

Article

How Do Continuous High-Resolution Models of Patchy Seabed Habitats Enhance Classification Schemes?

Gustav Kågesten ^{1,*}, Dario Fiorentino ^{2,3}, Finn Baumgartner ¹ and Lovisa Zillén ¹

¹ Geological Survey of Sweden, 752 36 Uppsala, Sweden; finn.baumgartner@sgu.se (F.B.); lovisa.zillen@sgu.se (L.Z.)

² Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, 27570 Bremerhaven, Germany; dario.fiorentino@awi.de

³ Helmholtz Institute for Functional Marine Biodiversity at the University Oldenburg (HIFMB), 26129 Oldenburg, Germany

* Correspondence: gustav.kagesten@sgu.se

Received: 23 March 2019; Accepted: 21 May 2019; Published: 23 May 2019



Abstract: Predefined classification schemes and fixed geographic scales are often used to simplify and cost-effectively map the spatial complexity of nature. These simplifications can however limit the usefulness of the mapping effort for users who need information across a different range of thematic and spatial resolutions. We demonstrate how substrate and biological information from point samples and photos, combined with continuous multibeam data, can be modeled to predictively map percentage cover conforming with multiple existing classification schemes (i.e., HELCOM HUB; Natura 2000), while also providing high-resolution (5 m) maps of individual substrate and biological components across a 1344 km² offshore bank in the Baltic Sea. Data for substrate and epibenthic organisms were obtained from high-resolution photo mosaics, sediment grab samples, legacy data and expert annotations. Environmental variables included pixel and object based metrics at multiple scales (0.5 m–2 km), which improved the accuracy of models. We found that using Boosted Regression Trees (BRTs) to predict continuous models of substrate and biological components provided additional detail for each component without losing accuracy in the classified maps, compared with a thematic model. Results demonstrate the sensitivity of habitat maps to the effects of spatial and thematic resolution and the importance of high-resolution maps to management applications.

Keywords: habitat mapping; HELCOM HUB; Natura 2000; Baltic Sea; spatial scale; seascape; machine learning; boosted regression trees; percent cover; substrate; biota; multibeam; backscatter

1. Introduction

Classification schemes are common tools to support spatial decision-making. In marine spatial planning and conservation, they help to inform and prioritize management actions including design of monitoring programs, risk assessments, placement of marine protected areas, etc. [1,2]. Furthermore, classification schemes can provide advances in transnational efforts towards common strategies for nature conservation and protection (see Strong et al. [3] for a review and comparison of several classification schemes). Building a classification scheme includes important decisions on spatial and thematic resolutions and the way to represent structures [4]. As exemplified in this paper, it is critical for mapmakers to understand and communicate the implications of these decisions in order to inform management actions and evaluations.

At the European level, the European Union Nature Information System (EUNIS) [5], the Baltic Marine Environment Protection Commission Underwater Biotope and Habitat Classification System [6]

(HELCOM HUB, hereafter referred to as HUB) and Natura 2000 [7] are common classification schemes used to report against the European Marine Strategy Framework Directive (MSFD) [8] and the Habitat Directive [9]. The intended use of these schemes is to classify marine areas through several steps. Classification is first done according to the environmental setting (e.g., vertical zone), then by the geological features (soft/hard substratum), and finally by the biological components (e.g., communities, dominant species).

Similarly, in Sweden, classification schemes that combine both geological and biological information are required to support ongoing marine spatial planning initiatives (national and regional), design of marine reserves and other ecosystem-based fisheries management actions. In particular, the Geological Survey of Sweden (SGU) and the Swedish Agency for Marine and Water Management (SwAM) cooperate to provide detailed HUB and Natura 2000 maps in priority areas in the Baltic Sea. Ideally, these products should be backward compatible with SGU legacy maps, and include percent coverage of substrate and benthic species used in the national marine spatial planning cumulative impact assessment tool *Symphony* [10]. The methods presented in this paper are grounded in the need to develop such multiuse products.

The HUB classification (Table 1), a relatively new classification scheme with few existing maps [11], has been developed to be compatible with EUNIS and is specifically designed to create a common understanding of the Baltic Sea habitats, biotopes and communities [6]. HUB is a hierarchical classification scheme that consists of six levels with biotopes defined by applying split rules similar to a dichotomous key. The first three levels describe habitats, whilst levels 4–6 describe biotopes. HUB codes contain a maximum of six characters with each character referring to a specific level where a split rule has been applied. Biological communities are characterized by HUB Level 5, which is most comparable to EUNIS Level 4. The spatial scale for HUB biotopes is not strictly defined. The guidance given in the technical report is that a biotope patch should be large enough to function as a biotope [6], implying that the community is a distinct unit with some distinct functions. For instance, the EUSeaMap project has used a practical application of a 5×5 m scale as a minimum patch size [12].

While the efforts at European level are a first attempt towards establishing a network of transnational nature conservation and protection areas, providing a platform to better cope with the challenges posed by climate change [13], there are a number of limitations of such broad scale classifications. For example, as noted by Galparsoro et al., [1] the EUNIS classification needs to include new habitat classes and develop the classifications of less well-represented areas (e.g., deep-sea areas). Regardless of how refined such classification schemes become, they will remain a simplification of nature, for example by neglecting transitional areas, reducing local complexities (and neglecting some important feature at a local scale) and creating artefacts [14].

This paper focuses on the advantages of using high-resolution continuous models of substrate and biota based on multibeam data, sediment samples and photos, to overcome some of the shortcomings in existing classification schemes. A recent study in the Baltic have looked at predicting percent coverage of substrate and biota using similar data [15], and found the method to be useful, but does not use the prediction to also provide thematic habitat maps. We present a novel approach to prepare and predict high resolution habitat maps comparing both a thematic model workflow with predefined HUB and Natura 2000 classes, and a continuous model workflow that allows classified maps to be created for multiple scales and definitions, as well as providing percent coverage maps of substrate and biological components for more detailed analysis. In addition, we estimate the importance of different predictor variables; for instance, the performance of the models using backscatter data versus a variety of metrics (pixel and object based) from high-resolution terrain models. Such information is quite relevant in Sweden where backscatter is not yet provided as a standard for multibeam surveys. We also investigate if the substrate models can be improved by supplementing the training data with legacy observations and expert annotations, and if the resulting substrate maps can help improve the biological models. Finally, we look closer at some of the end-user implications of using thematic and

continuous maps at different spatial scales. Throughout the paper, we refer to *continuous models* as the result of the percent cover of each thematic class (for example percent cover of the substrate class sand).

Table 1. The Baltic Marine Environment Protection Commission Underwater Biotope and Habitat Classification System [6] (HELCOM HUB, hereafter referred to as HUB) split rule summary. Examples of HUB level 1–6 biotopes are found in the in the Supplementary Materials (Legend S1, Table S3).

| Level | Description |
|-------------------------------------|--|
| 1 Baltic (Letter) | Baltic |
| 2 Vertical Zone (Letter) | Photic Aphotic |
| 3 Substrate (Letter) | Coverage of a specified substrate type $\geq 90\%$ Coverage $< 90\%$, Mixed |
| 4 Community Structure (Number) | Coverage of Macroscopic vegetation or sessile macroscopic epifauna $\geq 10\%$ Coverage $> 0\% < 10\%$, Sparse Coverage = 0%, No vegetation or macro fauna present |
| 5 Characteristic Community (Letter) | Coverage of a specified taxonomic group $\geq 10\%$ Coverage $\geq 10\%$ but not of a specified taxonomic group, Mixed community Coverage = 0%, No macroscopic community |
| 6 Dominating Taxa (Number) | Biomass/bio volume of some specified taxa $\geq 50\%$ |

2. Materials and Methods

The analytical workflow follows a logical sequence of steps: (1) data acquisition and processing; (2) processing of environmental predictors; (3) development of thematic and continuous models; (4) combining models into classified maps; and (5) map accuracy assessments.

2.1. Acquisition and Processing

2.1.1. Study Area

The Baltic Sea offshore area Hoburgs Bank (Figure 1) was selected as a pilot project for high resolution mapping in Sweden. The survey covered a 1344 km² area of the bank (10–60 m depth) with the seafloor containing a complex mix of substrates: mainly sand, gravel, hard clay, rock and boulder, shaped by glacial processes, waves and currents. Hoburgs Bank is part of a newly expanded Natura 2000 marine protected area, motivated by the population of seabirds and recent findings on the threatened Baltic Sea harbor porpoise [16]. The most apparent human-caused threat to the area is a major shipping lane that crosses the bank's northwestern part, and a second shipping lane that runs southeast of the bank.

2.1.2. Hydrographic Survey

In autumn 2016 and spring/summer 2017, we conducted a full-coverage hydrographic survey of the Hoburgs Bank study area using SGU's vessel S/V Ocean Surveyor that was equipped with a multibeam EM2040D (Kongsberg Maritime AS, Kongsberg) and a sub-bottom profiler (SBP) (Echoes 3500 T3, iXBlue, Natick). Hydrographic data collection was generally conducted from late afternoon to early morning, with sampling operations occurring during the daytime. We used a moving vessel profiler to collect sound speed data, and daily CTD casts to update the absorption coefficient for multibeam backscatter data in the acquisition software Seafloor Information System (Kongsberg Maritime AS, Kongsberg). Bathymetric data from the EM2040D multibeam were processed in CARIS HIPS and SIPS (v 10.3.1, Teledyne CARIS, Fredericton) using the Combined Uncertainty and Bathymetric Estimator (CUBE) surface generation method [17]. Seafloor backscatter data were processed by implementing the geocoder engine [18] in the Fledermaus Geocoder Toolbox (FMGT) (v7.7.6, QPS, Zeist). Both depth and backscatter were initially exported at 2-m resolution for use in on-board sampling design. Final post processing depth grids and backscatter mosaics were exported at 0.5-m resolution (Figure 2), together with depth uncertainty, survey track lines, as well as all angular range analysis (ARA) backscatter statistics available in FMGT. More details on the acquisition are available in SGU's survey report [19].

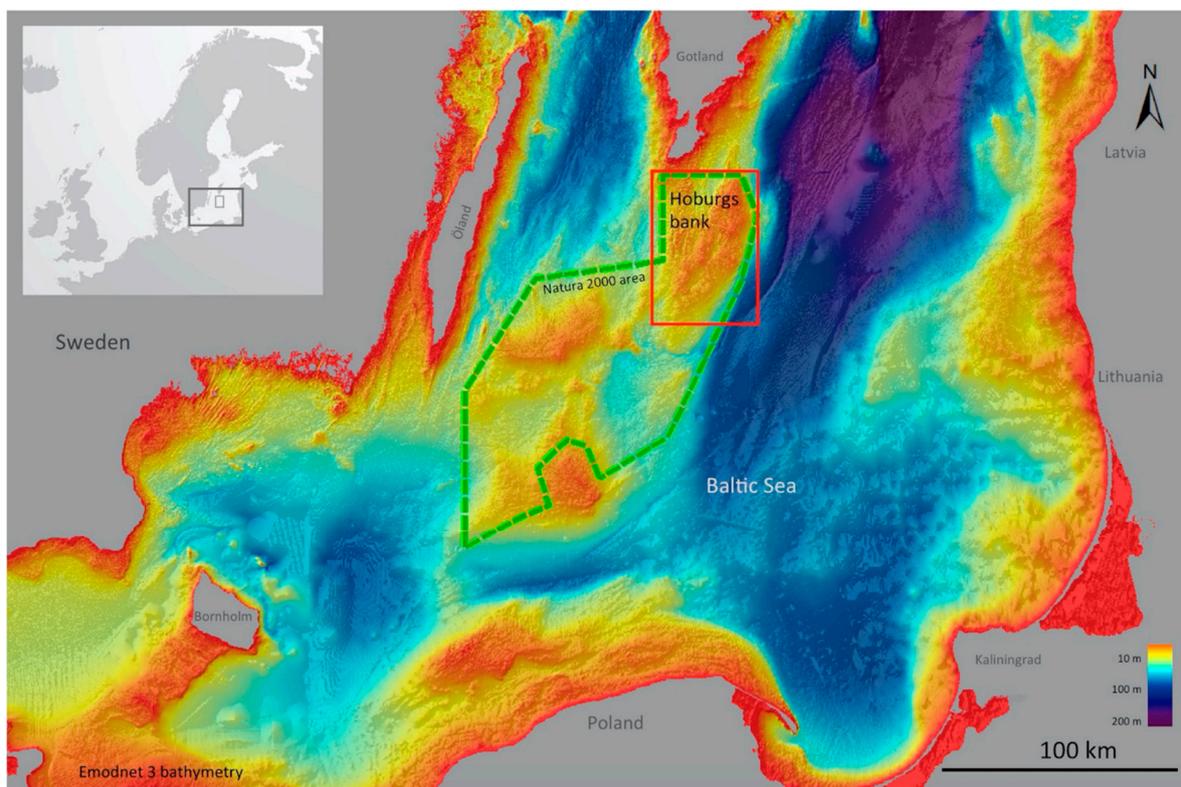


Figure 1. Map of the Hoburgs Bank study area.

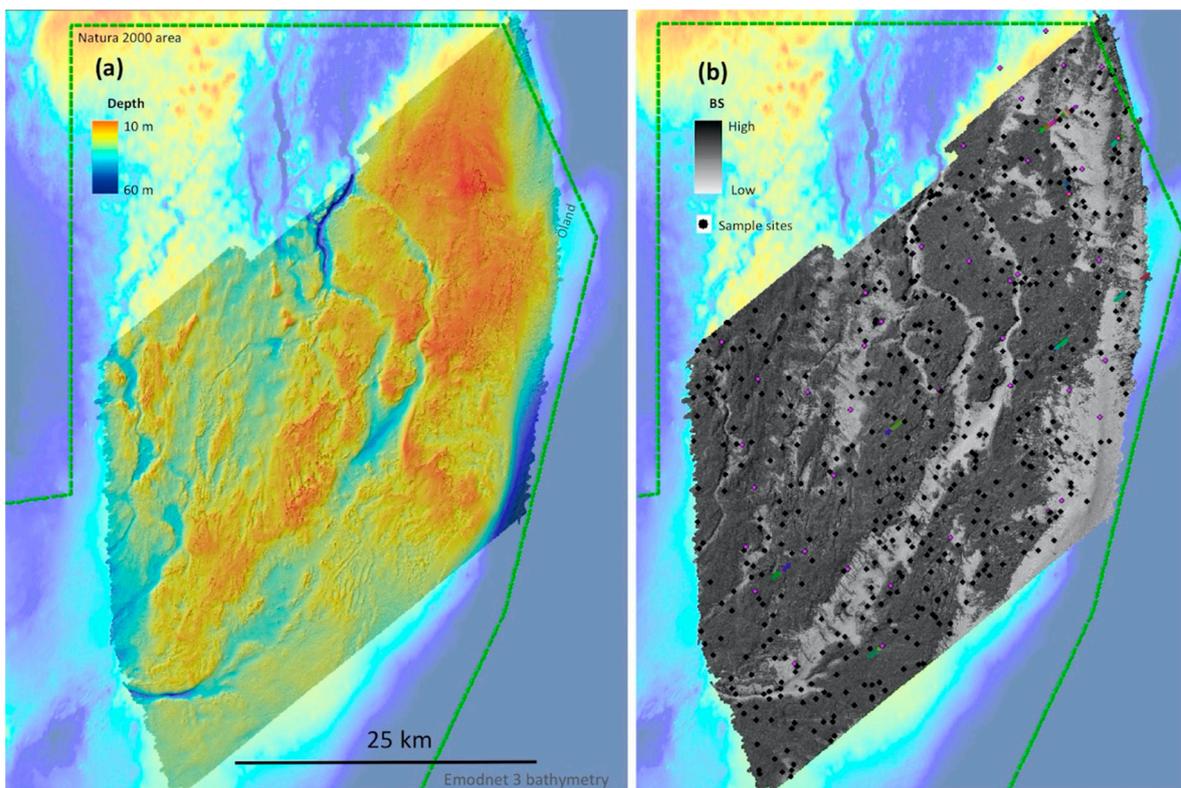


Figure 2. Hoburgs Bank study area: (a) multibeam depth grid, (b) multibeam backscatter mosaic and samples sites (black dots 2016–2017 survey; colored dots legacy data from 2005).

