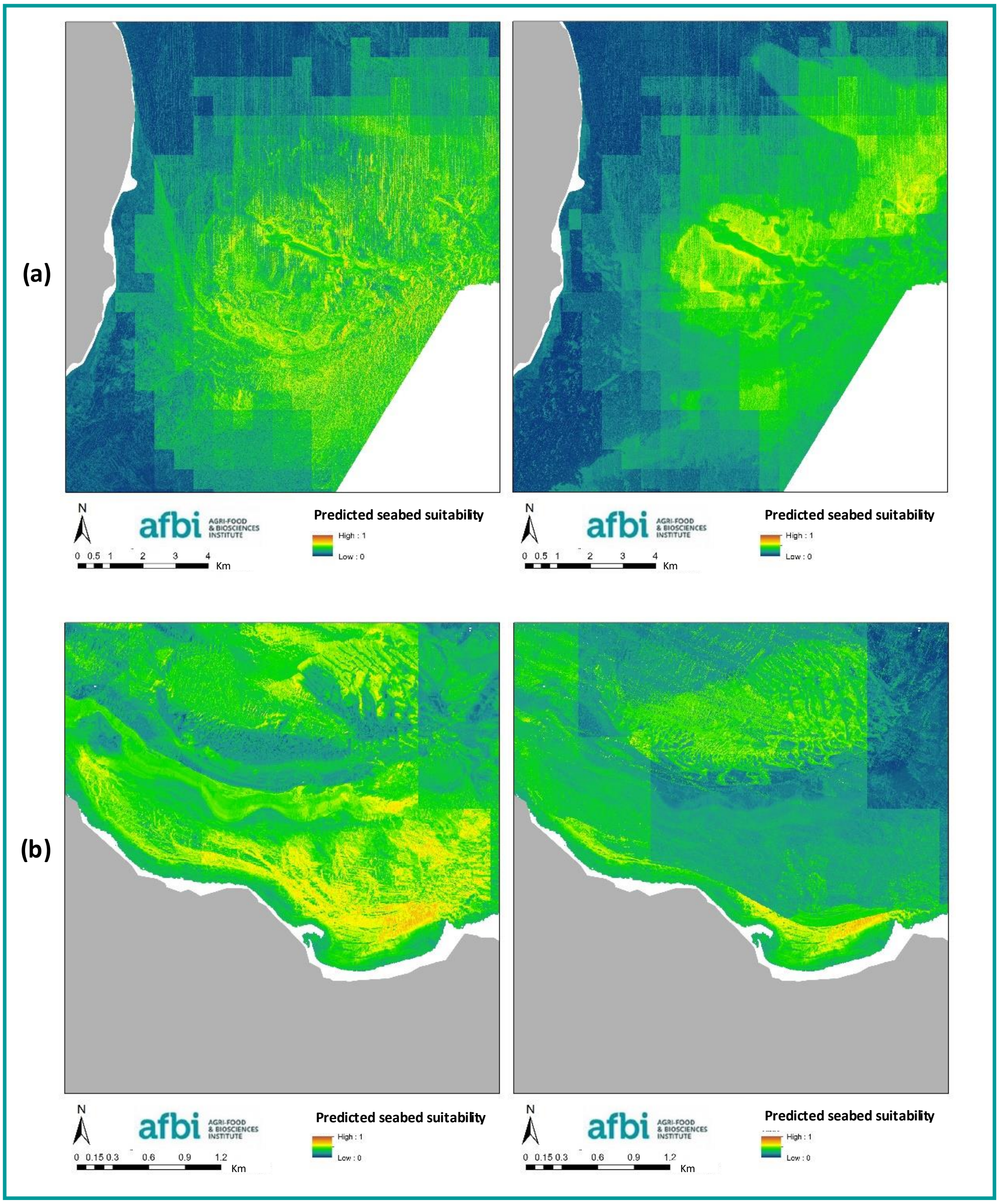


Figure 12. Images to show the increased predictive power of Model 2 and Model 3 in comparison with Model 1. (a) Shows a selected area with Model 1 prediction on the left and Model 2 on the right. (b) Shows a selected area with Model 1 prediction on the left and Model 3 on the right.

have failed to find significant correlations between temperature and sandeel abundance or condition (Wright & Bailey, 1996; Christensen *et al.*, 2008; Eliasen *et al.*, 2011; MacDonald *et al.*, 2019) and it is unlikely that temperature is a significant driver of sandeel habitation in the study area examined here.

Independent data from other AFBI FAEB surveys from the past 5 years showed Ammodytes are frequently caught just off Strangford Narrows which corresponds with an area where Model 2 predicted $\geq$ 50% probability of being suitable sandeel habitat. Similar can be seen to the south of Murlough Bay (Figure 7), reinforcing the possibility of sandeel habitat in these areas and therefore possibly requiring further investigation. These independent data help to verify the predictive capabilities of the model, even in areas where this study had limited or no samples and did not capture sandeels, highlighting the strength of the methods employed.

4.2 Existing protection of potential sandeel habitat

Within the area of Model 2, 98.2% (Table 5) of the areas identified as being ≥ 70% likely to be suitable sandeel habitat are within the boundary of the East Coast Marine pSPA (Figure 8). When applying a threshold of 60% probability of habitat suitability 91.3% (Table 5) of all areas identified as ≥ 60% probability within the Model 2 area are within the boundary of the East Coast Marine pSPA.

The two SPAs within the spatial extent of Model 3 currently overlap 0.1% of ≥ 70% probability sandeel habitat, which increases to 1.2% when the threshold is reduced to 60% (Table 6 and Figure 10). When examining SACs and MCZs within the area of the Model 3 spatial region we find 30.6% of the ≥ 70% potential sandeel habitat lies within the boundary of the network, which increases to 32.2% when the threshold is decreased to ≥ 60%. The majority of this is within the Skerries and Causeway SAC. This leaves a significant area of over 19,000 Ha which has ≥ 60% probability of being sandeel habitat outside current designated sites.

4.3 Potential seabird foraging areas

Identifying potential sandeel habitat is a vital component in the protection of certain seabird populations (Thaxter *et al.*, 2012), particularly species that have a limited foraging range and/or species that exhibit limited foraging strategies. Examples of the former would be cormorant and European shag while the latter include black-legged kittiwakes which are almost exclusively surface feeders (Harris & Riddiford, 1989; Furness & Tasker, 2000). Arctic, common, sandwich and roseate tern fall into both of these groups, with mean foraging ranges of less than 10 km and possessing little ability to dive, making them particularly vulnerable to changes in prey species population (Woodward *et al.*, 2019; Furness & Tasker, 2000). In a review of forage fish species and their predators in the North Sea, Engelhard *et al.*, (2014) identified sandwich terns, shags, great skuas, puffins, common guillemots, razorbills and kittiwakes as depending on sandeels for between 25% and 95% of their diet. These behaviours may be subject to localised differences, as seen by Chivers et al. (2012), however in the absence of similar reviews of the diet of different seabird species in Northern Ireland this section will examine a selection of these potentially vulnerable species protected at different sites.

Larne Lough, the Gobbins and Muck Island host significant numbers of razorbill, common guillemot, black-legged kittiwake, and three tern species – sandwich, common and roseate (JNCC Seabird Monitoring Programme). The potential sandeel habitat identified at the mouth of Belfast Lough, and patches to the east and south-east of The Maidens would be well within the mean foraging range for common guillemot, kittiwake and razorbill (Figure 13). The model shows the potential for a thin strip of sandeel habitat to run along the coastline in this area. Unfortunately ground-truthing data here is limited and high confidence cannot be applied to these data. Anecdotally, an incident was brought to the Department's and the author's attention on 6/8/2021 where a mass stranding of thousands of Ammodytes had taken place at Portmuck Harbour, on the north-eastern edge of Islandmagee. The drivers behind the stranding are currently unknown, but with the knowledge that *Ammodytes* spp. do not stray far from night time burrowing habitat (Engelhard *et al.*, 2008) it can be assumed a significant area of habitat lies within the vicinity of Portmuck.

The seabed around the Copeland Islands seems to have high potential for sandeel habitat (Figure 14) and the Islands are home to breeding colonies of Arctic terns and Manx shearwaters (NIEA, 2015; JNCC SMP). Both of these species could be very reliant on this adjacent sandeel habitat for prey, particularly the Arctic tern with their limited mean foraging range of under 10 km (Woodward *et al.*, 2019), and this habitat could play an important role in colony success for both these species. Further south, around The Feathers, another sizeable area of suitable sandeel habitat can be found as well as a long, thin strip of potentially suitable habitat running near the shore from The Feathers north to Skullmartin (Figure 14). Summer days see large numbers of feeding terns along this coastline, around Burial Island and The Feathers (author pers. obs.). Whether this area is within the foraging range of Strangford Lough based tern colonies or birds from other locations would, however, require investigation. Burial Island hosts large numbers of cormorants (BTO, 2021), a species that also has quite limited foraging range, and for which the surrounding potential sandeel habitat could be very important.

Sheep Island is an SPA on the north Antrim coast (Figure 15) designated for cormorants but also home to large populations of common guillemot, razorbill and black-legged kittiwake (JNCC SMP). A small number of shags are also found here and similar to cormorants they have quite a limited foraging range, with a mean limit of under 10 km (Woodward *et al.*, 2019). The surrounding area shows a large area with high probability of being sandeel habitat, particularly to the west of the island but all well within the foraging range of cormorants and shags. The Skerries is also a significant point for cormorants (BTO, 2021), and the surrounding sea shows a high probability of suitable sandeel habitat (Figure 16).

