RESEARCH ARTICLE

A multiscale approach to mapping seabed sediments

Benjamin Misiuk[1]*, Vincent Lecours[2], Trevor Bell[1]

1 Department of Geography, Memorial University of Newfoundland, St. John's, Newfoundland, Canada,
2 Department of Fisheries and Aquatic Sciences, School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, United States of America

* bmisiuk@mun.ca

Abstract

Benthic habitat maps, including maps of seabed sediments, have become critical spatial-decision support tools for marine ecological management and conservation. Despite the increasing recognition that environmental variables should be considered at multiple spatial scales, variables used in habitat mapping are often implemented at a single scale. The objective of this study was to evaluate the potential for using environmental variables at multiple scales for modelling and mapping seabed sediments. Sixteen environmental variables were derived from multibeam echosounder data collected near Qikiqtarjuaq, Nunavut, Canada at eight spatial scales ranging from 5 to 275 m, and were tested as predictor variables for modelling seabed sediment distributions. Using grain size data obtained from grab samples, we tested which scales of each predictor variable contributed most to sediment models. Results showed that the default scale was often not the best. Out of 129 potential scale-dependent variables, 11 were selected to model the additive log-ratio of mud and sand at five different scales, and 15 were selected to model the additive log-ratio of gravel and sand, also at five different scales. Boosted Regression Tree models that explained between 46.4 and 56.3% of statistical deviance produced multiscale predictions of mud, sand, and gravel that were correlated with cross-validated test data (Spearman's $\rho_{mud} = 0.77$, $\rho_{sand} = 0.71$, $\rho_{gravel} = 0.58$). Predictions of individual size fractions were classified to produce a map of seabed sediments that is useful for marine spatial planning. Based on the scale-dependence of variables in this study, we concluded that spatial scale consideration is at least as important as variable selection in seabed mapping.

Introduction

Marine ecosystems provide a broad range of services to humans, including food, extractive resources, and cultural identity [1,2]. These systems are now being threatened and profoundly impacted on local and global scales by a suite of anthropogenic stressors such as climate change, overfishing, and pollution [3,4]. As pressures on marine systems intensify, there is an urgent need to monitor and mitigate impacts to ensure ecosystem viability and sustainable ecosystem services. Despite the importance of marine ecosystems to human well-being, and

Competing interests: The authors have declared
that no competing interests exist.

the immediate threats they face, we often lack the necessary information to make informed
management decisions.

Seabed maps provide necessary information for a number of conservation and management applications. Habitat maps in particular are used to monitor anthropogenic impacts, to support government marine spatial planning, for marine protected area design, to generate knowledge about ecosystems and geology, and to assess seabed resources for economic and management purposes [5,6]. Benthic habitat mapping is defined broadly as "the use of spatially continuous data sets to represent and predict biological patterns on the seafloor" [7]. Habitat mapping can be applied to species, communities, or physical features of interest, but a fundamental requirement for generating useful maps in all cases is the availability of the appropriate high quality environmental spatial data.

Though benthic habitats are determined by a range of environmental variables, McArthur et al. [8] identified seabed substrate characteristics as the strongest independent predictors of benthic habitats. Sediment grain size is a particularly important substrate characteristic that can constrain the distribution of benthic habitats [9,10]. Along with other habitat-defining parameters, distribution maps of sediment grain size can thus serve as management tools for predicting the distribution of individual species and assemblages [11,12]. The increasing availability of accurate marine spatial data has improved our ability to map the distribution of seabed sediments. For instance, primary data collected from multibeam echosounders (MBES)–bathymetry (i.e. water depth) and backscatter (i.e. acoustic reflectivity)–and their derivatives can be used to delineate and model sediment grain size over large areas (thousands of km) at a high spatial resolution (metres) when coupled with ground-truth substrate samples (e.g. [12–14]). Quantitative predictions of sediment grain size can be used on their own as continuous explanatory variables in further analyses or can be classified for interpretation or use as categorical variables [15]. This quantitative predictive approach represents a departure from subjective expert-based interpretation towards more objective, repeatable methods [15,16].

Recent biological and geological modeling approaches have relied heavily on bathymetry-derived terrain variables (e.g. slope and rugosity) and backscatter-derived variables (e.g. hardness and heterogeneity) to predict the response of organisms (e.g. [17,18]), habitats (e.g. [19–21]), or sediment properties such as grain size or the presence of rock (e.g. [13,22]). Terrain variables can act as surrogates for patterns and processes on the seabed (e.g. seabed morphology, current dynamics, relative position) that may influence the distribution of sediments or biota [23]. While these processes are scale dependent (e.g. [24]), terrain variables are most often derived at the resolution of the primary data layers (bathymetry and backscatter) by default. The resolution of the primary data is selected by the data analyst, who must consider the specifications of the MBES system, the specific survey, and the quality of the data. Terrain variables are usually raster data products that are calculated using "focal" or "neighborhood" cell analyses on the primary layers. Deriving terrain variables at the resolution of the primary data imposes a spatial scale on them that may not be appropriate for representing the processes of interest [25–27]. To avoid the arbitrary selection of data scale it may be desirable to test at which scales explanatory variables have the greatest influence on the response variable. A solution that has been proposed in recent years is to move towards multiple scale or multiscale analyses [28]. *Multiple scale* analyses are those that consider data at multiple successive scales, and *multiscale* analyses are those that integrate information from multiple scales simultaneously [26]. Several terrestrial (e.g. [29–31]) and marine (e.g. [32–34]) studies have demonstrated that the use of data at different scales can affect results and interpretations. Since different environmental processes operate at different spatial scales [35], the adoption of a multiscale approach ensures that the relevant scale-dependent patterns and processes are captured [28].

The overall objective of this study was to evaluate the potential of multiscale approaches for predicting the distribution of sediment grain size for use in habitat mapping and marine spatial planning. Using a case study approach, we first examined whether using the default data resolution of terrain variables was optimal for predicting the distribution of sediment grain size. We then determined which terrain variables at which spatial scales most strongly influenced the distribution of sediment grain size. Finally, we applied this knowledge to predict distributions of mud, sand, and gravel fractions at optimal spatial scales in the study area.

Data and methods

Setting

This study was conducted in the coastal zone near the hamlet of Qikiqtarjuaq on the east-central coast of Baffin Island, Nunavut, Canada (Fig 1A). The surrounding terrain is mountainous, hosting upland ice caps and alpine glaciers, shaped by repeated glacial cycles during the Quaternary [36]. For example, deep valleys and fjords (> 300 m deep, such as the one south of Qikiqtarjuaq; Fig 1B) channeled glacial ice flowing from inland source areas onto the continental shelf. Coarse-grained glacial deposits mantle the coastal terrain and extend offshore. The north-south channel is relatively shallow (60–70 m deep), and is shallowest opposite Qikiqtarjuaq (16 m deep), where a tombolo may have joined the two islands during the postglacial sea-level lowstand [37]. Surface currents of 0.5–0.8 m/s in the channel winnow the local seabed [38].
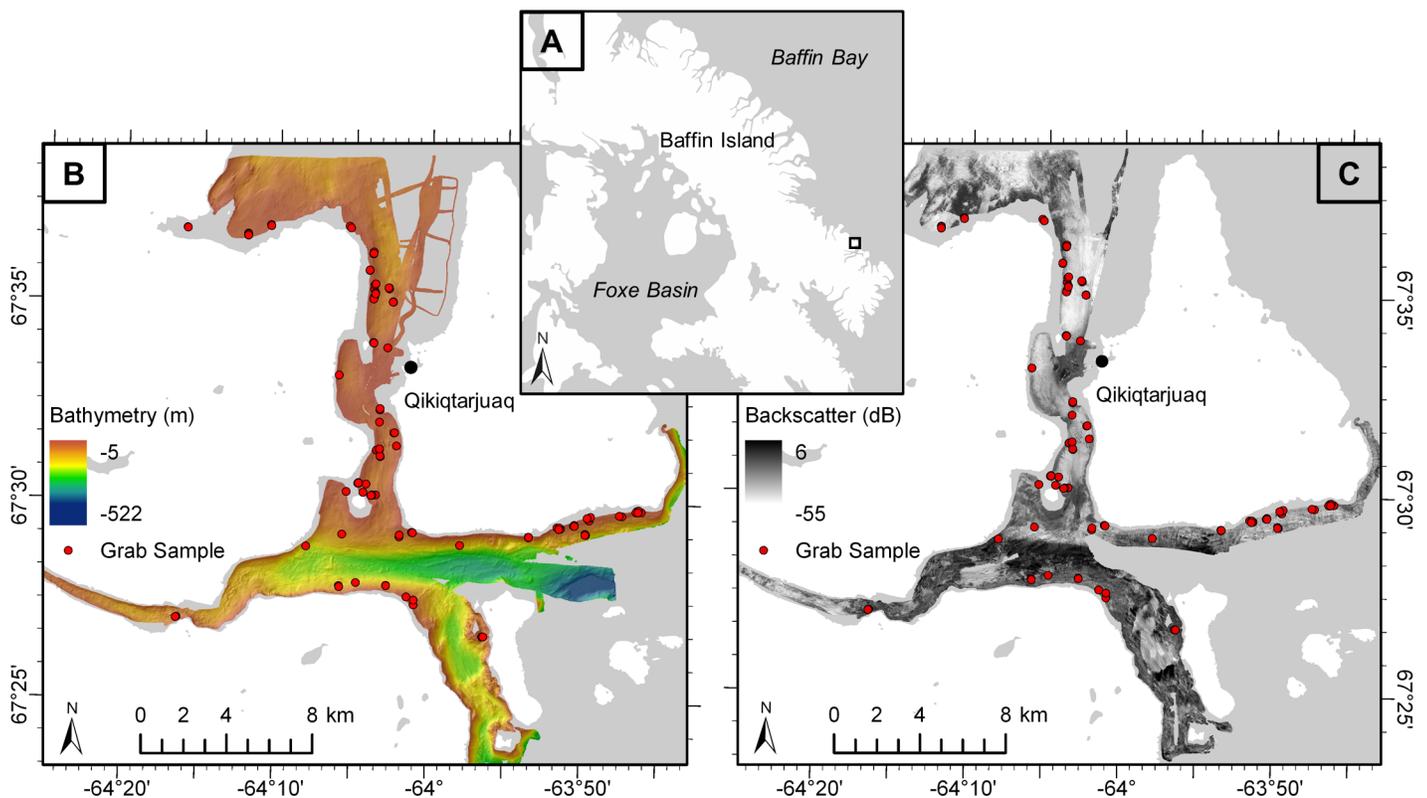
Fig 1. Study site. (A) Location of study site on east Baffin Island, NU, Canada. (B) Bathymetry data collected via MBES, with grab sample sites in red. (C) Backscatter data collected via MBES, with grab sample sites in red. (A) was modified from the USGS National Map, available under the public domain; basemap in (B) and (C) was obtained from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence–Canada.

https://doi.org/10.1371/journal.pone.0193647.g001

Primary data

Bathymetry and backscatter data were collected using a MBES over the course of five years in the coastal zone near Qikiqtarjuaq (Fig 1). The surveyed region can be morphologically separated into two broad areas: 1) the relatively shallow channel, oriented north-south, separating Qikiqtarjuaq from Baffin Island; and 2) a deeper fjord basin oriented east-west located south of Qikiqtarjuaq (Fig 1B). The *CCGS Amundsen* collected data in the deepest area (> 600 m) using a Kongsberg EM300 30 kHz (variable beam width) echosounder in 2007 [39]. The Government of Nunavut scientific research vessel *RV Nuliajuk* collected MBES data using a Kongsberg EM3002 300 kHz (1.5˚ x 1.5˚ beam width) echosounder in 2012–2013 and a Kongsberg EM2040C 200–400 kHz (1˚ x 1˚ beam width) echosounder in 2014–2015 [40,41]. Datasets from the different survey years were harmonized and used as single, continuous layers for analyses. Details of how datasets from multiple MBES systems were harmonized are provided in supporting information (S2 File). Depths mapped by the *RV Nuliajuk* were between 5 and 350 m, and up to 522 m by the *Amundsen*. The mapped area was approximately 135 km^2. Because the acoustic response of the seabed is dependent on MBES operating frequency, the 30 kHz *Amundsen* backscatter dataset, which differed substantially from the 300 and 200–400 kHz, was omitted, yielding an area of 112 km^2 analyzed.