2.1.3. Sampling and Underwater Observations

Stratified random sampling was used to select the locations for sediment grab samples and underwater observations (using a drop-camera) via the ArcGIS 10 Sampling Design Tool [20], based on (1) depth intervals, (2) substrate types (using backscatter intervals) and (3) complexity intervals (a combination of depth standard deviation and bathymetric position index). The classified depth, substrate and complexity raster were combined into polygons, and slivers were removed. The sampling design tool was used to randomly choose at least one of each thematically unique polygon found within a $\sim 3 \times 12$ km area (approximately what we could sample and survey during 24 h in optimal conditions), with the samples placed randomly within each polygon. The process was repeated for a total of 34 blocks that divided the survey area (see Supplementary materials Map S1). We visited 559 locations for the drop-camera, which also included sediment samples at 434 of the locations (some boulder sites were excluded from sediment sampling). Distribution of the drop-camera samples divided into HUB 3 classes and HUB 4–6 classes are available in Appendix A Tables A4 and A5.

Positioning was collected via a real-time kinematic (RTK) gps (Seapath 330 GNSS-RTK, Kongsberg Maritime AS, Kongsberg) with corrections from SWEPOS base stations, and adjusted to the cable breakpoint on the moonpool (grab-sampler) and A-frame winches (drop-camera), located in the middle and rear of the vessel respectively, about 15-m apart. A dynamic positioning system retained the vessel's position over the site.

While the position uncertainty of the grab samples and drop-camera varied with depth and currents, it was generally within ± 2 m when distinct features such as clay reefs spurs were visited. Fine sediment (from silt to gravel) was sampled with a Van Veen grab (0.1 m^2), whereas coarser sediment and hard substrates were sampled with an Orange Peel Bucket ($\sim 0.5 \text{ m}^2$). Onboard geologists interpreted all samples (origin, composition), and submitted a representative subset ($n = 117$) of fine sediment types (silt to gravel) from surface sediments ($\sim 0\text{--}5$ cm) to sieve grain size analysis.

Table 2. Grain size ranges for the surface substrate compositions, and their sampling methods.

| Size Range (mm) | Hard/Soft | Substrate | Grab-Sampler | Drop-Camera |
|-----------------|-------------|----------------------|----------------|------------------------------|
| | | Bedrock ¹ | | Point intercept |
| >600 | Hard bottom | Large boulders | | Point intercept |
| 200–600 | | Boulders | | Point intercept |
| 60–200 | | Large stones | | Point intercept |
| 20–60 | | Pebbles & stones | | Point intercept |
| 2–20 | | Gravel | | Point intercept |
| 0.6–2 | Soft bottom | Coarse sand | Sieve analysis | |
| 0.2–0.6 | | Medium sand | Sieve analysis | Point intercept ² |
| 0.006–0.2 | | Fine sand | Sieve analysis | |
| 0.002–0.06 | | Silt | Sieve analysis | Point intercept ² |
| <0.002 | | Soft clay | Sieve analysis | Point intercept ² |
| <0.002 | Hard bottom | Hard clay | | Point intercept |

¹ not found in the study area. ² clay, silt and sand fractions were classified according to the dominant grain size (always sand in the study area).

For the underwater observations, we used a custom-built drop-camera system with a 360° rotating and 90° tilting full frame Canon EOS 6D DSLR camera, ~ 75 cm above the seafloor. In order to capture images in all directions, we applied an automated photo and light protocol through which we were able to capture 100% sub-millimeter resolution coverage of $\sim 15 \text{ m}^2$ of the seafloor (see Figure 3 for example) and additional images of the surrounding seascape. We also used a GoPro Hero 4 camera to aid the photo interpretation with video, and two parallel red lasers placed 30 cm apart to provide a scale reference.

We analyzed the images using a point intercept method in order to obtain 100 points per site of substrate and biological coverage, as well as presence absence data. Presence of ripples, shells, crawl tracks, and fish were also noted. We relied on sediment samples, video imagery, as well as

high-resolution multibeam data to better annotate images when substrate components were covered with benthic organisms. The point intercept annotations could not differentiate between finer grain-sizes than the difference between gravel and sand; hence the sand class from the photomosaic also included finer sediment fractions when present (Table 2). More detailed information on fine grain sizes were obtained from the sieved surface sediment samples, which was a separate dataset (Table 3). While we did not conduct a statistical uncertainty estimate of the image analysis process, a second expert double-checked all interpreted sites in order to minimize human error.

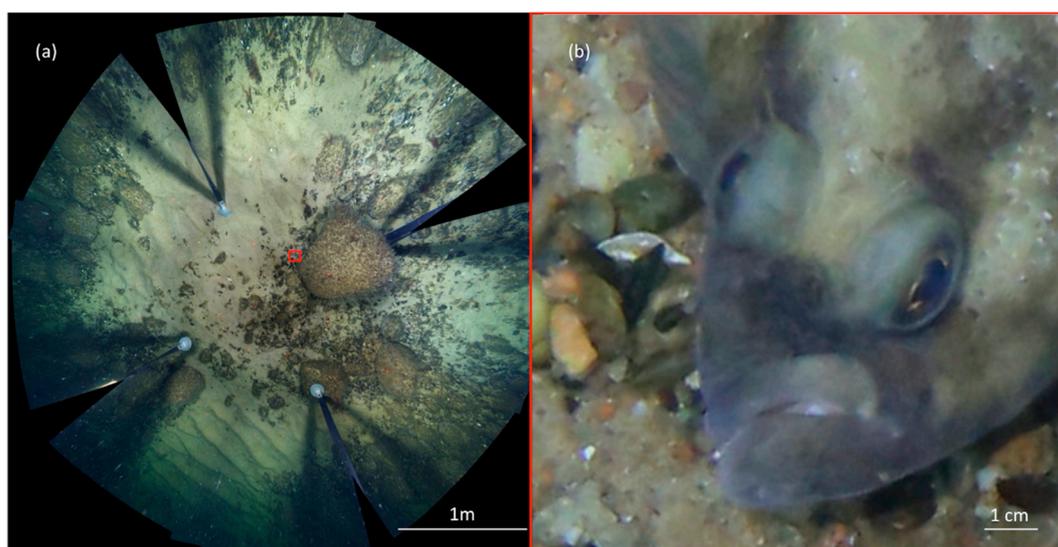


Figure 3. Example of a typical mixed hard and soft seafloor habitat on Hoburgs Bank. (a) Full drop-camera photomosaic covering 15m² seafloor. (b) Detail zoom showing a flatfish, mussels and substrate.

2.1.4. Legacy Data and Expert Annotations

In order to further improve the training dataset for the substrate models we used lower quality legacy data and expert annotations. Data collected during a survey in 2005 by SGU using drop-camera, samples and video transects [21] were processed and re-annotated with substrate information (449 sites). ArcGIS (ArcGIS Desktop v10.5.1, ArcGIS Pro v2.1.2) was used to display and interpret the samples/photos and video transects together with high-resolution backscatter and multibeam data in order to verify, and if needed modify, the location of some sites. For areas with a small number of samples, under-represented habitats and obvious artifacts in early model outputs, we also conducted additional expert annotations (550 sites) of substrate components based on the high-resolution multibeam data and available ground truthing. Distributions of the legacy and expert annotations divided into HUB 3 classes are available in Appendix A Table A4.

2.2. Environmental Setting

Multibeam bathymetry and backscatter were the backbone of the environmental variables that we developed (hereafter referred to as *predictors*), and we used both pixel metrics applied across the seascape at multiple spatial scales and OBIA to make the most of this information. We also included *in situ* oceanographic data collected during the survey (salinity, temperature, and current speed and direction near the seafloor), as well as interpreted SBP data (postglacial sand deposits). In addition, we included geographical coordinates (northing/easting), which have been found to account for autocorrelation in the model [22]. All predictors were resampled to the modeling resolution of 5 m, which roughly corresponded with the scale of the underwater observations ~15 m² while also providing a practical minimum mapping unit for the end users (See Supplementary materials Tables S1 and S2 for a complete list of the predictor variables).

Bathymetric and backscatter information were recalculated to include multiple scale resolutions relevant for the model building (0.5 m, 1 m, 2.5 m, 5 m, 20 m, 50 m, 100 m, 200 m, 500 m, 1 km and 2 km), hereafter referred to as *multiscale metrics*. To represent scales smaller than 5 m, we first developed terrain metrics or *morphometrics* [23] in ArcGIS (standard deviation, profile, planform, and standard curvature, slope, slope of the slope (a measure of the magnitude of slope change), terrain-surface roughness and surface area to planar area) from the high resolution grids (0.5 m, 1 m, 2.5 m) then aggregated the information to 5 m using minimum, maximum, range and mean values. The same metrics were also run using 5 m bathymetry and backscatter (with the addition of median value and bathymetric position index, BPI). To represent scales larger than 5 m, we generated a set of smoothed depth grids at 5-m resolution using a Gaussian low-pass filter to remove details at different neighborhoods (i.e., 20 m to 2 km), followed by calculations of selected terrain metrics (i.e., slope, slope of slope, BPI and slope direction). In order to adjust for potential angle artifacts, uncertainty data retrieved from multibeam bathymetry as well as the Euclidean distance to the survey routes were included as predictors.

In addition to the pixel based metrics, we used object based image analysis (OBIA) in ENVI Feature Extraction Module (v 5.4, Harris Geospatial Solutions, Broomfield) to generate segments with statistics from depth and backscatter data, a process similar to previous semi-automated coral reef habitat mapping work [24,25]. Segments were developed from a composite image (principal component compressed depth metrics in 5 bands and backscatter in one) at 2.5-m resolution. ENVI segmentation thresholds were set to segmentation 11 and merge 91, using an edge detection algorithm. The resulting segment statistics (shape, textural and spectral attributes, [26] were exported as raster images and aggregated to the modeling resolution of 5 m.

We also applied another alternative way for treating habitat features as objects by calculating the Euclidean distance to features (often consisting of moraine ridges) seen in fine scale BPI (inner radius 5 m, outer radius 25 m, $BPI \geq 2$), as well as distance to fine-medium sand patches (using a threshold reclassified backscatter mosaic).

In order to reduce collinearity and the amount of predictors, we grouped similar predictors and used principal component analysis (PCA) to extract the components that contributed to >5% of the total variance within each group (see supplementary Script S1). Selected predictor variables, such as depth, were excluded from the PCA process to better understand their importance in the models. Furthermore, expert evaluation in the model fitting stage helped in reducing the total number of predictors from 184 to 41.

2.3. Thematic and Continuous Models

One of the goals with the analysis was to compare thematic models (i.e., HUB 3–6, Natura 2000) and thematic maps classified from continuous models (hereafter referred to as reclassified maps), in this case we always used the same testing and training data, tuning grid and predictor variables.

For the general data preparation, we split the drop-camera data from 2016–2017 into 30% testing data ($n = 154$) and 70% training data ($n = 405$), using stratified random sampling based on HUB level 1–6 classes (reclassified from percent cover of substrate and biota). The same testing data was used for all models (substrate and biota), except models using sediment grab samples which used all data for training due to the low number of sieve-analyzed samples ($n = 117$). See Table 3 for an overview of the data and the targets they were used for. One of the strengths of our approach was the possibility to integrate different data sources, such as results of grain size analysis from samples (which captured fine grain size fractions not visible in photos) and the percent coverage from images, (which captured the distribution of coarser grain sizes, and the total coverage of fine grain sizes).

Table 3. List of sources of training and testing data according to purpose of application. Tables A4 and A5 in Appendix A show the distribution of HUB classes within the data.

| Data Source | N | Method | Target |
|--|-----|--------------------|---|
| Drop camera samples (2016–2017) ¹ | 559 | Point intercept | Percent coverage of benthic organisms. |
| Drop camera samples (2016–2017) ¹ | 559 | Point intercept | Percent coverage of substrate fractions (i.e., sand, gravel, pebbles, stones, boulders, large boulders and hard clay) |
| Legacy data (drop camera, video transects and sediment samples from 2005) ² | 449 | Estimated coverage | Percent coverage of substrate fractions. |
| Expert annotation (depth, backscatter) ² | 550 | Estimated coverage | Percent coverage of substrate fractions. |
| Grab samples (2016–2017) ² | 117 | Sieve analysis | Percent coverage of fine substrate fractions (i.e., soft clay, silt, fine - coarse sand) |
| Grab samples (2016–2017) ² | 434 | Expert | Thematic substrate classes (i.e., silty gravelly sand) |

¹ Testing (30%) and training (70%), split using stratified random sampling based on HUB level 1–6 classes. ² Training data only.

For the thematic models, we first reclassified the testing and training data into thematic classes for HUB and Natura 2000, before the models were run for each level (HUB level 3, HUB level 4–6, and Natura 2000 reef/sandbanks). For the continuous models, we prepared the training data by assigning 0.1% coverage to presence with <1% cover and 0.001% coverage to absence, followed by logit transformation (all individual models). We also evaluated continuous compositional models. The modeling framework of the latter approach explicitly takes into account that substrate components must sum up to 100 % at each location [27].

Model algorithms were run with R v3.4.1 [28] (see supplementary Script S2–S5). For both thematic and continuous models, we used boosted regression trees (BRTs) implemented by the *gbm* [29] and *caret* package [30]. BRTs is a flexible nonparametric statistical machine learning algorithm suitable for regression and classification problems that can handle nonlinear interactions between predictors [31–33], and has been successfully applied in different contexts including coral reefs [23,34] and marine protected area design [35,36]. High performance in BRTs is reached by assembling a large number of simple trees [37], starting from a tree produced from a random subset of the data and sequentially adding further trees to find the best model fit [31–33], as measured by loss in predictive performance (e.g., deviance) [29]. The final prediction is calculated as a weighted average across all trees [29]. We used the number of trees, interaction depth and shrinkage as parameters to tune the models [33], and relied on root mean squared error (continuous models) and overall accuracy (thematic models) to choose the best combination of parameters that reduce overfit as suggested by Elith et al. [29]. In fact, we relied on the recommendation by Hastie et al. (2009) that setting the shrinkage parameter in BRT “tends to eliminate the problem of overfitting, especially for larger data sets”. Final parameters used to tune the models are available in the Supplementary materials (Script S3–S5). Although overfitting has been recognized as an issue in machine learning techniques that reduce model’s generality [29,38], Zaki and Meira [39] stated that some overfit is actually helpful, especially for those classes that might go underrepresented. Furthermore, we followed Schiele et al. [11] and ran models separately to keep the number of classes at each run relatively low.

We applied BRTs (Gaussian distribution) to predict percent coverage of finer substrates (i.e., soft clay, silt, fine to coarse sand from sieved samples, see Table 2 for size ranges), percent coverage of coarser substrates (i.e., sand (which also included subfractions of silt and clay not visible in the imagery), gravel, pebbles, stones, boulders, large boulders and bedrock, see Table 2 for size ranges) and percent coverage of benthic organisms according to HUB level 5 and 6 groups (Table 4). We also applied BRTs (multinomial distribution) to predict thematic maps of HUB 3–6 and Natura 2000 classes directly from reclassified training data.

We assessed the outputs of the substrate models to spot obvious artifacts (e.g., due to survey routes), followed by expert annotation, which helped to improve the training dataset. The latter step was repeated until the draft maps were approved. The substrate maps were then used as part of the predictor variables for the biological models, as well as the fine sediment fraction models (i.e., in addition to the 41 predictor variables also used in the substrate models). Finally, we reran

several models with different predictor variables removed to better understand how different groups of the predictors affected map accuracy. BRTs, as an ensemble approach, use a high number of trees to turn possibly weak predictors into a strong prediction (Kuhn et al. 2013). Being quite robust to collinearity, BRTs handles many predictors as it ignores non-informative predictors when fitting trees [29].