Rathlin Island hosts the largest numbers of breeding seabirds in the study area, notably razorbill, kittiwake, puffin and guillemots (JNCC SMP) but the model does not predict high probability sandeel habitat in the immediate vicinity of the island. The area which does have $\geq$ 60% probability of sandeel habitat is to the north of the island, however all of this identified area is over 200 m deep. It seems unlikely this would be suitable for Ammodytes habitation as 150 m would be the lower limit of their depth range (Wheeler, 1978). Rathlin Island is still within 10 – 20 km of large areas of identified potential sandeel habitat and razorbill, kittiwake and guillemot are all thought to have foraging ranges of over 30 km (Woodward *et al.*, 2019). Also, the work of Chivers *et al.* (2012) should be noted here, where it was found that only 10% of kittiwakes breeding on Rathlin Island were found to contain sandeels, in comparison with 100% feeding on clupeids (herring and sprat). Although only a small sample size the study could be representative of Rathlin breeding colonies given that clupeids may be more readily available in the vicinity of the island than *Ammodytes* spp.

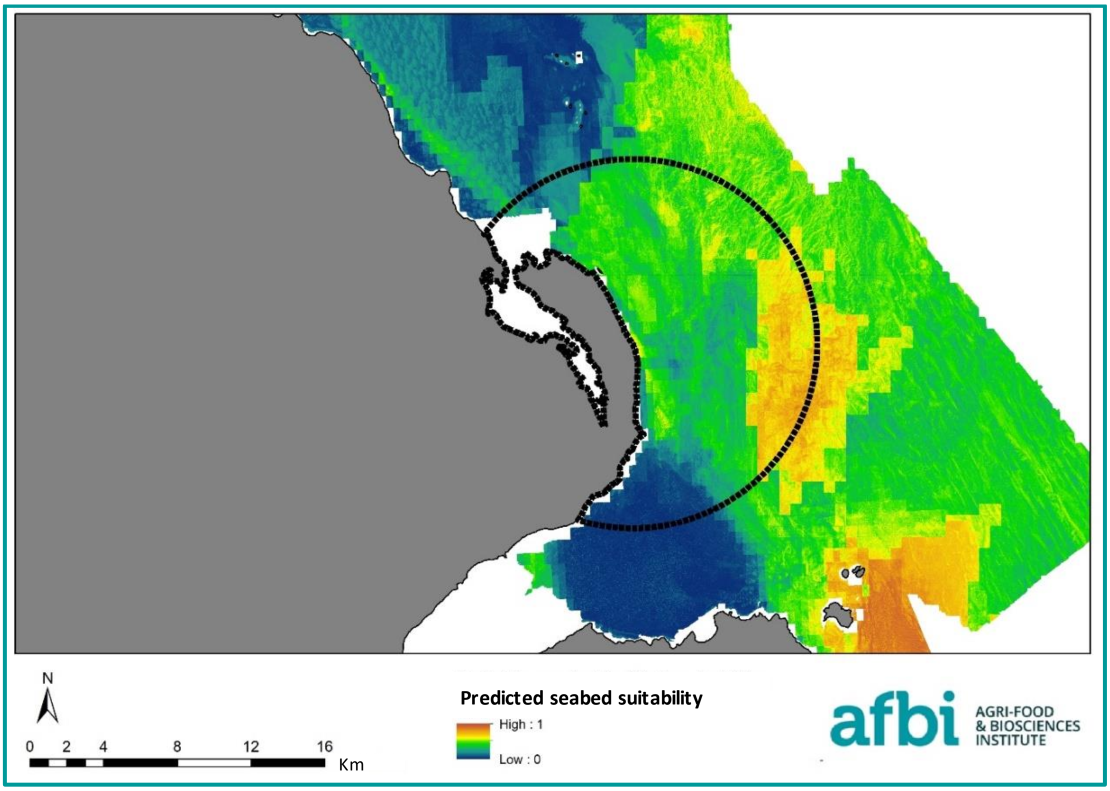


Figure 13. Predicted seabed suitability for sandeel surrounding Larne Lough/Gobbins area. Black dashed line represents 10 km range from Gobbins seabird colonies.

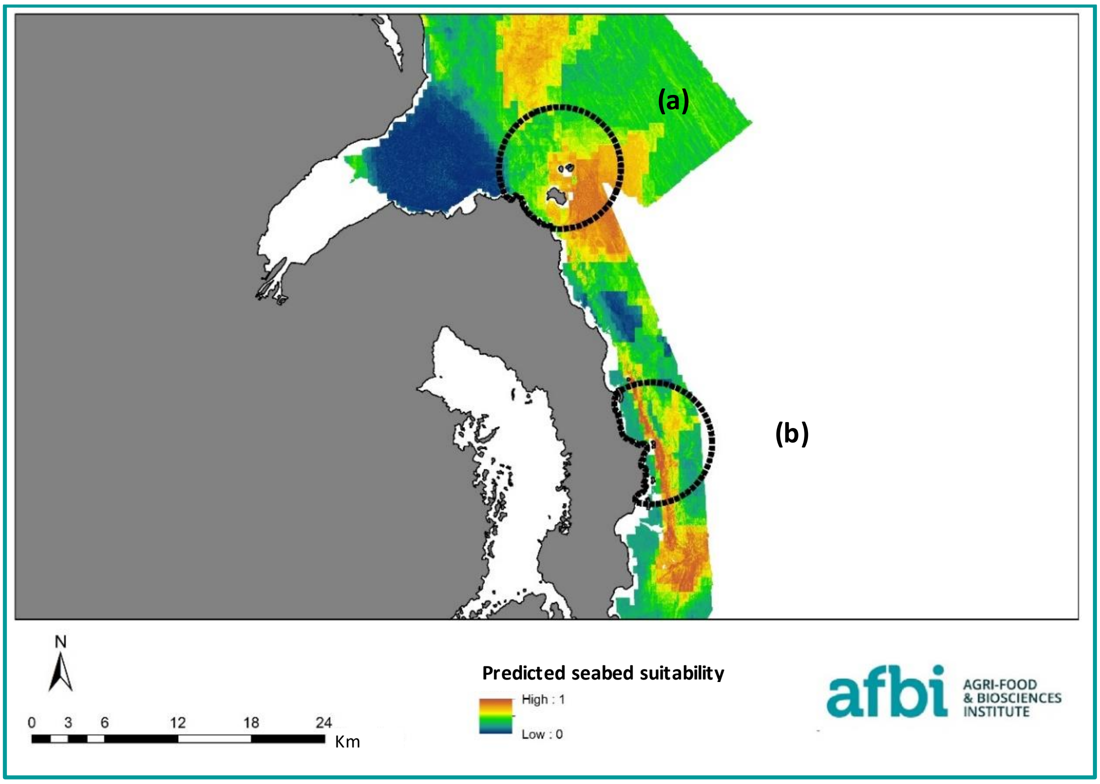


Figure 14. Predicted seabed suitability for sandeel surrounding the Ards Peninsula. Black dashed line represents 5 km range from (a) Copeland Islands SPA, and (b) Burial Island, important for cormorants.

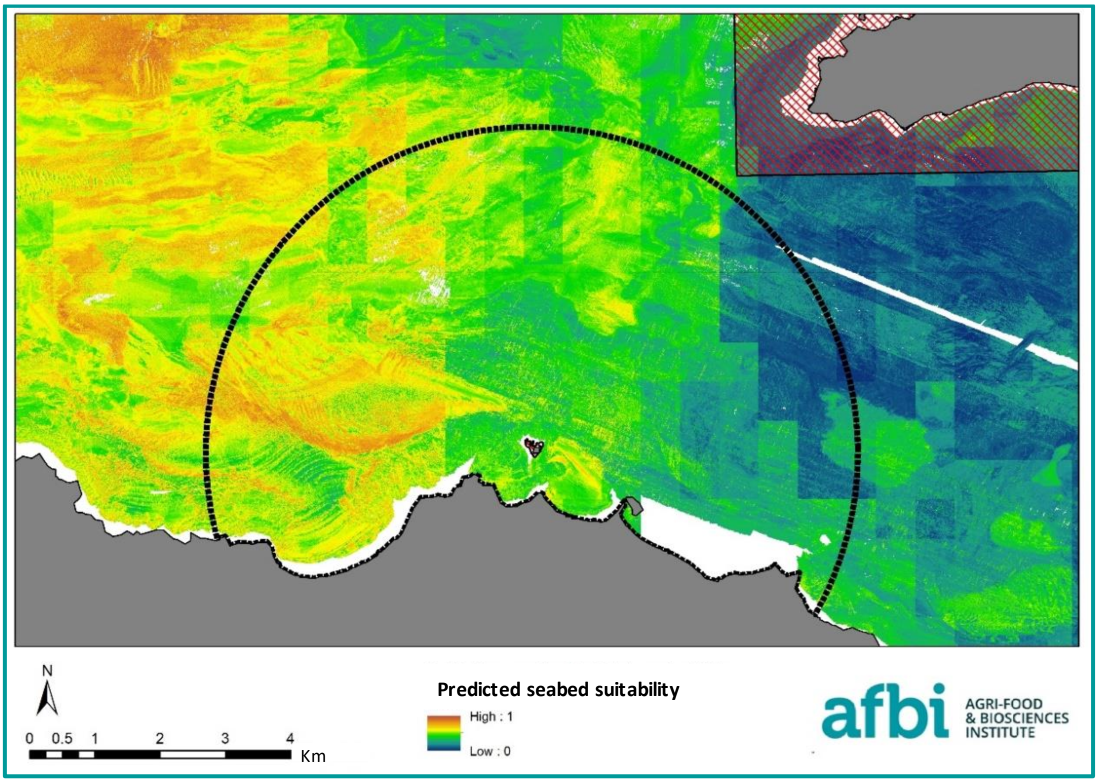


Figure 15. Predicted seabed suitability for sandeel around Sheep Island SPA. Black dashed line represents 5 km range from island and red hatch areas show SPAs (including the most proximal part of the Rathlin Island SPA in the north-east corner).

