Marine Environmental Research 201 (2024) 106706

Contents lists available at ScienceDirect

Marine Environmental Research

journal homepage: www.elsevier.com/locate/marenvrev

Spatial distribution of sandeel (*Hyperoplus lanceolatus)* and implications for monitoring marine protected sites

Robert Mzungu Runya [a,1,*], Chris McGonigle [a], Rory Quinn [a], Alexander Callaway [b]

[a] School of Geography & Environmental Sciences, Ulster University, Cromore Rd, BT52 1SA, Co. Derry, Northern Ireland, UK
[b] Agri-Food and Biosciences Institute, 18a Newforge Lane, BT9 5PX, Belfast, CO. Antrim, Northern Ireland, UK

ARTICLE INFO

Keywords:
Multi-frequency backscatter
Unsupervised classification
K-means clustering
Marine protected areas
Conservation
Management
Species

ABSTRACT

Increased human demand on the marine environment and associated biodiversity threatens sustainable delivery of ecosystem goods and services, particularly for shallow shelf-sea habitats. As a result, more attention is being paid to quantifying the geographical range and distribution of seabed habitats and keystone species vulnerable to human pressures. In this study, we develop a workflow based on unsupervised K-Means classification units and Generalized Linear Models built from multi-frequency backscatter analyses (95, 300 kHz), bathymetry and bathymetry derivatives (slope) to predict different levels of sandeel densities in Hempton's Turbot Bank Special Area of Conservation (SAC). For *Hyperoplus lanceolatus* densities, the performance of single frequency verses multi-frequency models is compared. Relatively high agreement between K-Means clustering outputs (from 95 kHz and multi-frequency models) and ground-truthed sandeel densities is noted. Moreover, Root Mean Squared Error (RMSE) values in this instance demonstrate that single-frequency models are favoured over the multi-frequency model in terms of predictive ability. This is mostly linked to the species strong affinity for sedimentary environments whose variability is better captured by the lower frequency system. Generally, these results provide important information about species-habitat relationships and pinpoint bedform features where sandeels are likely to be found and whose variability is potentially linked to the bathymetry domain. The workflow developed in this study also provides a proof of concept to support the design of a robust species-specific monitoring plan in marine protected areas. Most importantly, we highlight how decisions made during sampling, data handling, analysis could impact the final outputs and interpretation of Species Distribution Models and benthic habitat mapping.

1. Introduction

Human pressures, primarily coastal and tourism development, aggregate extraction, commercial fishing, and pollution are constantly threatening continental shelf ecosystems around the world (Sala et al., 2021; Jefferson et al., 2021; Halpern et al., 2019, McCauley et al., 2015). These adverse effects are being exacerbated by the rapid growth of the coastal population, which now accounts for approximately 5% of the world's population, placing a high demand on marine resources (Jefferson et al., 2021; Halpern et al., 2019). Marine shelf-sea environments represent ~10% of the world's seabed yet are among the most productive systems, delivering a diverse array of ecosystem goods and services (Spalding et al., 2007). This has prompted the need for a quantitative evaluation of the magnitude of the consequences and the identification of regions at risk of additional degradation, notably the very dynamic seabed habitats (Mitchell et al., 2019). Acoustic mapping is critical for providing broad-scale information on the geographical range and distribution of benthic habitats and features of interest (Brown et al., 2019; Costa, 2019). However, for decades, acoustic surveys utilising multibeam echosounders (MBES) have placed a premium on gathering and processing high-quality bathymetry data, placing less emphasis on the significance of backscatter analyses. These MBES bathymetry datasets have been beneficial for a variety of marine applications, including military applications (Blondel, 2000), underwater archaeology (Majcher et al., 2020), geology (Runya et al., 2021), informing marine policy (Howell et al., 2010), and ecological assessments (Costa, 2019).

The validity of backscatter data is constrained by a paucity of

* Corresponding author.
E-mail address: r.-mzungu-runya@ulster.ac.uk (R.M. Runya).
[1] Fisheries Ecosystems Advisory Services (FEAS), Marine Institute, Galway H91 R673, Ireland. robert.runya@marine.ie

https://doi.org/10.1016/j.marenvres.2024.106706
Received 22 March 2024; Received in revised form 26 July 2024; Accepted 23 August 2024
Available online 25 August 2024

temporally congruent biological ground truth data and a lack of an appropriate scale in which remote sensing and ground truth data are comparable (Brown et al., 2019; Costa, 2019; Lecours et al., 2015). Additionally, backscatter acquisition is often unsupervised, uncalibrated, and data are mostly obtained for opportunistic analyses in subsequent studies (Runya et al., 2021; Misiuk et al., 2020). Over the last decade, scientific interest in mapping geological and biological features on the seabed at a finer scale has intensified (Brown et al., 2019; Costa, 2019). With less than 30% of the seafloor well mapped (Seabed2030, 2024), the availability of multi-frequency MBES backscatter data and high-resolution bathymetry data allows for a comprehensive seafloor discrimination and mapping (Costa, 2019; Fakiris et al., 2018). Furthermore, the emerging techniques for segmenting MBES acoustic data has fundamentally changed how seabed habitats are characterised and mapped (Summers et al., 2023; Lecours et al., 2016; Calvert et al., 2015; McGonigle et al., 2009). Benthic habitat maps derived from MBES data are crucial for spatial planning and management of marine protected areas (Costa, 2019). Again, wider global initiatives such as the Seabed2030, an initiative between the Nippon Foundation and GEBCO (The General Bathymetric Chart of the Oceans) aim to generate 100% seabed coverage of the world's oceans. Similarly, the United Nations Decade of the Ocean has a vision that supports the collection of scientific information to aid the monitoring and management of the marine biodiversity (GEBCO, 2022; Wölfl et al., 2019). International, regional, and national policies, including the Convention on Biological Diversity (CBD), the European Marine Strategy Framework Directive (MSFD), and the EU Habitats and Council Directive 92/43/EEC, address the need for effective marine conservation and fisheries management Diesing et al. (2020); Hogg et al. (2018).

Mapping seabed habitats enables policymakers, managers, scientists, and direct resource users get a better understanding of the types and distribution of habitats and biological communities on the seafloor (Costa, 2019; Calvert et al., 2015). Besides, access to reliable seabed habitat maps helps to develop successful management plans and strategies for the marine environment (Costa, 2019). Acoustic surveyors may now distinguish seabed habitats and geomorphological features with varied acoustic signatures using multi-frequency backscatter data analysis (Runya et al., 2021; Fakiris et al., 2018; Montereale-Gavazzi et al., 2018). The use of multi-frequency backscatter is gaining wider scientific attention particularly for exploring the varied frequency responses of different sediment types (Menandro et al., 2022; Gaida et al., 2018). To derive ecologically meaningful information, Lecours et al. (2015) recommends that these mapping efforts consider the spatial scale between the acoustic remotely sensed data and available ground truth data, as this is likely to influence our inference of the spatial extent and distribution of benthic habitats and biological communities on the seafloor. Acoustic data are often utilised in conjunction with biological ground truth data as abiotic surrogates to aid the generation of spatially detailed benthic habitat maps (Costa, 2019; Calvert et al., 2015; Lucieer et al., 2013). These techniques are classified broadly as unsupervised and supervised classification (Misiuk and Brown, 2024; Brown et al., 2017; Calvert et al., 2015; McGonigle et al., 2009). Recent literature has emphasized and addressed specific strategies for quantitative integration of remotely sensed environmental variables with ground truth data (Langton et al., 2021; Costa, 2019; Brown et al., 2017; Calvert et al., 2015). However, most of this research is based on opportunistic analyses, often using data sets that are not only non-contemporaneous in time but were also acquired for other objectives (Brown et al., 2017; Clarke, 2015). Again, this complicates the task of comparing survey-specific data and diminishes their ecological validity, a difficulty recognised by the scientific community (Lamarche and Lurton, 2018; Lurton and Lamarche, 2015).

The Great sandeel (*Hyperoplus lanceolatus)* hereafter synonymised with "sandeel" is a species found largely inshore inhabiting depths of up to roughly 60 m, primarily between 6 and 30 m (Ruiz, 2008). It only occurs between 80°N and 37°N, 25°W and 36°E (OBIS, 2022) and is widely distributed in the Northeast Atlantic Ocean (UK and Ireland included). The species is highly linked with sandy substrates, with their abundance influenced by sediment type and particle size (Baker et al., 2022; Greene et al., 2020; Tien et al., 2017). They frequently spawn in the spring and summer with mature individuals reaching 40 cm in length (Tien et al., 2017). *H. lanceolatus* is an important forage species that contribute significantly to the diet of predatory seabirds, piscivorous commercial fish and marine mammals (Nadolna-Ałtyn et al., 2017), playing a significant role in distributing energy up the food chain (Wanless et al., 1998; Santos et al., 2004). Despite their small size, their ecological role has spurred scientific interest focussing on predicting their occurrence and spatial distribution (Langton et al., 2021; Hill et al., 2020; Matta and Baker, 2020; Baker et al., 2019), for conservation and impact assessment purposes. However, data on this keystone species is limited in our study area - the Hempton's Turbot Bank (HTB) Special Area of Conservation (SAC) and the wider North Atlantic Ocean with most records coming from grab samples, dredge and beam trawls (OBIS, 2022; Baker et al., 2019; Han et al., 2012). Similarly, the HTB SAC controls energy fluxes, fish larval dispersion, and contributes to total marine biodiversity (NPWS, 2024).

Our research develops a workflow that can assess the ecological validity of multi-frequency backscatter data analysis to examine the linkages between species occurrence and geomorphological characteristics that can help management and conservation efforts. The study particularly focusses on vulnerable but usually less studied habitat and species such as shallow sand banks and sandeels of the North Atlantic Ocean. Specifically, the study aims to use the developed workflow to demonstrate how we can; (i) define the optimal class size for unsupervised classification that better represents environmental heterogeneity, (ii) identify geomorphological features linked to high levels of sandeel densities, and (iii) assess the variability in sandeel densities as described by multi-frequency backscatter, bathymetry, and slope angle. Even though dual-frequency MBES backscatter data (95, 300-kHz) are utilised, this study was originally conceived as a multi-frequency backscatter investigation; therefore, the term "multi-frequency" will be used.

2. Materials and methods

2.1. Regional setting

MBES and sandeel data were gathered on Hempton's Turbot Bank (HTB hereafter) off the north coast of Ireland (Fig. 1), an Annex I Habitat under the EU Habitats Directive (Code 1110: sandbanks which are slightly covered by seawater all the time; Site Number 002999). Despite the current designation's assumption of homogeneity, the Special Area of Conservation (SAC) reflects a region of geomorphological and biological diversity (Evans, 2018; Picton and Costello, 1998). Resident burrowing species in the SAC include lesser and greater sandeels, polychaetes, platyhelminths, nematodes, nematodes among others.

The Bank's seabed substrates include coarse gravel, cobble pavements, coarse and fine sediments, and sand mixed with organic compounds (Evans, 2018). Specifically, the area of HTB (Fig. 1) investigated here is a sandbank comprising medium sand, coarse sand, very coarse sand and sandy gravel (Runya et al., 2021). It has a complex topography made up of asymmetric sand waves, ripples and areas of flat seabed (Evans et al., 2015). The Bank forms part of a larger mobile sand feature that is part of the shelf in the South Malin Sea (Fig. 1). HTB is approximately 8.6 km long, 2.3 km wide (6.91 km^2), and <80 m deep (Runya et al., 2021). Between Malin Head and Islay (Fig. 1) are strong hydrodynamic forces associated with the Islay front (a tidal mixing front) that contribute to the shelf-wide ecosystem's biological productivity (Simpson et al., 1979).