In addition to MBES data, 109 sediment grab samples were collected between 2014 and 2015 to measure the grain size of sediment (Fig 1B and 1C). Seabed sampling that impacted benthic fauna was permitted by Fisheries and Oceans Canada in 2014 and 2015 (license no. S-14/15-1041-NU and S-15/16-1010-NU-A1). Work in this region was conducted in collaboration with the Government of Nunavut, Department of Environment, Fisheries and Sealing Division in 2014, and was further permitted in 2015 by the Nunavut Research Institute (license no. 01 025 15N-M). Sample sites in 2014 were targeted to cover a previously completed shallow underwater image survey conducted between 0 and 40 m water depth. Data from this study were also appropriate for use in the current study. Sample sites in 2015 were selected randomly in the area of the MBES survey, stratified by bathymetry (up to 200 m depth), bathymetry-derived seabed slope, and backscatter, in order to obtain sediment samples at a range of these values. All sediment samples were collected using an 8.2 L Wildco® Ponar Grab.

Secondary data

While many terrain attributes can be derived from bathymetric data to describe seabed morphology, Lecours et al. [42] recommended using a specific combination of six attributes that together capture most of the topographic structure of a surface. The Qikiqtarjuaq bathymetric data were used with the Terrain Attribute Selection for Spatial Ecology (TASSE) toolbox [43] in ESRI ArcGIS v10.3.1 to compute values for those six terrain attributes using a default 3 x 3 window of analysis. The six terrain attributes include eastness and northness (unitless sin- and cosine-transformed measures of orientation or aspect), relative difference to the mean value (RDMV; a unitless measure of topographic position), standard deviation (a measure of terrain variability; metres), slope angle (degrees), and local mean (metres water depth). Local mean was strongly correlated with the input bathymetry layer and thus was not included in further analyses.

In addition to the attributes identified by Lecours et al. [42], we derived a set of variables that may apply specifically to the distribution of sediment grain size. Seabed curvature (degrees per metre) influences current regimes and can be used to identify landform boundaries [44], while measures of relative seabed position such as benthic position index (BPI; metres) identify topographic highs and lows that can affect bottom currents and sediment transport [45,46]. Rugosity (the ratio of surface area to planar area) and the vector ruggedness index (the

variability in surface orientation) are both unitless measures of terrain variability, which can describe seabed topography and substrate at appropriate scales (e.g. rough/rocky, flat/soft; [44]). Curvature measures were generated using the "Curvature" tool in ESRI ArcGIS v.10.3.1; BPI at broad and fine scales (scale factors of 100 and 250 metres, respectively; [47]), rugosity, and ruggedness were derived using the Benthic Terrain Modeler (BTM) toolbox [48]. Back-scatter heterogeneity (hereafter Δbackscatter; dB), which is useful for differentiating coarse and fine substrates [49], was derived from the primary backscatter layer using the same func-tion applied to calculate surface roughness (i.e., obtaining "minimum" and "maximum" 3 x 3 pixel neighborhood layers, then subtracting "maximum"—"minimum"; cf. "backscatter rough-ness"; [49]). ΔBackscatter was calculated using the "Focal Statistics" and "Raster Calculator" tools in ESRI ArcGIS v10.3.1. Last, distance from the coast (metres), a potential driver of grain size distribution [15], was calculated from a coastal polygon layer generated by Cowan [37] using Euclidean distance.

Data artefacts that were not visible in the bathymetry layer became apparent in some terrain attributes. These occurred most commonly at the interface between MBES datasets collected from different years, or near the depth limits of the MBES systems. Because terrain variable artefacts can affect habitat mapping results [50] we decided to exclude these areas from the analysis, resulting in several narrow data gaps (see supporting information; Fig D in S2 File).

All variables, except for distance from the coast (calculated independent of MBES data), were calculated at multiple scales by first deriving them from the original bathymetric and backscatter data at 5-m resolution then averaging them over increasing windows of analysis [26]. Variables were averaged over 3 x 3-, 5 x 5-, 9 x 9-, 13 x 13-, 21 x 21-, 35 x 35-, and 55 x 55-pixel neighborhoods using the "Focal Statistics" tool in ESRI ArcGIS v10.3.1. These neigh-borhoods followed the Fibonacci sequence (rounded up when even)—a convenient number series of increasing interval size [44,51]. This resulted in 129 potential variables for predicting the response of sediment grain size, at eight different spatial scales (Table 1).

Sediment grain size distribution modelling

Response variables. Ninety-eight grab samples were from locations within the MBES sur-vey, and were used to model the distribution of sediment grain size. Following recommenda-tions by Aitchison [52] and Stephens and Diesing [12], mud, sand, and gravel fractions were treated as compositional data that sum to 1 for each sample, allowing grain size classes to be considered concurrently. These data were transformed to an additive log-ratio (ALR) scale for use in modeling, resulting in values that are a ratio of two of the grain size classes, which can be back-transformed to yield predictions of mud, sand, and gravel after modeling [12]:

$$ALR_{ms} = \log(\frac{mud}{sand})$$

$$ALR_{gs} = \log(\frac{gravel}{sand})$$

Some sediment samples had mud or gravel fractions equal to zero, which may be due to one of the following: a) sample sites lacking sediment of a given size class; b) recovering amounts of a class that were too small to measure; or c) not retaining all size classes when sub-sampling sediment grabs (occurred occasionally in low-gravel areas). Since samples with zero values still provide valuable information on grain size composition, including those data points in the analysis was important. Since the log of zero is undefined, a replacement method was necessary for zero values to facilitate the grain size data transformation. The "simple

Table 1. Multiple scale explanatory variables selected for modeling sediment grain size.

Variable		Scales (m)	Calculation Method	Method Source
Primary	**Secondary**			
Bathymetry		5,15,25,45,65,105,175,275	-	-
	Eastness	5,15,25,45,65,105,175,275	TASSE	[43]
	Northness	5,15,25,45,65,105,175,275	TASSE	[43]
	RDMV	5,15,25,45,65,105,175,275	TASSE	[43]
	Standard Deviation	5,15,25,45,65,105,175,275	TASSE	[43]
	Slope	5,15,25,45,65,105,175,275	TASSE	[43]
	Fine BPI*	5,15,25,45,65,105,175,275	BTM	[48]
	Broad BPI*	5,15,25,45,65,105,175,275	BTM	[48]
	Curvature	5,15,25,45,65,105,175,275	Curvature Tool	-
	Profile Curvature	5,15,25,45,65,105,175,275	Curvature Tool	-
	Plan Curvature	5,15,25,45,65,105,175,275	Curvature Tool	-
	Area	5,15,25,45,65,105,175,275	BTM	[48]
	Rugosity	5,15,25,45,65,105,175,275	BTM	[48]
	Ruggedness	5,15,25,45,65,105,175,275	BTM	[48]
Backscatter		5,15,25,45,65,105,175,275	-	-
	ΔBackscatter	5,15,25,45,65,105,175,275	Focal Statistics	[49]
Distance from Coast		-	Euclidean Distance	-

See text for explanation and discussion of individual variables and calculation methods.

*Fine scale BPI calculated with inner radius of 1 and outer radius of 20; broad scale BPI calculated with inner radius of 15 and outer radius of 50. Scale factors of 100 m (fine BPI) and 250 m (broad BPI) averaged over the increasing window sizes result in scales of 100, 300, 500, 900, 1300, 2100, 3500, and 5500 m for fine; 250, 750, 1250, 2250, 3250, 5250, 8750, and 13750 m for broad.

https://doi.org/10.1371/journal.pone.0193647.t001

replacement" method (reviewed in [53]), which replaces zero values with measurements less than the minimum recorded for a given size class, was used in this study. Following Lark et al. [54] and Diesing et al. [13], we replaced the observed zero values with values below the level of precision of our scientific equipment (1×10^{-4}), maintaining the possibility that the size class did occur in trace amounts.

Statistical modeling. Boosted Regression Trees (BRT; [55,56]), a popular tree-based machine learning (ML) technique for ecological modeling [57], was applied to model the responses of ALR_{ms} and ALR_{gs} to the different explanatory variables at different scales (see Table 1). This non-parametric technique can accommodate large numbers of non-linear categorical and numerical explanatory variables simultaneously, while automatically modeling interaction between predictors [57,58]. BRTs were chosen because they commonly outperform other quantitative modeling methods [10,59–61], ignore unimportant variables, are insensitive to outliers, and tend to avoid over-fitting [55–57,59]. BRTs can provide plots of partial response, variable interaction, and measures of variable contribution, which allow the user to explore mechanistic relationships between explanatory and response variables [15,62]. Details on how BRTs work have been detailed by Friedman et al. [55], and with ecological examples by Elith et al. [62].

While BRTs can accommodate large numbers of explanatory variables and tend to ignore those that are not important, it is still desirable to limit the number of variables in a model to facilitate understanding of the scale-dependent mechanisms that control grain size distribution [15,63]. BRTs can also be used as an exploratory tool to perform this function, as they return information on the importance of explanatory variables in predicting the response [62,64]. Thus, we first fitted models of both ALR_{ms} and ALR_{gs} iteratively for each individual variable,

but at all scales (Fig 2, Step 4). These preliminary models provided two useful pieces of information: 1) a measure of which scales best explain the distribution of grain size for each variable, and 2) a ranked order of how well each variable, considered at multiple scales, explains the distribution of grain size. Percent relative importance was used to determine at which scales each variable performed best. Decision trees that comprised the BRT models were grown based on reducing the maximum amount of deviance in each model, thus models with the least residual deviance after fitting were considered best and were ranked as such. This methodology was applied separately for ALR_{ms} and ALR_{gs}, resulting in two sets of both ranked variables and information on the best-performing scales for each variable. All modeling was conducted in R v3.2.3, with code modified from that provided by Elith and Leathwick [65] and Ridgeway [66].