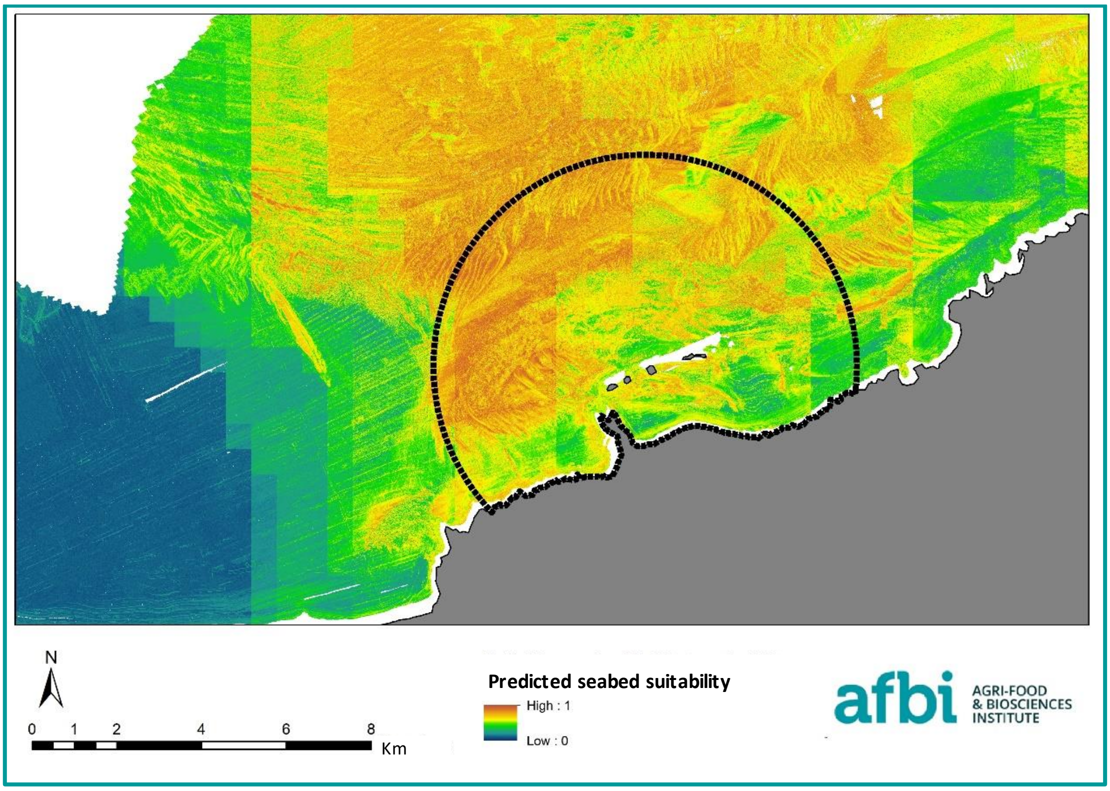


Figure 16. Predicted seabed suitability for sandeel around Skerries. Black dashed line represents 5 km range from midpoint of Skerries Rocks.

4.4 Conclusion and recommended work

This study presents the use of machine learning to predict habitat suitability for sandeel over a large area, showing its use and functionality for environmental management and decision making in the Northern Ireland inshore region. The substrate suitability for sandeel habitation across a large proportion of the Northern Ireland inshore region was predicted, to a fine-scale resolution (5 m x 5 m) appropriate for locating relatively small, scattered patches of important habitat that would otherwise be lost in broad-scale analyses. The rasters produced are applicable for policy makers and easily interpreted, clearly illustrating potential seabird foraging grounds. This method is repeatable and the results can be augmented as new datasets become available, strengthening the existing maps, expanding into other areas as well as opportunities for application to other similar species and habitats.

Due to the large extent of the target area of this study, regular, widespread coverage of ground-truthing data was not available. Even with the additional targeted survey effort undertaken since 2019, gaps still remain. The North Channel area requires investigation, so too the coast in the Inishowen/Magilligan area. Although descriptively valuable to identifying sandeel suitability, Day grab sample footprint is only 0.1 m^2 therefore larger scale surveying is essential.

With the addition of MBES surveys 3 – 12 nM off the Ards Peninsula and the Red Bay to Torr Head area a larger predictive map could be created. The value of MBES backscatter is seen in the variable importance plots, therefore processing of existing MBES backscatter using the harmonisation method applied to the North Coast region would be an important future undertaking. Through the application of moderate additional resource, a highly valuable and informative dataset would be made available for the identification of potential sandeel habitat on the east coast.

Similarly, with slight adjustment of the current data on the east coast, a stronger prediction could be achieved around Carlingford Lough and its approaches, where Model 2 produced a patchy image due to resolution issues. Strangford Lough could also be investigated further, with particular focus on the mouth of the Narrows where the model predicted an area of close to 60% probability of sandeel suitability. This prediction was made in the absence of a large number of samples in the area and was later strengthened when independent fisheries survey bycatch showed a consistent record of *Ammodytes* spp. at this location, a clear demonstration of the predictive capabilities of the models produced in this study. Further independent, supporting data could also be provided through the use of specific seabird colony studies, for example, tracking data and diet analyses.

Protection of seabird species was the primary theme of this study, but literature also suggests the importance of sandeels in the diet of certain cetaceans (Santos *et al.*, 2004; Sveegaard *et al.*, 2012) particularly harbour porpoise. This species is the target for protection in the North Channel SAC and harbour porpoise are a qualifying feature of Skerries and Causeway SAC. Preliminary data collected by AFBI FAEB suggests that harbour porpoise utilise Skerries and Causeway SAC for foraging activity. Applying a systematic approach to certain target areas such as the Skerries to achieve high resolution current data (through deployment of acoustic Doppler current profilers), MBES bathymetry and backscatter, and porpoise detection through moored acoustic receivers, could lead to greater understanding of the species' potential dependence on sandeels.

With the production of refined sandeel habitat maps, work could begin on estimating stock abundance by applying fisheries survey methods to the targeted areas. For example, acoustic surveys and demersal/pelagic trawls, which already have proven ability to provide information on prey species of seabirds. The knowledge gained through these surveys could be used to determine the prospective importance of an area for supporting food resources for protected seabird species and cetaceans.

5. References

AFBI, 2015. Bathymetric and Habitat Map for Strangford Lough (Special Area of Conservation and Marine Conservation Zone), Northern Ireland. Report to the Department of the Environment.

AFBI, 2015. Species and habitat data for Marine Conservation Zone Areas of Interest: Rathlin Island, Ballycastle Bay, Outer Belfast Lough. Report to the Department of the Environment.

AFBI, 2016. Special Area of Conservation Designation Assessment of Outer Ards *Modiolus modiolus* Biogenic Reef. Report to the Department of Agriculture, Environment and Rural Affairs, Northern Ireland.

AFBI, 2020. Habitat preferences of sandeels, *Ammodytes* spp. Report to the Department of Agriculture, Environment and Rural Affairs, Northern Ireland.

Bargiel, D. (2013) Capabilities of high resolution satellite radar for the detection of semi-natural habitat structures and grasslands in agricultural landscapes. *Ecological Informatics* **13:** 9-16.

Barrett, R. T., Røv, N., Loen, J., Montevecchi, W. A. (1990) Diets of shags *Phalacrocorax aristotelis* and cormorants *P. carbo* in Norway and possible implications for gadoid stock recruitment. *Marine Ecology Progress Series* **66**: 205-218.

Breiman, L. (2001) Random Forests. *Machine Learning* **45:** 5-32.

BTO 2021, Northern Ireland Seabird Report 2020. pp88. ISBN 978-1-912642-23-6 https://www.bto.org/sites/default/files/publications/northern_ireland_seabird_report_2020_web.pdf

Camphuysen, K. & Henderson, P (2017) North Sea Fish and their remains. Pisces Conservation Ltd.

Chivers, L. S., Lundy, M. G., Colhoun, K., Newton, S. F., Reid, N. (2012) Diet of Black-leged Kittiwakes (*Rissa tridactyla*) feeding chicks at two Irish colonies highlights the importance of clupeids. *Bird Study* **59**: 363-367.

Chivers, L. S., Lundy, M. G., Colhoun, K., Newton, S. F., Houghton, J. D. R., Reid, N. (2013) Identifying optimal feeding habitat and proposed Marine Protected Areas (pMPAs) for the black-legged kittiwake (*Rissa tridactyla*) suggests a need for complementary management approaches. *Biological Conservation* **164**: 73-81.

Christensen, A., Jensen, H., Mosegaard, H., St. John, M., Schrum, C. (2008) Sandeel (Ammodytes marinus) larval transport in the North Sea from an individual-based hydrodynamic egg and larval model. *Canadian Journal of Fisheries and Aquatic Sciences* **65**: 1498-1511.

Eliasen, K., Reinert, J., Gaard, E., Hansen, B., Jacobsen, J. A., Grønkjær, P., Christensen, J. T. (2011) Sandeel as a link between primary production and higher trophic levels on the Faroe shelf. *Marine Ecology Progress Series* **438**: 185-194.

Engelhard, G. H., van der Kooij, J., Bell, E. D., Pinnegar, J. K., Blanchard, J. L., Mackinson, S., Righton, D. A. (2008) Fishing mortality versus natural predation on diurnally migrating sandeels *Ammodytes marinus*. *Marine Ecology Progress Series* **369**: 213-227.

Engelhard, G. H., Peck, M. A., Rindorf, A., Smout, S. C., van Deurs, M., Raab, K., Andersen, K. H., Garthe, S., Lauerburg, R. A., Scott, F. and Brunel, T. (2014) Forage fish, their fisheries, and their predators: who drives whom? *ICES Journal of Marine Science: Journal du Conseil*, **71**(1): 90-104.

Frederiksen, M., Edwards, M., Richardson, A. J., Halliday, N. C., Wanless, S. (2006) From plankton to top predators: bottom-up control of a marine food web across four trophic levels. *Journal of Animal Ecology* **75**: 1259-1268.

Furness, R. W. (2002) Management implications of interactions between fisheries and sandeel-dependent seabirds and seals in the North Sea. *ICES Journal of Marine Science* **59**: 261–269.

Furness, R. W. & Tasker, M. L. (2000) Seabird-fishery interactions: quantifying the sensitivity of seabirds to reductions in sandeel abundance, and identification of key areas for sensitive seabirds in the North Sea. *Marine Ecology Progress Series* **202**: 253-264.

