ICES Journal of Marine Science

ICES Journal of Marine Science (2019), 76(7), 2349–2361. doi:10.1093/icesjms/fsz099

Original Article

Mapping Arctic clam abundance using multiple datasets, models, and a spatially explicit accuracy assessment

Benjamin Misiuk [1,*], Trevor Bell[1], Alec Aitken[2], Craig J. Brown [3], and Evan N. Edinger[1]

[1]*Department of Geography, Memorial University of Newfoundland, 232 Elizabeth Ave, St. John's, Newfoundland A1B 3X9, Canada*
[2]*Department of Geography & Planning, University of Saskatchewan, Kirk Hall Building 117 Science Place, Saskatoon, Saskatchewan S7N 5C8, Canada*
[3]*Applied Research, Nova Scotia Community College Ivany Campus, 80 Mawiomi Place, Dartmouth, Nova Scotia B2Y 0A5, Canada*

**Corresponding author: tel: +1 709 864 6127; e-mail: bmisiuk@mun.ca*

Misiuk, B., Bell, T., Aitken, A., Brown, C. J., and Edinger, E. N. Mapping Arctic clam abundance using multiple datasets, models, and a spatially explicit accuracy assessment. – ICES Journal of Marine Science, 76: 2349–2361.

Received 23 August 2018; revised 26 April 2019; accepted 5 May 2019; advance access publication 6 June 2019.

Species distribution models are commonly used in the marine environment as management tools. The high cost of collecting marine data for modelling makes them finite, especially in remote locations. Underwater image datasets from multiple surveys were leveraged to model the presence–absence and abundance of Arctic soft-shell clam (*Mya* spp.) to support the management of a local small-scale fishery in Qikiqtarjuaq, Nunavut, Canada. These models were combined to predict *Mya* abundance, conditional on presence throughout the study area. Results suggested that water depth was the primary environmental factor limiting *Mya* habitat suitability, yet seabed topography and substrate characteristics influence their abundance within suitable habitat. Ten-fold cross-validation and spatial leave-one-out cross-validation (LOO CV) were used to assess the accuracy of combined predictions and to test whether this was inflated by the spatial autocorrelation of transect sample data. Results demonstrated that four different measures of predictive accuracy were substantially inflated due to spatial autocorrelation, and the spatial LOO CV results were therefore adopted as the best estimates of performance.

Keywords: Arctic science, benthic habitat mapping, fisheries management, spatial autocorrelation, species distribution modelling

Introduction

Species distribution models (SDMs) have become important tools for the management of marine resources (Brown *et al.*, 2011; Hattab *et al.*, 2013). By exploring the relationships between an organism of interest and environmental variables, SDMs are used to predict presence, absence, and abundance of taxa (Franklin, 2009). In addition to predicting distributions, SDMs can be used to investigate the environmental conditions that meet a given species' habitat requirements. This information is essential to effectively manage marine ecosystems. A typical SDM workflow is to sample an organism across a range of environmental variables, use statistical relationships to create spatially continuous predictions of its distribution, and evaluate these predictions to provide estimates of model accuracy. There are many different SDM statistical methods and algorithms that have been thoroughly reviewed in the literature (Guisan and Zimmermann, 2000; Elith *et al.*, 2006; Franklin, 2009; Miller, 2010; Drew *et al.*, 2011).

In this study, we created an SDM to investigate the environmental drivers of soft-shell clam (*Mya* spp.; hereafter referred to as "*Mya*") distribution and to predict their abundance in support of community-based fisheries management near Qikiqtarjuaq, Nunavut (Arctic Canada). *Mya* are commonly harvested in the intertidal zone as a source of food in Inuit communities (Aitken *et al.*, 1988; Nunavut Department of Environment—Fisheries and Sealing Division, 2012). A small number of local SCUBA divers in Qikiqtarjuaq have had success over the past few decades in efficiently harvesting clams at depths where they are abundant (~20 m), and this has generated interest in formalizing a community-based fishery. High-resolution information on how the *Mya* population is distributed in this area can serve as an effective management tool for the sustainable development of this fishery.

Siferd (2005) surveyed the *Mya* population along the coasts near Qikiqtarjuaq as part of a Department of Fisheries and Oceans (DFO) stock assessment in zone CFZ3. The assessment

quantified *Mya* abundance using still image transects running parallel to shore at 10, 20, 30, and 40 m isobaths, and found populations >50 individuals/m^2 on average in many areas, and nearly 100 individuals/m^2 on average in the densest region. They estimated the total population in this area at over 1.5 billion individuals and modelled the effects of various fishing rates on that population. This assessment can be supplemented by an SDM that continuously predicts *Mya* abundance over the extent of the study area. Maps from SDMs are visually intuitive and useful to experts and non-experts alike. Furthermore, Smith *et al.* (2017) demonstrated that incorporating SDMs into fishery stock assessments introduces a spatial component that is important for maintaining the long-term viability of a fishery, since exploitation is non-uniform and tends to correspond with high levels of habitat suitability for the target species.