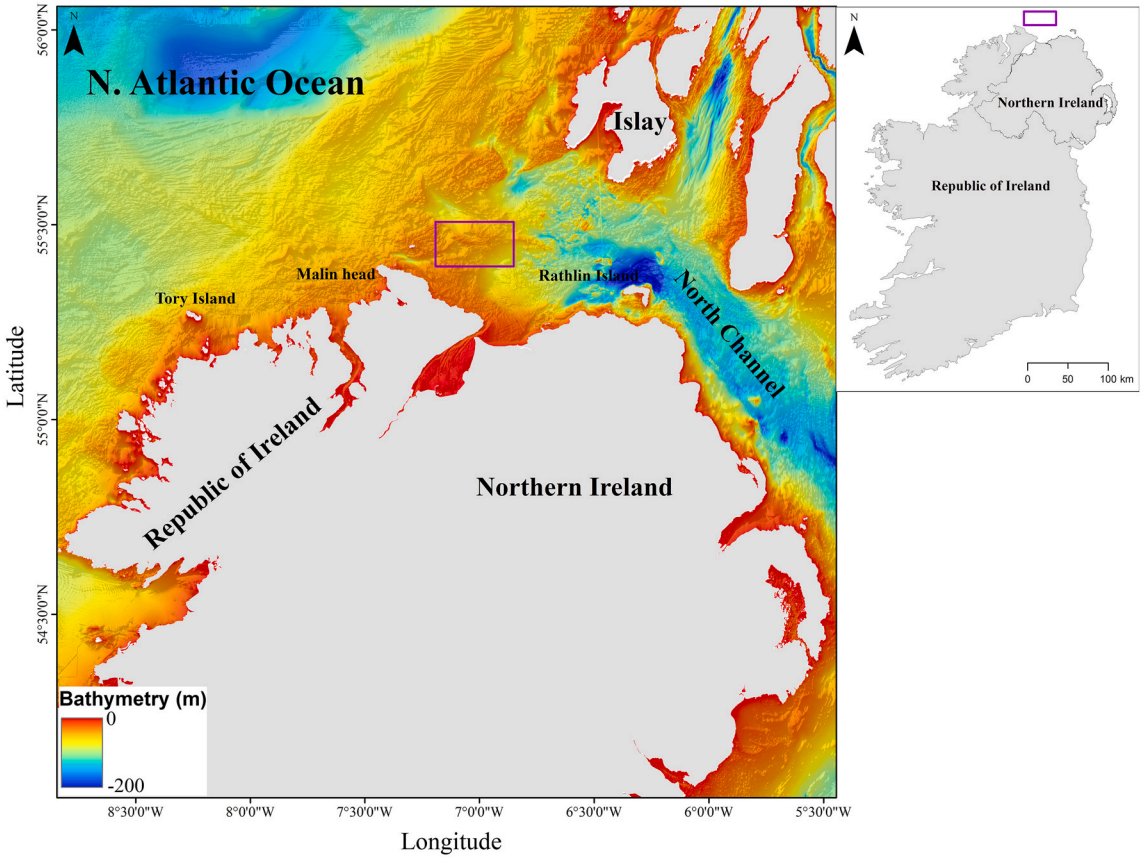

Fig. 1. Map of the study area and bathymetry of Hempton's Turbot Bank Special Area of Conservation (SAC). Inset map bathymetry source: GEBCO, 2019.

2.2. Data acquisition and processing

2.2.1. Environmental variables

MBES data were acquired during three research cruises: CE19007 in 2019 on RV *Celtic Explorer*, CV13030 in 2013 on RV *Celtic Voyager* and CE0402 in 2004 on RV *Celtic Explorer*. Kongsberg Simrad EM302 (30 kHz), dual-head Kongsberg Simrad EM3002 (300 kHz) and Kongsberg Simrad EM1002 (95 kHz) systems were used during the surveys. A Kongsberg SDP-10 (DP1) positioning system was used together with a motion reference unit (Kongsberg Simrad Seapath 200) to account for pitch, roll and heave. MBES data were acquired with Kongsberg's Seafloor Information System (SIS); and Sound Velocity Profiles (SVP) were measured using Conductivity, Temperature, Depth (CTD) casting to account for the variation in sound velocity and quality control of multibeam data (Montereale Gavazzi et al., 2016).

Backscatter data were processed using QPS Fledermaus Machine Geocoder Toolbox (FMGT v.7.8.9). The FMGT backscatter processing was implemented by applying an AVG (Angular Varying Gain) correction using an adaptive AVG band (300 window size) with mosaic parameters set at 50%-line blending, no nadir if possible 25% and a dB mean filter (Lurton and Lamarche, 2015). Meanwhile, the bathymetry data was processed using CARIS HIPS & SIPS v.9.1. The backscatter mosaics and bathymetric grid were exported as ASCII formatted grids at 0.5 m and 2 m spatial resolution (WGS1984, UTM zone 29N) respectively. These 32-bit floating point raster datasets were imported in ArcGIS v.10.6.5 and R (v.4.1.0) for further analysis. However, the spatial overlap among the three backscatter mosaics was limited, leading to the exclusion of the 30 kHz mosaic in subsequent analysis. Note that, the backscatter mosaics were resampled to 2 m spatial resolution to ensure consistence with the bathymetry grid.

2.2.2. Bathymetry derivatives and feature selection

Bathymetry data were used to generate terrain derivatives that are commonly used as abiotic proxies in benthic habitat mapping (Lecours et al., 2016; Calvert et al., 2015) using the Spatial Analyst and Benthic Terrain Modeler toolbox in ArcGIS v.10.6.5 (Walbridge et al., 2018). Apart from backscatter data, other initial variables considered were eastness, northness, curvature, Vector Ruggedness Measure (VRM), standardised fine- and broad-scale Bathymetric Position Index (BPI), and slope which were extracted from bathymetry data. The broad-scale BPI was calculated using an inner radius of 10 cells and an outer radius of 100 cells. Meanwhile, the fine-scale BPI was calculated using an inner radius of 1 cell and an outer radius of 20 cells (Plets et al., 2012). The inner radius of 1 cell was selected because it relates to the native 2 m × 2 m spatial resolution of the data and also to allow for delineation of fine-scale topographic variation of the seafloor, which is important for understanding and identifying subtle features linked to sandeel distribution. Slope was calculated using the planar method and represents the steepness or gradient of the seafloor area in degrees (Menandro et al., 2022). Meanwhile, VRM which is a measure of seafloor roughness was calculated with a neighbourhood size of 3 x 3 pixels and represents the ratio of surface area to planar area (Pearman et al., 2020). This approach supports a more detailed geomorphological characterisation of the seafloor as evidenced in literature (Lecours et al., 2016; Calvert et al., 2015; Plets et al., 2012).

Final features that were used as inputs in our classification and modelling were selected primarily based on a Variance Inflation Factor threshold (VIF >0.5; Zuur et al., 2010) to ensure they were not highly correlated and eliminate multi-collinearity. This was further supported by use of pairwise scatter plots and Pearson Correlation Coefficients ($r > 0.7$) to remove one variable in a pair of highly-correlated variables, as recommended in literature (Tong et al., 2023; Principe et al., 2021; Price et al., 2019; Lauria et al., 2017; Zuur et al., 2010). Variables with a VIF

>0.5, and one in a pair of correlated variables with r > 0.7 were excluded from the analysis. These thresholds were chosen to balance the need to avoid or minimize multi-collinearity and the inclusion of ecologically relevant variables. The variables that were qualified using these criteria were backscatter (95, 300-kHz), bathymetry and slope. Variables such as curvature, VRM, eastness, northness, BPI (fine scale and broad scale) were eliminated due to their high VIF and p values indicating significant multi-collinearity.

2.2.3. K-means clustering

The choice of K-Means clustering algorithm was informed by its speed and simplicity of implementation in identifying underlying patterns in large image data, including those acquired from MBES surveys. This algorithm is widely used in marine mapping particularly during image segmentation and pattern recognition (Alevizos et al., 2018; Raykov et al., 2016, Ismail et al., 2015; Ahmed and Demšar, 2013), making it fit for our study objective of carrying out an unsupervised classification on the abiotic variables to discriminate substrate characteristics and geomorphological variability of the seafloor, and then examine how well these align with the predicted sandeel densities. The segmentation based on K-Means clustering allow for characterisation of the environmental heterogeneity in the benthic environment. Such segmentation outputs can be used to provide initial insights into sandeel-environment linkages. A visual and statistical examination between clustering outputs and sandeel densities will help to provide initial insights of how sandeels interact with their-environment By minimising the within-cluster distance and maximising the between-cluster distance, this algorithm seeks to uncover underlying patterns in the data. In R software (v.4.1.0) cluster package (v.1.5.2), a K-Means clustering algorithm was used for seafloor segmentation on four variable combinations including backscatter (95, 300-kHz), bathymetry, and slope, to compare single frequency versus multi-frequency backscatter analysis.

The clustering process in this paper involved a two-hierarchical step approach (Menandro et al., 2022; Swanborn et al., 2022; Hogg et al., 2016). The first step of the segmentation process involved image clustering involving the entire dataset, and the second involved segmentation within an existing designation of the sandbank feature which appears visually homogeneous, to further uncover underlying geomorphological variability (Misiuk and Brown, 2024; Menandro et al., 2022; Trzcinska et al., 2021; Hogg et al., 2016). Generally, the K- Means clustering approach in this study follows four main steps: (i) data gathering and preparation, (ii) K-Means algorithm implementation, (iii) determination of optimal number of clusters also known as k and, lastly (iv) cluster validation to examine the level of agreement between clustering outputs at different frequencies. The first step of data gathering involved the use of backscatter, bathymetry, and slope variables used as K-Means clustering inputs. These variables were scaled between 1 and 0 to ensure that all features received equal weight (Lecours et al., 2016; Calvert et al., 2015). The second step of K-Means algorithm implementation begins with a random selection of initial cluster centres. In an n-dimensional space, it assigns data points to the cluster centres that are closest to them using Euclidean distance in order to minimize the within-cluster variance (Ismail et al., 2015; Hamerly and Elkan, 2002). The process is subjected to an iteration and repeated until no more data points fit into a new cluster.

Thirdly, a decision on the optimal number of clusters (k) must be made for a K-Means clustering exercise to be achieved. Several methods for determining the value of k in K-Means clustering have been proposed (Ismail et al., 2015; Calinski and Harabasz, 1974). Using the cluster package (v.1.5.2) in R, the Silhouette Index (SI) was determined in the clustering while altering the value of k from 2 to 20. The SI is a measure of how similar a data point is to its own cluster compared to other clusters. The maximum value of SI is commonly regarded as the optimal value of k. After determining the optimal value of k, a final clustering map was generated to show the cluster membership at each location (Raykov et al., 2016; Ismail et al., 2015).

According to the SI values, k = 2 provided the optimal fit. However, pixel-based classifiers, on the other hand, run the risk of clustering group of cells of heterogeneous features into one class that appear similar in terms of their grey level (Lucieer et al., 2013). While the SI calculation suggested k = 2 as the optimal value, we settled on k = 4 based on statistical evaluation and prior knowledge of the area's geomorphology and sedimentary characteristics. Specifically, the observed plateau in the SI values between k = 2 and k = 4 together with previously documented presence of at least four distinct sediment types (Runya et al., 2021), justified the use of k = 4 in our final clustering process. The segmentation based on the k = 4 allowed us to identify fine-scale variability in geomorphology and sediment characteristics which are potentially ecologically relevant to sandeels. Finally, cluster validation comparing the level of agreement between clustering outputs for different frequencies was determined using Adjusted Rand Index (ARI. The closer the ARI values are to one, the better the agreement (Saeid et al., 2016). Similarly, these clustering outputs allowed us to evaluate the relative contribution of different seafloor types and geomorphological gradients to the distribution and predicted sandeel densities.