Gibson, R. N., Robb, L., Burrows, M. T., Ansell, A. D. (1996) Tidal, diel and longer term changes in the distribution of fishes on a Scottish sandy beach. *Marine Ecology Progress Series* **130**: 1-17.

Greenstreet, S. P. R., Armstrong, E., Mosegaard, E., Jensen, H., Gibb, I. M., Fraser, H. M., Scott, B. E., Holland, G. J., and Sharples, J. (2006) Variation in the abundance of sandeels *Ammodytes marinus* off southeast Scotland: an evaluation of area-closure fisheries management and stock abundance assessment methods. *ICES Journal of Marine Science* **63**: 1530-1550.

Greenstreet, S. P. R., Holland, G. J., Guirey, E. J., Armstrong, E., Fraser, H. M., Gibb, I. M. (2010) Combining hydroacoustic seabed survey and grab sampling techniques to assess “local” sandeel population abundance. *ICES Journal of Marine Science* **67**: 971-984.

Harris, M. P. & Riddiford, N. J. (1989) The food of some young seabirds on Fair Isle in 1986–1988. *Scottish Birds* **15**: 119–125.

Harris, M. P., Wanless, S. (1991) The importance of the lesser sandeel *Ammodytes marinus* in the diet of the Shag *Phalacrocorax aristotelis*. *Scandinavian Journal of Ornithology* **22**: 375-382.

Holland, G. J., Greenstreet, S. P. R., Gibb, I. M., Fraser, H. M., Robertson, M. R. (2005) Identifying sandeel *Ammodytes marinus* sediment habitat preferences in the marine environment. *Marine Ecology Progress Series* **303**: 269–282.

ICES. 2019. Sandeel (*Ammodytes* spp.) in divisions 4.a–b, Sandeel Area 4 (northern and central North Sea). In Report of the ICES Advisory Committee, 2019. ICES Advice 2019, san.sa.4, https://doi.org/10.17895/ices.advice.4723

Jensen, H., Rindorf, A., Wright, P. J., Mosegaard, H. (2011) Inferring the location and scale of mixing between habitat areas of lesser sandeel through information from the fishery. *ICES Journal of Marine Science* **68**: 43−51.

JNCC Seabird Monitoring Programme database https://app.bto.org/seabirds/public/data.jsp

Jørgensen, M. G. P., van Deurs, M., Butts, I. A. E., Jørgensen, K., Behrens, J. W. (2017) PIT-tagging method for small fishes: A case study using sandeel (*Ammodytes tobianus*). *Fisheries Research* **193**: 95-103.

Jovanovic, B., Longmore, C., O’Leary, Á., Mariani, S. (2007) Fish community structure and distribution in a macro-tidal inshore habitat in the Irish Sea. *Estuarine, Coastal and Shelf Science* **75**: 135-142.

Kellnreitner, F., Pockberger, M., Asmus, H. (2011) Seasonal variation of assemblage and feeding guild structure of fish species in a boreal tidal basin. *Estuarine, Coastal and Shelf Science* **108**: 97-108.

MacDonald, A., Spiers, D. C., Greenstreet, S. P. R., Boulcott, P., Heath, M. R. (2019) Trends in Sandeel Growth and Abundance off the East Coast of Scotland. *Frontiers in Marine Science* **6**: 201.

Macer, C.T. (1966) Sandeels (Ammodytidae) in the south-western North Sea: their biology and fishery. MAFF Fishery Investigations London Series 2 **24** (6**)** 55pp.

NIEA, 2015. Copeland Islands Special Protection Area (SPA) Conservation Objectives.

Pearman, T. R. R., Robert, K., Callaway, A., Hall, R., Lo Iacono, C., Huvenne, V. A. I. (2020) Improving the predictive capability of benthic species distribution models by incorporating oceanographic data – Towards holistic ecological modelling of a submarine canyon. *Progress in Oceanography* **184**: 102338. doi: 10.1016/j.pocean.2020.102338

Pedersen, S. A., Lewy, P., and Wright, P. J. (1999) Assessments of the lesser sandeel (*Ammodytes marinus)* in the North Sea based on revised stock divisions. *Fisheries Research* **41**: 221–241.

Reay, P. J. (1970) Synopsis of biological data on North Atlantic sandeels of the genus *Ammodytes*. *Food and Agriculture Organisation of the UN Report 82*.

Santos, M. B., Pierce, G. J., Learmonth, J. A., Reid, R. J., Ross, H. M., Patterson, I. A. P., Reid, D. G., *et al*. (2004) Variability in the diet of harbor porpoises (Phocoena phocoena) in Scottish waters 1992–2003. *Marine Mammal Science* **20**: 1–27.

Sveegaard, S., Andreasen, H., Mouritsen, K. N., Jeppesen, J. P., Teilmann, J., Kinze, C. C. (2012) Correlation between the seasonal distribution of harbour porpoises and their prey in the Sound, Baltic Sea. *Marine Biology* **159**: 1029-1037.

Thaxter, C. B., Lascelles, B., Sugar, K., Cook, A.S.C.P., Roos, S., Bolton, M., Langston, R.H.W., Burton, N.H.K. (2012) Seabird foraging ranges as a preliminary tool for identifying candidate marine protected areas. *Biological Conservation* **156**: 53–61.

Thompson, Katherine Russel (1987) The ecology of the Manx shearwater *Puffinus puffinus* on Rhum, West Scotland. PhD thesis.

Tien, N. S. H., Craeymeersch, J., van Damme, C., Couperus, A. S., Adema, J., Tulp, I. (2017) Burrow distribution of three sandeel species relates to beam trawl fishing, sediment composition and water velocity, in Dutch coastal waters. *Journal of Sea Research* **127**: 194-202.

Torres, M. & Qiu, G. (2014) Automatic habitat classification using image analysis and random forest. *Ecological Informatics* **23:** 126-136.

Turner, J. A., Babcock, R. C., Hovey, R., Kendrick, G. A. (2018) Can single classifiers be as useful as model ensembles to produce benthic substratum maps? *Estuarine, Coastal and Shelf Science* **204**: 149-163.

van Deurs, M., Grome, T. M., Kaspersen, M., Jensen, H., Stenberg, C., Sørensen T.K., Støttrup, J., Warnar, T., Mosegaard, H. (2012) Short- and long-term effects of an offshore wind farm on three species of sandeel and their sand habitat. *Marine Ecology Progress Series* **458**: 169−180.

van der Kooij, J., Scott, B. E., Mackinson, S. (2008) The effects of environmental factors on daytime sandeel distribution and abundance on the Dogger Bank. *Journal of Sea Research* **60**: 201-209.

Walbridge, S.; Slocum, N.; Pobuda, M.; Wright, D.J. Unified Geomorphological Analysis Workflows with Benthic Terrain Modeler. *Geosciences* **2018**, *8*, 94. doi:10.3390/geosciences8030094

Winslade, P. R. (1974) Behavioural studies on the lesser sandeel, *Ammodytes marinus* (Raitt) III. The effect of temperature on activity and the environmental control of the annual cycle of activity. *Journal of Fish Biology* **6**: 587-599.

Wheeler, A. (1978) Key to the fishes of Northern Europe. Frederick Warne, London.

Woodward, I., Thaxter, C.B., Owen, E. & Cook, A.S.C.P. (2019). Desk-based revision of seabird foraging ranges used for HRA screening, Report of work carried out by the British Trust for Ornithology on behalf of NIRAS and The Crown Estate, ISBN 978-1-912642-12-0

Wright, P. J., Jensen, H., Tuck, I. (2000) The influence of sediment type on the distribution of the lesser sandeel, *Ammodytes marinus*. *Journal of Sea Research* **44**: 243-256.

Wright, P. J., Régnier, T., Gibb, F. M., Augley, J., Devalla, S. (2018) Identifying stock structuring in the sandeel, *Ammodytes marinus*, from otolith microchemistry. *Fisheries Research* **199**: 19-25.

Wright, P. J. & Bailey, M. C. (1996) Timing of hatching in *Ammodytes marinus* from Shetland waters and its significance to early growth and survivorship. *Marine Biology* **126**: 143-152.

APPENDIX I

HARMONIZING MBES BACKSCATTER REPORT

1. Datasets

Multibeam backscatter datasets are increasingly used as ingredients in support of seabed mapping campaigns. However, the lack of calibration between repeat and overlapping surveys and between multibeam echosounders (MBES) is a major challenge (Misiuk *et al*., 2021, Misiuk *et al*., 2020). Here we present a composite harmonized mosaic for three overlapping backscatter datasets (Appendix A) collected during the 2007 Joint Irish Bathymetric Survey (JIBS) using three different sensors mounted on three different vessels: *Meridian* (Reson 7125), *Jetstream* (Kongsberg EM3002) and *Victor Hensen* (Kongsberg EM710). System settings for the MBES are given in Table 1. We use .TIFF file formats due to their ability to handle the **no data** mask (.IMG file formats cannot). Harmonizing multiple backscatter layers makes use of mutual overlap between surveys (Figure 1), to implement a relative statistical calibration approach, also known as **bulk shift** (Misiuk *et al*., 2020; Hughes Clarke *et al*., 2008).