Fitting individual models of ALR_{ms} and ALR_{gs} for all scales of a given variable provided information on the relative contribution of each scale, and only scales that contributed $\geq 10\%$ to a given model were subsequently considered. Correlation between scales of a given variable was measured using Spearman's rank correlation coefficient (Fig 2, Step 5). Correlated scales of a given variable were removed if $\rho \geq 0.7$, giving preference to those that contributed more to a given model [22,67]. Once correlated scales of each variable were removed, correlation was assessed at all scales between different variables. Variables with correlation $\rho \geq 0.7$ were removed, giving preference to variables that had the least residual deviance after model fitting (Fig 2, Step 6). This methodology allowed for the selection of the most important, non-correlated scales of each variable for inclusion in modeling ALR_{ms} and ALR_{gs}. Models were then fitted for both ALR_{ms} and ALR_{gs} using the remaining non-correlated variables (Fig 2, Step 7).

ALR_{ms} and ALR_{gs} were back-transformed to produce predictions of mud, sand, and gravel fractions, individually (Fig 2, Step 8). Back-transformation was performed using the following functions in the "Raster Calculator" tool in ESRI ArcGIS v.10.3.1 [15]:

$$mud = \frac{Exp10(ALR_{ms})}{Exp10(ALR_{ms}) + Exp10(ALR_{gs}) + 1}$$

$$gravel = \frac{Exp10(ALR_{gs})}{Exp10(ALR_{ms}) + Exp10(ALR_{gs}) + 1}$$

$$sand = 1 - (mud + gravel)$$

These functions resulted in continuous predictions of mud, sand, and gravel fractions summing to 1 for all locations with MBES data. Following Stephens and Diesing [12], mud, sand, and gravel fractions were classified according to Long [68] to produce a single map of sediment distribution (Fig 2, Step 9). Long's scheme combines these fractions into the classes "mud and sandy mud", "sand and muddy sand", "mixed sediment" and "coarse sediment", allowing them to be represented simultaneously in a single map.

Model evaluation. Elith and Leathwick's [65] extension to the Generalized Boosted Regression Models (gbm) package in R [66] implements an *n*-fold cross-validation (CV) procedure for BRT model building. CV partitions the response data into *n* folds, *n*-1 of which are used to train a model that is evaluated using the excluded partition. This is repeated *n* times, and the results are averaged to produce the final model and evaluation statistics. CV within Elith and Leathwick's [65] code calculates average percent deviance explained over 10 model folds by default, which is useful for evaluating model fit [62,67]. An additional manual 10-fold CV was conducted to measure the average Spearman's rank correlation between back-transformed predictions (i.e., mud, sand, and gravel fractions) and observed values of the withheld

1. Derive variables (n = 16) that explain grain size distribution

16 variables

2. Average variables over eight spatial scales

5, 15, 25, 45, 65, 105, 175, 275 m

3. Model the response of ALR_{ms} and ALR_{gs} separately for all variables, at eight scales

(ALR_{ms}~8 scales)x16   (ALR_{gs}~8 scales)x16

4. Rank each scale of every variable on model contribution for ALR_{ms} and ALR_{gs}; remove scales that contribute < 10% (in gray)

e.g.
1. slope275
2. slope175
3. slope65
4. slope45   x 16
5. slope5
6. slope105
7. slope15
8. slope25

e.g.
1. rugg275
2. rugg5
3. rugg45
4. rugg175   x 16
5. rugg25
6. rugg15
7. rugg105
8. rugg65

5. Remove correlated scales if $\rho \geq 0.70$ (in gray); retain those of highest rank

e.g.   275
slope275   175 ρ=0.97
           65 ρ=0.87
           45 ρ=0.85

e.g.   275
rugg275   5 ρ=0.74

6. Remove correlated variables if $\rho \geq 0.70$ (in gray); retain those of highest rank

e.g.   curve5 ρ=-0.18
slope275 curve65 ρ=-0.22
          sd275 ρ=0.99
          bathy5 ρ=-0.35
          etc…

e.g.   b.bpi175 ρ=0.01
rugg275   east45 ρ=0.20
          s.area ρ=0.99
          slope275 ρ=0.97
          etc…

7. Model ALR_{ms} and ALR_{gs} using remaining non-correlated variables, at non-correlated scales

ALR_{ms}~slope275+curve5+ curve65+…

ALR_{gs}~rugg275+b.bpi175+ east45+…

8. Back-transform ALR_{ms} and ALR_{gs} model predictions to predict mud, sand, and gravel grain size fractions

mud, sand, gravel

9. Classify mud, sand, and gravel predictions according to Long (2006)

sediment classes

Fig 2. Sediment grain size modeling workflow. Procedure for selecting explanatory variables at multiple scales to model the response of ALR_{ms} and ALR_{gs} and predict the distribution of grain size classes.

https://doi.org/10.1371/journal.pone.0193647.g002

data partition. Spearman's correlation coefficient provides a non-parametric ranked measure of monotone relationship between predictions and observed values, providing an indication of the model's ability to predict grain size at new locations [69,70].

Results

Variable and scale selection

ALR_{ms} and ALR_{gs} responses to the explanatory variables tested demonstrated that the default 5-m scale of analysis was not necessarily the most appropriate scale for all variables. Out of 51 scale-specific variables considered for modeling ALR_{ms} (i.e. those that contributed $\geq 10\%$ to their respective single-variable model during testing) only 10 were at the default 5-m scale (Fig 3). Similarly, out of 55 scale-specific variables considered for modeling ALR_{gs}, 11 were at the default 5-m scale. All scales tested were selected for modeling at least once, yet 5, 175, and 275 m were most common for both ALR_{ms} and ALR_{gs} (Fig 3).

Eleven variables were ultimately selected to model the response of ALR_{ms} (mud and sand) at five different scales (Fig 4): broad BPI (175 m), eastness (5 m), backscatter (45 m), plan curvature (5 m), rugosity (275 m), northness (275 m), Δbackscatter (5 m), Δbackscatter (105 m), distance from the coast, plan curvature (105 m), and plan curvature (275 m). ALR_{ms} was most strongly influenced by broad BPI at 175-m scale, the eastness component of aspect at 5-m scale, and backscatter at 45-m scale, together which contributed over 73% to model building. Partial dependence plots showed a strong negative trend between broad BPI and ALR_{ms}. The ratio of mud to sand was generally higher on west-facing slopes and in areas of low backscatter response.
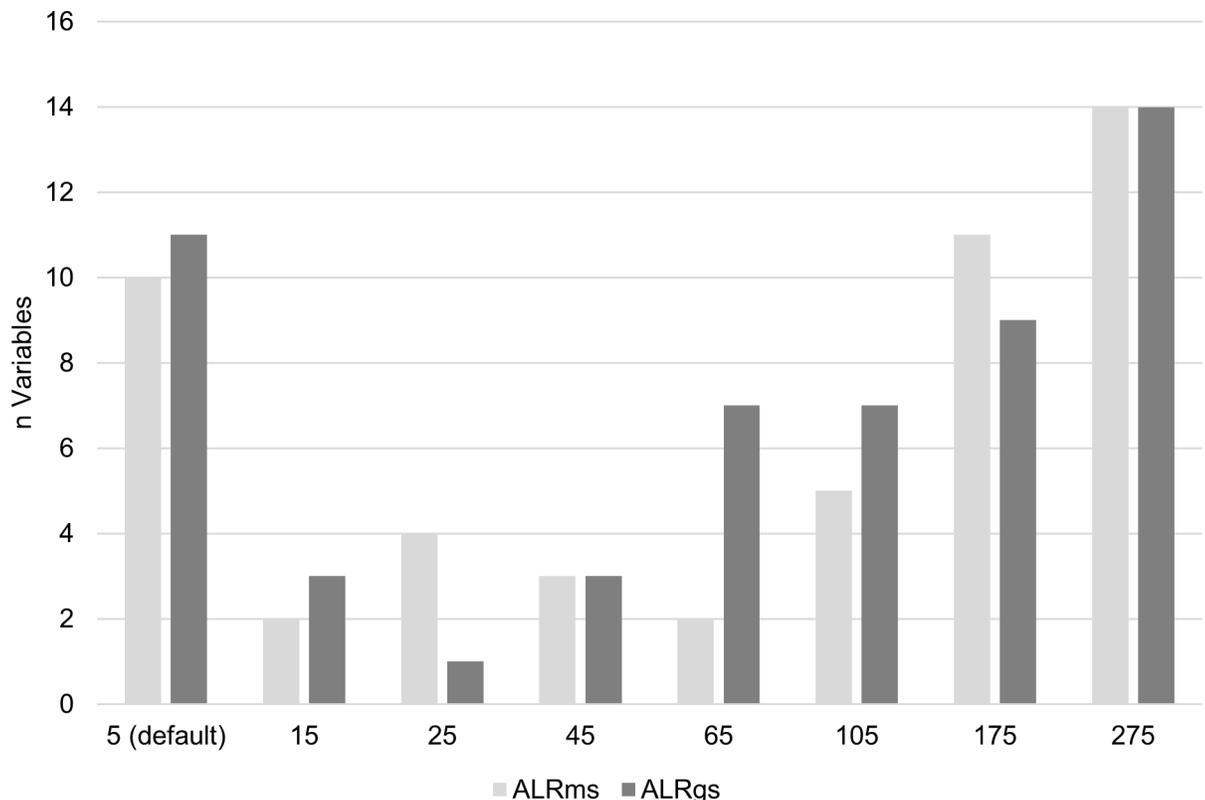
Fig 3. Scales selected for modeling. Number of times each scale contributed ≥ 10% to test models, and was selected for modeling.
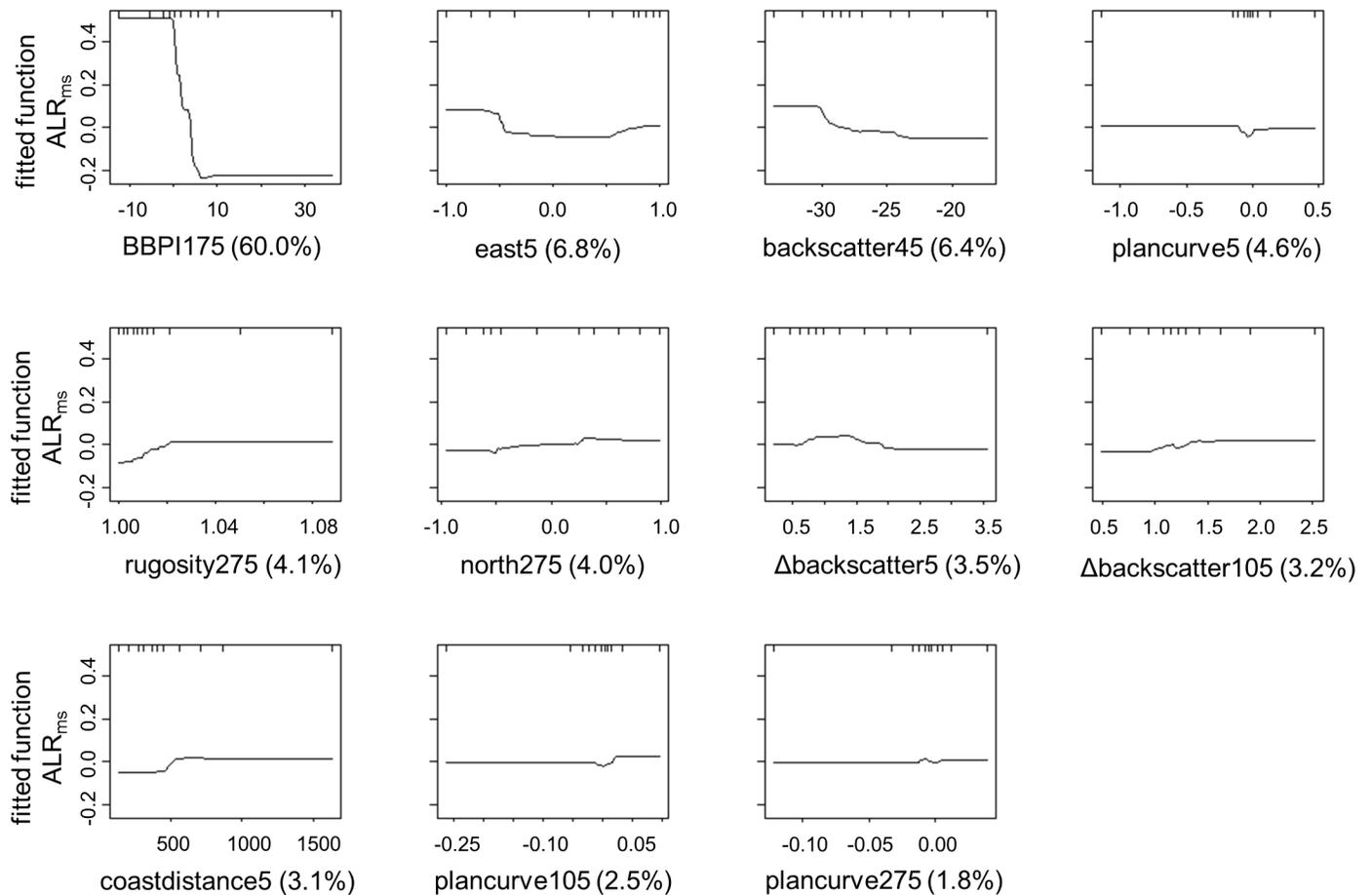
Fig 4. Variables selected to model ALR$_{ms}$. Partial dependence plots for multiple scale variables selected to model ALR$_{ms}$ with percent contribution to the model and data deciles on the upper x-axis.