Table 4. List of HUB level 5 and 6 groups in the study area, and the species included.

| HUB Level 5 and 6 | Species included in Group |
|--|--|
| Characterized by annual algae | <i>Ceramium tenuicorne</i> , <i>Ceramium</i> sp., <i>Chorda filum</i> , <i>Dictyosiphon/Stictyosiphon</i> (complex), <i>Ectocarpus/Pylaiella</i> (complex), Filamentous Phaeophyceae, <i>Haliosiphon tomentosus</i> |
| Characterized by perennial algae | <i>Battersia arctica</i> , <i>Coccolytus/Phyllophora</i> (complex), <i>Delesseria sanguinea</i> , Filamentous Rhodophyceae, <i>Furcellaria lumbricalis</i> , <i>Polysiphonia/Rhodomela</i> (complex), Unidentified Rhodophyceae. |
| Dominated by perennial filamentous algae | <i>Battersia arctica</i> , Filamentous Rhodophyceae, <i>Polysiphonia/Rhodomela</i> (complex). |
| Characterized by epibenthic bivalves | <i>Mytilus</i> spp. |
| Dominated by Mytilidae | <i>Mytilus</i> spp. |
| Characterized by cnidarians | Hydrozoa (<i>Cordylophora caspia</i>) |
| Characterized by moss animals | <i>Electra</i> sp. |

2.4. Combining Models into Classified Maps

To combine the models into classes, we adjusted all substrate and biological components to ensure that the model outputs corresponded with the rules of the point intercept analysis. The substrate components had to sum to 100% in each pixel, which was realized by dividing each component by the total sum of all components. Similarly, the fine fraction components modeled from grab samples (soft clay to coarse sand) were adjusted to fit the sand model, realized by multiplying the proportion of each fine fraction component with the sand component. Biological components were adjusted in a similar way; uncolonized substrate and biological cover had to sum to 100% or more, and no single biological component when summed with uncolonized substrate could be >100%.

Next, we reclassified the outputs of the percent coverage models using conditional statements according to HUB (e.g., as seen in Table 1) [6], Natura 2000, and the SGU substrate scheme [40] (R script available, see supplementary S6–S8). For cases with unclear definitions as percent coverage or terrain metrics (i.e., Natura 2000 subtypes, SGU substrate scheme), we relied on text descriptions to define a conditional statement. Since scale was not clearly defined for any of the classification schemes, we used the full resolution of the models (5 m), but also created classified maps in other resolutions up to 250 m pixels by aggregating the percent coverage models before the conditional statements were run.

2.5. Accuracy Assessment

We validated all maps using a stratified random (HUB 1–6) subset of data retrieved from the high-resolution drop-camera (154 testing locations extracted from the original 559 sites). The substrate models were also evaluated with training data using bootstrap prior to the spatial prediction.

Thematic models were assessed with confusion matrices of observed *vs.* predicted [41] for each level of HUB and Natura 2000 maps. Furthermore, Kendall rank and correlation coefficients (tau) were calculated [42].

Continuous models were assessed using R^2 , root mean square error (RMSE), mean absolute error (MAE) and mean error (ME, observed—predicted), as well as regression of observed *vs.* predicted values [41], which included 95% confidence band between observed and predicted coverage. In addition, the predicted values were grouped into intervals (0–10%, 10–50%, 50–90%, 90–100%) and the observed *vs.* predicted for each group was assessed using ME and standard deviation (SD). Finally, the same groups, as well as absence/presence were evaluated as classes using the same confusion matrix as the thematic models (overall accuracy and Tau). The continuous models of fine fractions were only

evaluated using bootstrap prior to the spatial prediction due to the relatively low number of sieved grab samples.

Spatial patterns for model uncertainty was estimated combining all model outputs (generated using different predictor variables and training data in the development phase) for our different substrate models (10 different models for each component, which included both individual and compositional model runs) and all outputs for the biological models (5 different models for each component using different predictor variables) and produced mean standard deviation maps for substrate and biological models in ArcGIS.

3. Results

Below we describe the results from the study structured along two main themes: (1) technical section with modeled maps and accuracy (Sections 3.1–3.4); and (2) End-user examples (Section 3.5), including the effect of scale and predefined thresholds. In addition to the images shown in the result, a geospatial pdf showing maps and survey data in more detail are provided in the Supplementary materials (Map S1).

3.1. Thematic vs. Reclassified

When comparing thematic maps (i.e., thematic models) and the reclassified thematic maps (i.e., from the continuous workflow) our results indicated that the accuracy was similar across all levels of the classification schemes regardless of which BRT workflow we used (Table 5). This observation was further strengthened by studying the spatial output of the maps together with high-resolution backscatter and bathymetry data, taking visual aspects such as survey artifacts into account.

Table 5. Observed vs. predicted accuracy from 154 testing sites for thematic maps (thematic distribution BRT models) and reclassified thematic maps (from continuous BRT models).

| | Overall Accuracy (%) | | Tau | |
|--------------------|----------------------|--------------|----------|--------------|
| | Thematic | Reclassified | Thematic | Reclassified |
| HUB 3 | 77.9 | 80.5 | 0.72 | 0.76 |
| HUB 4 | 79.9 | 81.8 | 0.70 | 0.72 |
| HUB 4–5 | 63.0 | 62.3 | 0.55 | 0.55 |
| HUB 4–6 | 55.8 | 53.2 | 0.49 | 0.47 |
| Natura 2000 | 87.7 | 87.7 | 0.75 | 0.75 |
| Mean | 72.9 | 73.1 | 0.64 | 0.65 |

Map accuracy was highest for the maps with fewer classes (i.e., Natura 2000) with decreasing accuracy as the number of classes increased (Tables 5 and 6). The number of classes in the reclassified map products was generally higher than the original classes identified in the training data, while the thematic maps were restricted to the number of classes in the training data set. For the complete HUB map level 1–6 the number of classes doubled in the reclassified map from 29 to 59 classes, however 8 of these HUB 1–6 classes had fewer than 10 predicted pixels (Table 6), of which some could be the result of pixel mismatch between the combined models. Very few pixels had obviously illogical combination of classes (based on our observations and knowledge of the survey area), the only clear example was 2 pixels classified as sand dominated by *Mytilus* spp. (full list of classes and pixel count for HUB 1–6 available in Supplementary materials Table S3).

Table 6. Number of unique classes found in the reclassified thematic maps versus the thematic maps (same number of classes as in the training data). Number of classes with less than 10 predicted pixels are put in brackets (full list of classes and pixel count for HUB 1–6 available in Supplementary Materials Table S3).

| | HUB 3 | HUB 4 | HUB 4–5 | HUB 4–6 | HUB 1–6 | N2000 | N2000 Subtypes |
|-------------------|-------|-------|---------|---------|---------|-------|----------------|
| Thematic maps | 5 | 3 | 7 | 9 | 29 | 2 | - |
| Reclassified maps | 5 | 3 | 8 | 12 | 59 (8) | 2 | 12 |

3.2. Predictors and Training Data

We investigated how different types of predictors and training data contributed to the thematic HUB 3 substrate models (Table 7). Models trained using the same predictors, but only the point intercept dataset for training, showed a large decline in accuracy (testing data overall accuracy 77.9% vs. 58.9%, while bootstrap of training data showed overall accuracy 77% vs. 73%, indicating overfitting in the model with only survey data included). These results highlight that legacy data and expert annotations were important to the model fit. Expert annotations took approximately 5 min per site, legacy data approximately 10 min, while substrate and biota point intercept annotations of the drop camera data (2016–2017) averaged around one hour (annotations of 100 points/site ranged from 5 min to 2 h depending on site complexity).

The highest accuracy was observed for the model using all training data (including legacy and expert annotations) and all types of predictor variables (including backscatter, multiscale metrics 0.5 m–2 km and OBIA). When testing how OBIA, multiscale metrics and backscatter affected model accuracy (in addition to 5 m depth metrics which was always included), we observed that models that included only a backscatter mosaic had relatively high accuracy (73.4%), showing that backscatter is an important predictor and that relatively simple models with 5 m depth and backscatter metrics can perform well. The accuracy improved further (+4.5%) when OBIA and multiscale (0.5 m–2 km) metrics were also included, with OBIA having the highest percent contribution in the models (25.8%, Table S2). Performance reduced in all models when backscatter was not included, even if more advanced depth metrics were used. However, some of the loss in accuracy was regained by including multiscale (0.5 m–2 km) depth metrics (OBIA was not tested due to the inclusion of backscatter in the segmenting process). The percent contribution for all predictors used in the thematic models (HUB 3–6) is available in Appendix A Table A1 (summary) and Supplementary Materials Table S2.

Table 7. Comparison of HUB 3 classification model overall accuracy (OA), Cohen’s kappa and Tau, when using different predictor variables and training data. The accuracy was calculated from the testing data (30% of 2016–2017 drop camera samples), and additionally by bootstrapping the training data in the model fitting stage (may show inflated accuracy numbers in case of overfitting). Predictor variables (all groups contained northing/easting): (1) Depth metrics 0.5 m–2 km, OBIA metrics and backscatter metrics. (2) Depth metrics 0.5 m–2 km, backscatter. (3) Depth metrics 5 m, backscatter mosaic. (4) Depth metrics 0.5 m–2 km. (5) Depth metrics 0.5 m–5 m. (6) Depth metrics 5 m.

| Predictors | Training Data | Bootstrap (Training Data) | Validation (Testing Data) |
|------------|------------------------|---------------------------|---------------------------|
| 1) All | Survey, legacy, expert | OA 77%, kappa 0.68 | OA 77.9%, Tau 0.72 |
| 2) Reduced | Survey, legacy, expert | OA 76%, kappa 0.66 | OA 77.2%, Tau 0.71 |
| 3) Reduced | Survey, legacy, expert | OA 70%, kappa 0.58 | OA 73.4%, Tau 0.67 |
| 4) Reduced | Survey, legacy, expert | OA 68%, kappa 0.55 | OA 64.3%, Tau 0.55 |
| 5) Reduced | Survey, legacy, expert | OA 64% kappa 0.49 | OA 61.0%, Tau 0.51 |
| 6) Reduced | Survey, legacy, expert | OA 63% kappa 0.47 | OA 55.2%, Tau 0.44 |
| 1) All | Survey | OA 73% kappa 0.60 | OA 58.9%, Tau 0.49 |

Comparisons between biological models that included substrate models as predictors and those that did not, indicated that the inclusion of substrate models developed using a larger training data set (i.e., survey, legacy and expert annotations) could be used to increase the accuracy of biological models. For example, accuracy of *Mytilus* spp. models improved with the addition of substrate models, and the total mussel cover increased from 8% to 10.7% (Table 8). The total contribution of substrate predictors in the model was 75% (Appendix A Table A1, Supplementary materials Table S2), further highlighting the strong link between substrate and *Mytilus* spp. A visual comparison between the outputs showed that the greatest change in mussel cover occurred in flat areas composed of hard substrates, and areas with coarse soft substrates where it was difficult to differentiate between the two types of environments in the depth and backscatter data. When testing the same models but also removing backscatter the performance decreased further and the mean mussel cover changed to 4.4% (Table 8), showing the importance of backscatter as a predictor variable for the biological models. For the thematic models of biological HUB classes, the overall accuracy increased slightly for all levels when substrate models were included as predictors (HUB 4 +0.7%, HUB 4–5 +1.3%, HUB 4–6 +1.9 %).

Table 8. Accuracy table and mean percent cover for *Mytilus* spp. models with different predictor variables. Accuracy was measured (testing data, observed *vs.* predicted) with R^2 , root mean square error (RMSE), mean absolute error (MAE) and mean error (ME). In addition, the predicted values were grouped into intervals (0–10%, 10–50%, 50–90%, 90–100%) and evaluated with overall accuracy (OA) as classes and absence—presence. Predictors (1) all refer to both substrate model outputs (developed using all available training data and predictor variables) and survey metrics (Supplementary materials Table S2), (2) same as (1) but excluding substrate model outputs, (3) same as (2) excluding all backscatter derived predictors.

| Predictors | R^2 | RMSE | MAE | ME | OA Classes | OA Abs-pres. | Cover (Mean) |
|-------------------|-------|-------|-------|------|------------|--------------|--------------|
| 1) All | 0.62 | 15.53 | 9.10 | 2.30 | 75 % | 89 % | 10.7 % |
| 2) No substrate | 0.60 | 16.69 | 9.84 | 5.61 | 69 % | 88 % | 8.0 % |
| 3) No backscatter | 0.49 | 19.60 | 11.83 | 7.96 | 60 % | 75 % | 4.4 % |

3.3. Continuous Percent Coverage Models

3.3.1. Substrate Components

Initially, two different ways to predict substrate components using BRTs were explored i.e., individual and compositional models. The two modeling approaches showed similar overall accuracy, with some slight differences, e.g., small patches with a high cover of hard clay were under predicted in the compositional model, which was not the case for the individual model. Therefore, we present the simpler individual continuous modeling approach (hereafter referred to as continuous models). Furthermore, most individual models improved slightly when adjusting the models to sum to 100% (by dividing with the total sum). A comparison between individual and compositional substrate models can be found in Appendix A, Table A2 where we also combined the two approaches with some success.

The overall spatial patterns of the modeled substrate components (Figure 4) were as expected based on knowledge from the survey. For example, the model of hard clay successfully mapped clay features also seen in the multibeam data and the large boulder model captured pixels where distinct boulders could be identified in the multibeam data. Survey artifacts (mainly due to sound velocity errors in the outer multibeam swath, as well as nadir artifacts in the backscatter signal) were suppressed in the models, but remained a challenge (seen as striping in SW/NE direction, Figure 4i,j), especially in flat areas of mixed substrates where backscatter was observed to be less useful to differentiate between coarse sand-gravel and low relief hard bottom areas.

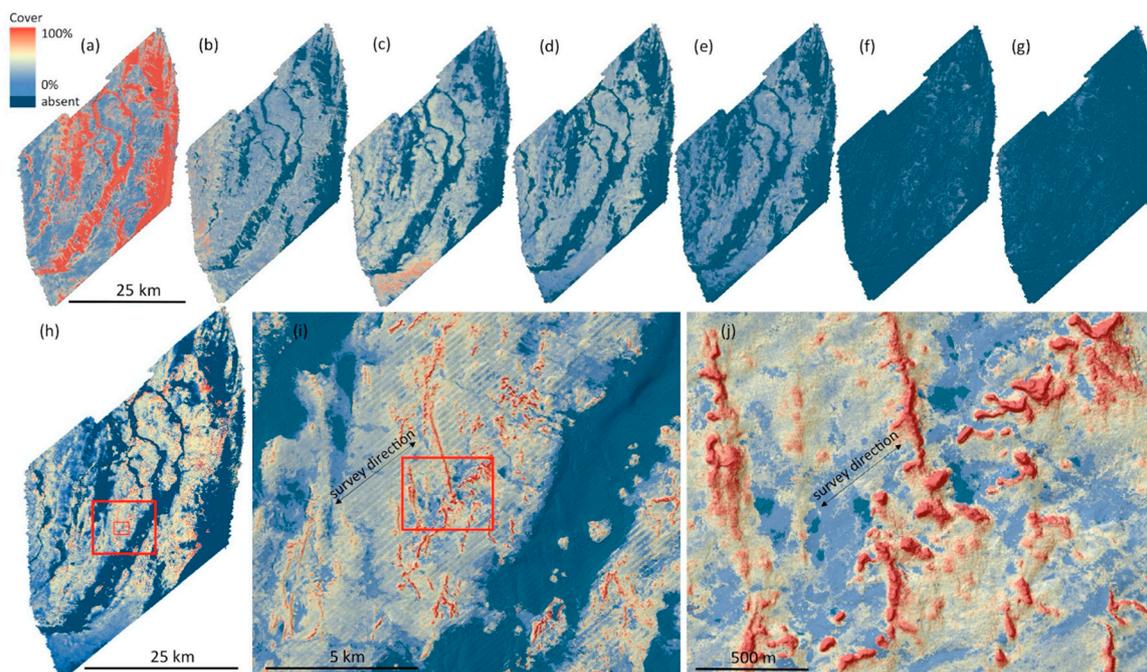


Figure 4. Coverage maps (%) of substrate components on Hoburgs Bank. (a) sand (0.006–0.2 cm), (b) gravel (0.2–2 cm), (c) pebbles (2–6 cm), (d) large stones (6–20 cm), (e) boulders (20–60 cm), (f) large boulders (>60 cm), (g) hard clay, (h) hard bottom, (i) hard bottom zoom 1 with hillshade, (j) hard bottom zoom 2 with hillshade. Absence is based on modeled threshold (<0.01% cover).