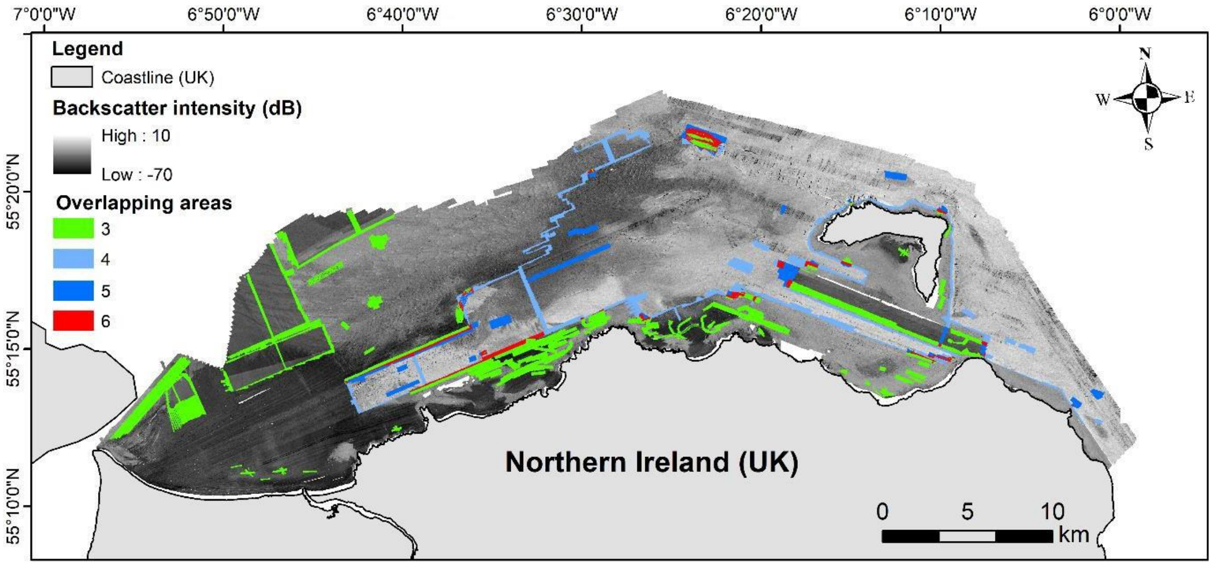


Figure 1: Composite backscatter mosaic comprising three MBES datasets and overlap between them represented in green, light blue, blue and red respectively as; 3 (*Jetstream* + *Meridian* datasets), 4 (*Jetstream* + *Victor Hensen* datasets), 5 (*Meridian* + *Victor Hensen* datasets) and 6 (all three datasets overlap).

Table 1: MBES systems settings for JIBS 2007 surveys (spatial resolution of 1m)

Vessel	**System**	**Sounding Mode**	**Frequency (kHz)**	**Pulse Length (µsec)**
Jetstream	Kongsberg EM3002	1 (shallow)	293	149
Meridian	Reson 7125	-	200/400	10-300
Victor Hensen	Kongsberg EM710	0 (Very shallow)	71-97-83	206
		1 (Shallow)	71-83-77	500
		2 (Medium)	71-77-74	2000

2. Harmonization

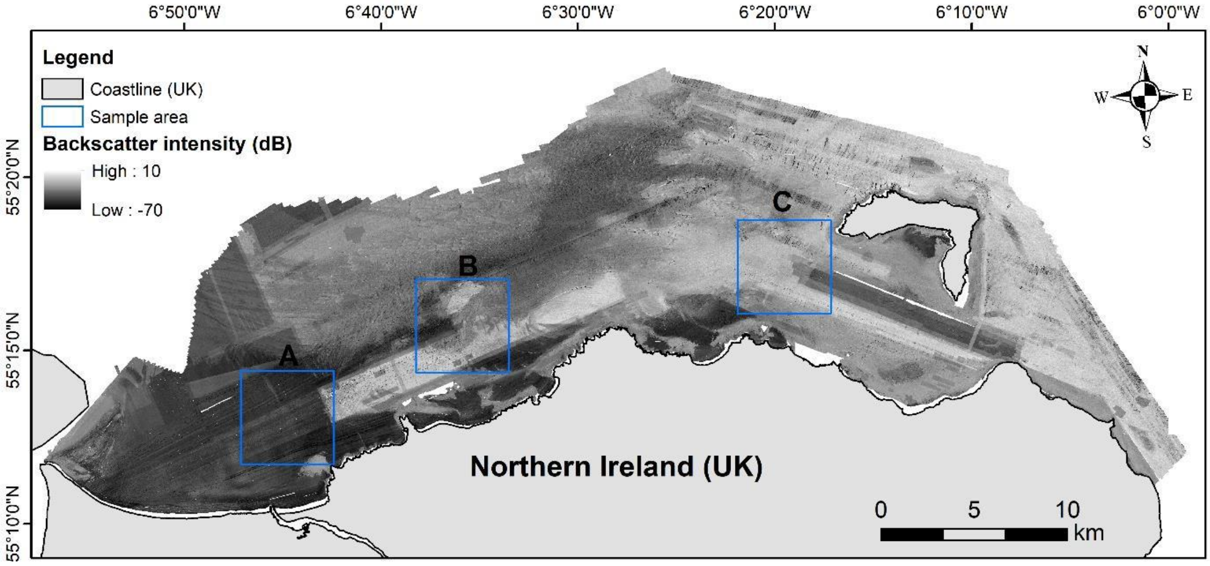


Figure 2: Composite backscatter mosaic comprising the three MBES datasets used; showing three 5 x 5 km sample areas used for evaluating the performance of the harmonization method: **A** (Portsewart), **B** (The Skerries) and **C** (Rathlin Island).

The Bulkshift () function in R was used to implement the backscatter harmonization (Misiuk *et al.*, 2020). Available at https://github.com/benjaminmisiuk/bulkshift, the package contains a series of statistical functions (**Bulkshift approaches**) which are used here, with the results presented below. The *Jetstream* dataset was used as the "**target**" dataset due to its wide coverage and high signal-to-noise ratio (SNR). The "target" represents the reference backscatter layer, and the *Meridian* and *Victor Hensen* datasets were the "**shift**" datasets corrected before harmonization was conducted. We first individually corrected the *Meridian* and *Victor Hensen* datasets as a function of the *Jetstream* dataset and subsequently harmonized all the three datasets; target (*Jetstream*), shift layer 1 (*Meridian*) and shift layer 2 (*Victor Hensen*). The JIBS bathymetry layer (Appendix B) was included as an additional covariate during the bulkshift process. This resulted in two harmonized backscatter mosaics (Figure 3); H1 (using linear model) and H2 (using MEAN function).

According to Misiuk *et al.* (2020), harmonization depends on three things: (i) the amount of overlap between datasets, (ii) the bulkshift approach (statistical model) used, and (iii) the inclusion of an additional covariate like bathymetry which might improve results. Overlap between surveys provides a mutual area to compare backscatter intensity values and derive statistical relationships between datasets. Based on these factors, several bulkshift approaches have been proposed and successfully tested (Misiuk *et al.*, 2020). The "**MEAN**" Bulkshift approach which adds the mean of the backscatter error to the shift layer is recommended for harmonizing multiple datasets with minimal overlap like in our case. This function can run with additional covariates like bathymetry but will not influence results (Misiuk *et al.*, 2020). Other simple statistical models like "**linear model**" have shown to be sufficient in harmonizing backscatter datasets yielding good results across the entire dataset (Misiuk *et al.*, 2020).

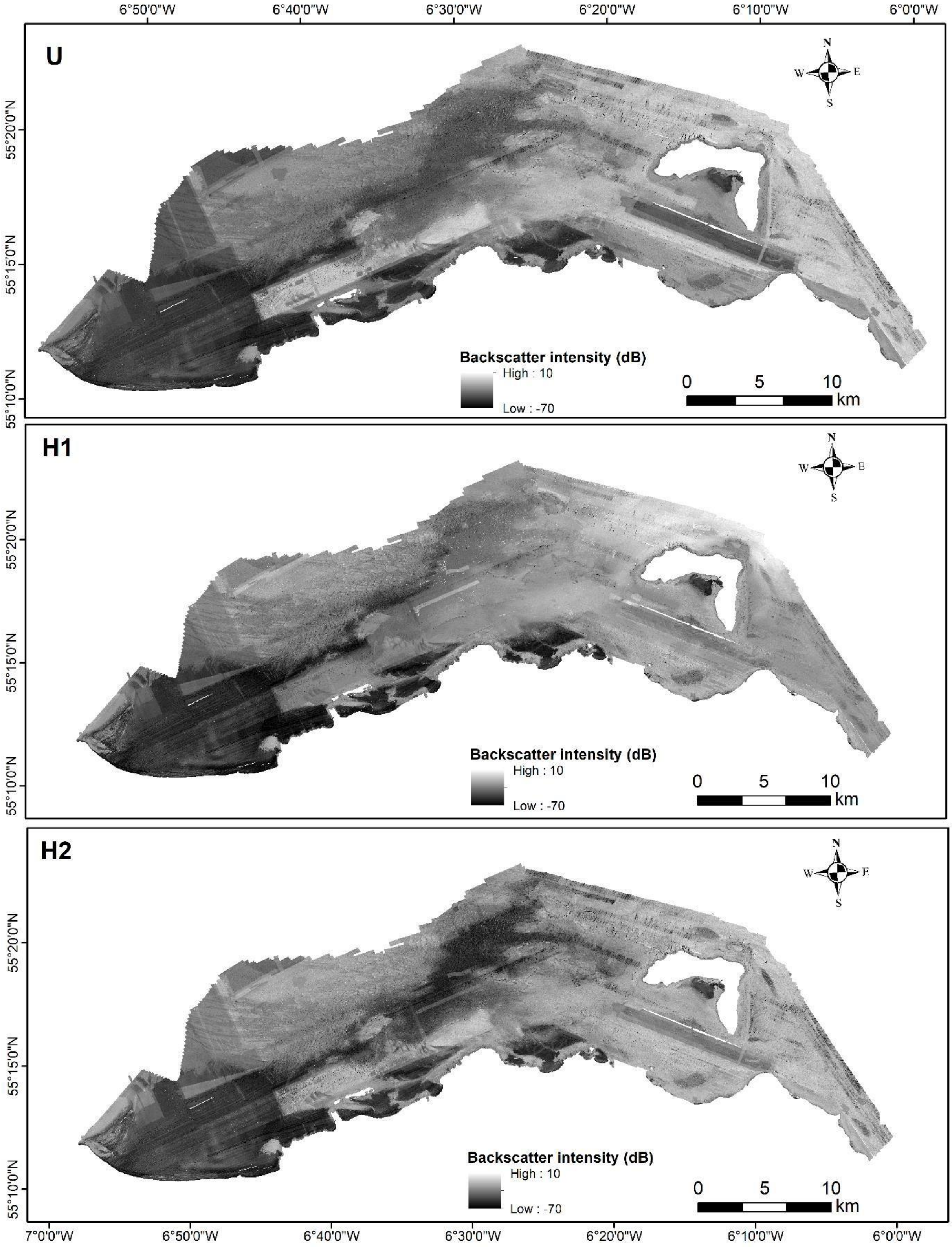


Figure 3: Comparisons between the backscatter mosaics; **U** (uncorrelated mosaic), **H1** (Harmonized mosaic using linear model with bathymetry) and **H2** (Harmonized with MEAN function with bathymetry).