Fifteen variables were selected to model ALR$_{gs}$ (gravel and sand) at five different scales (Fig 5): backscatter (175 m), bathymetry (5 m), eastness (45 m), Δbackscatter (275 m), plan curvature (65 m), surface area (275 m), northness (275 m), plan curvature (175 m), curvature (5 m), profile curvature (105 m), curvature (65 m), plan curvature (275 m), Δbackscatter (5 m), distance from the coast, and plan curvature (105 m). ALR$_{gs}$ was most strongly influenced by backscatter at 175-m scale, bathymetry at 5-m scale, and eastness at 45-m scale, together which contributed over 68% to model building. Partial dependence plots suggested a positive trend between backscatter response and ALR$_{gs}$, a decrease of ALR$_{gs}$ shallower than ~60 m depth, and a lower ALR$_{gs}$ on east-facing slopes (Fig 5).

Prediction

Back-transformed additive log-ratios produced continuous predictions of mud, sand, and gravel fractions over the area of environmental data coverage (Fig 6). Sand was the dominant size fraction, comprising between 50.9 and 98.8% of sediment composition, with a mean of 82.3%. Sand was most abundant in the north-south oriented channel, and less so in the deeper fjord basin to the south (Fig 6B). Gravel comprised between 0 and 45.1% of sediment composition with a mean of 10.8% and was most abundant in the deepest waters to the south, where the east-west oriented fjord empties into Baffin Bay (Fig 6C). Mud was the least abundant size
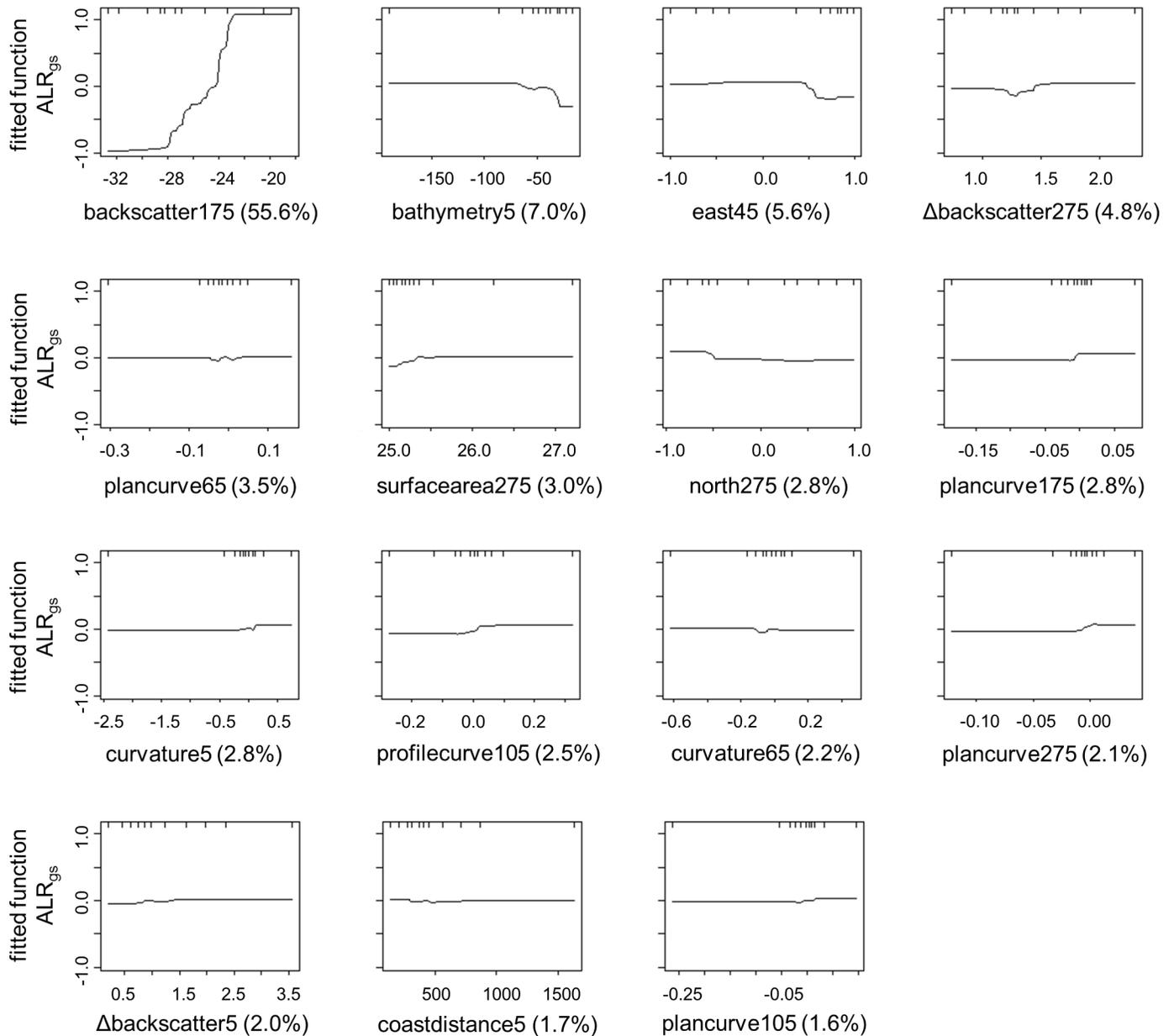
Fig 5. Variables selected to model ALR$_{gs}$. Partial dependence plots for multiple scale variables selected to model ALR$_{gs}$, with percent contribution to the model and data deciles on the upper x-axis.

fraction, comprising between 0.8 and 21.6% of sediment composition with a mean of 6.9%. Mud was generally more abundant farther from shore, and was most common in patches north of the community of Qikiqtarjuaq (Fig 6A).

Model evaluation

On average over 10 folds, 56.3 and 46.4% of the statistical deviance was explained by the ALR$_{ms}$ and ALR$_{gs}$ models, respectively. Mud predictions had the highest average Spearman's rank correlation (ρ_{mud} = 0.772), followed by sand (ρ_{sand} = 0.712) and gravel (ρ_{gravel} = 0.578). Ten-fold CV also produced a map of standard deviation for each grain size fraction, indicating
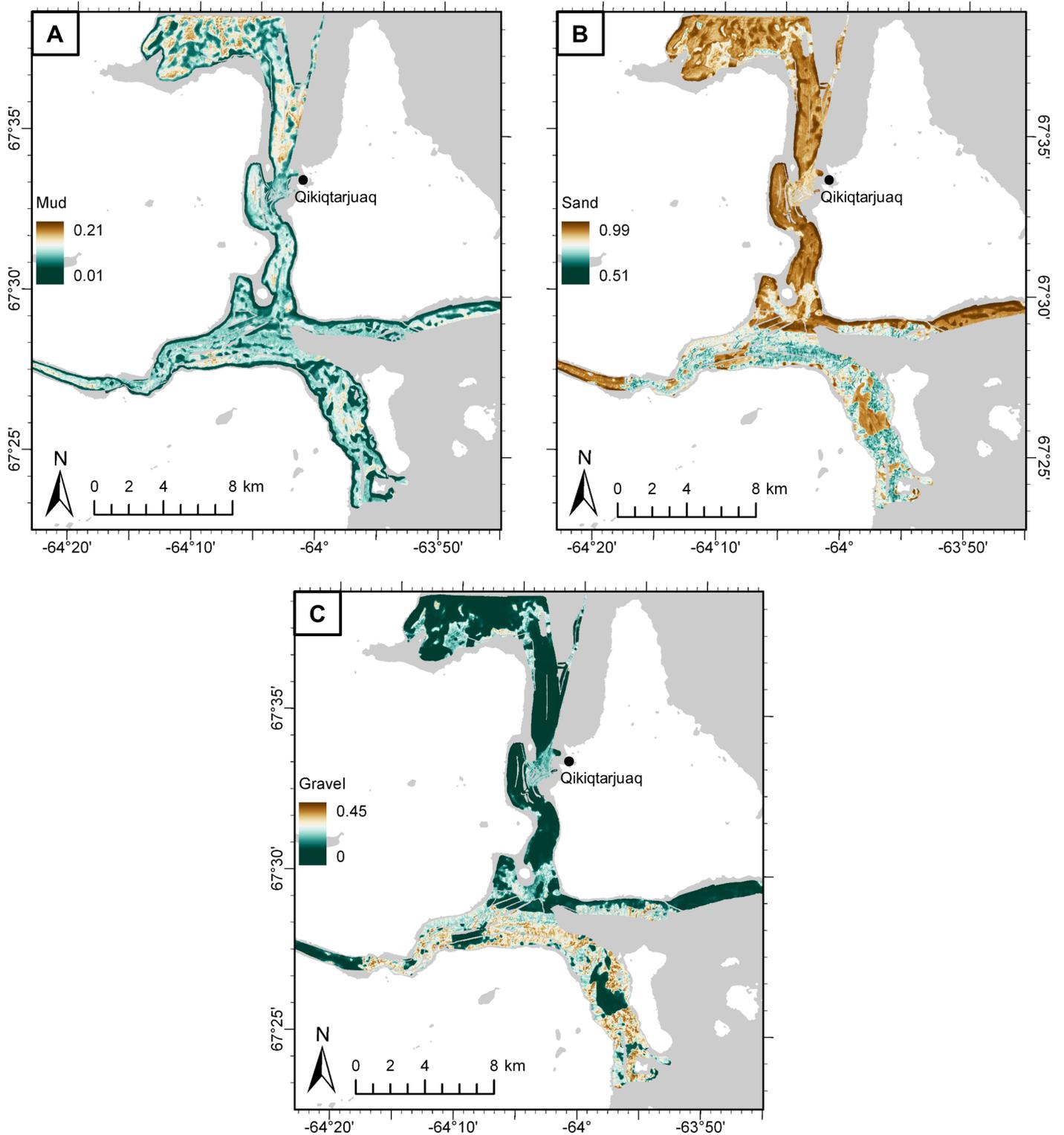
Fig 6. Potential grain size distribution of seabed substrate. Predicted proportions of A) mud, B) sand, and C) gravel fractions. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence–Canada.

https://doi.org/10.1371/journal.pone.0193647.g006

areas of high and low model consensus (Fig 7). Mud predictions had the lowest mean standard deviation (σ_{mud} = 0.007), with the greatest model consensus in areas of low mud proportion, near the coasts (Fig 7A). Sand and gravel had similar mean standard deviations (σ_{sand} = 0.021, σ_{gravel} = 0.019) that were also distributed similarly in space (Fig 7B and 7C). Standard deviations for sand and gravel were highest in areas of high predicted gravel proportion and lowest in sandier areas.