Regression analysis of observed and predicted substrate cover showed varying degrees of fit dependent on the substrate component with R^2 ranging from 0.32 (large stones) to 0.83 (sand) (Figure 5, Table 9). MAE ranged from 1.8% (hard clay) to 10.1% (sand) across all substrate component models, however the low MAE for hard clay was likely a result of a low occurrence of hard clay in the area. On average there didn't appear to be a substantial amount of bias in the models with ME ranging from -2.3% (gravel) to 3.4% (pebbles). However, all models, to varying degrees, tended to underestimate toward the lower end of the cover scale and overestimate towards the upper end. This was exemplified by positive ME in the 0–10% class and a predominance of increasingly negative ME values in classes $>10\%$ (Table 10). Furthermore, regression lines tended to be above the 1:1 line at the lower end of the scale and below the 1:1 line at the upper end indicating the aforementioned effect (Figure 5). Overall accuracy when predicted values were reclassified into classes 0–10%, $>10\text{--}50\%$, $>50\text{--}90\%$ and $>90\text{--}100\%$ (OA classes) or presence $>0.001\%$ (OA abs-pres.) were generally high ranging from 72.1% (sand) to 91.6% (hard clay), and 77.9% (gravel) to 92.9% (hard substrates) respectively (Table 9).

For sand, which was the dominating substrate type on the bank, the highest accuracy was found in the 0–10% interval (which in most cases equals 90–100% hard bottom) and the 90–100% interval (i.e., 0–10% hard bottom), showing higher uncertainty in the mixed hard/soft bottom areas. The individual hard bottom components, which rarely occurred in the 90–100% interval, had the highest accuracy in the 0–10% interval, increasing uncertainty of predictions $>10\%$.

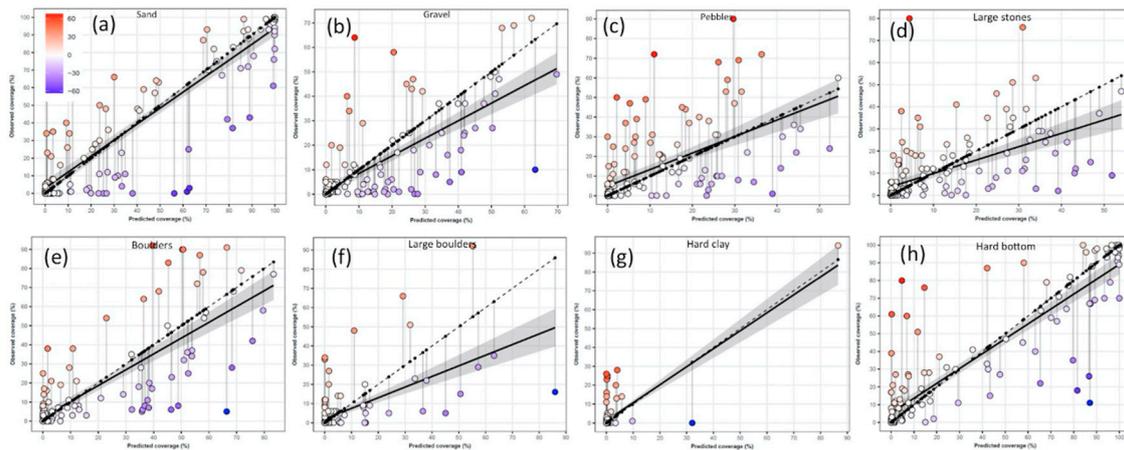


Figure 5. Relationship between observed coverage values (%) from testing data (n = 154) on the y-axis, and corresponding predicted coverage values (%) from substrate component models on the x-axis. (a) Sand (0.02–0.2 cm), (b) gravel (0.2–2 cm), (c) pebbles (2–6 cm), (d) large stones (6–20 cm), (e) boulders (20–60 cm), (f) large boulders (60 cm), (g) hard clay, (h) hard bottom. The solid black line corresponds to the linear regression and 95% confidence band (grey shade) between observed and predicted coverage. The dashed line corresponds to a theoretical 1:1 line where observed and predicted values are equal (i.e., $y = x$, y -intercept = 0, slope = 1), black points along this line represent the specific predicted coverage values. Point colors represent the magnitude of difference between an observed value and its corresponding predicted value on the 1:1 line (black point).

Table 9. Accuracy of percent coverage substrate components (individual models adjusted to sum to 100%). Distribution of the ground-truth dataset divided into HUB 3 classes are available in Table A4. Accuracy was measured (testing data, observed *vs.* predicted) with R^2 , root mean square error (RMSE), mean absolute error (MAE) and mean error (ME), in addition, the predicted values were grouped and evaluated with overall accuracy (OA) as classes (0–10%, 10–50%, 50–90%, 90–100%) and absence - presence (<absence threshold, >absence threshold-100%).

| | R^2 | RMSE | MAE | ME | OA Classes | OA Abs-pres. |
|-----------------------|-------|------|------|------|------------|--------------|
| Sand | 0.83 | 17.9 | 10.1 | −1.4 | 72.1 | 84.4 |
| Gravel | 0.53 | 12.6 | 7.4 | −2.3 | 74.0 | 77.9 |
| Pebbles | 0.34 | 15.7 | 9.3 | 3.4 | 75.3 | 81.2 |
| Large stones | 0.32 | 13.4 | 7.4 | −0.2 | 77.9 | 80.5 |
| Boulders | 0.61 | 15.4 | 8.7 | −0.4 | 74.7 | 85.1 |
| Large boulders | 0.36 | 11.5 | 4.7 | −0.2 | 83.1 | 87.7 |
| Hard clay | 0.61 | 5.8 | 1.8 | 1.1 | 91.6 | 92.2 |
| Hard bottom | 0.79 | 18.2 | 9.4 | 0.3 | 75.3 | 92.9 |

Table 10. Accuracy of percent coverage substrate components (individual models adjusted to sum to 100%), divided into groups. Accuracy was measured (testing data, observed *vs.* predicted) with mean error (ME) and standard deviation (SD).

| | 0–10% | | | 10–50% | | | 50–90% | | | 90–100% | | |
|-----------------------|-------|-----|------|--------|-------|------|--------|-------|------|---------|------|-----|
| | n | ME | SD | n | ME | SD | n | ME | SD | n | ME | SD |
| Sand | 53 | 5.6 | 15.3 | 33 | −2.1 | 20.8 | 20 | −15.5 | 27.3 | 48 | −2.7 | 7.2 |
| Gravel | 95 | 1.9 | 7.8 | 50 | −9.1 | 14.2 | 9 | −8.5 | 21.6 | 0 | - | - |
| Pebbles | 100 | 4.2 | 10.0 | 52 | 2.3 | 22.2 | 2 | −11.4 | 24.0 | 0 | - | - |
| Large stones | 102 | 3.0 | 9.9 | 50 | −5.7 | 16.4 | 2 | −24.9 | 25.3 | 0 | - | - |
| Boulders | 105 | 2.0 | 5.9 | 29 | −6.3 | 24.6 | 20 | −4.6 | 27.2 | 0 | - | - |
| Large boulders | 134 | 1.3 | 5.5 | 15 | −5.2 | 21.8 | 5 | −25.1 | 38.6 | 0 | - | - |
| Hard clay | 152 | 1.3 | 5.0 | 1 | −32.1 | - | 1 | 7.5 | - | 0 | - | - |
| Hard bottom | 86 | 4.8 | 13.3 | 21 | 3.6 | 22.4 | 24 | −14.7 | 26.6 | 23 | −4.4 | 8.9 |

3.3.2. Substrate Fine Fractions

The maps of sieved fine fractions adjusted to the main sand model (which included finer subfractions as well) are shown in Figure 6, with accuracy assessments using bootstrap of the sieved training data shown in Table 11. In a visual assessment of the outputs, we noticed that silt and fine sand accumulated in areas where post-glacial sand deposits were seen in the sediment echo sounder profiles. This observation often also corresponded with sediments with no ripples and the occurrence of *Macoma balthica* shell aggregations. The fine fraction models did not have very high accuracy overall and would benefit from a larger training dataset to describe such substrate features. However, the accuracy of the adjusted maps shown in Figure 6 is likely higher due to the adjustment to the main sand model (Tables 9 and 10).

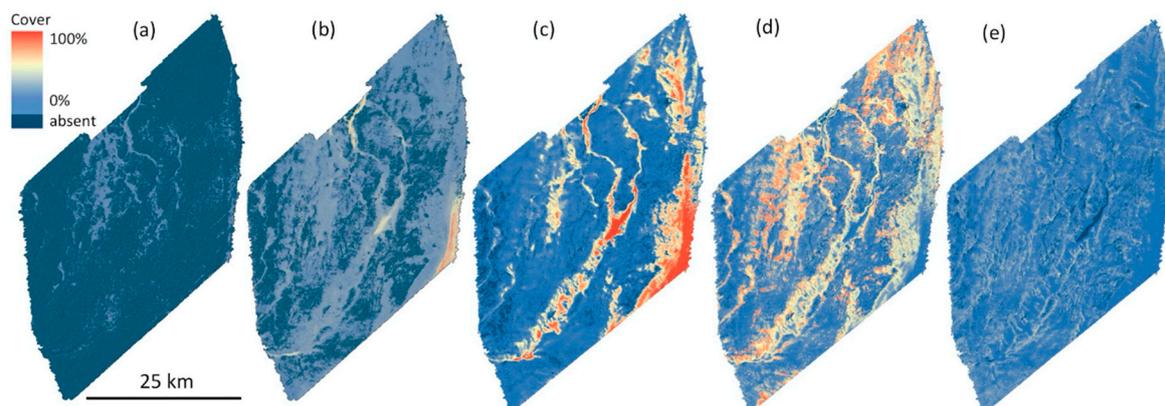


Figure 6. Coverage maps (%) of fine substrate components on Hoburgs Bank modeled from sieved grab samples, then adjusted to the main sand model. (a) Clay (<0.002 mm), (b) silt (0.002–0.06 mm), (c) fine sand (0.06–0.2 mm), (d) medium sand (0.2–0.6 mm), (e) coarse sand (0.2–2 mm)). For clay and silt the colors are stretched between 0–10% instead of 0–100%. Absence is based on modeled threshold (<0.01% cover).

Table 11. Accuracy statistics (using bootstrap of training data, n = 25) for fine fraction proportions predicted from sieved sediment samples. R², root mean square error (RMSE), mean absolute error (MAE).

| | R ² | RMSE | MAE |
|-------------|----------------|------|------|
| Clay | 0.03 | 2.02 | 0.92 |
| Clay-silt | 0.27 | 0.71 | 0.54 |
| Silt | 0.32 | 0.66 | 0.53 |
| Fine sand | 0.76 | 1.07 | 0.85 |
| Medium sand | 0.79 | 0.76 | 0.59 |
| Coarse sand | 0.58 | 1.41 | 1.02 |

3.3.3. Biological Components

The individual outputs and corresponding accuracy assessment from the biological models included in the HUB 5 level are shown in Figure 7. Similar to the substrate components some survey artifacts carried over to the biological models (seen as striping in SW/NE direction along survey lines in Figure 7g–i). However most of the artifacts occurred in mixed sediment areas with low relief, while sand areas and more distinct reef features showed very few visual artifacts.

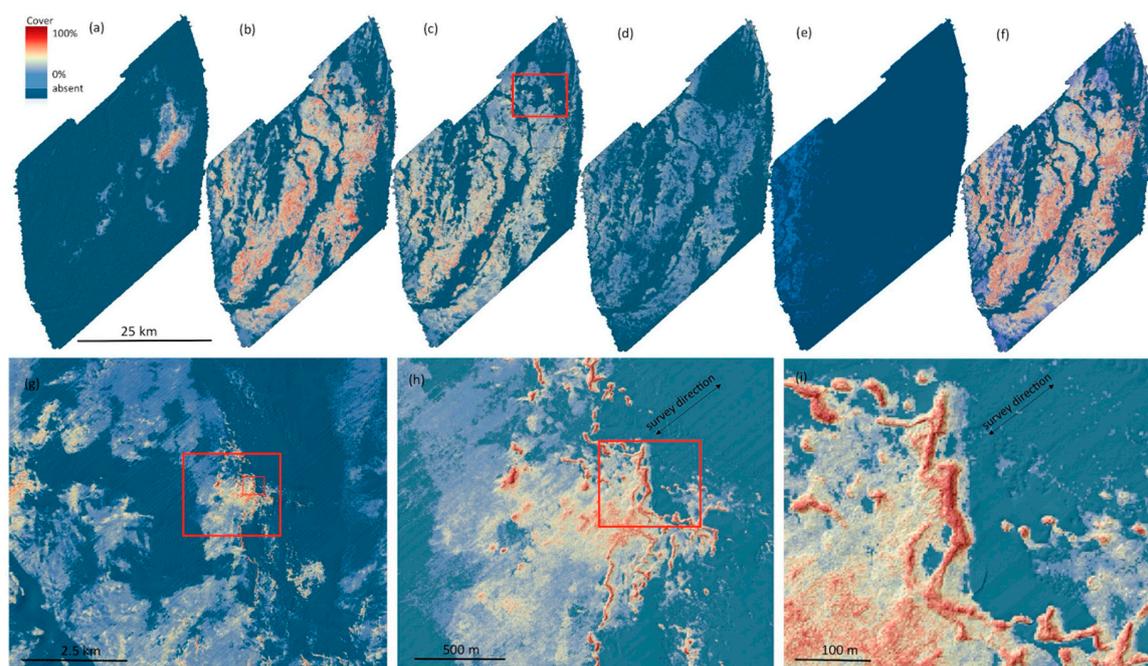


Figure 7. Percent coverage of HUB level-5 groups. (a) Annual algae, (b) perennial algae, (c) epibenthic bivalves (*Mytilus* spp.), (d) cnidarians, (e) epibenthic moss animals (*Bryozoans*), (f) colonized substrate, (g) *Mytilus* spp. with hillshade 5 m zoom level 1, (h) *Mytilus* spp. with hillshade 1 m zoom level 2, (i) *Mytilus* spp. with hillshade 1 m zoom level 3. Absence is based on modeled threshold (<0.01% cover).

Regression analysis of observed and predicted HUB level-5 biological cover showed varying degrees of fit dependent on the biological component with R^2 ranging from 0.01 (epibenthic moss animals (*Bryozoans*)) to 0.84 (colonized substrate) (Figure 8, Table 12). MAE ranged from 0.2% (moss animals) to 13% (perennial algae) across all HUB level-5 biological component models, however the low MAE for moss animals was likely a result of extremely low cover of moss animals in the area. On average, bias in the models was relatively low with the exception of uncolonized substrate where ME was -6.7% (Table 12). Similar to substrate components, models for perennial algae, epibenthic bivalves, and epibenthic cnidarians, to varying degrees, tended to underestimate toward the lower end of the cover scale and overestimate toward the upper end (Table 13 and Figure 8b–d). The annual algae model did not behave in the same manner tending to show an underestimation of annual algae cover, however this was likely because 149 of 154 of the predicted cover values were $<2\%$ making it difficult to assess. The epibenthic moss animal model was difficult to assess as all predicted values were $<0.9\%$ and observed coverage was generally low ($<6\%$, Figure 8e). The uncolonized substrate model, on the other hand, tended to overestimate the cover of uncolonized substrate. This was exemplified by negative ME values across all classes (Table 13) as well as the regression line below the 1:1 line (Figure 8f). Overall accuracy where predicted values were reclassified into classes 0–10%, >10–50%, >50–90% and >90–100% (OA classes) were generally high ranging from 68.2% (perennial algae) to 100% (epibenthic moss animals) (Table 8). However, high values for epibenthic moss animals and annual algae were not surprising as few to no observed values were $>10\%$ (Table 8). Overall accuracy where predicted values were reclassified to presence (OA abs-pres.) was high, ranging from 79.9% (Cnidarians) to 97.4% (Table 8).

Both substrate and biological component generally showed higher uncertainty in the mixed and often patchy hard/soft bottom areas (Figure 5, Figure 8 and Table 10, Table 13). This was further confirmed by the maps of standard deviation between all model outputs (Figure 9).