3. Evaluation of performance

In this section we present harmonization outputs for the bulkshift approaches used (**linear model** and **MEAN** functions) with bathymetry as a covariate and compare these with the uncorrected mosaic of the three datasets (Figure 3). Minimal variation in the mean backscatter intensity between the uncorrelated backscatter mosaic (U) and the two harmonized backscatter mosaics (H1, H2) is observed: -25.32±7.2 (U), -26.18±5.5 (H1) and -27.33±6.0 (H2) respectively (Table 2).

Three difference layers between the uncorrelated composite backscatter mosaic and each of the harmonized backscatter mosaic were generated (Figure 4) in the raster calculator tool of ESRI's ArcMap V.10.5.1. These difference layers indicate a deviation of backscatter values between uncorrelated and harmonized mosaics. These provide a visual and quantitative assessment of where the harmonization process worked vis-à-vis where it failed. Evidently, the harmonization of the three MBES datasets worked in areas covered by the target layer and within the overlapping areas.

We select three areas of interest marked A (Portsewart), B (The Skerries) and C (Rathlin Island) to evaluate the performance of harmonization (Figure 4). These sample areas were selected based on the proximity to Special Area of Conservations (Inishtrahull Sound and Hempton's Turbot bank SACs), along the overlapping coverage and consideration on whether the harmonization either worked or failed. These three areas vary significantly in terms of their mean relative backscatter intensity in the range of -34.87±4.0dB (Portsewart) to -20.4±4.8dB (The Rathlin Island). The variation in backscatter responses within each site and between the harmonization outputs (H1, H2) in comparison to the target layer (U) is minimal (Table 2). Texturally, the deviation of backscatter values of the harmonized layers (H1, H2) from the uncorrelated mosaic (U) within each of the sampled areas vary significantly in the magnitude of -0.2±1.3dB to 2.5±3.7dB (H1) and -0.14±1.0dB to -3.5±2.8dB (H2) as given in Figure 5. Similarly, a significant textural variation between H1 and H2 is noted particularly for Skerries and The Rathlin Island (Figure 5). A statistical evaluation yields lower Mean Absolute Error (MAE) value for H2 than H1 at 2.664 and 6.387 respectively. Based on these MAE values between uncorrelated composite mosaic (U) and the harmonized backscatter mosaics (H1, H2), the results indicate that the **MEAN** bulkshift function provides better harmonization results for our datasets than the linear model function.

Generally, the process of harmonizing multiple backscatter datasets from different surveys considers two main assumptions: (i) that backscatter values of each dataset is a function of same substrates (internally consistent) and (ii) a reasonable temporal homogeneity between datasets exists. Besides, harmonizing datasets of different frequencies depends on the magnitude of difference between the frequencies (Misiuk *et al.*, 2020). Therefore, for successful harmonization to be achieved, we recommend that sufficient overlap between datasets of at least one full survey line exists.

In terms of future work, to further improve the harmonization of the JIBS backscatter data, additional MBES backscatter surveys could be conducted to cover at least the entire dynamic range of the backscatter and bathymetry values, and benthic conditions from previous surveys (Misiuk *et al.*, 2020). A narrower dynamic range in backscatter values between shift and target backscatter layers are likely to generate harmonized mosaics with little spectral detail (Misiuk *et al.*, 2020). If new surveys are carried out to supplement the JIBS data, they should take into account the survey geometry and processing parameters to ensure internal consistency with previous datasets used in this report. If these recommendations are followed, the use of straightforward statistical methods like **Linear Model** would probably be sufficient due to its simplicity and its robustness against dynamic range compression due to inconsistent relationships between backscatter datasets.

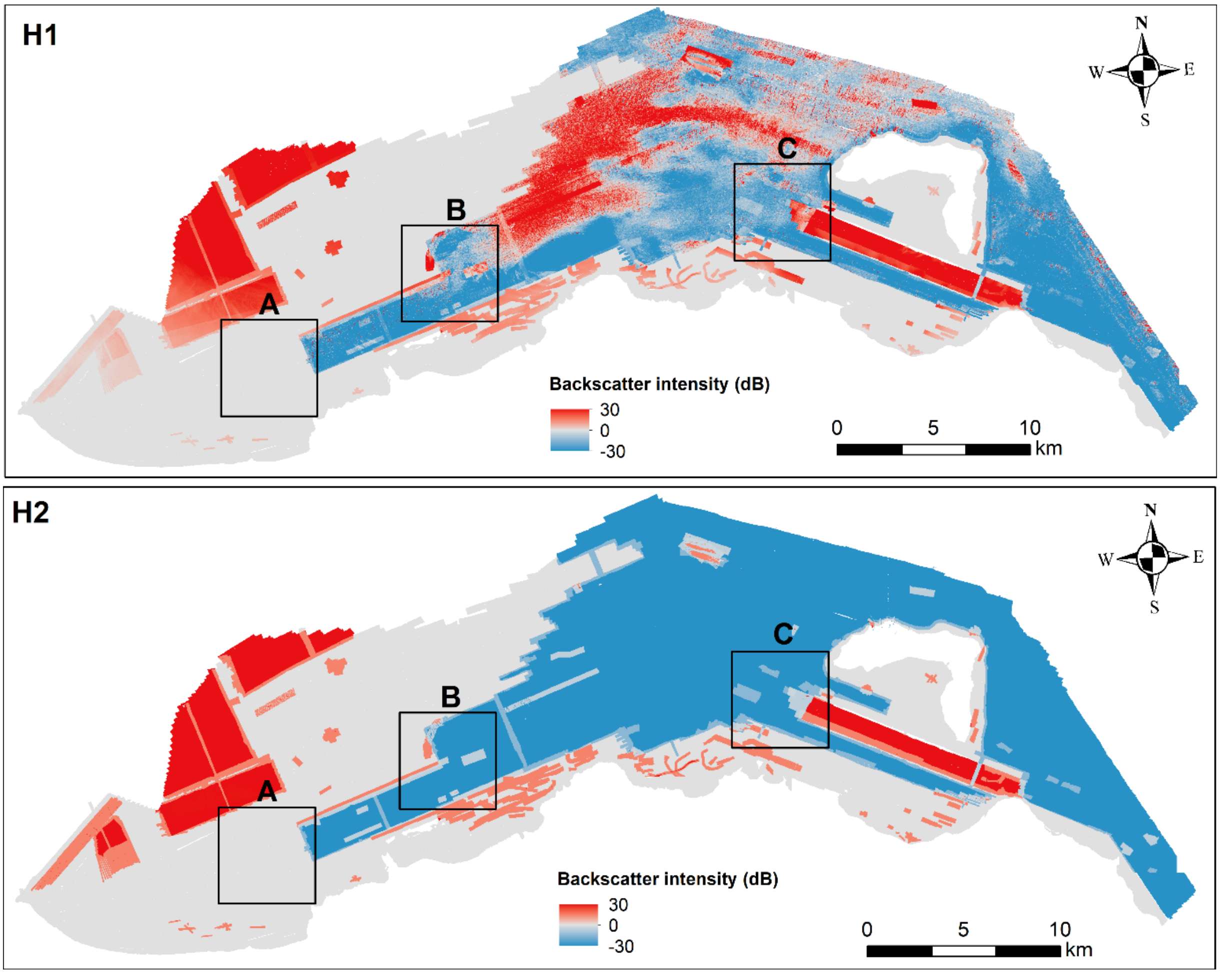


Figure 4: Difference layers from harmonized mosaics (H1 and H2) showing the magnitude of change between uncorrelated backscatter mosaic and the harmonized mosaics.