2.2.4. Ground truth data

During the Agri-Food Bioscience Institute's (AFBI) research cruise CO3420 in October 2020 (Autumn), a 2 m wide scientific beam trawl with a dual tickler chain and an 8 mm blinder in the cod end was deployed. The beam trawl was selected because of its ability to permit quantitative and repeatable sampling, as well as its ease of penetration into a wide variety of substrate types (Reiss et al., 2006). The abundance of sandeels was recorded in 14 stations (over a 10-min trawl with varying tow speed and tow length). Other taxa were noted but did not form the primary focus of this study and will not be included in the analysis. These sampling stations were selected based on an evaluation of high-medium-low MBES backscatter segmentation to identify distinct substrate and geomorphological characteristics associated with sandeels, based on previous sampling using Day grab and pipe dredge (Langton et al.,2021). The high-medium-low backscatter ranges are indicative of hard/coarser-mixed sediments with varying degrees of sand and gravel – soft sediments respectively. This approach allowed for a representative coverage of different substrate types to be captured. We also sampled areas where predicted absences had not been validated by ground truthing. However, the sampling did not consider transect selection and transect crossing through different substrate classes.

2.3. Sandeel modelling

MBES data and derivative (final variable selected were: backscatter, bathymetry, and slope) were used to fit a Generalized Linear Model (GLM) to spatially predict the density of sandeels. The density of sandeels (response) was calculated based on the number of individuals captured per unit area of the tow footprint as numbers per square metre. This footprint area covered by a tow was calculated as a product of the tow length (distance covered from start to end by each tow) and the area of the beam trawl used (2 m by 2 m). However, this approach assumes that the tow was conducted in a straight line, and thus fails to account for the effect of drag. The decision to use density instead of presence data as a response variable was made to compensate for the lack of precise spatial information of where the sandeels were captured within a tow. Again use of abundance data might be prone to overdispersion. Three GLM models were built in the sdm package (v.1.2.1) in R software: 95 kHz, 300 kHz, and multi-frequency models following a similar approach to Runya et al. (2021). The GLM models took the form of a Gaussian distribution with an identity link function. To ensure validity of our models, the data was subjected to assumption testing for: (i) linearity using scatter plots of predictors against the residuals, (ii) homoscedasticity using plots of the residuals against fitted values and examining the patterns that would show the lack of it, (iii) normality of residuals using

Q-Q plots where no deviations from normality were noted, and lastly (iv) the independence of residuals which was checked by evaluating the autocorrelation plots of the residuals. The GLM was used to generate models depicting geomorphological features and areas linked with high sandeel densities, and the analysis workflow shown in Fig. 2.

First, the correlation between individual predictor variables against sandeel densities was determined using the Pearson Correlation Coefficient, which indicates the direction and strength of association between the response and the explanatory variables (Runya et al., 2021). The predictive capacity of the GLM models was then assessed using the Root Mean Square Error (RMSE) and R^2 values at the statistical significance level of $p < 0.05$. The Leave One Out Cross-Validation (LOOCV) method with a random seed number of 250 was used, which is recommended for small sample sizes. This method allows for the use of all samples in both training and testing (Runya et al., 2021). We also determined the most important variable driving the modelling outputs based on the p-values. Finally, a visual demonstration and comparison of three tow stations with distinct geomorphological patterns is made and presented in the results section (stations 4, 12 and 13). These were separated into 100 m segments based on their tow footprint area. Backscatter, bathymetry, slope, K-Means classes and predicted sandeel density are presented and compared for each of these tow per segment.

3. Results

3.1. Multibeam echosounder data

In general, the bathymetry of the study area ranges between 19.2 and 51.3 m. Compared to the region bounded by the SAC, the south-eastern part of the study area is the deeper (Fig. 3). The slope values represent a variable seafloor gradient, with values ranging from between 0 and 45° (Mean: 2.1 ± 2.9°). The bathymetry at each sampling station ranges from 32.3 m at the shallowest station (11) to 45.8 m at the deepest station (13) (Fig. 3; Table 1). Again, the mean slope values at each sampling station are low (between 0.5 and 6.9°).

The mean backscatter of the data (95 and 300 kHz) ranges between 0.23 dB and −70 dB. The backscatter intensity for 95 kHz is relatively higher (Mean: 25.2 ± 6.8 dB) than the 300 kHz data (Mean: 26.7 ± 3.3 dB). The acoustic response of the 14 samples stations varies slightly (Fig. 3; Table 1). The stations with the highest mean backscatter intensity are 13 (95 kHz) and 14 (300 kHz), with respective values of −20.7 dB and −24 dB.

3.2. K-Means clustering

The k = 4 produced four distinct clusters for the 95 kHz, multi-frequency models, whereas for the 300 kHz model, three clear clusters are delineated (Fig. 4). The 95 kHz result of four distinct clusters demonstrates its ability to discriminate more variability in substrate types and geomorphology, possibly due to its deeper penetration. Nevertheless, the 300 kHz only provided three meaningful clear clusters with the fourth one (cluster 3) not providing additional interpretative value. We found a higher level of agreement (ARI = 0.92) between 95 kHz and the multi-frequency model, followed by 300 kHz and multi-frequency (ARI = 0.74), and the lowest for 300 kHz and the 95 kHz K-Means model (ARI = 0.72). Large sand waves, particularly on the SE section of the feature, are clearly identified in all three models, indicating a wide range of geomorphological differences between large and small features. The four clusters identified are generally in agreement with areas previously identified with different substrate types comprising of; sand (medium to coarse sand), gravelly sand (medium to coarse sand with shell fragments), sandy gravel (medium pebbles, medium sand and shell materials) and gravel (Runya et al., 2021). Large mobile sand waves consisting of coarse sand can also be found (Evans, 2018).

3.3. Sandeel data

Individual counts of sandeels from beam trawl samples reveals 72 *H. lanceolatus* individuals in 14 different locations (Fig. 3; Table 1). Other taxa recorded in the samples include; tunicates, rhodophytes, poriferas, ophiuroids, gastropods, bivalves and polychaetes. However, these additional taxa are not the primary focus of this research and their

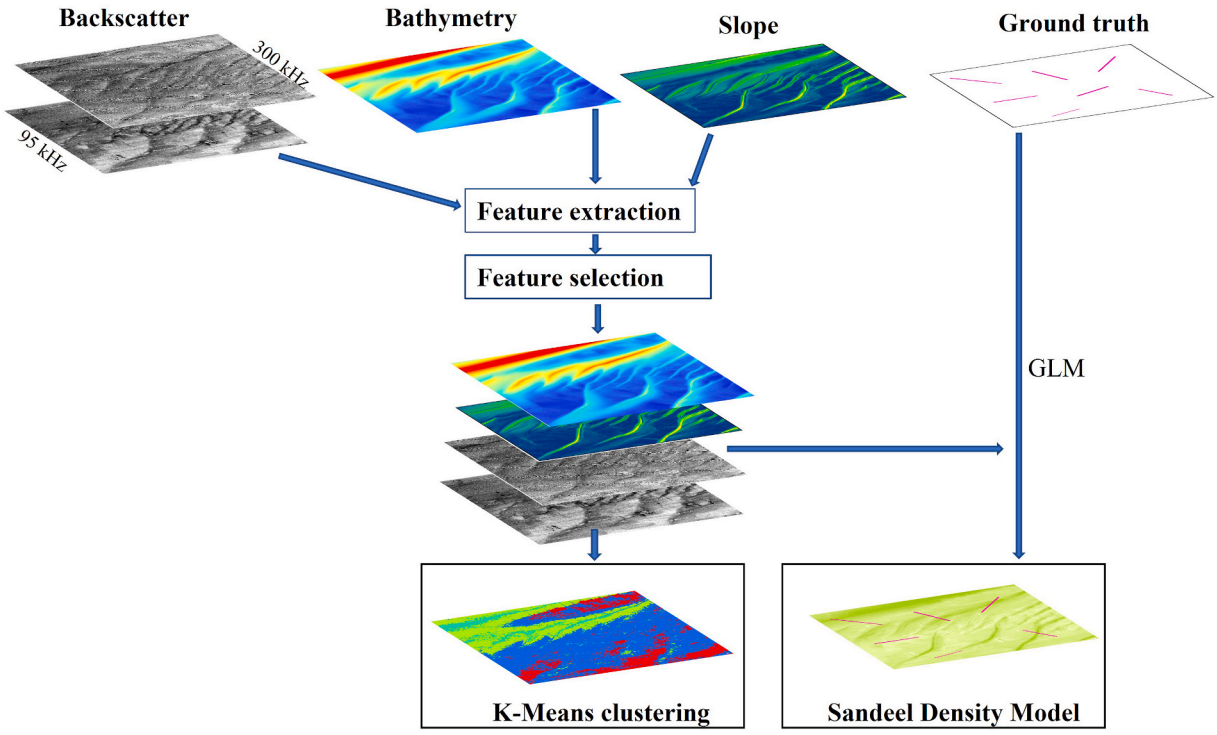

Fig. 2. Flow chart highlighting the variables and analytical workflow used in this study.

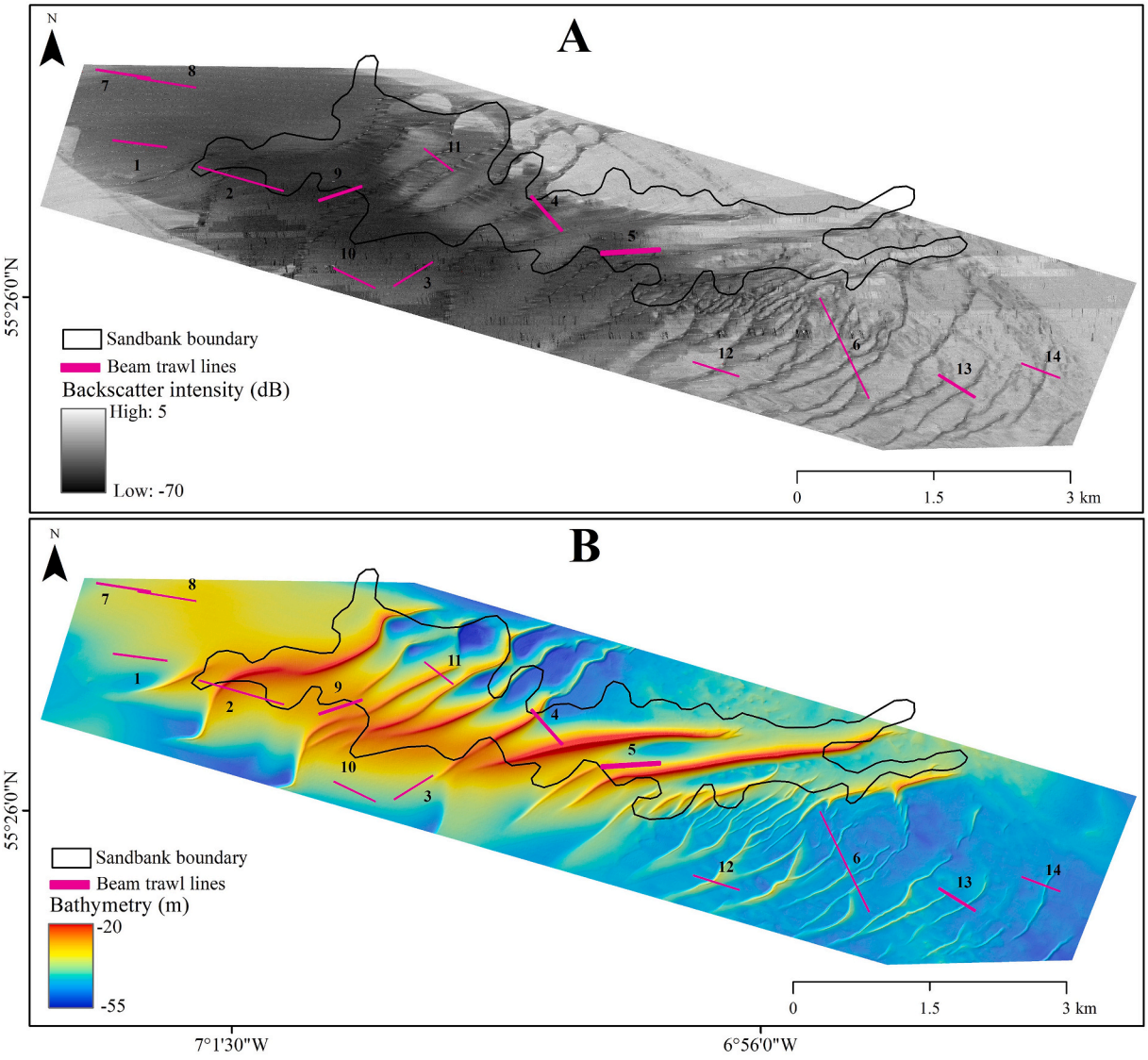

Fig. 3. Beam trawl footprint (20 m wide) plotted on (A) the 95 kHz backscatter mosaic and (B) bathymetry for the 14 ground truth stations (Table 1). The thickness of the beam trawls is a relative demonstration of the abundance values for each trawl station (the thicker the line the higher the abundance for that trawl station).