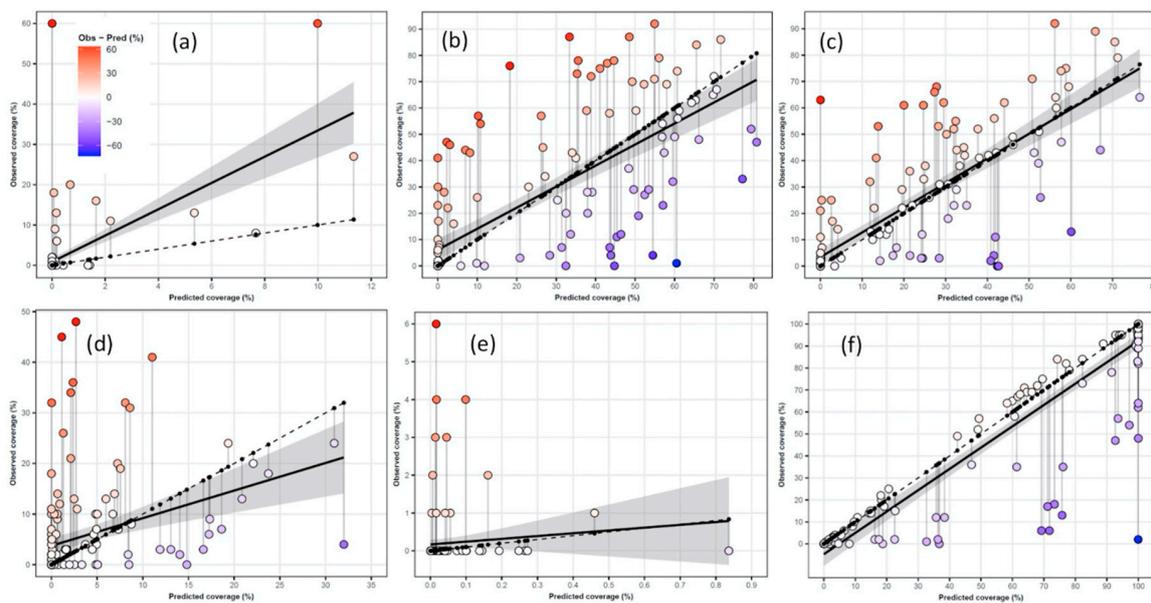


Figure 8. Relationship between observed coverage values (%) from testing data (n = 154) on the y-axis, and corresponding predicted coverage values (%) from biological component models on the x-axis. (a) Annual algae, (b) perennial algae, (c) epibenthic bivalves (*Mytilus* spp.), (d) cnidarians, (e) epibenthic moss animals (*Bryozoans*), (f) colonized substrate. The solid black line corresponds to the linear regression and 95% confidence band (grey shade) between observed and predicted coverage. The dashed line corresponds to a theoretical 1:1 line where observed and predicted values are equal (i.e., y = x, y-intercept = 0, slope = 1), black points along this line represent the specific predicted coverage values. Point colors represent the magnitude of difference between an observed value and its corresponding predicted value on the 1:1 line (black point).

Table 12. Accuracy of percent coverage biological components (HUB level-5 groups). Accuracy was measured (testing data, observed vs. predicted) with R², root mean square error (RMSE), mean absolute error (MAE) and mean error (ME), in addition, the predicted values were grouped and evaluated with overall accuracy (OA) as classes (0–10%, 10–50%, 50–90%, 90–100%) and absence - presence (<absence threshold, >absence threshold 100%).

| | R ² | RMSE | MAE | ME | OA Classes | OA Abs-pres. |
|---------------------|----------------|------|------|------|------------|--------------|
| Annual algae | 0.37 | 7.1 | 1.5 | 1.5 | 94.8 | 91.6 |
| Perennial algae | 0.52 | 20.4 | 13.0 | 1.8 | 68.2 | 91.6 |
| <i>Mytilus</i> spp. | 0.62 | 15.5 | 9.1 | 2.3 | 75.3 | 89.0 |
| Cnidarians | 0.12 | 9.7 | 4.7 | 2.4 | 77.3 | 79.9 |
| Moss animals | 0.01 | 0.8 | 0.2 | −0.2 | 100 | 90.3 |
| Colonized substrate | 0.84 | 17.9 | 8.0 | −6.7 | 80.5 | 97.4 |

Table 13. Accuracy of percent coverage biological components (HUB level-5 groups) divided into map percent cover intervals. Distribution of the ground-truth dataset divided into HUB 4–6 classes are available in Table A5. Accuracy was measured (testing data, observed vs. predicted) with mean error (ME) and standard deviation (SD).

| | 0–10% | | | 10–50% | | | 50–90% | | | 90–100% | | |
|---------------------|-------|------|------|--------|------|------|--------|-------|------|---------|------|------|
| | n | ME | SD | n | ME | SD | n | ME | SD | n | ME | SD |
| Annual algae | 153 | 1.4 | 6.9 | 1 | 15.7 | - | 0 | - | - | 0 | - | - |
| Perennial algae | 84 | 4.3 | 10.9 | 39 | 5.0 | 29.5 | 31 | −8.8 | 23.0 | 0 | - | - |
| <i>Mytilus</i> spp. | 80 | 2.6 | 8.4 | 56 | 2.5 | 20.7 | 18 | 0.6 | 20.5 | 0 | - | - |
| Cnidarian | 139 | 3.4 | 8.5 | 15 | −7.1 | 12.4 | 0 | - | - | 0 | - | - |
| Moss animals | 154 | 0.2 | 0.8 | 0 | - | - | 0 | - | - | 0 | - | - |
| Colonized | 27 | −0.8 | 1.8 | 22 | −9.8 | 14.2 | 23 | −13.7 | 27.6 | 82 | −5.9 | 15.3 |

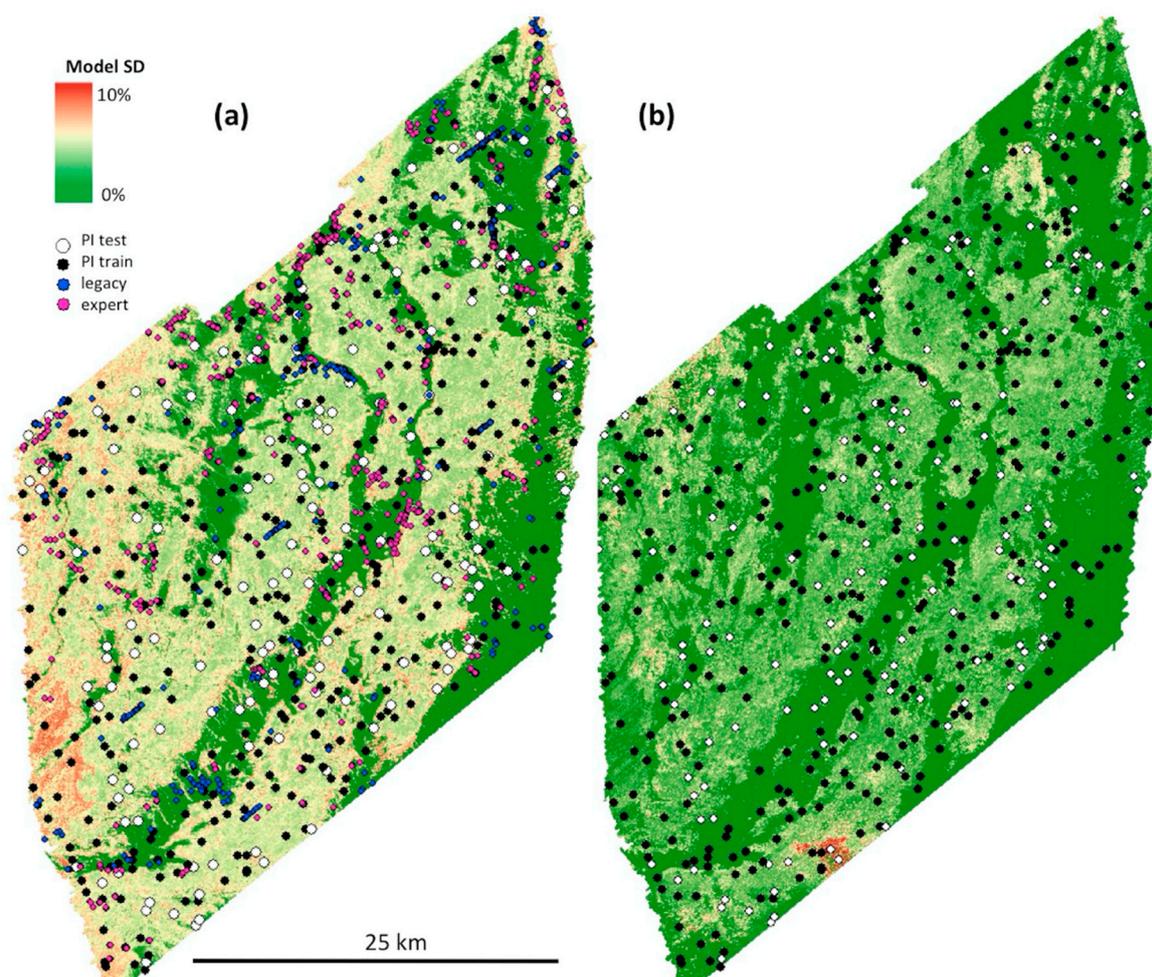


Figure 9. Locations of testing (PI stands for point intercept data from the 2016–2017 survey) and training data used for the modeling, as well as mean standard deviation for all continuous models tested for (a) substrate components (10 models of each component) and (b) biological components (5 models for each component). The maps are only intended to capture spatial patterns and cannot be used to compare the values between (a) and (b).

3.4. Thematic Maps

The reclassified thematic maps according to HUB, Natura 2000, and SGU substrate maps at 5-m resolution are shown in Figures 10–12. The accuracy of the HUB and Natura 2000 maps had overall accuracy ranging from 53.2% (HUB level 4–6) to 87.7% (Natura 2000). More details are available in Table 5, as well as in the user-producer matrix of HUB 3 (Appendix A Table A3). No accuracy assessment was available for the SGU substrate maps and Natura 2000 subtypes, since the testing data did not include these classes due to the included terrain metrics (Natura 2000 reef subtypes) and sediment samples (i.e., fine sand class in SGUs substrate maps).

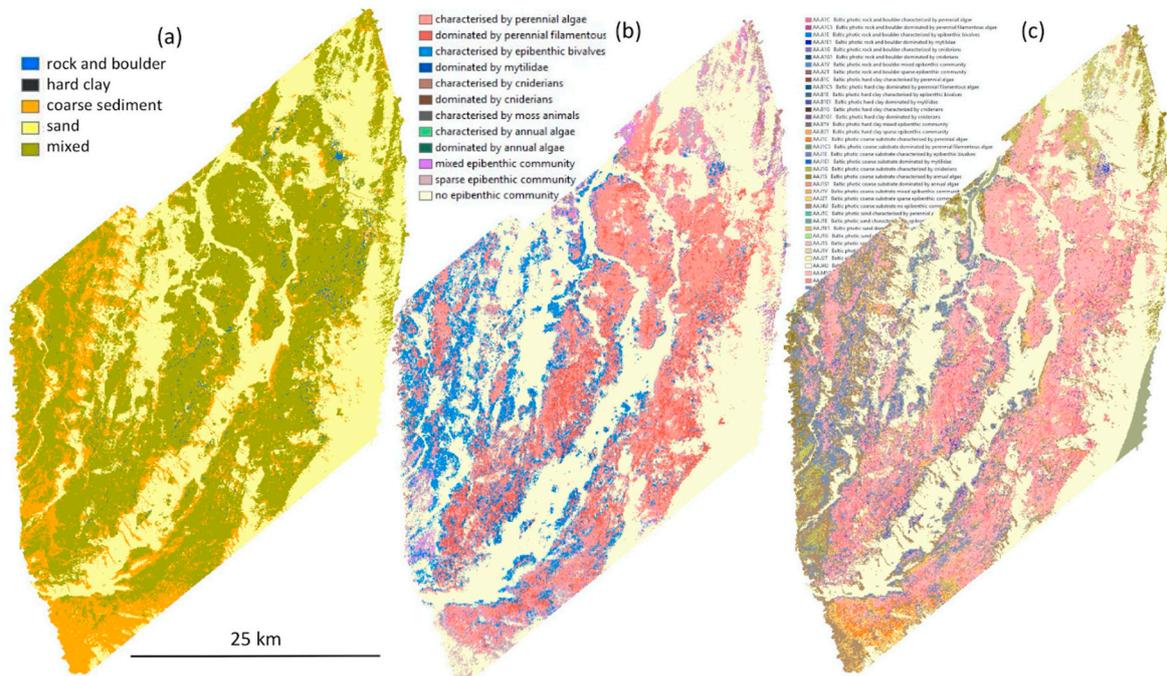


Figure 10. (a) HUB level 3, (b) HUB 4–6 (12 classes found), (c) HUB 1–6 (59 classes found, see Supplementary materials Legend S1, Table S3 and Map S1 for enlarged legend, map and more details). HELCOM HUB is a hierarchical classification scheme that consists of six levels with biotopes defined by applying split rules similar to a dichotomous key. The first 3 levels describe habitats (region, vertical zone, substrate) whilst levels 4–6 describe biotopes (community structure, characterizing community, dominating taxa).

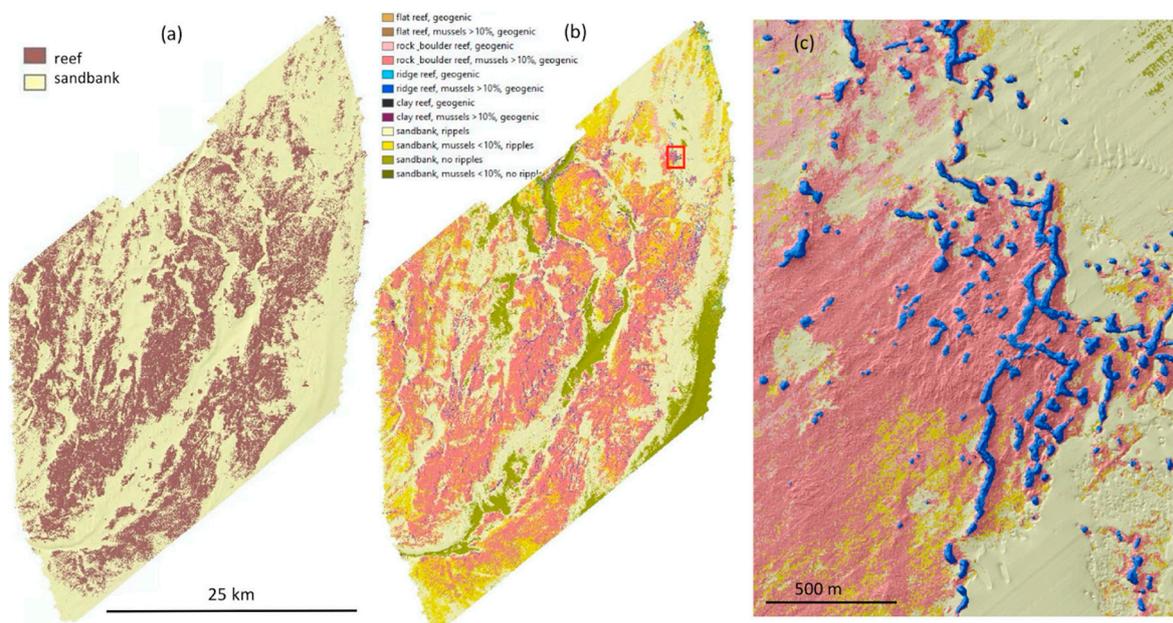


Figure 11. (a) Natura 2000 reefs and sandbanks, (b) Natura 2000 subtypes according to the Geological Survey of Sweden (SGU) implementation, (c) zoom of (b). The Natura 2000 subtypes were developed through a combination of continuous substrate and biological predictions, terrain metrics (i.e., fine-scale bpi) and thematic models (i.e., ripples or no ripples).

The Natura 2000 subtype map captured eight reef classes and four sandbank classes, revealing features and seascape patterns not seen in the main Natura 2000 map (two classes, sandbanks and reefs),

including an extensive pattern of ridges (Figure 11). Similarly, the reclassified SGU surface substrate map captured 6 classes and showed considerably more detail than the legacy surface substrate map, especially for the areas mapped with 1:500 000 scale (Figure 12).