Classification

Following Stephens and Diesing [12], predictive maps of mud, sand, and gravel distribution were classified according to Long's [68] modification of Folk's [71] original scheme in order to facilitate interpretation and application as a management tool (Fig 8). This system was developed for use in the European Nature Information System (EUNIS) habitat classification, but we found it convenient for application to this study–it is simple and easy to interpret. A natural neighbor interpolation was applied to the individual grain size predictions in ESRI ArcGIS v.10.3.1 prior to classification to fill data gaps that were removed due to acoustic artefacts (see supporting information; Fig D in S2 File). The resulting map shows that the north-south oriented channel is composed primarily of "sand and muddy sand" except for the area proximal to Qikiqtarjuaq, which is "coarse". "Sand and muddy sand" were predicted north of Qikiqtarjuaq with "coarse" patches at scales from 100s of metres to kilometres, with small patches of "mixed" sediment at scales from 10s to 100s of metres. "Sand and muddy sand" were predicted directly south of Qikiqtarjuaq, eventually coarsening farther south where the north-south oriented channel meets the east-west oriented fjord. "Coarse" patches in this area were predicted to occur over the scale of kilometres, with finer-scale patches of "mixed" sediment occurring over 10s to 100s of metres. "Coarse" and "mixed" substrates were predicted in the deep, high-relief, southernmost portion of the study area, with the exception of a ~3 x 1 km "sand and muddy sand"-filled basin in the middle of the channel. "Mud and sandy mud" occurred in such small quantities that the class was excluded from the map.

Discussion

Scale selection

The variable and scale selection process (Fig 2) demonstrated that the default scale of explanatory terrain variables was not necessarily the best option for modeling the distribution of sediment grain size, confirming previous observations that terrain variables are scale-dependent (e.g. [26,32,35]). Our methodology sought to identify the optimal scales at which explanatory variables influenced sediment grain size distributions. The default was sometimes identified as the best scale for an explanatory variable, but this was not always the case, and found to be not applicable for most variables. This has broader implications for management and habitat mapping efforts that use such modeling predictions. Since the use of the same environmental variables at different scales will ultimately produce different modeling results [44,45,64,72,73], the effects of scale selection will be propagated throughout the modelling process, and will impact the results of these efforts. For instance, ALR_{gs} models indicated that variables representing influences of bottom currents (e.g. aspect, curvature) were most appropriate at scales between 45 and 275 m (Fig 5). Variables were selected at these scales and used in the models to produce maps of grain size distribution. Map products would differ had they included these variables at the default resolution.

The broadest scales (175 and 275 m) were commonly selected for predictor variables, but fine and intermediate scales also contributed to the final models of ALR_{ms} and ALR_{gs}. During step 6 of the modeling process (removing correlated variables; Fig 2) more than half of the
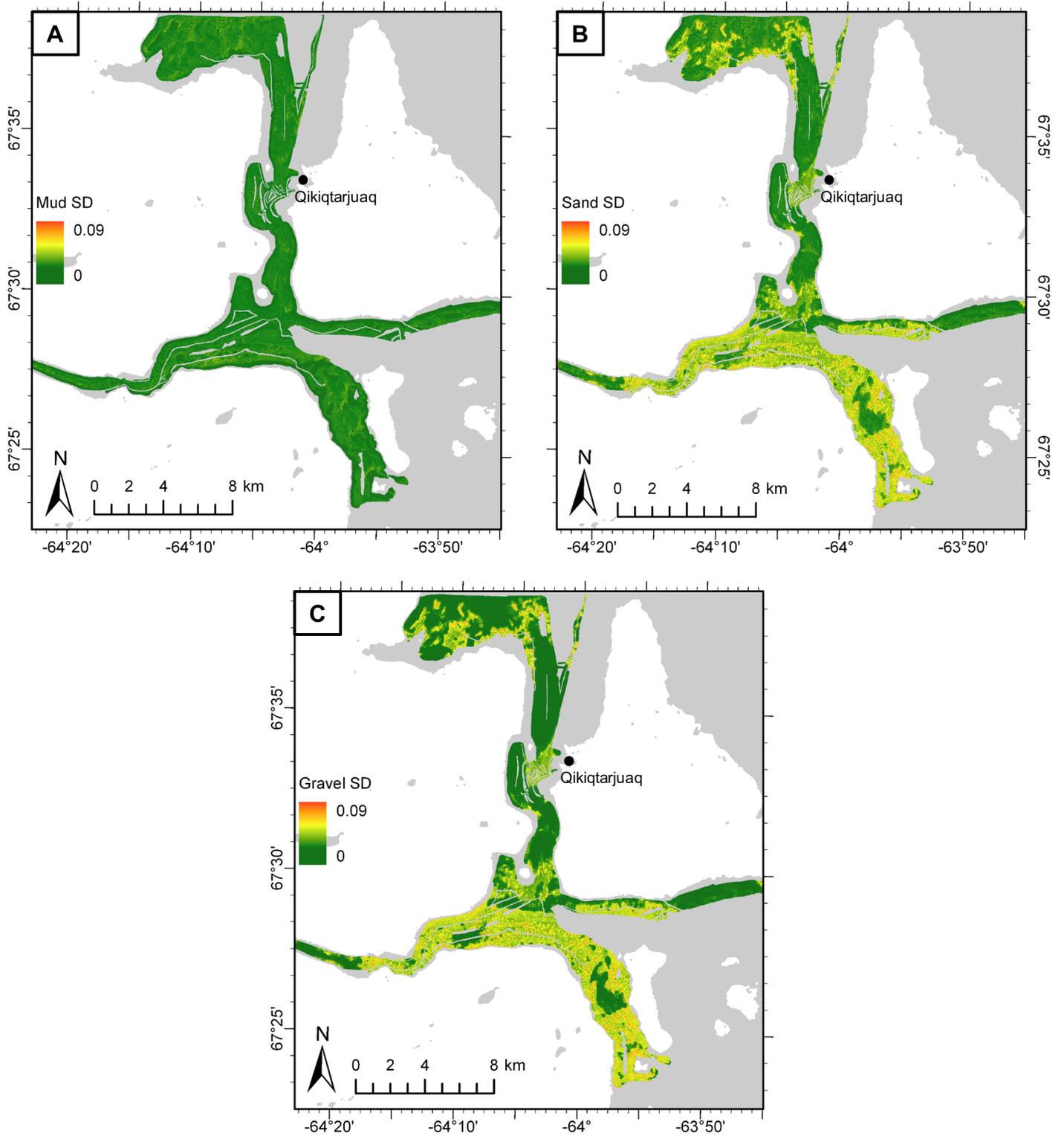
Fig 7. Grain size predictive uncertainty. Ten-fold CV standard deviations (SD) for A) mud, B) sand, and C) gravel predictions. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence–Canada.

https://doi.org/10.1371/journal.pone.0193647.g007
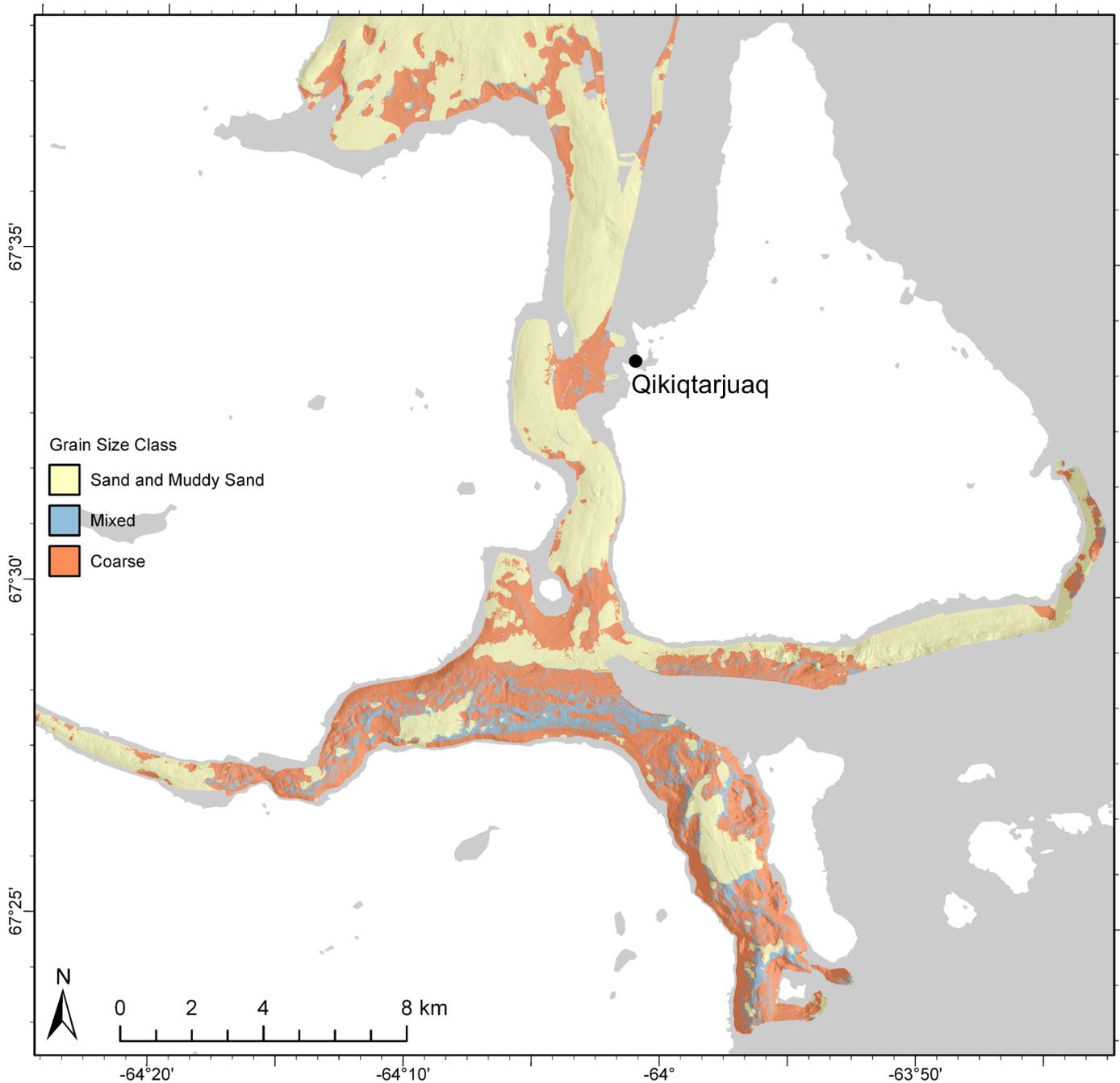
Fig 8. Grain size classification. Predictions of mud, sand, and gravel classified according to Long's [68] modification of Folk's [71] original classification scheme. See text for discussion. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence–Canada.

https://doi.org/10.1371/journal.pone.0193647.g008

variables selected at the coarsest resolution (275 m) were correlated with other variables, and were removed. Other authors (e.g. [45,74,75]) have noted the effect of "coarse-graining", in which coarsening data resolution reduces the range of values, causing them to converge upon a mean. This effect may have caused increased correlation between broad scale variables in our study. Consequently, final models included fewer broad scale variables than were originally selected (Figs 4 and 5).