Table 1
Ground truth observations showing the number of sandeels and associated taxonomic data in 14 sampling stations.

Stations	Tow Length (m)	Abundance (#)	Density (No. per metre sq. x 10^{-3})	Sandeels Lengths (cm)	Average Backscatter (95 kHz)	Average Slope angle (deg.)	Average Bathymetry (m)
1	603.4	3	1.243	24.33 ± 9.5	−30.82	1.7048	−41.48
2	977.0	1	0.256	16 ± 0	−37.57	1.0333	−35.53
3	504.9	2	0.99	19 ± 5.7	−36.50	1.4448	−36.21
4	507.1	13	6.409	26.08 ± 3.1	−32.30	2.6999	−41.26
5	632.9	26	10.27	23.81 ± 5.9	−26.65	6.8663	−34.60
6	1230.5	0	0	0	−22.20	4.4841	−45.68
7	607.9	3	1.234	19.33 ± 2.5	−21.15	1.2446	−42.39
8	658.0	1	0.38	20 ± 0	−29.94	0.5026	−36.58
9	495.9	11	5.545	19.18 ± 2.1	−37.40	1.2369	−32.66
10	514.0	0	0	0	−36.66	1.3216	−40.07
11	403.2	0	0	0	−24.11	1.9343	−32.34
12	533.7	2	0.937	27 ± 0	−25.89	2.5159	−41.69
13	465.5	9	4.834	25.56 ± 6.2	−20.74	0.5464	−45.84
14	455.3	1	0.549	15 ± 0	−20.83	1.7149	−45.50
Average		5.14 ± 7.4	2.332 ± 3.2	21.3 ± 4.4			

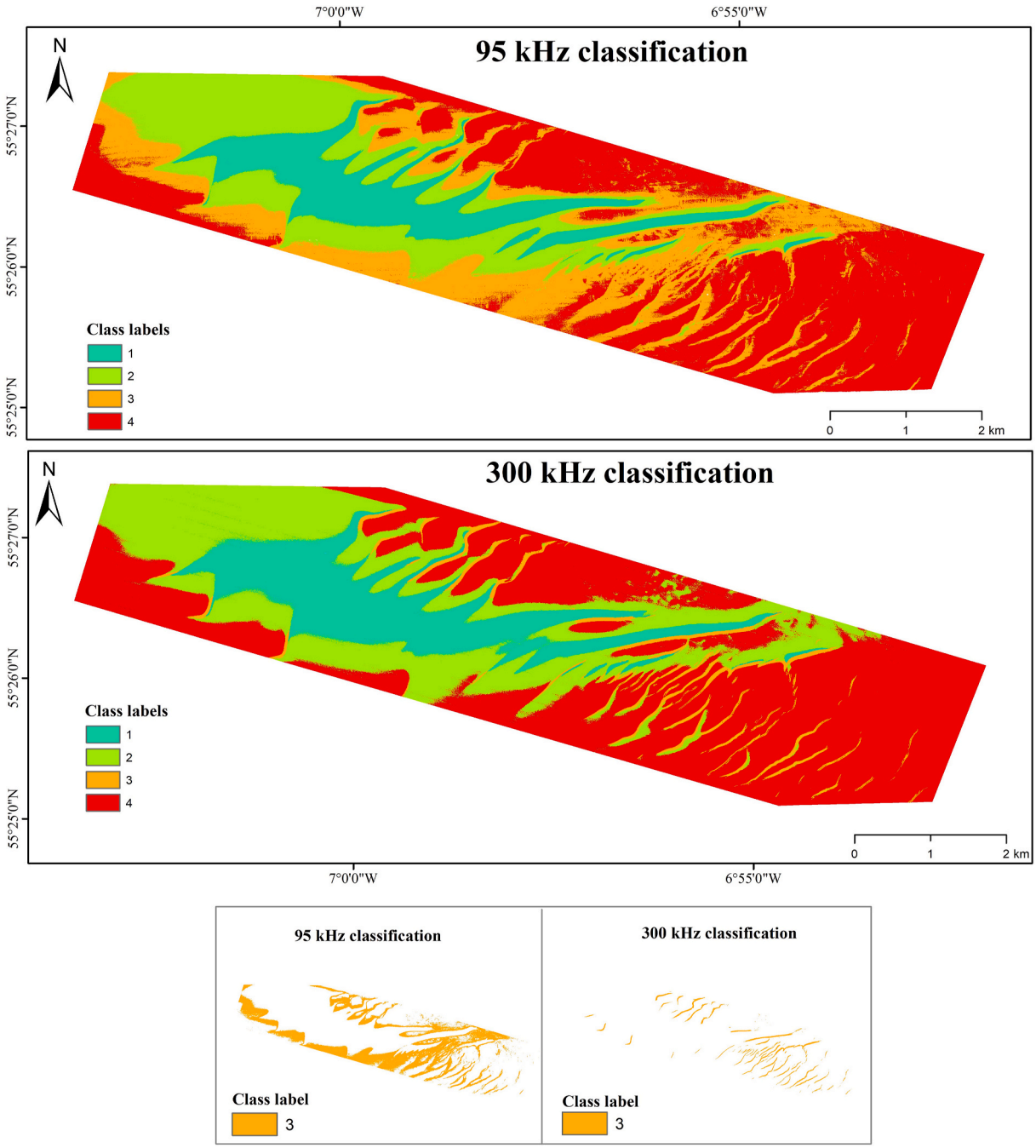

Fig. 4. K-means clustering maps for k = 4, comparing the 95 kHz and 300 kHz models. Inset shows the comparison for class 3 between 95 kHz and 300 kHz models.

mention is aimed at providing an ecological context of the area. These samples are dominated by medium and coarse sediments with a relatively low backscatter intensity (Fig. 3). Station 5 has the greatest number of sandeels (n = 38), followed by stations 9 (n = 11) and 13 (n = 9). Stations 6, 10, and 11 recorded n = 0 sandeels at water depths of 45 m, 39 m, and 41 m, respectively. Meanwhile, the sandeels in these samples are mature individuals ranging in length from 15 to 27 cm (Table 1). A wide variability is also noted in the sandeel densities across all the 14 stations (Table 1).

With respect to the clustering outputs, cluster 4 has the highest density of sandeels (2.6×10^{-3} m^{-2}) for 95 kHz, cluster 3 for the 300 kHz (1.0×10^{-3} per m^{-2}), and cluster 1 for the multi-frequency output (0.5×10^{-3} m^{-2}). Meanwhile, in all three clustering outputs, lower sandeel densities are recorded in cluster 2 ($<0.5 \times 10^{-3}$ per m^{-2}), with a greater range in the sandeel densities in the 300 kHz output. The inclusion of the segmentation outputs provided an additional layer of information that highlights the influence of geomorphological and sediment characteristics on sandeel densities.

3.4. Modelling of sandeel density

The correlation between the response (sandeel density) and the predictor variables (Backscatter, bathymetry, slope) was found to be weak but positive. Specifically, sandeel density display a weak correlation with slope (R = 0.37, p > 0.05), a very weak correlation with backscatter at 300 and 95 kHz, (R = 0.14 and 0.19 respectively; p > 0.05), and a negligible correlation with bathymetry (R = 0.05, p > 0.05). Similarly, the test for assumptions as needed for using Gaussian GLM

showed that the relationship between the predictor variables and sandeel densities was not only linear but also maintained homoscedasticity with residuals being normally distributed and independent. No significant spatial autocorrelation was noted.

Regardless of these weak correlations, we produced three spatial distribution models of sandeel density using two single-frequency (95, 300 kHz) and one multi-frequency model (Fig. 5). These models depict areas where relatively higher densities of sandeels are likely to occur, especially in the central section and along sand waves with low backscatter responses and coarse sandy sediments. A visual interpretation of these models at the scale of the individual beam trawls indicates a certain level of agreement especially between sandeel densities and, both slope and bathymetry as demonstrated with stations 4, 12 and 13 (Figs. 6–8). These trawl stations indicated a finer-scale variability, with sand waves having relatively higher slope values. In general, station 13 has the highest mean backscatter intensity (−20.7 dB) and a slightly higher mean water depth (45.8 m) of the three stations (Figs. 6–8). Station 4 has the lowest mean backscatter intensity (−32.3 dB) and the shallowest (41.3 m). Station 12 and 13 are dominated by cluster 4 and have less of a presence of cluster 2 and 3 respectively (95 kHz) with no representation for cluster 1. Meanwhile, seafloor types at station 4 were proportionally distributed for clusters 1, 2 and 3 and an observed fine-scale variability in the slope gradient which is consistent with bathymetry and a more homogenous backscatter profiles (Figs. 6–8).

With regards to the model performance from the GLM analysis, the RMSE values for the 300 kHz (0.007) and 95 kHz (0.0072) models were slightly better than the multi-frequency model (RMSE = 0.0074). However, all models had a marginally low predictive capacity, with adjusted R^2 values ranging from 0.32 (300 kHz and multi-frequency models) to 0.26 (95 kHz model). The models predicted mean sandeel