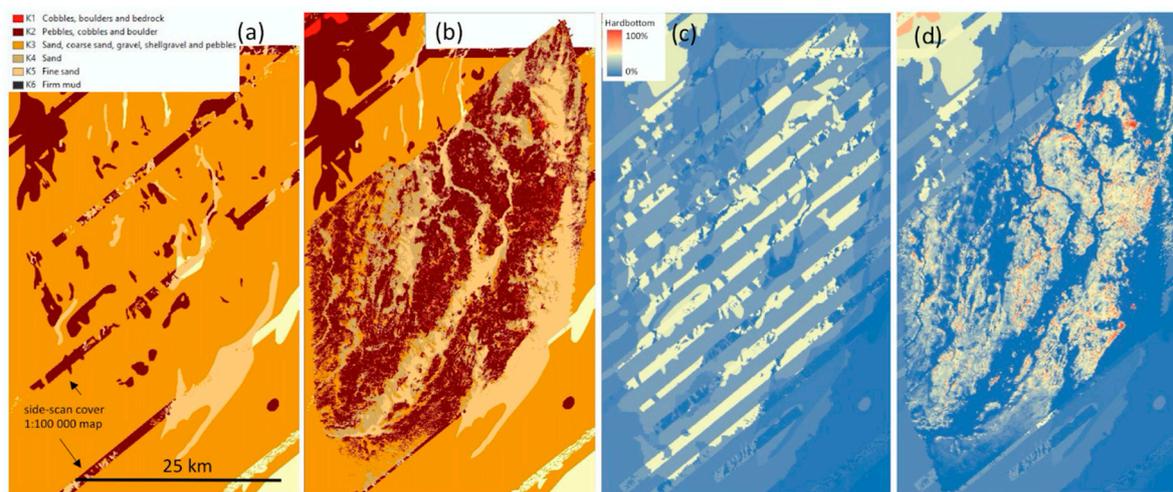


Figure 12. Before and after comparison between existing thematic map products of surface substrate and the new continuous models overlaid. (a) SGU legacy surface substrate maps (combination of 1:500 000 and 1:100 000 map), (b) same as (a) with the reclassified substrate models map overlaid, (c) percent hard bottom map developed by SGU from the legacy thematic maps (same as a) for the marine spatial planning cumulative impact assessment tool Symphony [10], (d) same as (c) with the new percent hard bottom model overlaid.

3.5. End-User Applications

In the following sections we investigate some of the end user applications of reclassified continuous maps.

3.5.1. Impact of Scale

The spatial resolution used when reclassifying continuous maps affected the percent coverage of the resulting thematic classes. When the HUB level 3 classification was applied to a growing-size window, small distinct features declined in cover and were replaced by more general classes (Figure 13, Table 14).

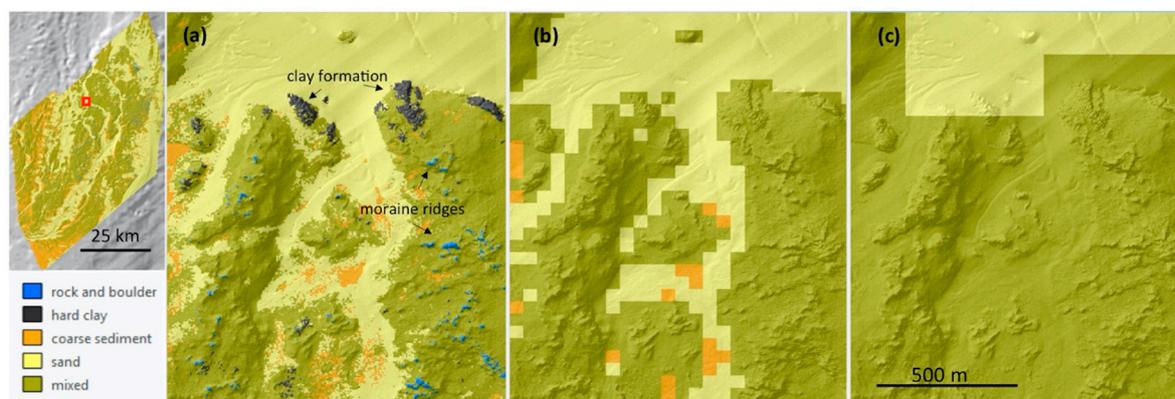


Figure 13. Comparison of HUB 3 substrate maps created using different scales; as the analysis window grows smaller features quickly give way to more general classes such as mixed sediments. (a) 5-m resolution, (b) 50-m resolution, (c) 250-m resolution.

Table 14. Percent coverage changes for HUB level 3 classes as the analysis window changes from 5 m to 250 pixels. Significant changes of cover for the rarer classes occur already at the 10-m pixels.

| | 5 m | 10 m | 25 m | 50 m | 250 m |
|-------------------------|-----------------------------|-----------------------------|------------------------------|----------------------------|-----------------------------|
| Sand | 36.4%, 469 km ² | 35.2%, 472 km ² | 35%, 465 km ² | 33.9%, 456 km ² | 31.3%, 420 km ² |
| Coarse | 16.2%, 217 km ² | 14.3%, 192 km ² | 13%, 179 km ² | 12.4%, 167 km ² | 10.0%, 133 km ² |
| Mixed | 46.4%, 623 km ² | 50.0%, 672 km ² | 51.8%, 696 km ² | 53.5%, 719 km ² | 58.7%, 789 km ² |
| Rock and boulder | 0.74%, 10 km ² | 0.56%, 7.5 km ² | 0.32%, 4.3 km ² | 0.10%, 1.4 km ² | 0.01%, 0.12 km ² |
| Hard clay | 0.06%, 0.81 km ² | 0.02%, 0.22 km ² | 0.003%, 0.04 km ² | 0 | 0 |

In the Natura 2000 maps (reef/sandbanks), the reef class increased in cover as the window-size increased (Figure 14), moving towards 100% cover of reef when using mean cover for the whole bank. Predicted mean cover of hard bottom (19.3%) was below the threshold for reef (>50%) but *Mytilus* in the study area was above the threshold (>10% mussel cover), which then classified it as reef. Conversely, if the mean cover for *Mytilus* would have been < 10%, the mean cover of the study area would have been classified as sandbank, reversing the pattern of the growing analysis window.

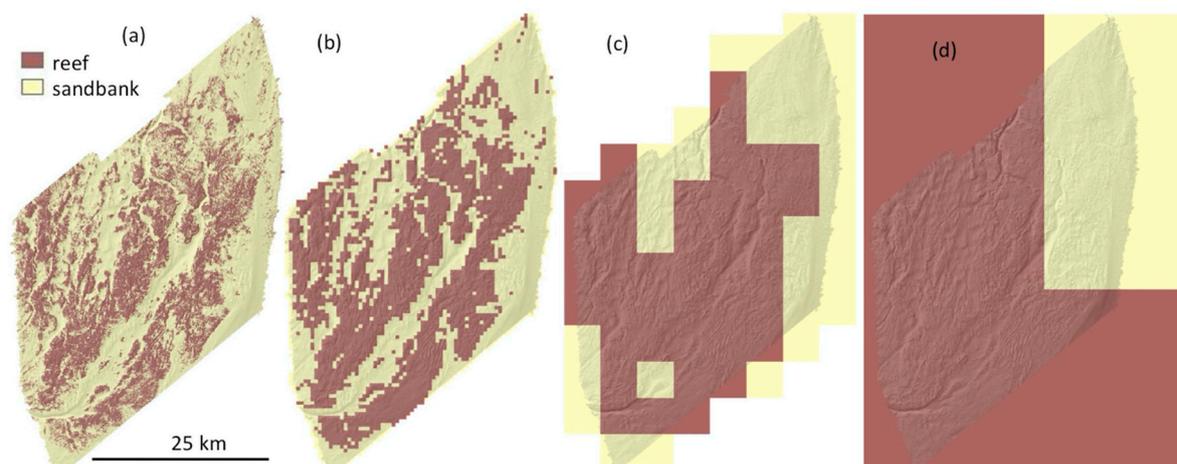


Figure 14. Comparison of Natura 2000 maps created using different scales from 5 m to 25 km analysis window. The definition for “reef” is >50% hard bottom (the mean hard bottom cover on the bank was 19.3%) or >10% epibenthic bivalves (the mean *Mytilus* cover on the bank was 10.7%). The choice of scale changes the cover of each class. (a) 5-m resolution (reef 44% coverage), (b) 500-m resolution (reef 51% coverage), (c) 5 km resolution (reef 55% coverage), (d) 25 km resolution (reef 67% coverage). At the scale of the whole bank it would be 100% cover of reef.

High-resolution habitat maps generally allow the user to choose across many different options, therefore, aggregating the information to the appropriate-management scale. Applying a management scale of 250 m pixels to the HUB 3 classification made the classes rock and boulder and hard clay disappear (0.01% and 0%). However, if instead all 250 m pixels containing these classes at the 5 m scale were mapped, rock and boulder covered 32% and hard clay 17.8% (Figure 15).

3.5.2. Effect of Predefined Thresholds in HUB

The areas classified as characterized by epibenthic bivalves or dominated by *Mytilidae* (i.e., $\geq 10\%$ and $\geq 50\%$ mussel cover respectively) in the reclassified HUB map were 161 km² (12%) and 11.6 km² (0.9% of the area) respectively. The area using the same thresholds but calculated directly from the percent coverage map of *Mytilus* was 479 km² (36%) and 19.8 km² (1.5%) (Figure 16). Due to overlap with other benthic organisms with higher cover (primarily perennial algae), the HUB map was not accurate when analyzing percent cover of individual organisms.

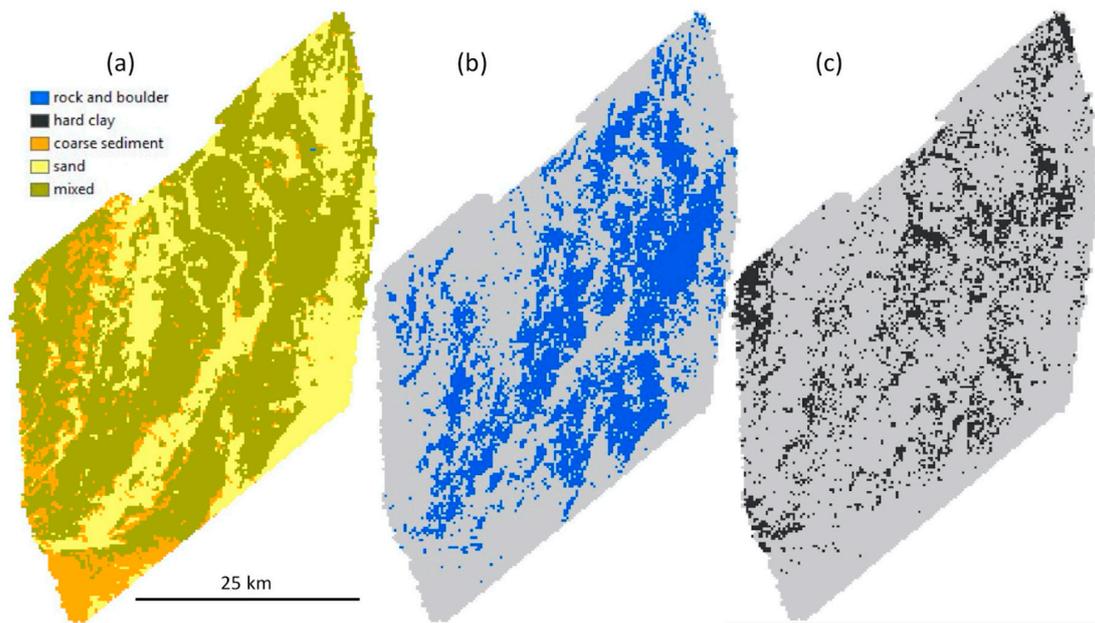


Figure 15. (a) HUB 3 substrate map at the 250 m scale, (b) 250 m pixels that contain HUB 3 rock and boulder class (blue) at the 5-m scale, (c) 250 m pixels that contain HUB 3 hard clay class (black) at the 5-m scale.

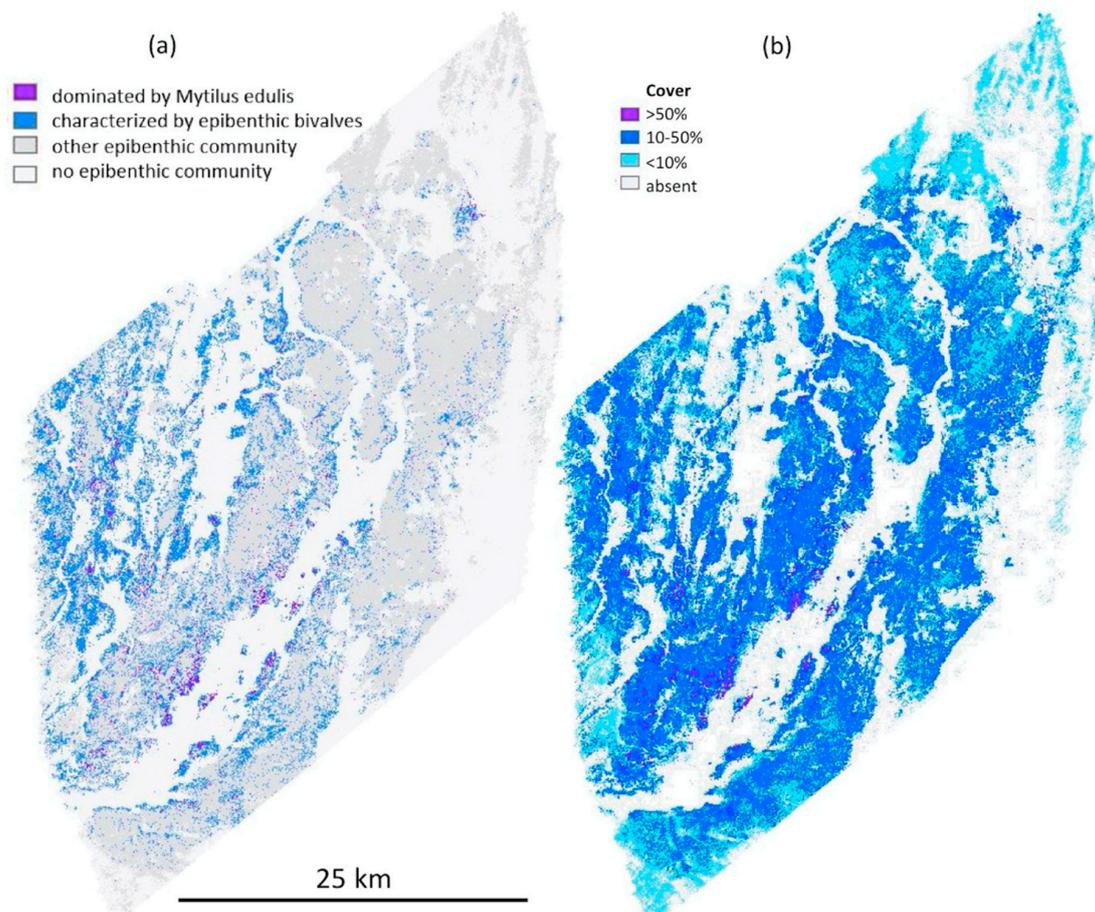


Figure 16. Comparison of *Mytilus* spp. maps from HUB classes versus the continuous prediction (a) *Mytilus* classes from HUB level 4–6 map (characterized by = 10–50%, dominated by = >50% cover), (b) *Mytilus* classes extracted from percent coverage prediction (absent, 0–10%, 10–50% and >50%).

4. Discussion

High-resolution continuous and thematic models can help to overcome some of the limitations of classification schemes. We found that the accuracy (overall accuracy and Tau of observed vs. predicted) of classified HUB and Natura 2000 maps were similar regardless of whether we used classification models directly, or reclassified models of continuous percent coverage. However, we also found that producing continuous maps and then applying classifications had a greater value due to the added possibilities, such as simple application of thresholds (conditions) or spatial analysis directly from continuous variables (e.g., through raster calculation).