It is possible that importance of broad scale variables was not due entirely to their scale-dependent relationship to the response. Data coarsening can reduce the effects of ground truth locational inaccuracy, which was not quantified while grab sampling, but which could potentially affect model performance [67]. Noise present in the primary bathymetry and backscatter data layers can also be propagated, and even amplified in derivative layers [75], which can also affect model performance [50]. A decrease in spatial resolution can reduce data noise in derivative layers [46,67]. By averaging derivative data layers over an increasing area, noise can be smoothed out and made less distinct (see [32] for a comprehensive analysis of the effects of data coarsening on bathymetric derivatives). We found that data noise was less distinct at broader scales in this study. Though this may have produced a slight increase in the predictive ability of coarser scale data layers, the effect was not overwhelming—indicated by the frequent importance of fine scale variables identified during variable testing (Fig 3). A better understanding of the effects of error propagation in bathymetric data could clarify the impact of data coarsening on scale selection and model performance.

Variable selection

Results from the variable selection process suggest that the morphology of the seabed strongly influenced ALR_{ms}. The broad scale response of mud and sand to BPI (Fig 4) shows that the relative abundances of these size fractions were sensitive to topography over scales of 100s of metres to kilometres, with finer grain sizes increasing at topographic lows. During the variable and scale selection process, BPI was found to correlate with bathymetry ($\rho = 0.79$), yet the former was a stronger predictor of ALR_{ms} and was selected for the final model. This implies that broad BPI may have acted as a surrogate for bathymetry in the model, especially close to shore where BPI exhibited an edge effect caused by "no data" points outside the area of MBES coverage. This nearshore area of the BPI layer mimicked the shallowing of the bathymetry layer, which elicited a strong response in mud and sand predictions (Fig 6). Thus, we consider bathymetry to also be an important factor influencing the distribution of grain size, for which BPI was a surrogate. ALR_{ms} increased with distance from the coast (Fig 4), describing the increased transport of finer sediments.

Bottom currents transport sediments and control rates of erosion and deposition, making them one of the strongest drivers of sediment distribution [32,46,76]. Morphology influences the speed and orientation of currents, and also describes the exposure of seabed to them. Variables that describe seabed morphology, including bathymetry, eastness, northness, curvature, and slope, together can serve as surrogates for bottom currents. The importance of eastness at the 5-m scale in our study is potentially a result of its surrogacy for current information [34,77]. The moderate response of ALR_{ms} to backscatter at an intermediate scale (45 m) suggests that sand was slightly more acoustically reflective than mud (Fig 4). Backscatter information was most useful for the ALR_{ms} model when averaged over 45 m. MBES data collection was not optimized for backscatter data quality; averaging these data may have smoothed noise that was present in the data, which was impacting model performance [67]. It is possible that the selection of 45-m backscatter indicates a scale-dependence with ALR_{ms}, yet, assuming backscatter is a proxy for substrate hardness, it is unclear why this relationship would be most apparent at the 45-m scale. Though slightly weaker predictors, plan curve and Δbackscatter variables had multiple non-correlated scales that contributed to the ALR_{ms} model. This suggests that these variables did not capture the same terrain information at different scales, and can be considered concurrently.

ALR_{gs} responded most strongly to backscatter averaged over 175 m, confirming that gravel was more acoustically reflective than sand (Fig 5). This corroborates findings by other authors

(e.g. [78]) and highlights the usefulness of backscatter as a surrogate for bottom substrate properties (e.g. hardness, roughness)–it contributed over 55% of the information used to train the ALR_{gs} model. The broad scale relationship between backscatter and ALR_{gs} suggests that backscatter was a useful predictor of grain size averaged over a large area–or potentially for larger patches of sediment. This relationship also may have been affected by the noise reduction of the backscatter layer after averaging. The heterogeneity of backscatter over a broad area, represented by the Δbackscatter variable at 275-m scale, was useful for predicting ALR_{gs}. A positive trend between these variables suggests that extensive gravelly areas caused increased variability in backscatter return, compared to sandy areas. Bathymetry and eastness were important predictor variables at fine scales, reinforcing the importance of high-resolution data in habitat mapping [7,34,45,79]. Measures of curvature (i.e., curvature, plan, profile) were weaker predictors of ALR_{gs}, but were non-correlated at multiple scales, allowing each to be included in the model.

Model prediction and evaluation

Sand was predicted to be the most abundant grain size fraction in the area, and gravel was predicted to occur in higher proportions than mud, but was less widespread across the study area (Fig 6). This is not surprising given the local igneous and metamorphic bedrock geology (granites and gneisses), which is scoured by glacial processes and overlain by sandy till veneer [80,81]. Although field observations and underwater video also suggested that sand was the dominant size fraction there is the possibility of sampling bias, which may have influenced these results. Generally, grab sampling was most successful in sandy areas. Compact muddy sediments and gravel both occasionally limited the penetration of the grab sampler, and high-gravel proportions were typically not captured in the grab, or caused sample loss. For instance, the field team noted some areas composed almost entirely of clasts ranging from pebbles to cobbles, yet the lowest predicted proportion of sand was 0.51. Fewer successful grab samples in gravelly areas may have contributed to its lower predicted abundance compared to sand. Despite the potential for bias, these results seem to accurately represent most of the study site. For instance, grab sampling, underwater video, and field observation all suggested that the north-south oriented channel was primarily sand, with little mud or gravel, except near Qikiqtarjuaq (Fig 8).

Percent deviance explained, calculated internally using withheld data over the 10 CV model folds for ALR_{ms} and ALR_{gs}, provided a measure of quality for model fit (i.e., calibration [62,82,83]), and Spearman's correlation coefficient calculated for predictions of mud, sand, and gravel indicated the model's ability to quantitatively predict grain size fractions in un-sampled locations (i.e., discrimination [83]). Each modeling scenario is different, and there is no objective threshold of percent deviance explained at which a model is considered "well-calibrated". Regardless, the percent deviance explained in our models of ALR_{ms} and ALR_{gs} on withheld CV data (56.3% and 46.4%, respectively) compare favorably with the literature (e.g. [67,83,84]). This metric was also useful in the variable and scale selection process because it provided a relative indication of goodness-of-fit, allowing for comparison between prospective models (i.e. Fig 2, step 4). Spearman's correlation coefficient for 10-fold CV predictions of mud (ρ_{mud} = 0.772), sand (ρ_{sand} = 0.712), and gravel (ρ_{gravel} = 0.578) indicated a strong positive association between predicted and observed grain size fractions. Gravel had the lowest correlation score, with high-gravel areas (Fig 6C) displaying the most variability between the 10 model folds, as measured by standard deviation (Fig 7). We suspect that the difficulty in predicting the gravel fraction was largely due to the bias in grab sampling.

Maximum grain size retained was also a limiting factor to this study. The field team noted difficulties in retaining sediment grains > 4000 μm, effectively limiting the ability to model sediments larger than small pebbles. This means that large gravel was not predicted. Thus, the map of the gravel fraction (Fig 6C) represents the distribution of gravel ≤ 4000 μm, and the "mixed" and "coarse" classes (Fig 8) only include substrates up to this size. For instance, the presence of cobbles was obvious in the area near Qikiqtarjuaq from underwater video and field observations, yet model outputs were simply classified as "coarse" (Fig 9), which represents the substrate component surrounding larger clasts. These predictions are valid and useful from an ecological perspective, but it is important to understand their limitations. Future work could investigate methods for integrating larger clasts observed in underwater video with sediment grain size predictions (e.g. [85]).

Sampling was limited to 200 m depth by equipment performance, making validation of predictions in the deepest parts of the study area impossible. For example, high backscatter response in the deep east-west oriented fjord (Fig 1C) suggests a coarser grain size, yet these predictions could not be validated. Lack of sample sites in this area likely contributed to a higher standard deviation between model predictions (Fig 7). Despite this limitation, manual inspection suggests that predictions were largely based on high backscatter return (Fig 1C). Backscatter was the single most important variable in predicting ALR_{gs}, accounting for over 55% of explained deviance in the model (Fig 5), providing some confidence that these predictions are well-founded.

Classification

Long's [68] simplification of Folk's [71] scheme was chosen to classify sediment grain size predictions because of its generality. The objective in classifying grain size predictions was to create an accessible resource for scientists and managers, yet Folk's classification, which places grain size into one of 15 categories, is complex and less accessible to non-experts. Long's modification places samples into one of only four groups, which uses simpler terminology (e.g., "gravelly muddy sand", "muddy sandy gravel", "muddy gravel", and "gravelly mud" are grouped into the class "mixed"). Each of the four classes occurred in this study area, but the class "mud and sandy mud" was very rare, and was ultimately excluded because it did not add meaningful information to the map (Fig 8). The quantitative predictions of mud, sand, and gravel produced in this study can be readily classified into any other sediment grain size scheme based on user need [15].

Conclusions

Results of this study demonstrate that the default data resolution of each terrain variable was not necessarily at the appropriate scale for explaining the distribution of grain sizes for seabed sediment. Terrain variables acting as surrogates for seabed morphology and hydrodynamics (e.g. BPI, bathymetry, and aspect), implemented at broad and fine scales, were the most important variables for differentiating the mud and sand fractions. Broad scale backscatter was the most important variable for distinguishing gravel from sand; terrain variables were of secondary importance. Multiple scale models were used to predict the distribution of the different sediment grain sizes, avoiding the arbitrary selection of spatial scale for explanatory variables. The results of this analysis can be used quantitatively in subsequent habitat mapping studies, or can easily be reclassified based on management need.