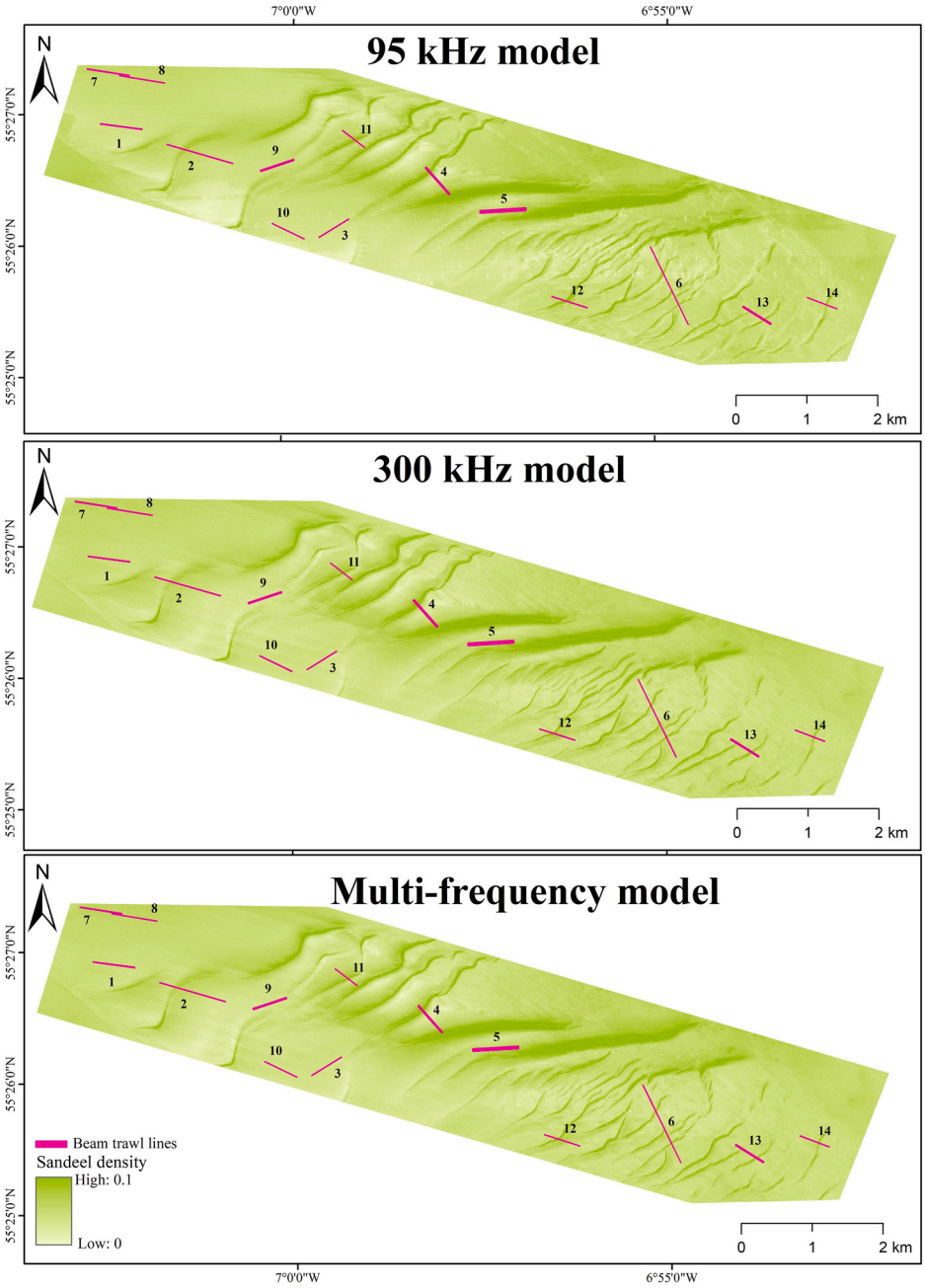

Fig. 5. Predicted sandeel densities for single-frequency (95, 300-kHz) and multi-frequency models. The beam trawl footprint is overlayed on top. To demonstrate the visual distinction of these three models, trawl 4, 12 and 13 as an illustration are further shown in Figs. 6–8).

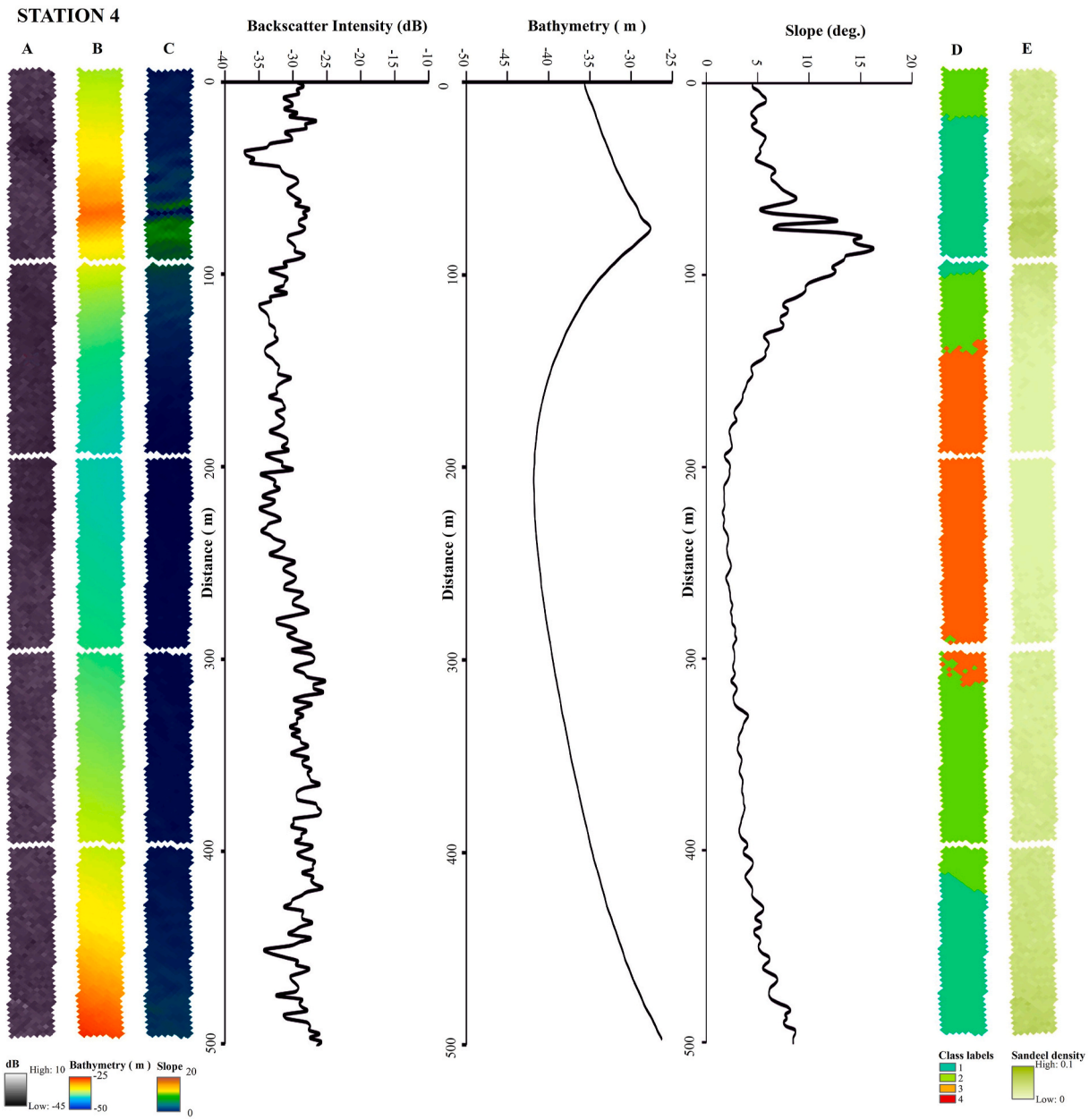

Fig. 6. Fine scale variability of the seafloor for trawl station 4 showing variation in backscatter intensity(A), bathymetry (B), slope angle (C), K-Means clustering (D) and predicted sandeel density (E).

densities of $0.0055 \pm 0.0.07$ per m^2 (95 kHz model) and 0.0054 ± 0.008 per m^2 (multi-frequency and 300 kHz models). Overall, the single-frequency and multi-frequency models provided relatively similar predictions of sandeel densities with minimal variation.

4. Discussion

This research demonstrates a workflow for examining the utility of K-Means classification and GLM based on multi-frequency backscatter, bathymetry, and slope data to discriminate different levels of sandeel densities. In general, the data demonstrates that K-Means clustering may be used to distinguish geomorphological features and sediment characteristics associated with sandeel densities. Furthermore, the study provides key lessons and recommendations for optimising the use of these data and modelling techniques to predict the spatial distribution of sandeels. The emphasis for future research is placed on having a robust approach, comprising the initial decisions made during sampling design, data handling and processing to the final modelling and analysis of data. Such a robust approach is needed to make better-informed decisions for the monitoring and management of *H. lanceolatus* species.

4.1. K-Means clustering

The K-Means clustering outputs distinguish between seabed characteristics and bedform features that can support the presence of sandeel *H. lanceolatus*. This clustering approach has been extensively employed in the discrimination of seabed habitats (Alevizos et al., 2018; Ismail et al., 2015; Ahmed and Demšar, 2013). We examine single frequency and multi-frequency backscatter with respect to sandeel densities using the outputs of $k = 4$ clustering. A distinct separation of clusters is accomplished in all three models where three clear segments are evident especially for the 300 kHz output. The clustering outputs clearly distinguish between three areas covered by sandy gravel (medium pebbles, medium sand and shell materials), sand (medium to coarse

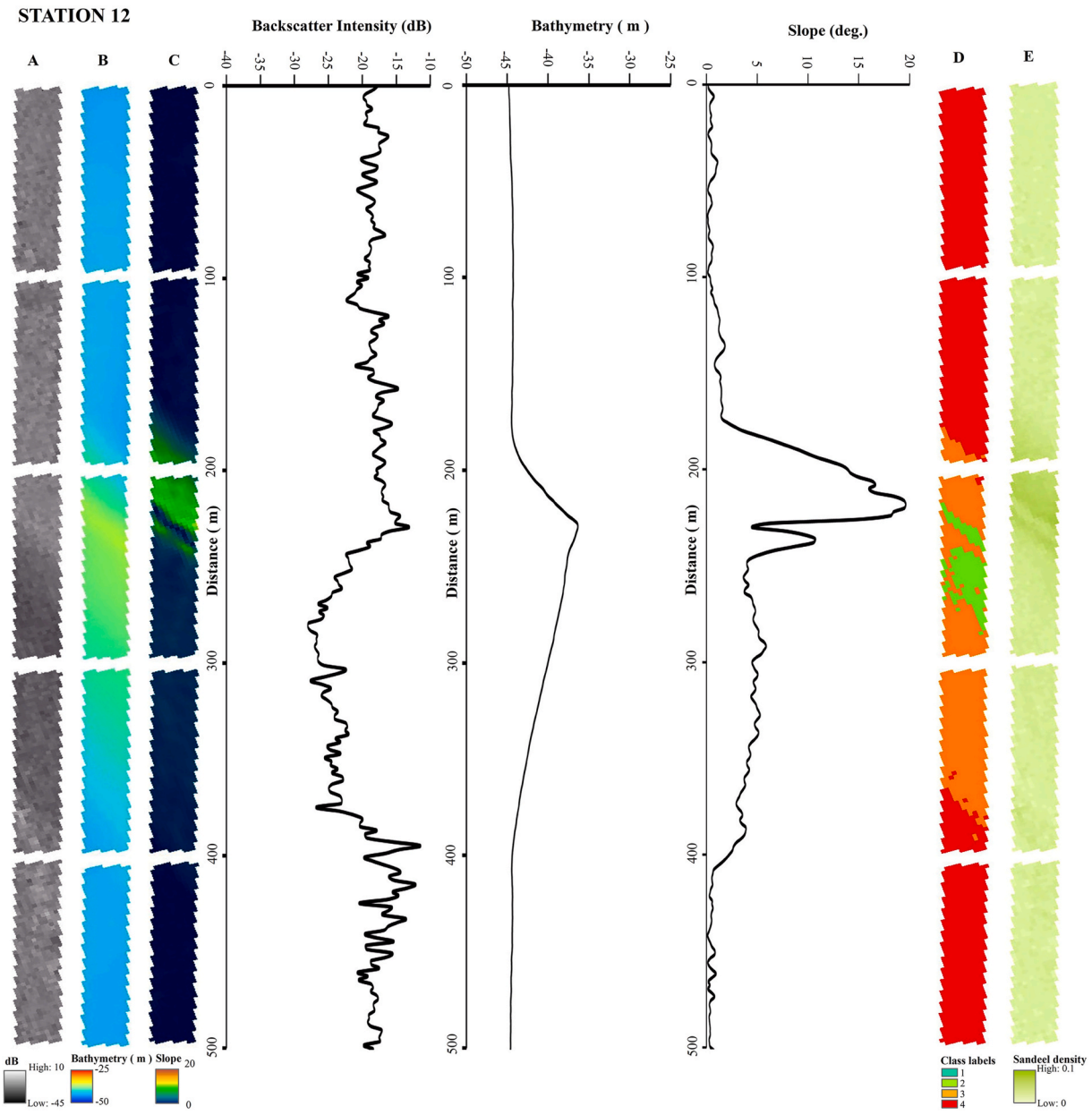

Fig. 7. Fine scale variability of the seafloor for trawl station 12 showing variation in backscatter intensity (A), bathymetry (B), slope angle (C), K-Means clustering (D) and predicted sandeel density (E).

sand) and gravelly sand (medium to coarse sand with shell fragments) (Appendix 1) which is consistent with earlier classification of substrates (Runya et al., 2021). The difference in the number of clear segments delineated by 95 kHz (four) and 300 kHz (three, with cluster 3 not providing much meaningful information and obscured in the main segmentation output) underscores the need to consider the unique characteristics of different frequencies during acoustic surveys campaigns. The 95 kHz data, with its deeper penetration ability delineated more clear substrate and geomorphological variations than that of the 300 kHz data (higher frequency) which discriminated fewer meaningful clusters due to its shallower penetration.