In the first part of the discussion below, we discuss how the use of continuous maps affects the possibilities for end users. In the second part, we focus on technical aspects of this study and the accuracy of the maps we produced.

4.1. User Applications

Given that benthic habitat maps are often supported by similar types of information (e.g., substrate and benthic organisms), it is possible to map different classification schemes that share the same continuous components (but perhaps not the same combination of thresholds), without having to rerun any models. For the case where maps include components that are not suitable for continuous models, the classification can still use a combination of percent coverage predictions and terrain metrics (e.g., BPI, slope) to map geomorphological features such as ridges and walls (in this study exemplified by the ridge class in the SGU implementation of Natura 2000 subtypes, mapped from a combination of percent hard bottom cover prediction and BPI thresholds), or use additional thematic models for some specific features (in this study exemplified by the sand ripple class in the SGU implementation of Natura 2000 subtypes).

Moreover, the thematic maps that are based on continuous models are not limited to one spatial scale but can easily be adjusted to any window larger than the minimum mapping unit. Therefore, we provided all HUB and Natura 2000 maps at multiple scales to accommodate different definitions and applications. The scale and the minimum mapping unit can dramatically change the occurrence of different thematic classes [43,44], especially in patchy environments like our study area as shown in Figure 13 and Table 14. Though some guidelines of scale often do exist [3], clear directions for how to address the problem of thematic changes at different scales are missing from many schemes, including EUNIS, HUB and Natura 2000. The importance of such schemes makes this a priority to address. When the mapmakers and end-users are free to decide on the scales relevant for a specific habitat and biotopes of interest it could come with unwanted effects. Assuming, for example, that different legislation is tied to the Natura 2000 reef and sandbank classes on Hoburgs Bank, it would then be important to also specify the scale, since the bank can be classified as a sandbank, a reef, or both depending on the size of the analysis window. Similarly, the scale also needs to be considered when looking to monitor habitat (and ecological status) change. A focus on conservation features (e.g., reefs) can often overlook the ecological importance of connected patches types such as sand or seagrasses in the seascape, which can also be critical resources for mobile species [45]. Furthermore, when transnational classification schemes are used in multiple mapping projects, the lack of a clear definition on scale can hamper the overarching goal of creating a unified map. Transnational maps can be seen as a common governance effort across countries borders as discussed by Pecl et al., [13] for instance when considering habitat maps as tools to manage effects of climate change to ecosystems.

In this study we show that an important advantage of continuous high-resolution maps is that they provide more options when moving information across different scales, for example from mapping to the management scale. In Figure 15, we recalculated the HUB level-3 substrate map at 250-m resolution, which is the resolution used by SwAM to assess current and future cumulative impact on the marine environment, and to evaluate the effect of the proposed national marine spatial plan [10]. At this scale, the less common classes hard clay and rock and boulder disappeared into the more commonly occurring HUB classes sand, coarse sediments and mixed. In fact, the coverage of the hard clay and

rock and boulder classes started to decline already at the 10 m scale. If small but distinct features associated with hard clay and rock and boulder classes have important ecosystem functions, despite their coverage, it could be key to include this information at the management scale. For example, SGU (observation by Gustav Kågesten during field season 2018 in the north middle-seabank) and Beisigel et al. [46] found that such features are used as spawning locations by herring and as shelter by cod. On the contrary, if only commonly occurring classes are of interest perhaps a coarser scale is appropriate as well. When high-resolution maps are available there are alternative ways to include such specific features at the management scale, for example by mapping management units that contain a presence of the features of interest. The HUB rock and boulder class in our study area covered nearly a third of the bank, when testing this approach for 250 m pixels (Figure 15). This could be essential information in a marine spatial plan, for example assuming that the mere presence of large boulders in a planning unit provides grounds for protection from bottom trawling.

Maps of continuous variables can have a broad variety of applications where classified maps may not be as suitable. An obvious example in the study area is to estimate mussel biomass and related ecosystem services [47] using the continuous prediction, thereby providing the percent mussel coverage with greater certainty than thematic maps, which rely on intervals of mussel coverage. However, if for instance a bird conservation action is being planned based on the total area suitable for birds feeding on blue mussels, and assuming that birds seek areas with at least 50% mussel cover according to their feeding strategy, one may wrongly think that the HUB classification can help in the conservation plan. Though classes at level 4 to 6 capture locations characterized by an organism (if it has more than 10% coverage) or locations dominated by an organism (if it has more than 50% coverage), a location would only be considered dominated by an organism if no other organism had a greater coverage in that cell (since holdfast and canopy count as overlay in our point intercept data the total cover can exceed 100%). In our study area, we observed that perennial algae often covered *Mytilus*, and this caused the HUB classification to be labeled as dominated by perennial algae instead of mussels (which often still had a coverage of >50%, though not captured in the HUB map). In this last case, the locations will incorrectly be considered as unsuitable for the bird conservation action. According to our assumptions, the total area of suitable bird habitat almost doubles when basing the calculation on the continuous mussel prediction instead of the HUB map (Figure 16). If a different coverage of mussels is set (e.g., 10–50%), the suitable habitat increased threefold when using the mussel percent cover map instead of the HUB map. Finally, if the suitable feeding location threshold is a number larger than 50% of mussel cover, the HUB map will not provide any suitable area with confidence.

Differences in the representation of the relative abundance of habitat types in the seascape can have important implications for management activities including identifying and quantifying spatial change, valuing natural capital and protecting resource designations such as essential fish habitat. Furthermore, when habitat maps are used as spatial predictors in species distribution models or to quantify spatial patterns the issue of spatial and thematic resolution must be considered specifically as a key variable and carefully evaluated [34,48,49].

4.2. Technical Applications

Our use of legacy data and expert annotations proved, as in previous studies [50], to have a high value in the models. For example, their use increased the HUB 3 thematic substrate model accuracy (overall accuracy observed vs. predicted) from 59% to 78% (Table 7). The additional training data concerning substrate components also indirectly improved the biological models, using substrate model outputs as predictor variables. The accuracy of the *Mytilus* spp. model (R^2 , RMSE, MAE and ME of observed vs. predicted) and HUB level 4–6 maps improved when the substrate predictions were included as predictor variables (Table 7), even when high resolution bathymetry and backscatter were available, which to some degree can act as surrogate for substrate grain size [51]. We also noticed fewer survey artifacts in the biological models when substrate models were included as predictor variables. The full combination of substrate and biological models into HUB biotopes level 1–6

(see Supplementary materials Legend S1 or Table S3 for all 59 resulting HUB classes), are likely to have high uncertainties at the pixel level for the end user (generally, the more thematic classes combined the lower the accuracy for each combination as shown in Table 5), but also generated very few pixels containing obviously rogue classes (such as an illogical combination of soft bottom dominated by a hard bottom species). We suspect that this was due to the use of substrate outputs in the biological models, and from using the same set of predictors. Our results showed the possibilities of including ground truthing data from several quality levels (i.e., high resolution survey, legacy data and expert annotations), that data from sediment samples and images can be combined using continuous models, and the benefits of integrated geological and biological surveys and models.

The high resolution of our survey enabled us to capture features such as sand ripples and individual boulders, using a combination of backscatter statistics, multiscale bathymetry and OBIA metrics (which had the highest overall contribution in the models, Table A1, Table S2). While models using the full range of predictor variables performed best and is the approach we recommend for future studies prioritising high accuracy, we concluded that a simpler combination of bathymetry metrics and a backscatter mosaic at resolution of 5 m would have been sufficient to create our models (accuracy for HUB level 3 decreased from 78% to 73%). The combination of 5 m bathymetry metrics and the backscatter mosaic outperformed any further combination of multiscale metrics that did not include backscatter, indicating the importance of backscatter as a predictor to capture sediment characteristics. This is not surprising given that backscatter has long been recognized as a key predictor variable [52,53]. When backscatter was not included, we observed that the multiscale metrics were increasingly important (OBIA metrics was not tested in this case, but it is likely it would be useful as well considering that it was important in the models that included backscatter), highlighting that some accuracy can be gained using high-resolution (≤ 1 m) bathymetry metrics and rescaling techniques.

Though we provided several different statistical measures describing the accuracy of our predictions, covering the gap between reality and the accuracy assessment of our predicted maps remains a challenge [14]. Combining continuous models into thematic maps, as well as using substrate models to train biological models made techniques such as cross-validation impractical. Hence, we used the same testing data for all models, which to some degree lowers the confidence in the accuracy statistics produced. Figure 9 confirms that error is higher in patchy areas and lower in homogeneous areas for both substrate and biological models. The manual point intercept annotations of high-resolution drop-camera imagery were a source of unknown error, and also a major part of the post processing effort. Due to the inclusion of lower quality expert annotations and legacy data in the training data set, we were able to lower the per site annotation effort for the substrate models, with a positive spillover effect to some of the biological models. Even so, we recognize that the time and cost needed to implement continuous percent coverage annotations may be a limiting factor for mappers, compared to thematic annotations. We believe that further development and application of computer vision technologies that have potential to support expert annotations of ground truth data and estimates of uncertainty [54–57] will be an important field to lower the costs and further improve the approach presented in this study.

Our maps of Hoburgs Bank are intended to be living rather than static products that can be enhanced as soon as, for instance, more or better training data becomes available. However, even with future improvements (e.g., related to data quality, reduction of human error, and more sophisticated models and techniques), we will always live with some degree of uncertainty due to natural variability that our modeling will not be able to incorporate [14,58].

5. Conclusions

Maps with continuously varying data on substrate and benthic organisms added more valuable information than more simplified patch-mosaic thematic maps. Continuous maps provide more flexibility in application and can be scale-adjusted and resampled more easily to address a wider range of research questions and management applications. Our findings provide a clear example of

how high-resolution continuous maps can be successfully generated by combining multibeam depth, backscatter, drop-camera and sediment samples, and that additional classified maps can be generated from this workflow without the loss of accuracy compared with thematic models. While we recognize the value and convenience of thematic maps, relying solely on such classification schemes can result in the neglect of habitats or features with great environmental value. It is therefore essential to understand the limitations of any classified maps for marine spatial planning and decision-making particularly with regard to scale and schema dependence and the associated errors that may be propagated through the mapping process.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-3263/9/5/237/s1>, Scripts: Legend S1: HELCOM HUB 1–6; Map S1: survey data and model outputs.pdf; Table S1: all predictors.pdf; Table S2: final predictors.pdf; Table S3: HELCOM HUB 1–6.ods; Script S1: predictor prep pca.R; Script 2: S2 model prep.R; Script 3: multinomial models.R; Script 4: continuous individual models.R; Script 5: continuous compositional models.R; Script 6: reclassify hub3.R; Script 7: reclassify hub1–6.R; Script 8: reclassify Natura2000.R; Script 9: accuracy continuous maps.R; Script 10: accuracy thematic maps.R

Author Contributions: Conceptualization, G.K., D.F. and L.Z.; methodology, G.K., D.F. and F.B.; validation, G.K. and F.B.; formal analysis, G.K. and D.F.; writing—original draft preparation, G.K. and D.F.; writing—review and editing, G.K., D.F., F.B. and L.Z.; visualization, G.K.; supervision, G.K.; project administration, L.Z.; funding acquisition, L.Z.

Funding: This research was funded by the Swedish Agency for Marine and Water Management, grant number 2234-16 and 3901-16.

Acknowledgments: The authors wish to thank Francis Freire, Annika Dahlgren at the Geological Survey of Sweden for assisting the multibeam processing (backscatter and bathymetry) and annotations, Anna Svensson, Liselott Wilin, Caroline Bringensparr and Pär Nordgren for additional support with data processing, and the crew on R/V Surveyor for many long hours at sea and technical support. We also thank Markus Diesing at the Geological Survey of Norway for assisting with the sampling design. Further, many thanks to the team at the NOAA National Centers for Coastal Ocean Science: Arliss Winship and Matthew Poti for their modeling support, Ken Buja and Charles Menza for support with the sampling design tool, and Bryan Costa and the late Brian Kinlan for providing scripts (Confusion matrix and multiscale Gaussian function) and inspiration. Special thanks also to Anna Kågesten, Simon Pittman and Jay Lazar for reviewing and editing the manuscript draft.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Percent contribution of predictors (grouped by theme, see Supplementary Materials, Table S2 for more detailed breakdown) in thematic BRTs with and without substrate (sub) models included in the predictor variables.

| Predictor Group | HUB3 % contr. | HUB4–6 no sub % contr. | HUB4–6 with sub % contr. | <i>Mytilus</i> no sub % contr. | <i>Mytilus</i> with sub % contr. | Average no sub% contr. | Average with sub % contr. |
|--|------------------|------------------------------|--------------------------------|--------------------------------------|--|------------------------------|---------------------------------|
| BS mosaic metrics 0.5 m–1 m | 3.1 | 4.3 | 3.0 | 3.3 | 1.2 | 3.5 | 2.1 |
| BS mosaic metrics 5 m | 10.0 | 4.6 | 2.0 | 3.8 | 0.7 | 6.1 | 1.3 |
| BS ARA | 2.8 | 4.9 | 1.4 | 3.1 | 0.6 | 3.6 | 1.0 |
| BS mosaic multiscale 20 m–2 km | 4.0 | 4.1 | 2.6 | 3.2 | 1.1 | 3.8 | 1.9 |
| Pg.s sediment thickness | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 |
| Depth 5 m | 4.4 | 3.8 | 3.7 | 2.1 | 0.8 | 3.4 | 2.3 |
| Terrain metrics 0.5 m–1 m | 8.5 | 9.4 | 5.6 | 10.1 | 2.7 | 9.3 | 4.1 |
| Terrain metrics 5 m | 5.4 | 7.7 | 5.3 | 6.7 | 2.4 | 6.6 | 3.9 |
| Terrain metrics multiscale 20 m–2 km | 16.3 | 20.6 | 15.5 | 14.7 | 7.3 | 17.2 | 11.4 |
| Distance to reefs and sand | 9.6 | 7.4 | 3.2 | 4.9 | 1.4 | 7.3 | 2.3 |
| OBIA based on terrain metrics and BS 2.5 m | 25.9 | 19.7 | 7.3 | 40.0 | 3.0 | 28.5 | 5.1 |
| Oceanography | 2.4 | 2.5 | 2.2 | 1.3 | 0.6 | 2.1 | 1.4 |
| Uncertainty (depth, backscatter) | 2.0 | 3.1 | 2.3 | 2.0 | 1.3 | 2.4 | 1.8 |
| Northing/easting | 4.3 | 3.5 | 3.1 | 1.7 | 0.9 | 3.1 | 2.0 |
| Substrate models | - | - | 40.2 | - | 75.2 | - | 57.7 |

Table A2. Substrate component accuracy assessment using the validation data set (n = 154) for four model types, individual models (ind), individual models adjusted to 100% sum (i.adj), compositional models (comp), and combined compositional and individual models (hard clay component) adjusted to sum to 100% (comb). All models were run using Gaussian distribution boosted regression tree models with the same training data and predictor variables.