These findings highlight the importance of considering variables at multiple scales for seabed mapping. By failing to test for scale-dependence of explanatory variables in predicting the response we risk creating less realistic maps. Multiscale and multiple scale analyses should not
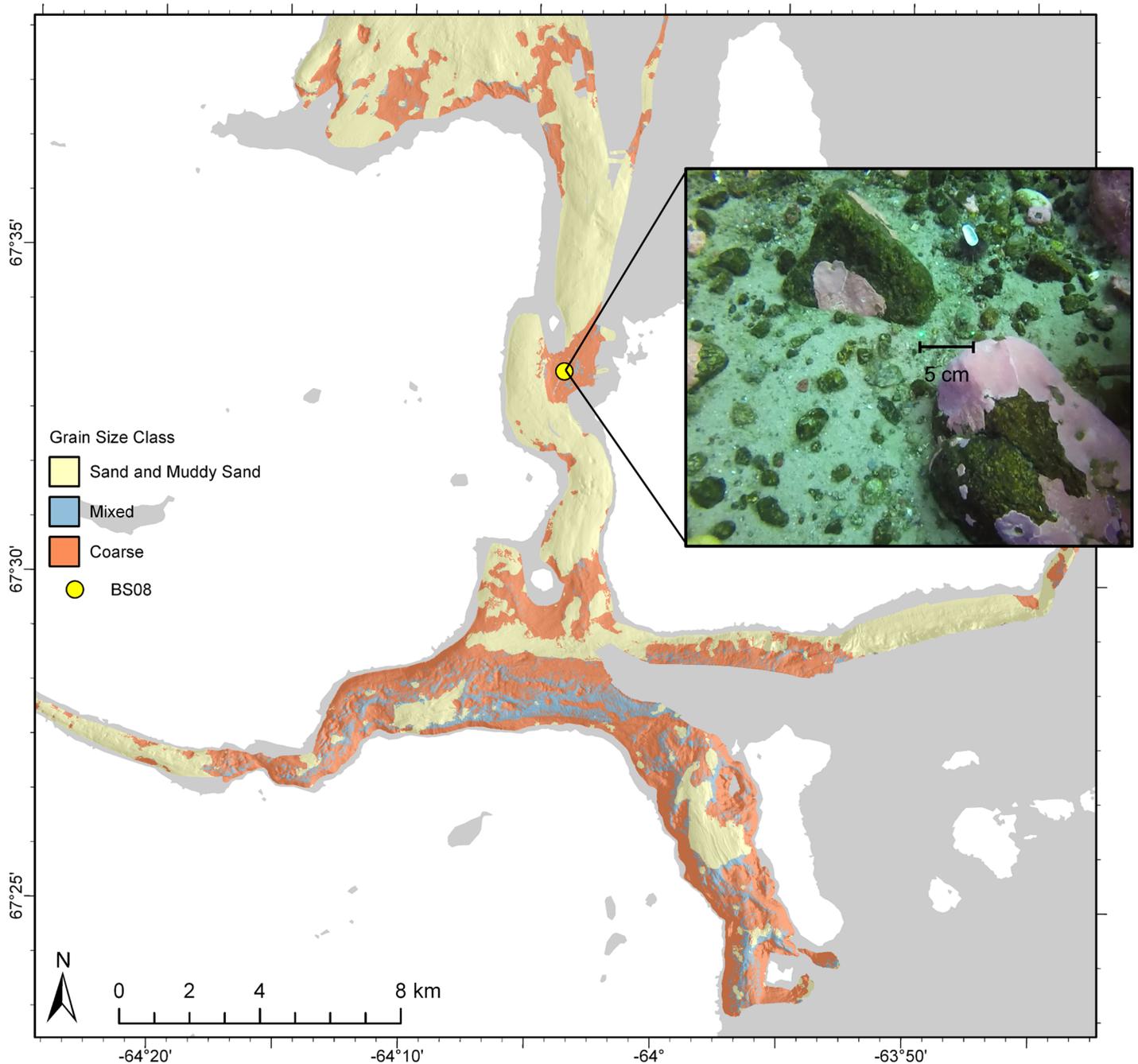
Fig 9. Large pebble and cobble observation. Clasts too large to sample in an area classified as "coarse", with 5-cm scale lasers. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence–Canada.

be considered a specialized form of analysis. We recommend that scale be considered an integral part of any benthic habitat mapping procedure–at least as important as variable selection.

Future work on the difference between mapping products as a result of multiscale derivation method (see [26]) could elucidate the importance of choosing one derivation method over another. Though there is strong evidence that different methods of deriving variables at

multiple scales produce different products, it is not clear how the end products of a study may differ based on multiscale method.

Supporting information

S1 File. Sediment grain size data. Mud, sand, and gravel size fraction data for grab samples collected near Qikiqtarjuaq, NU.
(XLSX)

S2 File. MBES data harmonization methodology. Detailed description of methods used to combine MBES data from multiple survey years and systems.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Benjamin Misiuk, Vincent Lecours.

Data curation: Benjamin Misiuk.

Formal analysis: Benjamin Misiuk.

Funding acquisition: Trevor Bell.

Investigation: Benjamin Misiuk, Trevor Bell.

Methodology: Benjamin Misiuk.

Project administration: Trevor Bell.

Resources: Trevor Bell.

Supervision: Vincent Lecours, Trevor Bell.

Validation: Benjamin Misiuk, Vincent Lecours, Trevor Bell.

Visualization: Benjamin Misiuk.

Writing – original draft: Benjamin Misiuk, Vincent Lecours, Trevor Bell.

Writing – review & editing: Benjamin Misiuk, Vincent Lecours, Trevor Bell.

References

1. Thurber AR, Sweetman AK, Narayanaswamy BE, Jones DOB, Ingels J, Hansman RL. Ecosystem function and services provided by the deep sea. Biogeosciences. 2014; 11: 3941–3963.

2. Galparsoro I, Borja A, Uyarra MC. Mapping ecosystem services provided by benthic habitats in the European North Atlantic Ocean. Front Mar Sci. 2014 Jul 20; 1: 23–36.

3. Myers RA, Worm B. Rapid worldwide depletion of predatory fish communities. Nature. 2003 May 15; 423: 280–283. https://doi.org/10.1038/nature01610 PMID: 12748640

4. Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C, et al. A global map of human impact on marine ecosystems. Science. 2008 Feb 15; 319(5865): 948–952. https://doi.org/10.1126/science.1149345 PMID: 18276889

5. Harris PT, Baker EK. Why map benthic habitats? In: Harris PT, Baker EK, editors. Seafloor geomorphology as benthic habitats: GeoHab atlas of seafloor geomorphic features and benthic habitats. Amsterdam: Elsevier; 2012. pp. 3–22.

6. Harris PT, Baker EK. GeoHab atlas of seafloor geomorphic features and benthic habitats: Synthesis and lessons learned. In: Harris PT, Baker EK, editors. Seafloor geomorphology as benthic habitats: GeoHab atlas of seafloor geomorphic features and benthic habitats. Amsterdam: Elsevier; 2012. pp. 871–890.

7. Brown C, Smith SJ, Lawton P, Anderson JT. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuar Coast Shelf Sci. 2011; 92: 502–520.

8. McArthur MA, Brooke BP, Przeslawski R, Ryan DA, Lucieer VL, Nichol S, et al. On the use of abiotic surrogates to describe marine benthic biodiversity. Estuar Coast Shelf Sci. 2010 Jun 10; 88: 21–32.

9. Ysebaert T, Meire P, Herman PMJ, Verbeek H. Macrobenthic species response surfaces along estuarine gradients: Prediction by logistic regression. Mar Ecol Prog Ser. 2002 Jan 11; 225: 79–95.

10. Reiss H, Birchenough S, Borja A, Buhl-Mortensen L, Craeymeerch J, Dannheim J. Benthos distribution modelling and its relevance for marine ecosystem management. ICES J Mar Sci. 2015 Jun 19; 72(2): 297–315.

11. Coggan R, Barrio Froján CRS, Diesing M, Aldridge J. Spatial patterns in gravel habitats and communities in the central and eastern English Channel. Estuar Coast Shelf Sci. 2012 Oct 1; 111(7): 118–128.

12. Stephens D, Diesing M. Towards quantitative spatial models of seabed sediment composition. PLoS ONE. 2015 Nov 23; 10(11): e0142502. https://doi.org/10.1371/journal.pone.0142502 PMID: 26600040

13. Diesing M, Green SL, Stephens D, Lark RM, Stewart HA, Dove D. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Cont Shelf Res. 2014 Aug 1; 84: 107–119.

14. Stephens D, Diesing M. A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. PLoS ONE. 2014 Apr 3; 9(4): e93950. https://doi.org/10.1371/journal.pone.0093950 PMID: 24699553

15. Diesing M. Case study: Quantitative spatial prediction of seabed sediment composition. Centre for Environment Fisheries & Aquaculture Science; 2015 Oct 7. EMODnet-Geology Phase II Report No.: C5818.

16. Lark RM, Marchant BP, Dove D, Green SL, Stewart H, Diesing M. Combining observations with acoustic swath bathymetry and backscatter to map seabed sediment texture classes: The empirical best linear unbiased predictor. Sed Geo. 2015 Oct; 328: 17–32.

17. Brown CJ, Sameoto JA, Smith SJ. Multiple methods, maps, and management applications: Purpose made seafloor maps in support of ocean management. J Sea Res. 2012; 72: 1–13.

18. Bučas M, Bergström U, Downie AL, Sundblad G, Gullström M, Von Numers M, et al. Empirical modelling of benthic species distribution, abundance, and diversity in the Baltic Sea: Evaluating the scope for predictive mapping using different modelling approaches. ICES J Mar Sci. 2013 May 19; 70(6): 1233–1243.

19. Calvert J, Strong JA, Service M, McGonigle C, Quinn R. An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data. ICES J Mar Sci. 2014 Dec 12; 72(5): 1498–1513.

20. Copeland A, Edinger E, Devillers R, Bell T, LeBlanc P, Wroblewski J. Marine habitat mapping in support of marine protected area management in a subarctic fjord: Gilbert Bay, Labrador, Canada. J Coast Conserv. 2013 Jun 1; 17(2): 225–237.

21. Dunn DC, Halpin PN. Rugosity-based regional modeling of hard-bottom habitat. Mar Ecol Prog Ser. 2009 Feb 26; 377: 1–11.

22. Downie AL, Dove D, Westhead RK, Diesing M, Green S, Cooper R. Semiautomated mapping of rock in the North Sea. Peterborough (UK): Joint Nature Conservation Committee. 2016 Jul. Report No.: 592.

23. Lecours V, Dolan MFJ, Micallef A, Lucieer VL. A review of marine geomorphometry, the quantitative study of the seafloor. Hydrol Earth Sys Sci. 2016; 20: 3207–3244.

24. Eidens C, Hauffe T, Bayraktarov E, Wild C, Wilke T. Multi-scale processes drive benthic community structure in upwelling-affected coral reefs. Front Mar Sci. 2015 Feb 11; 2: 1–11.

25. Gambi C, Danovaro R. A multiple-scale analysis of metazoan meiofaunal distribution in the deep Mediterranean Sea. Deep Sea Res Part 1 Oceanogr Res Pap. 2006 Jul; 53(7): 1117–1134.

26. Dolan MFJ. Calculation of slope angle from bathymetry data using GIS—effects of computation algorithms, data resolution and analysis scale. Trondheim (Norway): Geological Survey of Norway. 2012 Aug 7. Report No.: 2012.041.

27. Bradter U, Kunin WE, Altringham JD, Thom TJ, Benton TG. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. Methods Ecol Evol. 2013; 4: 167–174.

28. Lecours V, Devillers R, Schneider DC, Lucieer VL, Brown CJ, Edinger EN. Spatial scale and geographic context in benthic habitat mapping: Review and future directions. Mar Ecol Prog Ser. 2015 Sep 15; 535: 259–284.