While the sand bank feature is considered largely homogeneous (NPWS, 2024), the fine scale variability indicates that geological characteristics such as substrate types are likely to influence clustering outputs (Evans, 2018). Additionally, a detailed visual examination of the beam trawl footprints (Figs. 6–8) indicates that bathymetry and slope have the most variability with a higher discriminatory power than backscatter. Areas with relatively low slope seem to be homogeneous in our segmentation findings as shown in other studies (Lecours et al., 2016; Calvert et al., 2015). Clustering results discriminated bedform features that are likely to support the presence of higher sandeel densities especially crests of sand waves. Crests have characteristic coarser sediments and due to the strong tidal activity around these features, aggregation of food materials makes them suitable for higher sandeel densities to occur (Greene et al., 2020). Potentially, the activity of tidal streams in this site controls the formation of these bedform features which provide suitable coarser substrates (Fig. 8) and moving currents that define the niche for this species. The ARI values indicate that the clustering outputs of the 95 kHz and multi-frequency outputs are in significantly better agreement (ARI = 0.92) meaning the variability in the multi-frequency model is mainly explained by the 95 kHz model. Visual detection of geomorphological features and substrates that could be linked to sandeels' occurrence and habitat preference using these differences is possible.

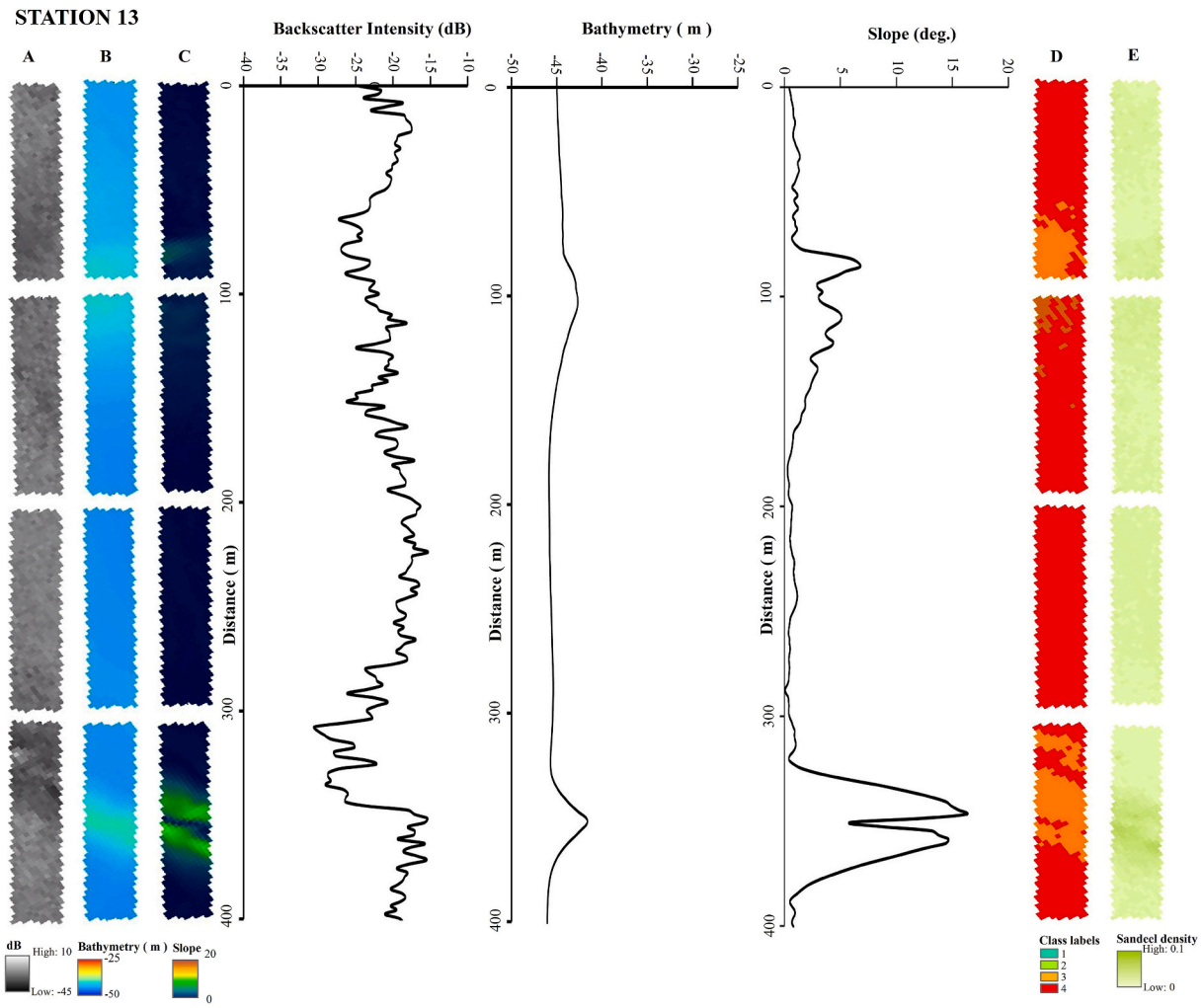

Fig. 8. Fine scale variability of the seafloor for trawl station 13 showing variation in backscatter intensity (A), bathymetry (B), slope angle (C), K-Means clustering (D) and predicted sandeel density (E).

4.2. Modelling of sandeel density

Meanwhile, between mean backscatter and sandeel densities, a weak correlation is observed based on Pearson correlation coefficients. However, despite these weak correlations, the findings from assumptions testing suggest that the predictor variables employed in the models are potentially suitable for explaining the variability in sandeel densities. Again, the findings of GLM models indicate that there is no significant difference in discriminating sandeel densities between single frequency and multifrequency models. However, based on RMSE values, a marginal gain in predictive capability is shown when single-frequency models (95, 300 kHz) are used. The high RMSE values compared to the predicted average sandeel densities generally depicts low predictive capacity of our models. This indicates the presence of an inherent variability in sandeel distribution coupled with the potential influence of other confounding environmental factors. Therefore, future research should attempt to integrate a wide-range of environmental factors and explore the potential utility of other modelling techniques to enhance the predictive capacity. A visual examination of these full-extent models reveals this minimal variation (Fig. 5). However, at the scale of the beam trawl footprint (Figs. 6–8), a somewhat higher sandeel density associated with areas comprising of coarser sediments is shown which demonstrates a geological control in sandeel distribution as observed in previous studies (Greene et al., 2020; Tien et al., 2017). The poor explanatory power of the models is extremely probable related to an imperfect beam trawl detection where precise spatial information of sandeel individuals cannot be ascertained, thus the generalizability of the final model results is limited (Zuur et al., 2010).

Again, it could also mean that sandeels are not utilising the sediments' environment during this time of the year, hence the rather weak association with the variables used (Baker et al., 2022). Besides, the lack of consideration for transect selection and transect crossing through different types of substrates and/or catchability of the beam trawl could have resulted in sub-optimal sampling. Sediment type and composition are inextricably related to the benthic communities that live on and in the seafloor (Pearman et al., 2020; Kostylev et al., 2001). Sandeels have an affinity for sandy substrates, particularly coarse sand with a low silt content (Baker et al., 2022; Greene et al., 2020; Tien et al., 2017; Holland et al., 2005). To some extent, the ease of penetration into the sediments and the capacity to allow exchanges in water, oxygen, and nutrients have been identified as the primary reason why sandeels prefer areas with coarser sediments to those with a high silt content or gravelly sediments (Endo et al., 2019; Holland et al., 2005; Wright et al., 2000).

A combination of water depth, residual flow rate, substrates and availability of food materials are strongly considered to impact local variation in sandeel densities (Jensen et al., 2011). The mechanical disturbance of sediments from the strong hydrodynamic conditions that characterises Hempton's Turbot Bank (Evans, 2018), is likely to result in sandeels following the preferred substrate around the cycle of sediment transport, more so within the crests-troughs footprint. This may help to

explain the patterns in sandeel densities predicted by the models which derives a relatively greater variability within the bathymetry domain (Figs. 6–8). Besides, past research has highlighted the effect of climate change and heavy exploitation of the sandeel fishery in characterising the regional variability and population structure of sandeels. Hempton's Turbot Bank (HTB) SAC is a protected area under the EU Habitats Directive because it has sandbanks, a refuge to a critically important sandeel species (NPWS, 2024). Sandeels contribute significantly to the diets of marine mammals, marine carnivores, and sea birds, and therefore the interactions between *H. lanceolatus* (prey) and its predators is also likely to influence its population dynamics (Langton et al., 2021; Hill et al., 2020; Wilson and Hammond, 2019; Tien et al., 2017; Anderwald et al., 2012).

Similarly, existing resource use conflicts between commercial fishing and seabird colonies that rely on sandeels provide an intriguing viewpoint and a challenge for sandeel fishery management (Furness, 2003; Wright, 1996). HTB SAC is not only a habitat for sandeels but also hosts several species of marine birds and marine mammals whose protection within this site is afforded on the mere basis of its geology (NPWS, 2024). Legally, more emphasis is given to charismatic megafauna while ignoring the ecological role of these small keystone species such as sandeels (Sibarani et al., 2019). However, developing methods for accurately monitoring the geographical range of sandeel species (that would have otherwise been considered less significant) and their habitats in SACs is an important aspect provided for by various regional policy frameworks targeting to enhance the conservation and management of marine biodiversity.

Modelling sandeel distribution in the Malin-Hebrides sea is important in developing effective conservation and management strategies for MPAs. Such an exercise is relevant for biodiversity conservation (identifying key habitats and biodiversity hotspots), identifying areas that have special importance to life history stages (e.g. feeding grounds), provides insights into ecosystem functioning (e.g. trophic dynamics), supports fisheries management (e.g. inform strategies for sustainable fishing practices), and being a keystone species, insights about their distribution over time can provide information about the impacts of human pressures and climate change. Lastly, the spatial dynamics of sandeels can form an important input to marine spatial planning processes thus promoting sustainable use of marine resources.

4.3. Limitations

Models can be used to predict the distribution of species such as sandeels beyond the geographic range of sampled locations, however this study acknowledges a number of limitations. Firstly, accurate prediction of species distribution is limited by the paucity of relevant ground truth data (Langton et al., 2021; Qiao et al., 2019). This limitation is linked to the use of legacy data in later studies with less consideration of its quality or intended application (Runya et al., 2021). As is the case with this research, we utilised a variety of acoustic datasets (to derive predictors) that are not consistent in time and were acquired for objectives other than ecological investigations. Secondly, the beam trawl's catchability is restricted in our study by the penetration depth, which varied based on the type of sediments (Reiss et al., 2006). The penetration depth of a commercial beam trawl can range between 8 cm in muddy sediments to 3 cm in sandy sediments (Reiss et al., 2006). Factors including beam trawl design, mesh size, and the presence of an 8 mm blinder further limits catchability (Holland et al., 2005; Freeman et al., 2004), thus leading to low capture rates and subsequent biased estimates of the actual spatial patterns and structure of sandeels in our study area. The study highlights the trade-off between gear catchability and the need to minimize environmental impacts during deployment. Also, employing an appropriate sampling strategy and gathering relevant data is emphasized. Thus the results are interpretable within the context of these constraints and the workflow can be used as a proof of a concept to provide better insights into spatial ecology of sandeels and inform the management of marine protected areas.

4.4. Future research

While our paper has successfully developed a workflow for modelling sandeels, there remains a significant opportunity for improvement in the methodology. The improvement includes the acquisition of specific datasets tailored around the individual species' life history strategies to fully account for its spatial dynamics. Integration with other surrogates such as hydrodynamic and oceanographic data can generate better information on the spatial range and behaviour of this species (e.g. Misiuk and Brown, 2024). To address the limitations highlighted in this study including detectability bias, a refined sampling design is needed, including careful transect selection and crossing through different substrate types. The segmentation of multibeam echosounder backscatter and bathymetry data can support transect selection and crossing to account for geomorphological variability (e.g. Summers et al., 2023), that can influence the spatial distribution of sandeels. Similarly, grab sampling can be used to provide spatially-explicit information of sandeels' occurrence, beyond the limitations of beam trawls. Overall, these recommendations provide an opportunity for the development of a robust, species-specific monitoring programme, that further enhances sandeel fishery research and management.