| | R ² | | | | RMSE | | | | MAE | | | | Bias | | | |
|------------|----------------|-------|------|------|------|-------|------|------|------|-------|------|------|------|-------|------|------|
| | ind | i.adj | comp | comb | ind | i.adj | comp | comb | Ind | i.adj | comp | comb | ind | i.adj | comp | comb |
| Sand | 0.80 | 0.83 | 0.81 | 0.81 | 20.4 | 17.9 | 19.2 | 19.3 | 11.8 | 10.1 | 10.0 | 10.0 | 5.7 | -1.4 | 0.9 | 0.9 |
| Gravel | 0.46 | 0.53 | 0.49 | 0.49 | 12.6 | 12.6 | 12.6 | 12.7 | 6.9 | 7.4 | 7.1 | 7.1 | 2.8 | -2.3 | 0.4 | 0.3 |
| Pebbles | 0.23 | 0.34 | 0.28 | 0.28 | 17.7 | 15.7 | 18.2 | 18.2 | 10.1 | 9.3 | 10.6 | 10.5 | 6.0 | 3.4 | -0.6 | -0.6 |
| L stones | 0.25 | 0.32 | 0.32 | 0.33 | 13.5 | 13.4 | 16.4 | 16.2 | 7.3 | 7.4 | 9.4 | 9.4 | 4.1 | -0.2 | -3.6 | -3.5 |
| Boulders | 0.44 | 0.61 | 0.55 | 0.56 | 18.8 | 15.4 | 16.8 | 16.7 | 9.8 | 8.7 | 8.7 | 8.5 | 5.6 | -0.4 | 0.4 | 0.6 |
| L boulders | 0.44 | 0.36 | 0.62 | 0.64 | 9.7 | 11.5 | 7.8 | 7.6 | 4.0 | 4.7 | 3.2 | 3.1 | 1.1 | -0.2 | 1.5 | 1.6 |
| Hard clay | 0.66 | 0.61 | 0.23 | 0.36 | 5.7 | 5.8 | 9.2 | 8.8 | 1.8 | 1.8 | 2.3 | 2.0 | 1.5 | 1.1 | 1.1 | 0.7 |
| Mean | 0.44 | 0.51 | 0.51 | 0.52 | 15.4 | 13.2 | 15.2 | 15.1 | 8.3 | 7.1 | 8.2 | 8.1 | 4.2 | 0 | -0.2 | -0.1 |
| SD | 0.20 | 0.19 | 0.21 | 0.19 | 5.3 | 3.9 | 4.5 | 4.6 | 3.6 | 1.8 | 3.3 | 3.4 | 2.1 | 1.8 | 1.7 | 1.7 |

Table A3. Observed vs. predicted confusion matrix for reclassified HUB 3 map using the validation data set (n = 154).

| | Coarse Substrate | Hard Clay | Mixed | Rock & Boulder | Sand | Sum (n) | User Accuracy % |
|---------------------|------------------|-----------|-------|----------------|------|---------|-----------------|
| Coarse substrate | 12 | 0 | 10 | 0 | 0 | 22 | 54.55 |
| Hard clay | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mixed | 3 | 1 | 42 | 3 | 0 | 49 | 85.71 |
| Rock & boulder | 0 | 0 | 3 | 16 | 0 | 19 | 84.21 |
| Sand | 3 | 0 | 7 | 0 | 54 | 64 | 84.38 |
| Sum (n) | 18 | 1 | 62 | 19 | 54 | 154 | 0 |
| Producer accuracy % | 66.67 | 0 | 67.74 | 84.21 | 100 | 0 | 80.52 |

Table A4. Distribution of the ground-truth dataset divided into HUB 3 classes.

| HUB 3 | UV obs. ¹ | Legacy ² | Expert ³ | Total |
|------------------|----------------------|---------------------|---------------------|-------|
| Rock and boulder | 67 | 34 | 16 | 117 |
| Hard clay | 5 | 0 | 83 | 88 |
| Coarse sediment | 72 | 89 | 138 | 299 |
| Sand | 195 | 159 | 252 | 606 |
| Mixed | 220 | 167 | 61 | 448 |
| Total | 559 | 449 | 550 | 1558 |

¹ Drop camera samples (2016–2017), testing (30%) and training (70%); ² Drop camera, video transects, sediment samples (2005), training only; ³ Expert annotation (based on depth and backscatter), training only.

Table A5. Distribution of the ground-truth dataset divided into HUB 4–6 classes.

| HUB Level 4–6 | UV obs. ¹ |
|--|----------------------|
| Characterised by perennial algae | 73 |
| Dominated by perennial filamentous algae | 94 |
| Characterised by epibenthic bivalves | 42 |
| Dominated by <i>Mytilus</i> spp. | 59 |
| Characterised by cnidarians | 25 |
| Characterised by annual algae | 2 |
| Mixed epibenthic community | 16 |
| Sparse epibenthic community | 79 |
| No epibenthic community | 169 |
| Total | 559 |

¹ Drop camera samples (2016–2017), testing (30%) and training (70%).

References

- Galparsoro, I.; Connor, D.W.; Borja, A.; Aish, A.; Amorim, P.; Bajjouk, T.; Chambers, C.; Coggan, R.; Dirberg, G.; Ellwood, H.; et al. Using EUNIS habitat classification for benthic mapping in European seas: Present concerns and future needs. *Mar. Pollut. Bull.* **2012**, *64*, 2630–2638. [CrossRef] [PubMed]
- Pittman, S.; Connor, D.W.; Radke, L.; Wright, D. Application of Estuarine and Coastal Classifications in Marine Spatial Management. *Treatise Estuar. Coast. Sci.* **2011**, *1*, 163–205.
- Strong, J.A.; Clements, A.; Lillis, H.; Galparsoro, I.; Bildstein, T.; Pesch, R. A review of the influence of marine habitat classification schemes on mapping studies: Inherent assumptions, influence on end products, and suggestions for future developments. *ICES J. Mar. Sci.* **2019**, *76*, 10–22. [CrossRef]
- Wedding, L.M.; Lepczyk, C.A.; Pittman, S.J.; Friedlander, A.M.; Jorgensen, S. Quantifying seascape structure: Extending terrestrial spatial pattern metrics to the marine realm. *Mar. Ecol. Prog. Ser.* **2011**, *427*, 219–232. [CrossRef]
- Davies, C.E.; Moss, D. *EUNIS Habitat Classification Marine Habitat Types: Revised Classification and Criteria*; C02492NEW; Centre for Ecology and Hydrology: Dorset, UK, 2004.
- HELCOM. *Technical Report on the HELCOM Underwater Biotope and Habitat Classification*; 139; Helsinki Commission: Helsinki, Finland, 2013; p. 96.
- European Commission. Natura 2000. Available online: http://ec.europa.eu/environment/nature/natura2000/index_en.htm (accessed on 23 March 2019).
- European Parliament Council. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). *Off. J. Eur. Commun.* **2008**, *164*, 19–40.
- Council of the European Union. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. *Off. J. Eur. Commun.* **1992**, *206*, 7–50.
- Havs- och Vattenmyndigheten. *Symphony. Integrerat Planeringsstöd för Stalig Havsplanering Utifrån en Ekosystemsansats*; Havs- och Vattenmyndigheten: Göteborg, Sweden, 2018; p. 72.
- Schiele, K.S.; Darr, A.; Zettler, M.L.; Friedland, R.; Tauber, F.; von Weber, M.; Voss, J. Biotope map of the German Baltic Sea. *Mar. Pollut. Bull.* **2015**, *96*, 127–135. [CrossRef]
- Cameron, A.; Askew, N. *EUSeaMap-Preparatory Action for Development and Assessment of a European Broad-Scale Seabed Habitat Map Final Report*; 2011. Available online: <http://jncc.gov.uk/euseamap> (accessed on 23 March 2019).
- Pecl, G.T.; Araujo, M.B.; Bell, J.D.; Blanchard, J.; Bonebrake, T.C.; Chen, I.C.; Clark, T.D.; Colwell, R.K.; Danielsen, F.; Evengard, B.; et al. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **2017**, *355*. [CrossRef]
- Fiorentino, D.; Lecours, V.; Brey, T. On the Art of Classification in Spatial Ecology: Fuzziness as an Alternative for Mapping Uncertainty. *Front. Ecol. Evol.* **2018**, *6*. [CrossRef]
- Herkül, K.; Peterson, A.; Paekivi, S. Applying multibeam sonar and mathematical modeling for mapping seabed substrate and biota of offshore shallows. *Estuar. Coast. Shelf Sci.* **2017**, *192*, 57–71. [CrossRef]
- Carlén, I.; Thomas, L.; Carlström, J.; Amundin, M.; Teilmann, J.; Tregenza, N.; Tougaard, J.; Koblit, J.C.; Sveegaard, S.; Wennerberg, D.; et al. Basin-scale distribution of harbour porpoises in the Baltic Sea provides basis for effective conservation actions. *Biol. Conserv.* **2018**, *226*, 42–53. [CrossRef]
- Calder, B.R.; Mayer, L.A. Automatic processing of high-rate, high-density multibeam echosounder data. *Geochem. Geophys. Geosyst.* **2003**, *4*, 1048. [CrossRef]
- Fonseca, L.; Calder, B. Geocoder: An Efficient Backscatter Map Constructor. In Proceedings of the US Hydrographic Conference, San Diego, CA, USA, 29–31 March 2005.
- Freire, F.; Kågesten, G.; Baumgartner, F.; Dahlgren, A. *Bio-Geophysical Survey Methods for Habitat Mapping of Hoburgs Bank 2019*; Geological Survey of Sweden: Uppsala, Sweden, 2019; p. 43.
- Buja, K. Sampling Design Tool for ArcGIS. Available online: <https://coastalscience.noaa.gov/project/sampling-design-tool-arcgis/> (accessed on 23 March 2019).
- Naturvårdsverket. *Inventering av Marina Naturtyper på Utsjöbankar. Rapport 5566*; Naturvårdsverket: Stockholm, Sweden, 2006; p. 79.
- Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.; Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **2018**, *6*, e5518. [CrossRef] [PubMed]

23. Pittman, S.J.; Costa, B.M.; Battista, T.A. Using Lidar Bathymetry and Boosted Regression Trees to Predict the Diversity and Abundance of Fish and Corals. *J. Coast. Res.* **2009**, *2009*, 27–38. [[CrossRef](#)]
24. Costa, B.; Battista, T. The semi-automated classification of acoustic imagery for characterizing coral reef ecosystems. *Int. J. Remote Sens.* **2013**, *34*, 6389–6422. [[CrossRef](#)]
25. Kågesten, G.; Sautter, W.; Edwards, K.A.; Costa, B.M.; Kracker, L.M.; Battista, T.A. *Shallow-Water Benthic Habitats of Northeast Puerto Rico and Culebra Island*; NOAA: Silver Spring, MD, USA, 2015.
26. ENVI. *Feature Extraction Module Version 4.6. User's Guide*. 2008. Available online: http://www.harrisgeospatial.com/portals/0/pdfs/envi/feature_extraction_module.pdf (accessed on 23 March 2019).
27. Stephens, D.; Diesing, M. Towards Quantitative Spatial Models of Seabed Sediment Composition. *PLoS ONE* **2015**, *10*, e0142502. [[CrossRef](#)]
28. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
29. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)]
30. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
31. De'ath, G. Boosted Trees for Ecological Modeling and Prediction. *Ecology* **2007**, *88*, 243–251. [[CrossRef](#)]
32. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
33. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA; Heidelberg, Germany; Dordrecht, The Netherlands; London, UK, 2013; p. 574.
34. Pittman, S.J.; Brown, K.A. Multi-Scale Approach for Predicting Fish Species Distributions across Coral Reef Seascapes. *PLoS ONE* **2011**, *6*, e20583. [[CrossRef](#)] [[PubMed](#)]
35. Leathwick, J.; Moilanen, A.; Francis, M.; Elith, J.; Taylor, P.; Julian, K.; Hastie, T.; Duffy, C. Novel methods for the design and evaluation of marine protected areas in offshore waters. *Conserv. Lett.* **2008**, *1*, 91–102. [[CrossRef](#)]
36. Stamoulis, K.; Delevaux, J.; Williams, I.; Poti, M.; Lecky, J.; Costa, B.; Kendall, M.; Pittman, S.; Donovan, M.; Wedding, L.; et al. Seascape Models Reveal Places to Focus Coastal Fisheries Management. *Ecol. Appl.* **2018**, *28*, 910–925. [[CrossRef](#)]
37. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
38. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Series in Statistics; Springer: New York, NY, USA, 2009.
39. Zaki, M.; Meira, W., Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms*; Cambridge University Press: Cambridge, UK, 2014. [[CrossRef](#)]
40. Hallberg, O.; Nyberg, J.; Elhammer, A.; Erlandsson, C. *Ytsubstratklassning av Maringeologisk Information*; Sveriges Geologiska Undersökning: Uppsala, Sweden, 2010.
41. Piñeiro, G.; Perelman, S.; Guerschman, J.P.; Paruelo, J.M. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Model.* **2008**, *216*, 316–322. [[CrossRef](#)]
42. Ma, Z.; Redmond, R.L. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogramm. Eng. Remote Sens.* **1995**, *61*, 435–439.
43. Lecours, V.; Devillers, R.; Schneider, D.C.; Lucieer, V.L.; Brown, C.J.; Edinger, E.N. Spatial scale and geographic context in benthic habitat mapping: Review and future directions. *Mar. Ecol. Prog. Ser.* **2015**, *535*, 259–284. [[CrossRef](#)]
44. Kendall, M.; Miller, T. The Influence of Thematic and Spatial Resolution on Maps of a Coral Reef Ecosystem. *Mar. Geod.* **2008**, *31*, 75–102. [[CrossRef](#)]
45. Sheehan, E.; Cousens, S.; Nancollas, S.; Stauss, C.; Royle, J.; Attrill, M. Drawing lines at the sand: Evidence for functional vs. visual reef boundaries in temperate Marine Protected Areas. *Mar. Pollut. Bull.* **2013**, *76*, 194–202. [[CrossRef](#)]
46. Beisiegel, K.; Tauber, F.; Gogina, M.; Zettler, M.L.; Darr, A. The potential exceptional role of a small Baltic boulder reef as a solitary habitat in a sea of mud. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2019**, *29*, 321–328. [[CrossRef](#)]

47. Rönnbäck, P.; Kautsky, N.; Pihl, L.; Troell, M.; Söderqvist, T.; Wennhage, H. Ecosystem Goods and Services from Swedish Coastal Habitats: Identification, Valuation, and Implications of Ecosystem Shifts. *AMBI* **2007**, *36*, 534–544. [[CrossRef](#)]
48. Rengstorf, A.; Grehan, A.; Yesson, C.; Brown, C. Towards High-Resolution Habitat Suitability Modeling of Vulnerable Marine Ecosystems in the Deep-Sea: Resolving Terrain Attribute Dependencies. *Mar. Geod.* **2012**, *35*, 343–361. [[CrossRef](#)]
49. Sandman, A.; Wikström, S.; Blomqvist, M.; Kautsky, H.; Isaeus, M. Scale-dependent influence of environmental variables on species distribution: A case study on five coastal benthic species in the Baltic Sea. *Ecography* **2012**, *35*, 001–010.
50. Peters, D.; Okin, G. A toolkit for ecosystem ecologists in the time of Big Science. *Ecosystems* **2016**, *20*, 259–266. [[CrossRef](#)]
51. Kågesten, G. Geological Seafloor Mapping with Backscatter Data from a Multibeam Echo Sounder. Master's Thesis, Department of Earth Sciences Uppsala University, Uppsala, Sweden, 2008; p. 32.
52. Kostylev, V.; Todd, B.; Fader, B.; Courtney, R.; Cameron, G.; Pickrill, R. Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs. *Mar. Ecol. Prog. Ser.* **2001**, *219*, 121–137. [[CrossRef](#)]
53. Brown, C.J.; Blondel, P. Developments in the application of multibeam sonar backscatter for seafloor habitat mapping. *Appl. Acoust.* **2009**, *70*, 1242–1247. [[CrossRef](#)]
54. Beijbom, O.; Edmunds, P.J.; Kline, D.I.; Mitchell, B.G.; Kriegman, D. Automated annotation of coral reef survey images. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1170–1177.
55. González-Rivero, M.; Beijbom, O.; Rodríguez-Ramírez, A.; Holtrop, T.; González-Marrero, Y.; Ganase, A.; Roelfsema, C.; Phinn, S.; Hoegh-Guldberg, O. Scaling up Ecological Measurements of Coral Reefs Using Semi-Automated Field Image Collection and Analysis. *Remote Sens.* **2016**, *8*, 30. [[CrossRef](#)]
56. Dumke, I.; Purser, A.; Marcon, Y.; Nornes, S.M.; Johnsen, G.; Ludvigsen, M.; Søreide, F. Underwater hyperspectral imaging as an in situ taxonomic tool for deep-sea megafauna. *Sci. Rep.* **2018**, *8*, 12860. [[CrossRef](#)]
57. Berthold, T.; Leichter, A.; Rosenhahn, B.; Berkahn, V.; Valerius, J. Seabed sediment classification of side-scan sonar data using convolutional neural networks. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017.
58. Rocchini, D.; Hortal, J.; Lengyel, S.; Lobo, J.M.; Jiménez-Valverde, A.; Ricotta, C.; Bacaro, G.; Chiarucci, A. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Prog. Phys. Geogr.* **2011**, *35*, 211–226. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).