29. Dixon B, Earls J. Resample or not?! Effects of resolution of DEMs in watershed modeling. Hydrol Process. 2009 Apr 20; 23(12): 1714–1724.

30. Seo C, Thorne JH, Hannah L, Thuiller W. Scale effects in species distribution models: Implications for conservation planning under climate change. Biol Lett. 2009; 5: 39–43. https://doi.org/10.1098/rsbl. 2008.0476 PMID: 18986960

31. Wolff C, Vafeidis AT, Lincke D, Marasmi C, Hinkel J. Effects of scale and input data on assessing the future impacts of coastal flooding: An application of DIVA for the Emilia-Romagna Coast. Front Mar Sci. 2016; 3: 41. https://doi.org/10.3389/fmars.2016.00041

32. Dolan MFJ, Lucieer VL. Variation and uncertainty in bathymetric slope calculations using geographic information systems. Mar Geod. 2014; 37: 187–219.

33. Ross LK, Ross RE, Stewart HA, Howell KL. The influence of data resolution on predicted distribution and estimates of extent of current protection of three "listed" deep-sea habitats. PLoS One. 2015 Oct 23; 10: e0140061. https://doi.org/10.1371/journal.pone.0140061 PMID: 26496639

34. Miyamoto M, Kiyota M, Murase H, Nakamura T, Hayashibara T. Effects of bathymetric grid-cell sizes on habitat suitability analysis of cold-water Gorgonian corals on seamounts. Mar Geod. 2017; 40(4): 205–223.

35. Lechner AM, Langford WT, Jones SD, Bekessy SA, Gordon A. Investigating species–environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. Eco Com. 2012 Sep; 11: 91–102.

36. Margareth A, Gosse JC, Dyke AS. Wisconsinan and early Holocene glacial dynamics of Cumberland Peninsula, Baffin Island, Arctic Canada. Quat Sci Rev. 2017 Jul 15; 168: 79–100.

37. Cowan B. Shorelines beneath the sea: Geomorphology and characterization of the postglacial sea-level lowstand, Cumberland Peninsula, Baffin Island, Nunavut [dissertation]. St. John's (NL): Memorial University of Newfoundland; 2015.

38. Gilbert R. Environmental studies in Maktak, Coronation, and North Pangnirtung fiords, Baffin Island, N. W.T. Final report. Kingston (ON): Queen's University; 1980 Feb. Sponsored by Petro-Canada Exploration Inc.

39. Muggah J. 2007 multibeam sonar data collected from the CCGS Amundsen. 2011 Sep 3 [cited 2017 May 25]. OMG Arctic Metadata. Available from: http://www.omg.unb.ca/Projects/Arctic/metadata/ Amundsen_2007_MB_General.html.

40. Muggah J. 2013 multibeam sonar data collected from the M/V Nuliajuk. 2014 Mar 21 [cited 2017 May 25]. OMG Arctic Metadata. Available from: http://www.omg.unb.ca/Projects/Arctic/metadata/Nuliajuk_ 2013_MB_General.html.

41. Muggah J. 2014 multibeam sonar data collected from the M/V Nuliajuk. 2015 Jan 26 [cited 2017 May 25]. OMG Arctic Metadata. Available from: http://www.omg.unb.ca/Projects/Arctic/metadata/Nuliajuk_ 2014_MB_General.html.

42. Lecours V, Devillers R, Simms AE, Lucieer VL, Brown CJ. Towards a framework for terrain attribute selection in environmental studies. Environ Model Softw. 2017 Mar; 89: 19–30.

43. Lecours V. Terrain attribute selection for spatial ecology. Version 1.0 [software]. 2015 [cited 2017 May 25]. https://doi.org/10.13140/RG.2.2.35147.18728.

44. Wilson MFJ, O'Connell B, Brown C, Guinan JC, Grehan AJ. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Mar Geod. 2007; 30(1–2): 3–35.

45. Rengstorf AM, 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(4): 343–361.

46. Tong R, Purser A, Guinan J, Unnithan V, Yu J, Zhang C. Quantifying relationships between abundances of cold-water coral Lophelia pertusa and terrain features: A case study on the Norwegian Margin. Cont Shelf Res. 2016 Mar 15; 116: 13–26.

47. Lundblad ER, Wright DJ, Miller J, Larkin EM, Rinehart R, Naar DF, et al. A benthic terrain classification scheme for American Samoa. Mar Geod. 2006; 29(2): 89–111.

48. Wright DJ, Pendleton M, Boulware J, Walbridge S, Gerlt B, Eslinger D, et al. ArcGIS benthic terrain modeler (BTM). Version 3.0 [software]. 2012 [cited 2017 May 25]. Available from: http://esriurl.com/ 5754.

49. Diesing M, Stephens D. A multi-model ensemble approach to seabed mapping. J Sea Res. 2015; 100: 62–69.

50. Lecours V, Devillers R, Edinger EN, Brown CJ, Lucieer VL. Influence of artefacts in marine digital terrain models on habitat maps and species distribution models: A multiscale assessment. Remote Sens Ecol Conserv. 2017 July 18. https://doi.org/10.1002/rse2.49

51. Giusti M, Innocenti C, Canese S. Predicting suitable habitat for the gold coral Savalia savaglia (Bertoloni, 1819) (Cnidaria, Zoantharia) in the South Tyrrhenian Sea. Cont Shelf Res. 2014 Jun 15; 81: 19–28.

52. Aitchison J. The statistical analysis of compositional data. London: Chapman and Hall; 1986.

53. Martín-Fernández JA, Thió-Henestrosa S. Rounded zeros: Some practical aspects for compositional data. In: Buccianti A, Mateu-Figueras G, Pawlowsky-Glahn V, editors. Compositional data analysis in the geosciences: From theory to practice. London: Geological Society; 2006. pp. 191–204.

54. Lark RM, Dove D, Green SL, Richardson AE, Stewart H, Stevenson A. Spatial prediction of seabed sediment texture classes by cokriging from a legacy database of point observations. Sediment Geol. 2012 Dec 15; 281: 35–49.

55. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: A statistical view of boosting. Ann Stat. 2000; 28(2): 337–407.

56. Breiman L. Statistical modeling: The two cultures. Statist Sci. 2001; 16(3): 199–231.

57. Olden JD. Machine learning methods without tears: A primer for ecologists. Q Rev Biol. 2008 Jun; 83 (2): 171–193. PMID: 18605534

58. Franklin J. Mapping species distributions: Spatial inference and prediction. New York: Cambridge University Press; 2009.

59. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002 Dec; 2–3: 18–22.

60. Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, et al. Novel methods improve prediction of species' distributions from occurrence data. Ecography. 2006 April; 29(2): 129–151.

61. Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT. What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? Ecol Monogr. 2007 Nov 1; 77 (4): 615–630.

62. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008 Jul; 77 (4): 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x PMID: 18397250

63. Evans JS, Murphy MA, Holden ZA, Cushman SA. Modeling species distribution and change using Random Forest. In: Drew CA, Wiersma YF, Huettmann F, editors. Predictive species and habitat modeling in landscape ecology: Concepts and applications. New York: Springer; 2011. pp. 139–159.

64. Hopkins RL II. Use of landscape pattern metrics and multiscale data in aquatic species distribution models: A case study of a freshwater mussel. Landsc Ecol. 2009 Aug; 24(7): 943–955.

65. Elith J, Leathwick JR. Tutorial for running boosted regression trees [Elith et al. 2008 –Appendix S3]. J Anim Ecol. 2008 July; 77(4): 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x PMID: 18397250

66. Ridgeway G. gbm: Generalized Boosted Regression Models (R package). Version 2.1.1 [software]. 2015 Mar 11 [cited 2017 May 30]. Available from: https://CRAN.R-project.org/package=gbm.

67. Gottschalk TK, Aue B, Hotes S, Ekschmitt K. Influence of grain size on species-habitat models. Ecol Modell. 2011 Sep 24; 222(18): 3403–3412.

68. Long D. BGS detailed explanation of seabed sediment modified Folk classification. 2006.

69. Guisan A, Zimmermann NE. Predictive habitat distribution models in ecology. Ecol Modell. 2000 Dec 5; 135(2–3): 147–186.

70. Potts JM, Elith J. Comparing species abundance models. Ecol Modell. 2006 Nov 16; 199(2): 153–163.

71. Folk RL. The distinction between grain size and mineral composition in sedimentary-rock nomenclature. J Geol. 1954 Jul; 62(4): 344–359.

72. Dixon Hamil KA, Iannone III BV, Huang WK, Fei S, Zhang H. Cross-scale contradictions in ecological relationships. Landsc Ecol. 2016 Jan; 31(1): 7–18.

73. Albani M, Klinkenberg B, Andison DW, Kimmins JP. The choice of window size in approximating topographic surfaces from digital elevation models. Int J Geogr Inf Sci. 2004 Sep 1; 18(6): 577–593.

74. Chow TE, Hodgson ME. Effects of lidar post-spacing and DEM resolution to mean slope estimation. Int J Geogr Inf Sci. 2009 Sep 23; 10: 1277–1295.

75. Lecours V, Devillers R, Lucieer VL, Brown CJ. Artefacts in marine digital terrain models: A multiscale analysis of their impact on the derivation of terrain attributes. IEEE Trans Geosci Remote Sens. 2017 Jun 14; 55: 1–16.

76. Tong R, Purser A, Unnithan V, Guinan J. Multivariate statistical analysis of distribution of deep-water gorgonian corals in relation to seabed topography on the Norwegian Margin. PLoS ONE. 2012 Aug 7; 7 (8): e43534. https://doi.org/10.1371/journal.pone.0043534 PMID: 22912887

77. Tong R, Purser A, Guinan J, Unnithan V. Modeling the habitat suitability for deep-water gorgonian corals based on terrain variables. Ecol Inf. 2013; 13: 123–132.

78. Goff J, Olson H, Duncan C. Correlation of side-scan backscatter intensity with grain-size distribution of shelf sediments, New Jersey margin. Geo Mar Lett. 2000 Jul; 20(1): 43–49.

79. Rengstorf AM, Yesson C, Brown C, Grehan AJ. High-resolution suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea. J Biogeogr. 2013; 40(9): 1702–1714.

80. Wheeler JO, Hoffman PF, Card KD, Davidson A, Sanford BV, Okulitch AV, et al. Geological map of Canada [map]. Geological Survey of Canada; 1996. https://doi.org/10.4095/208175.

81. Fulton RJ. Surficial materials of Canada [map]. Geological Survey of Canada; 1995. https://doi.org/10. 4095/205040.

82. Eigenbrod F, Hecnar SJ, Fahrig L. The relative effects of road traffic and forest cover on anuran populations. Biol Conserv. 2008 Jan; 141(1): 35–46.

83. Elith J, Graham CH. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography. 2009 Feb 27; 32: 66–77.

84. Nielsen SE, Johnson CJ, Heard DC, Boyce MS. Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. Ecography. 2005 April; 28(2): 197–208.

85. Diesing M, Green SL, Stephens D, Cooper R, Mellett CL. Semi-automated mapping of rock in the English Channel and Celtic Sea. Peterborough (UK): Joint Nature Conservation Committee. 2015 Jul. Report No.: 569.