5. Conclusion

The research demonstrates a workflow for examining the utility of K-Means classification and Generalized Linear Modelling based on multi-frequency backscatter, bathymetry, and slope data to discriminate different levels of sandeel densities. The outputs of single and multi-frequency clustering are quite similar and have a relatively high agreement with predicted sandeel densities. Although the models demonstrated a generally weak predictive power, single frequency models are favoured for discriminating different levels of sandeels density over the multi-frequency model due to the complexity and high cost involved in multifrequency data acquisition, processing and analysis. This study also establishes the linkage between higher sandeel densities and bedform features comprising of coarser sediments. The workflow herein provides a proof of concept for developing a robust sampling strategy and monitoring program for sandeels needed to enhance fisheries and MPA management. However, we emphasize that interpretations and conclusions drawn from these results should take into account the uncertainty of the models. Most importantly, we highlight how decisions made during sampling, data handling, analysis could impact the final outputs and interpretation of Species Distribution Models and habitat mapping.

Funding

This research was funded by the Marine Institute (MI) under the Marine Research Programme by the Irish Government Cruises CE19007, CV13030, CE0402 and CO3420 surveys. Staffing was supported through the Marine Protected Area Monitoring and Management (MarPAMM) project, which is supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPM) with matching funding from the Government of Ireland, the Northern Ireland Executive, and the Scottish Government. This work was also carried out as part of Robert's PhD research funded through the Vice Chancellor Research Scholarship of Ulster University (U.K.).

CRediT authorship contribution statement

Robert Mzungu Runya: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Chris McGonigle:** Writing – review & editing, Writing – original draft, Validation, Supervision,

Methodology, Funding acquisition, Conceptualization. **Rory Quinn:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Conceptualization. **Alexander Callaway:** Writing – review & editing, Validation, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors would like to express their deepest appreciation to the financial supporters of this research. Additionally, we appreciate the institutional assistance provided by UU, AFBI, and MI in ensuring the success of this study. The authors would like to express their appreciation to the scientific and technical staff at UU, AFBI, and MI for their support and involvement in cruise planning, hydrographic surveys, data processing, and analysis, as well as final manuscript writing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marenvres.2024.106706.

References

Ahmed, K.I., Demšar, U., 2013. Improving seabed classification from Multi-Beam Echo Sounder (MBES) backscatter data with visual data mining. J. Coast Conserv. 17 (3), 559–577. https://doi.org/10.1007/s11852-013-0254-3.

Alevizos, E., Snellen, M., Simons, D., Siemes, K., Greinert, J., 2018. Multiangle backscatter classification and sub-bottom profiling for improved seafloor characterization. Mar. Geophys. Res. 39, 289–306. https://doi.org/10.1007/s11001-017-9325-4.

Anderwald, P., Evans, P.G.H., Dyer, R., Dale, A., Wright, P.J., Hoelzel, A.R., 2012. Spatial scale and environmental determinants in minke whale habitat use and foraging. Mar. Ecol. Prog. Ser. 450, 259–274.

Baker, M.R., De Robertis, A., Levine, R.M., Cooper, D.W., Farley, E.V., 2022. Spatial distribution of arctic sand lance in the chukchi sea related to the physical environment. Deep Sea Res. Part II Top. Stud. Oceanogr. 206, 105213 https://doi.org/10.1016/j.dsr2.2022.105213.

Baker, M.R., Matta, M.E., Beaulieu, M., Paris, N., et al., 2019. Intra-seasonal and inter-annual patterns in the demographics of sand lance and response to environmental drivers in the North Pacific. Mar. Ecol. Prog. Ser. 617–618, 221–244.

Blondel, P., 2000. Automatic mine detection by textural analysis of COTS sidescan sonar imagery. Int. J. Rem. Sens. 21 (16), 3115–3128. https://doi.org/10.1080/01431160050144983.

Brown, C., Beaudoin, J., Brissette, M., Gazzola, V., 2017. Setting the stage for multi-spectral acoustic backscatter research. In: Proceedings of the United States Hydrographic Conference. Galveston, TX, USA, 26 September 2017.

Brown, C.J., Beaudoin, J., Brissette, M., Gazzola, V., 2019. Multispectral multibeam echo sounder backscatter as a tool for improved seafloor characterization. Geosciences 9 (2019), 126.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. Theor. Methods 3 (1), 1–27.

Calvert, J., Strong, J.A., Service, M., McGonigle, C., Quinn, R., 2015. An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data. ICES (Int. Counc. Explor. Sea) J. Mar. Sci. 72 (5), 1498–1513. https://doi.org/10.1093/icesjms/fsu223.

Clarke, J.E.H., 2015. Multispectral acoustic backscatter from multibeam. Improved Classif. Potential. U.S. Hydro 2015 Conf. 1 (506), 19.

Costa, B., 2019. Multispectral acoustic backscatter: how useful is it for marine habitat mapping and management? J. Coast Res. 35 (5), 1062–1079. https://doi.org/10.2112/JCOASTRES-D-18-00103.1.

Diesing, M., Mitchell, P.J., O'keeffe, E., Gavazzi, G.O.A.M., Bas, T. Le, 2020. Limitations of predicting substrate classes on a sedimentary complex but morphologically simple seabed. Rem. Sens. 12 (20), 1–23. https://doi.org/10.3390/rs12203398.

Endo, A., Iwasaki, N., Shibata, J.Y., Tomiyama, T., Sakai, Y., 2019. The burrowing sand lance Ammodytes japonicus (Teleostei, Ammodytidae) prefers benthic sediments of low shear strength. J. Ethol. 37, 213–219.

Evans, W., 2018. Hydrodynamic Modelling of Sediment Transport and Bedform Formation on the NW Irish Shelf [Ulster University]. www.ulster.ac.uk.

Evans, W., Benetti, S., Sacchetti, F., Jackson, D.W.T., Dunlop, P., Monteys, X., 2015. Bedforms on the northwest Irish Shelf: indication of modern active sediment transport and over printing of paleo-glacial sedimentary deposits. J. Maps 11 (4), 561–574. https://doi.org/10.1080/17445647.2014.956820.

Fakiris, E., Zoura, D., Ramfos, A., Spinos, E., Georgiou, N., Ferentinos, G., Papatheodorou, G., 2018. Object-based classification of sub-bottom profiling data for benthic habitat mapping. Comparison with sidescan and RoxAnn in a Greek shallow-water habitat. Estuar. Coast Shelf Sci. 208 (April), 219–234. https://doi.org/10.1016/j.ecss.2018.04.028.

Freeman, S., Mackinson, S., Flatt, R., 2004. Diel patterns in the habitat utilisation of sandeels revealed using integrated acoustic surveys. J. Exp. Mar. Biol. Ecol. 305, 141–154.

Furness, R.W., 2003. Impacts of fisheries on seabird communities. Sci. Mar. 67 (S2), 33–45.

Gaida, T.C., Ali, T.A.T., Snellen, M., Amiri-Simkooei, A., van Dijk, T.A.G.P., Simons, D.G., 2018. A multispectral bayesian classification method for increased acoustic discrimination of seabed sediments using multi-frequency multibeam backscatter data. Geosciences 8 (12). https://doi.org/10.3390/geosciences8120455.

GEBCO, 2022. General Bathymetric Chart of the Oceans. https://www.gebco.net/about_us/seabed2030_project/. (Accessed 4 February 2022).

GEBCO (General Bathymetric Chart of the Oceans), 2019. General bathymetric chart of the oceans 30 arc-second global grid of elevations 2019. https://gebco.net. (Accessed 3 February 2019).

Greene, H.G., Baker, M., Aschoff, J., 2020. A dynamic bedforms habitat for the forage fish Pacific sand lance, San Juan Islands, WA, United States. In: Harris, P.T., Baker, E. (Eds.), Seafloor Geomorphology as Benthic Habitat. Elsevier, Amsterdam, pp. 267–279.

Halpern, B.S., Frazier, M., Afflerbach, J., Lowndes, J.S., Micheli, F., O'Hara, C., Scarborough, C., Selkoe, K.A., 2019. Recent pace of change in human impact on the world's ocean. Sci. Rep. 9 (1), 1–8. https://doi.org/10.1038/s41598-019-47201-9.

Hamerly, G., Elkan, C., 2002. Alternatives to the k-means algorithm that find better clusterings. International Conference on Information and Knowledge Management. Proceedings, pp. 600–607. https://doi.org/10.1145/584792.584890.

Han, Z., Yanagimoto, T., Zhang, Y., Gao, T., 2012. Phylogeography study of *Ammodytes personatus* in Northwestern Pacific: pleistocene isolation, temperature and current conducted secondary contact. PLoS One 7 (5). https://doi.org/10.1371/journal.pone.0037425.

Hill, S.L., Hinke, J., Bertrand, S., Fritz, L., et al., 2020. Reference points for predators will progress ecosystem-based management of fisheries. Fish Fish. 21, 368–378.

Hogg, O.T., Huvenne, V.A.I., Griffiths, H.J., Linse, K., 2018. On the ecological relevance of landscape mapping and its application in the spatial planning of very large marine protected areas. Sci. Total Environ. 626 https://doi.org/10.1016/j.scitotenv.2018.01.009.

Hogg, O.T., Huvenne, V.A.I., Griffiths, H.J., Dorschel, B., Linse, K., 2016. Landscape mapping at sub-antarctic south Georgia provides a protocol for underpinning large-scale marine protected areas. Sci. Rep. 6 (1), 33163 https://doi.org/10.1038/srep33163.

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. Mar. Ecol. Prog. Ser. 303, 269–282.

Howell, K.L., Davies, J.S., Narayanaswamy, B.E., 2010. Identifying deep-sea megafaunal epibenthic assemblages for use in habitat mapping and marine protected area network design. J. Mar. Biol. Assoc. U. K. 90 (1), 33–68.

Ismail, K., Huvenne, V.A.I., Masson, D.G., 2015. Objective automated classification technique for marine landscape mapping in submarine canyons. Mar. Geol. 362, 17–32. https://doi.org/10.1016/j.margeo.2015.01.006.

Jefferson, T., Costello, M.J., Zhao, Q., Lundquist, C.J., 2021. Conserving threatened marine species and biodiversity requires 40% ocean protection. Biol. Conserv. 264, 109368 https://doi.org/10.1016/j.biocon.2021.109368.

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 J. Mar. Sci. 68, 43–51.

Kostylev, V.E., Todd, B.J., Fader, G.B.J., Courtney, R.C., Cameron, G.D.M., Pickrill, R.A., 2001. Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs. Mar. Ecol. Prog. Ser. 219, 121–137.

Lamarche, G., Lurton, X., 2018. Recommendations for improved and coherent acquisition and processing of backscatter data from seafloor-mapping sonars. Mar. Geophys. Res. 39 (1–2), 5–22. https://doi.org/10.1007/s11001-017-9315-6.

Langton, R., Boulcott, P., Wright, P., 2021. A verified distribution model for the lesser sandeel *Ammodytes marinus*. Mar. Ecol. Prog. Ser. 667, 145–159. https://doi.org/10.3354/meps13693.

Lauria, V., Garofalo, G., Fiorentino, F., Massi, D., Milisenda, G., Piraino, S., Russo, T., Gristina, M., 2017. Species distribution models of two critically endangered deep-sea octocorals reveal fishing impacts on vulnerable marine ecosystems in central mediterranean sea. Sci. Rep. 7 (1), 8049. https://doi.org/10.1038/s41598-017-08386-z.

Lecours, V., Devillers, R., Schneider, D.C., Lucieer, V.L., Brown, C.J., Edinger, E.N., 2015. Spatial scale and geographic context in benthic habitat mapping: review and future directions. Mar. Ecol. Prog. Ser. 535, 259–284. https://doi.org/10.3354/meps11378.