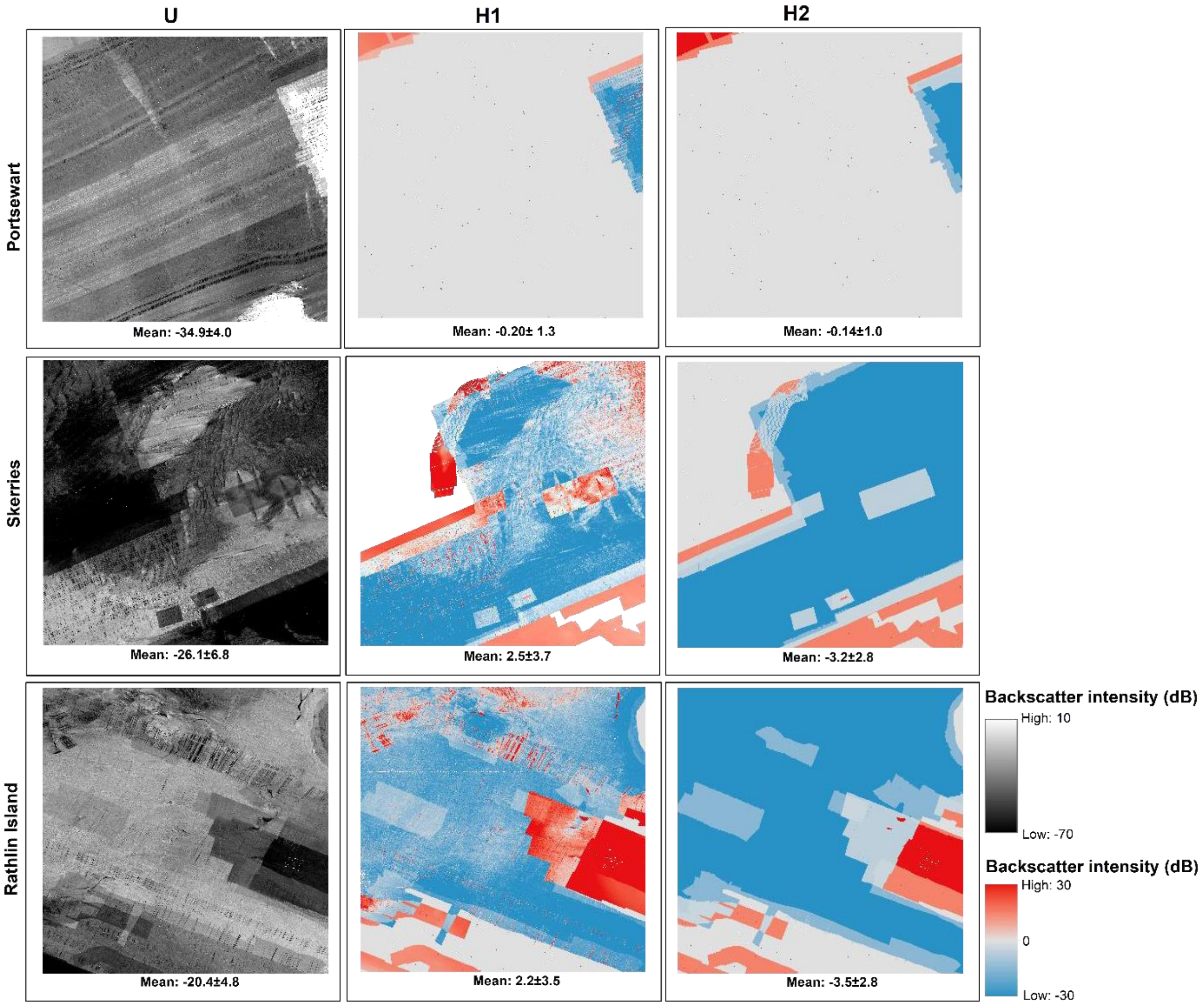


Figure 5: Sample areas comparing the deviation of backscatter intensity between harmonized backscatter (H1 and H2) from uncorrelated backscatter mosaic (U) for Portsewart, Skerries and The Rathlin Island. H1 and H2 represent the harmonized mosaics implemented by linear model and mean bulkshift functions respectively.

Table 2: Mean backscatter intensity (dB) for uncorrelated composite backscatter mosaic (U) and three harmonized backscatter mosaics (H1, H2) across the full extent and three sampled areas.

	U	H1	H2
Composite mosaic	-25.32±7.2	-26.18±5.5	-27.33±6.0
Portsewart	-34.87±4.0	-35.07±3.4	-35.02±3.6
The Skerries	-26.06±6.8	-28.57±4.0	-29.29±5.3
Rathlin Island	-20.37±4.8	-22.56±2.4	-23.88±3.8

4. References

Hughes Clarke, J.E.; Iwanowska, K.K.; Parrott, R.; Du_y, G.; Lamplugh, M.; Gri_n, J. Inter-calibrating multi-source, multi-platform backscatter data sets to assist in compiling regional sediment type maps: Bay of Fundy. In Proceedings of the Canadian Hydrographic Conference and National Surveyors Conference **2008**, Victoria, BC, Canada, 5–8 May 2008; p. 22.

Misiuk, B.; Brown, C.J.; Robert, K.; Lacharité, M. Harmonizing Multi-Source Sonar Backscatter Datasets for Seabed Mapping Using Bulk Shift Approaches. *Remote Sens.* **2020**, *12*, 601. https://doi.org/10.3390/rs12040601

Misiuk, B.; Lacharité, M.; Brown, C.J. Assessing the use of harmonized multisource backscatter data for thematic benthic habitat mapping, Science of Remote Sensing, Volume 3, **2021**, 100015, ISSN 2666-0172

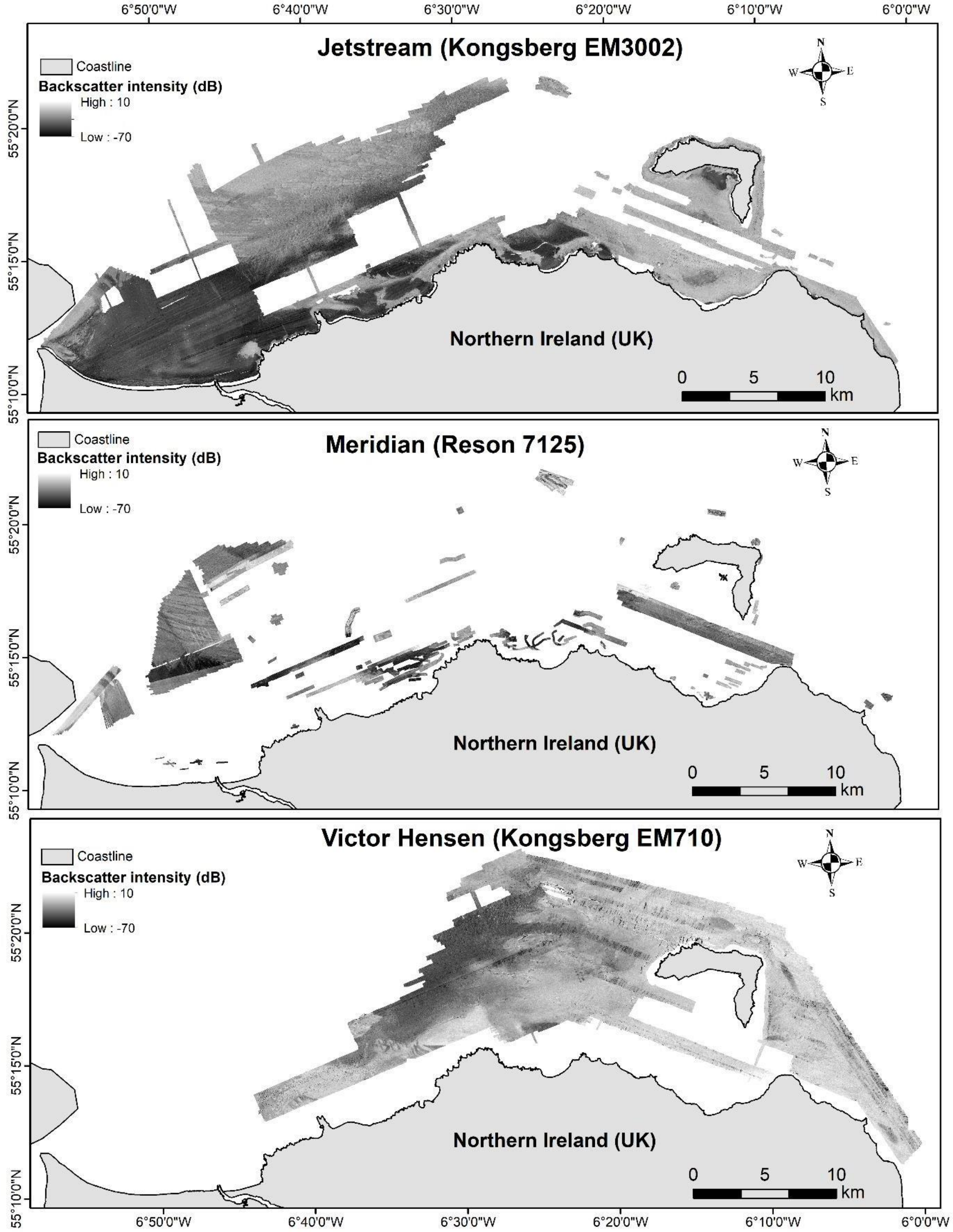
5. R-Script

Harmonization of backscatter implemented in RStudio via the High-Performance Computing (HPC) environment. See script used below:

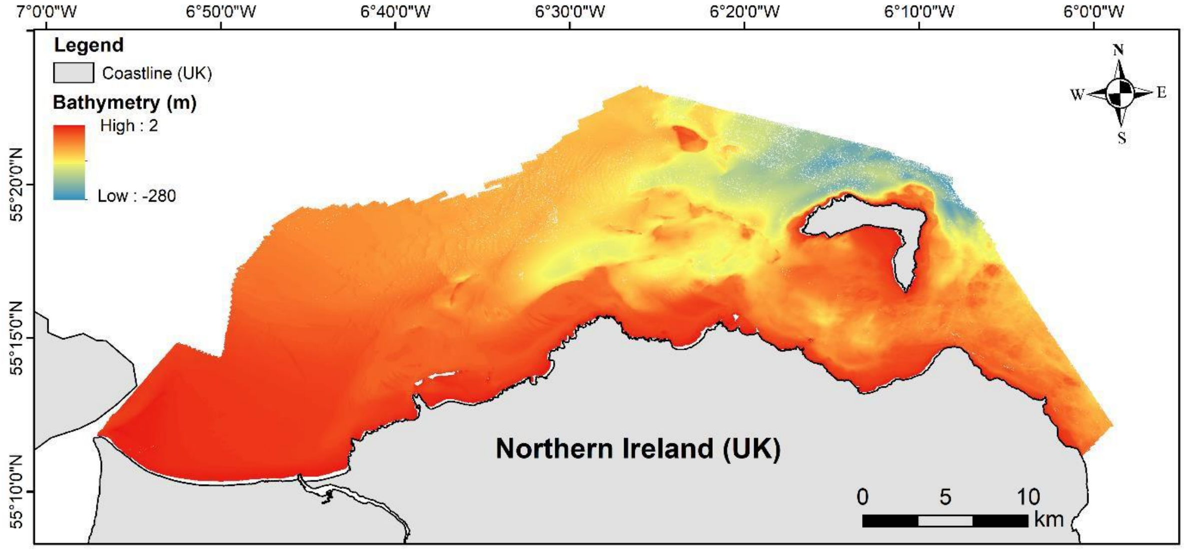
```
setwd("/home/robert/Harmon/") #setting working
directory source ("bulkshift. R") #source Bulkshift
functions args(bulkshift)#bulkshift arguments

view(bulkshift)#view Bulkshift functions ##Libraries
used

library(sp)
library(raster)
library(rgdal)
library(dismo)
library(ggplot2)
library(stringi)
library(reshape)
library(cowplot)


###HARMONIZATION

R1<-raster("/home/robert/Harmon/JS.tif") #target dataset
R2<-raster("/home/robert/Harmon/VH.tif") #shift dataset
1 Bathy<-
raster("/home/robert/Harmon/NI_1m.tif")#covariate   R3<-
raster("/home/robert/Harmon/MD.tif")  #shift dataset 2
##set same projection

projection(R1)

"+proj=utm +zone=29 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0"
projection(R2)

"+proj=utm +zone=29 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0"
##check if the projections match

projection(R2)==projection(R1)
projection(R3)

"+proj=utm +zone=29 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0"
projection(R3)==projection(R1)


##correcting the Victor Hensen dataset

bulk.shift <- bulkshift(R2, R1, predicts= Bathy, shift.method="mean", mosaic = TRUE,  mosaic.method="bilinear") #can be
        changed to lm function

##We can then export the corrected backscatter layer to the working directory for use in other GIS
applications writeRaster(bulk.shift$shifted, filename = "R2_corrected_mean", format="GTiff")#corrected VH
layer ##writeRaster(bulk.shift$mosaic, filename = "R2_R1_mosaic_mean", format="GTiff") #harmonized
JS+VH


##correcting the Meridian dataset

bulk.shift <- bulkshift(R3, R1, predicts=Bathy, shift.method="mean", mosaic = TRUE,  mosaic.method="bilinear") #using lm
        function

##We can then export the corrected backscatter layer to the working directory for use in other GIS
applications writeRaster(bulk.shift$shifted, filename = "R3_corrected_mean", format="GTiff")#corrected MD
layer writeRaster(bulk.shift$mosaic, filename = "R3_R1_mosaic_mean", format="GTiff") #harmonised JS+MD

##mosaicking all three datasets
```

```
RC3<-
raster("/home/robert/Harmon/R3_corrected_mean.tif")
RC2<-
raster("/home/robert/Harmon/R2_corrected_mean.tif")
R1<-raster("/home/robert/Harmon/JS.tif")

Harmonised<-mosaic(R1, RC3, RC2, fun=mean)#mosaicking all three datasets

writeRaster(Harmonised, filename = "Full_harmonised_mean", format="GTiff")#harmonized mosaic can be plotted and
        exported
```