Lecours, V., Dolan, M.F.J., Micallef, A., Lucieer, V.L., 2016. A review of marine geomorphometry, the quantitative study of the seafloor. Hydrol. Earth Syst. Sci. 20 (Issue 8) https://doi.org/10.5194/hess-20-3207-2016.

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. Estuar. Coast Shelf Sci. 117, 94–106. https://doi.org/10.1016/j.ecss.2012.11.001.

Lurton, X., Lamarche, G.(Eds) (2015). Backscatter measurements by seafloor mapping sonars. Guidelines and Recommendations.200p.http://geohab.org/wp-content/uploads/2014/05/BSWG-REPORT-MAY2015.pd.

Majcher, J., Plets, R., Quinn, R., 2020. Residual relief modelling: digital elevation enhancement for shipwreck site characterisation. Archaeol. Anthropol. Sci. 12 (6) https://doi.org/10.1007/s12520-020-01082-6.

Matta, M.E., Baker, M.R., 2020. Age and growth of pacific sand lance (ammodytes personatus) at the latitudinal extremes of the gulf of Alaska large marine ecosystem. Northwest. Nat. 101 (1), 34. https://doi.org/10.1898/1051-1733-101.1.34.

McCauley, D.J., Pinsky, M.L., Palumbi, S.R., Estes, J.A., Joyce, F.H., Warner, R.R., 2015. Marine defaunation: animal loss in the global ocean. Science 347 (6219).

McGonigle, C., Brown, C., Quinn, R., Grabowski, J., 2009. Evaluation of image-based multibeam sonar backscatter classification for benthic habitat discrimination and mapping at Stanton Banks, UK. Estuar. Coast Shelf Sci. 81 (3), 423–437. https://doi.org/10.1016/j.ecss.2008.11.017.

Menandro, P.S., Bastos, A.C., Misiuk, B., Brown, C.J., 2022. Applying a multi- method framework to analyze the multispectral acoustic response of the seafloor. Front. Remote Sens. 3, 860282 https://doi.org/10.3389/frsen.2022.860282.

Misiuk, B., Brown, C.J., 2024. Benthic habitat mapping: a review of three decades of mapping biological patterns on the seafloor. Estuar. Coast Shelf Sci. 296, 108599 https://doi.org/10.1016/j.ecss.2023.108599.

Misiuk, B., Brown, C.J., Robert, K., Lacharité, M., 2020. Harmonizing multi-source sonar backscatter datasets for seabed mapping using bulk shift approaches. Rem. Sens. 12 (4) https://doi.org/10.3390/rs12040601.

Mitchell, P.J., Aldridge, J., Diesing, M., 2019. Legacy data: how decades of seabed sampling can produce robust predictions and versatile products. Geosciences 9 (4). https://doi.org/10.3390/geosciences9040182.

Montereale Gavazzi, G., Madricardo, F., Janowski, L., Kruss, A., Blondel, P., Sigovini, M., Foglini, F., 2016. Evaluation of seabed mapping methods for fine-scale classification of extremely shallow benthic habitats - application to the Venice Lagoon, Italy. Estuarine. Coast. Shelf Sci. 170, 45–60. https://doi.org/10.1016/j.ecss.2015.12.014.

Montereale-Gavazzi, G., Roche, M., Lurton, X., Degrendele, K., Terseleer, N., Van Lancker, V., 2018. Seafloor change detection using multibeam echosounder backscatter: case study on the Belgian part of the North Sea. Mar. Geophys. Res. 39 (1–2), 229–247. https://doi.org/10.1007/s11001-017-9323-6.

Nadolna-Ałtyn, K., Podolska, M., Szostakowska, B., 2017. Great sandeel (Hyperoplus lanceolatus) as a putative transmitter of parasite *Contracaecum osculatum* (Nematoda: Anisakidae). Parasitol Res 116, 1931–1936. https://doi.org/10.1007/s00436-017-5471-5.

NPWS, 2024. National Parks and Wildlife Service, Conservation Objectives Series. https://www.npws.ie/protected-sites/sac/002999. (Accessed 23 January 2024).

OBIS, 2022. Ocean biogeographic information system). In: Global Map of Species Distribution Using Gridded Data. Ocean Biogeographic Information System. https://iobis.org. (Accessed 6 February 2022).

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. Prog. Oceanogr. 2020 (184), 102338.

Picton, B., Costello, M.J., 1998. BioMar biotope viewer: a guide to marine habitats, fauna and flora of Britain and Ireland. In: Environmental Sciences Unit. Trinity College, Dublin. ISBN 0 9526 735 4.

Plets, R., Clements, A., Quinn, R., Strong, J., Breen, J., Edwards, H., 2012. Marine substratum map of the causeway coast, northern Ireland. J. Maps 8 (1), 1–13. https://doi.org/10.1080/17445647.2012.661957.

Price, D.M., et al., 2019. Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. Coral Reefs 38, 1007–1021.

Principe, S.C., Acosta, A.L., Andrade, J.E., Lotufo, T.M.C., 2021. Predicted shifts in the distributions of Atlantic reef-building corals in the face of climate change. Front. Mar. Sci. 8 https://doi.org/10.3389/fmars.2021.673086.

Qiao, H., Feng, X., Escobar, L.E., Peterson, A.T., Soberón, J., Zhu, G., Papeş, M., 2019. An evaluation of transferability of ecological niche models. Ecography 42, 521–534.

Raykov, Y.P., Boukouvalas, A., Baig, F., Little, M.A., 2016. What to do when K-means clustering fails: a simple yet principled alternative algorithm. PLoS One 11 (9), 1–28. https://doi.org/10.1371/journal.pone.0162259.

Reiss, H., Kröncke, I., Ehrich, S., 2006. Estimating the catching efficiency of a 2-m beam trawl for sampling epifauna by removal experiments. ICES (Int. Counc. Explor. Sea) J. Mar. Sci. 63 (8), 1453–1464. https://doi.org/10.1016/j.icesjms.2006.06.001.

Ruiz, A., 2008. Hyperoplus lanceolatus Greater sandeel. In: Tyler-Walters, H., Hiscock, K. (Eds.), Marine Life Information Network: Biology and Sensitivity Key Information Reviews, Plymouth: Marine Biological Association of the United Kingdom. https://www.marlin.ac.uk/species/detail/2176.

Runya, R.M., McGonigle, C., Quinn, R., Howe, J., Collier, J., Fox, C., Dooley, J., O'loughlin, R., Calvert, J., Scott, L., Abernethy, C., Evans, W., 2021. Examining the links between multifrequency multibeam backscatter data and sediment grain size. Rem. Sens. 13 (8) https://doi.org/10.3390/rs13081539.

Saeid, Soheily-Khah, Ahlame, Douzal-Chouakria, Eric, Gaussier, 2016. Generalized k-means-based clustering for temporal data under weighted and kernel time warp. In: Journal of Pattern Recognition Letters.

Sala, E., Mayorga, J., Bradley, D., Cabral, R.B., Atwood, T.B., Auber, A., Cheung, W., Costello, C., Ferretti, F., Friedlander, A.M., Gaines, S.D., Garilao, C., Goodell, W., Halpern, B.S., Hinson, A., Kaschner, K., Kesner-Reyes, K., Leprieur, F., McGowan, J., Lubchenco, J., 2021. Protecting the global ocean for biodiversity, food and climate. Nature 592 (7856), E25. https://doi.org/10.1038/s41586-021-03496-1 (Nature, (2021), 592, 7854, (397-402), 10.1038/s41586-021-03371-z).

Santos, M.B., Pierce, G.J., Learmonth, J.A., Reid, R.J., et al., 2004. Variability in the diet of harbour porpoise (Pho- coena phocoena) in Scottish Waters 1992–2003. Mar. Mamm. Sci. 20, 1–27.

Seabed2030, 2024. https://seabed2030.org/2024/06/21/seabed-2030-announces-latest-progress-on-world-hydrography-day/. (Accessed 25 July 2024).

Sibarani, M.C., Di Marco, M., Rondinini, C., Kark, S., 2019. Measuring the surrogacy potential of charismatic megafauna species across taxonomic, phylogenetic and functional diversity on a megadiverse island. J. Appl. Ecol. 56 (5), 1220–1231. https://doi.org/10.1111/1365-2664.13360.

Simpson, J.H., Edelsten, D.J., Edwards, A., Morris, N.C.G., Tett, P.B., 1979. The Islay Front: physical structure and phytoplankton distribution. Estuar. Coast Mar. Sci. 9, 713–726.

Spalding, D.M., Fox, E.H., Allen, R.G., Davidson, N., Ferdaña, A.Z., Finlayson, M., Halpern, S.B., Jorge, A., Lombana, A., Lourie, A.S., Martin, D.K., McManus, E., Molnar, J., Recchia, A.C., Robertson, J., 2007. Marine ecoregions of the world: a bio-regionalisation of coastal and shelf areas. Bioscience 57, 573–583. https://doi.org/10.1641/B570707 (2007).

Summers, G., Lim, A., Wheeler, A.J., 2023. Multi resolution appraisal of cork harbour estuary: an object based image analysis approach. Geomorphology 439, 108851. https://doi.org/10.1016/j.geomorph.2023.108851.

Swanborn, D.J.B., Huvenne, V.A.I., Pittman, S.J., Woodall, L.C., 2022. Bringing seascape ecology to the deep seabed: a review and framework for its application. Limnol. Oceanogr. 67 (1), 66–88. https://doi.org/10.1002/lno.11976.

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. J.

Tong, R., Yesson, C., Yu, J., Luo, Y., Zhang, L., 2023. Key factors for species distribution modeling in benthic marine environments. Front. Mar. Sci. 10, 1222382 https://doi.org/10.3389/fmars.2023.1222382.

Trzcinska, K., Tegowski, J., Pocwiardowski, P., Janowski, L., Zdroik, J., Kruss, A., Rucinska, M., Lubniewski, Z., Schneider Von Deimling, J., 2021. Measurement of seafloor acoustic backscatter angular dependence at 150 kHz using a multibeam echosounder. Rem. Sens. 13 (23), 4771. https://doi.org/10.3390/rs13234771.

Walbridge, S., Slocum, N., Pobuda, M., Wright, D., 2018. Unified geomorphological analysis workflows with benthic terrain modeler. Geosciences 8 (3), 94. https://doi.org/10.3390/geosciences8030094.

Wanless, S., Harris, M.P., Greenstreet, S.P.R., 1998. Summer sandeel consumption by seabirds breeding in the Firth of Forth, south-east Scotland. ICES J. Mar. Sci. 55, 1141–1151.

Wilson, L.J., Hammond, P.S., 2019. The diet of harbour and grey seals around Britain: examining the role of prey as a potential cause of harbour seal declines. Aquat. Conserv. 29, 71–85.

Wölfl, A.C., Snaith, H., Amirebrahimi, S., Devey, C.W., Dorschel, B., Ferrini, V., Huvenne, V.A.I., Jakobsson, M., Jencks, J., Johnston, G., Lamarche, G., Mayer, L., Millar, D., Pedersen, T.H., Picard, K., Reitz, A., Schmitt, T., Visbeck, M., Weatherall, P., Wigley, R., 2019. Seafloor mapping - the challenge of a truly global ocean bathymetry. Front. Mar. Sci. 6 (JUN), 1–16. https://doi.org/10.3389/fmars.2019.00283.

Wright, P.J., 1996. Is there a conflict between sandeel fisheries and seabirds? A case study at Shetland. In: Green- stree, S.P.R., Tasker, M.L. (Eds.), Aquatic Predators and Their Prey. Fishing News Books, Blackwell Science, Oxford, pp. 154–165.

Wright, P.J., Jensen, H., Tuck, I.D., 2000. The influence of sediment type on the distribution of the lesser sandeel, *Ammodytes marinus*. J. Sea Res. 44, 243–256.

Zuur, A.F., Leno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. 1 (1), 3–14.