APPENDEX A (JIBS backscatter layers)



APPENDEX B (JIBS bathymetry layer)



APPENDIX II

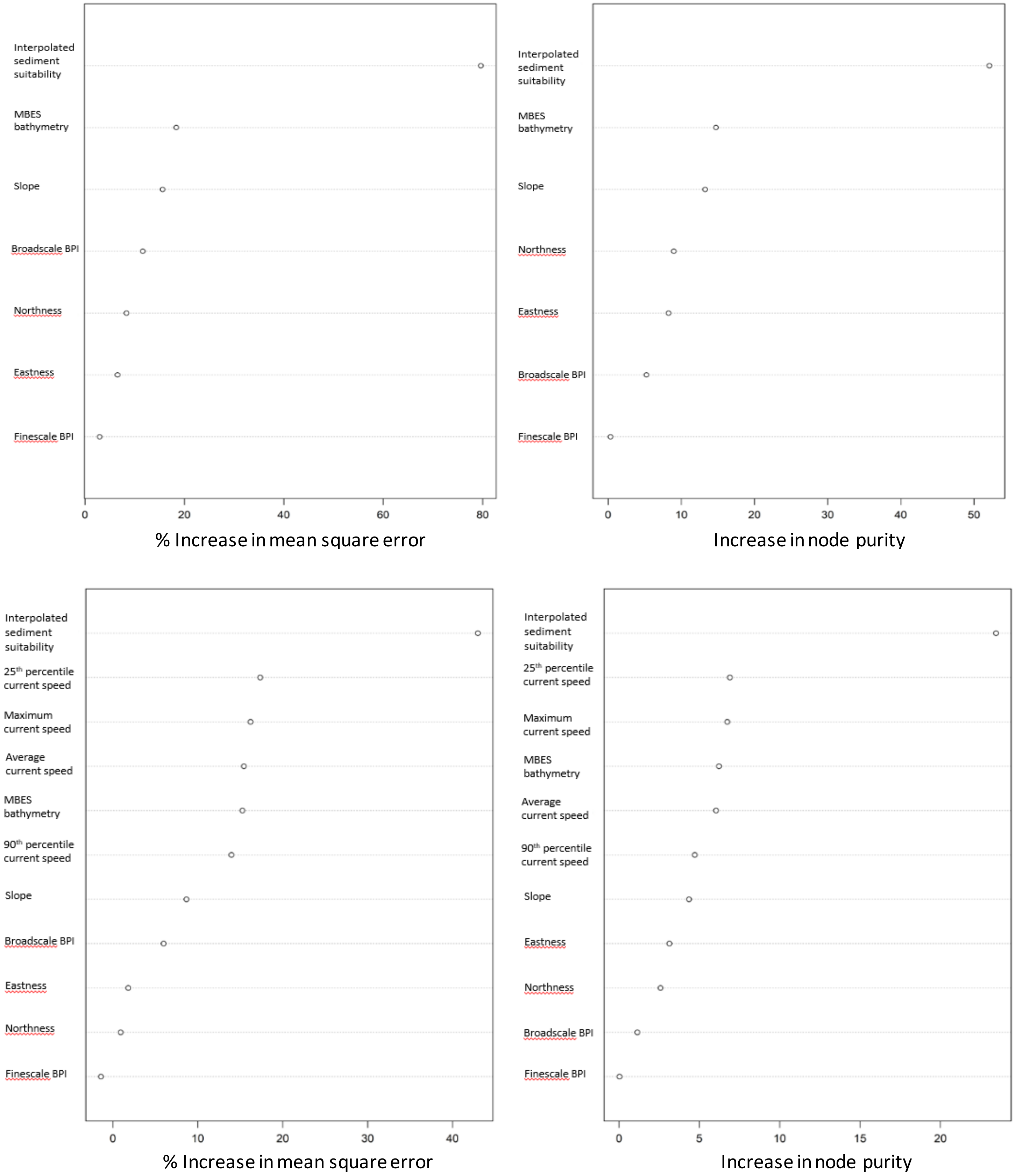


Figure 1. Variable importance plots for Model 1 (top) and Model 2 (bottom)

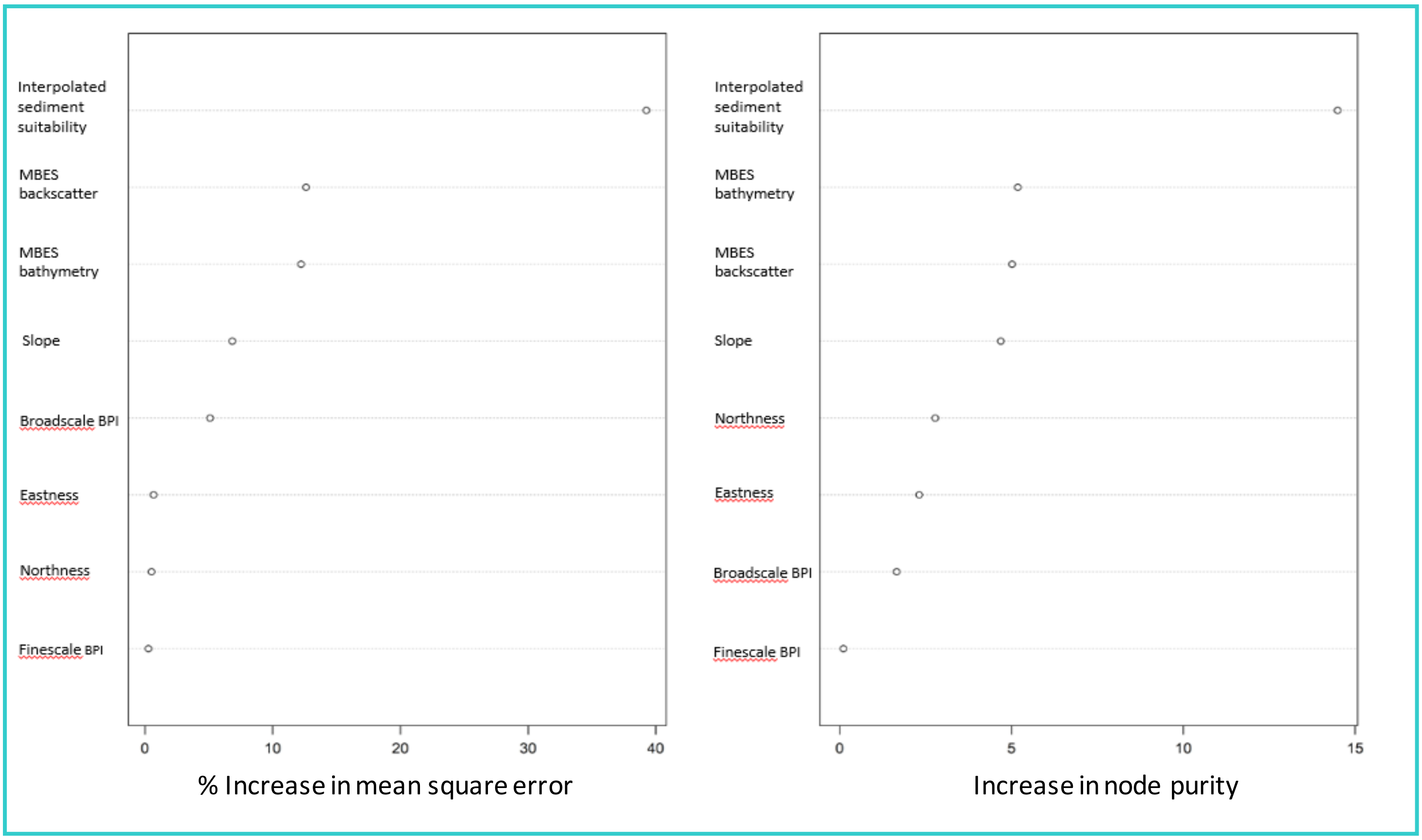


Figure 2. Variable importance plots for Model 3