Extensive benthic species abundance datasets are a valuable resource—the *Mya* image dataset can be put to further use as part of an SDM. This requires consideration of qualities of the dataset that complicate statistical modelling though. For instance, samples were only collected near the coast, and up to 40 m water depth. *Mya* likely inhabit environments outside these conditions, and it is desirable to sample across the full range of their habitat preference. In addition to informing habitat suitability, this allows for a better estimation of what habitats are unsuitable for *Mya*, and where they are likely to be absent. Relatedly, the second issue is that modelling species absence is not always straightforward with an abundance model (Ridout *et al.*, 1998; Martin *et al.*, 2005), often requiring more flexible approaches that can accommodate zero values. Third, images within sample transects, and potentially the transects themselves, are likely to be spatially autocorrelated, which may introduce bias in statistical models (Segurado *et al.*, 2006). Unchecked, bias can violate model assumptions (Legendre, 1993) and potentially inflate estimates of model performance (Bahn and McGill, 2013).

To better inform *Mya* models, we conducted additional surveys near Qikiqtarjuaq to supplement Siferd's (2005) data. The goal was to sample *Mya* over a greater spatial and environmental range (e.g. >40 m water depth). Because clams were still observed abundantly at the maximum sampling depth in Siferd's study (~40 m), it was important to investigate the maximum depths that they inhabit. In addition, most of Siferd's (2005) samples were nearshore, yet it is possible that more distal locations contain different topographic and substrate characteristics that influence the suitability of *Mya* habitat.

Sampling across a greater range of environmental variables can provide information on conditions that are unsuitable for *Mya*, yet this still may not result in reliable predictions of absence using an abundance model. Modelling datasets with zero values potentially requires data transformation or methods that allow for over-dispersion (Warton, 2005), or combined modelling approaches (e.g. hurdle or mixture models; Mullahy, 1986; Welsh *et al.*, 1996). The latter allow for the possibility that the environmental drivers of species occurrence (i.e. habitat suitability) are not necessarily the same that drive abundance (Clark *et al.*, 2014). Thus, these two characteristics of a species' spatial distribution may require separate modelling procedures—one that models the presence or absence of the species, and one that determines its abundance, conditional on presence (Welsh *et al.*, 1996).

Once modelled, predictions of species distribution require estimates of accuracy to indicate their performance (Franklin, 2009), yet these may be compromised when model training and test data are not independent. Ideally, test data would be collected separately from training data to ensure independence, yet this is often not feasible in marine science, and especially in the Arctic, where ship time and sampling season are limiting factors. Training data are therefore commonly subsampled to test model performance, yet this can result in biased evaluation if the data are not independent (Hijmans, 2012). Transect sampling is used in marine science to obtain many samples at a single location, or continuously over some distance (Foster *et al.*, 2014), and is likely to produce non-independent data. Furthermore, data collection for purposes other than modelling, such as the DFO survey that produced the dataset used here, may place greater emphasis on obtaining many samples than on ensuring their independence. In such cases it is necessary to account for spatial autocorrelation when evaluating statistical models. There are several methods for this, including designating entire sample transects as test or training data (Brown *et al.*, 2012; Porskamp *et al.*, 2018), spatial blocking (Roberts *et al.*, 2017), spatial subsampling (Kendall *et al.*, 2005; Segurado *et al.*, 2006; Veloz, 2009), and geostatistical methods (Li *et al.*, 2017).

The goal of this study was to predict the distribution of *Mya* to support sustainable development of the clam fishery near Qikiqtarjuaq, NU. Specifically, we set out to (i) supplement Siferd's (2005) survey data by sampling a broader range of environmental variables to determine the extent of *Mya* habitat, (ii) model *Mya* abundance using this combined dataset, including predictions of absence where habitat is unsuitable, and (iii) estimate the magnitude of inflation caused by spatial autocorrelation in the sample data to provide accurate estimates of model performance.

Study area and species

Qikiqtarjuaq is located on the west coast of Broughton Island, off eastern Baffin Island, Nunavut, Canada (Figure 1). The community is set across from the shallowest part of a sheltered, north–south-oriented channel that is seasonally impacted by sea ice, which modifies the seabed and coastline (Forbes and Taylor, 1994). The relief in this part of the channel is gradual compared to the surrounding terrain—much of the Baffin Island coast is characterized by steep topography above and below the waterline (Brigham, 1983). Glaciers flowing from the Penny Ice Cap carved deep valleys and fjords during the Quaternary Period, producing a distinct glacial landscape (Dyke *et al.*, 1982). These processes have scoured the local bedrock over repeated glacial cycles, producing a sandy till veneer that overlies granitic and gneissic bedrock (Dyke *et al.*, 1982; Brigham, 1983; Fulton, 1995; Wheeler *et al.*, 1996). The surficial seabed substrate is correspondingly sandy in much of the study area, with large patches of mixed and coarse sediments near Qikiqtarjuaq, in the nearby fjords, and to the south of the community (Misiuk *et al.*, 2018).

The sandy and mixed substrates near Qikiqtarjuaq form suitable habitat for soft-shelled clams, while accelerated currents in the north–south-oriented channel may increase food transport, supporting dense populations. Siferd's (2005) assessment covered fishing zone CFZ3 (Figure 1), and suggested abundances were greatest at 30–35 m water depth—beyond the range of local SCUBA harvesters. The survey also suggested that Arctic *Mya* near Qikiqtarjuaq take ~10 years to mature and can live up to 60 years (cf. 40 years; Hewitt and Dale, 1984). Previous surveys (Petersen, 1978; Abraham and Dillon, 1986; Siferd, 2005) have also suggested that *Mya* prefer shallow depths, and unconsolidated substrates that

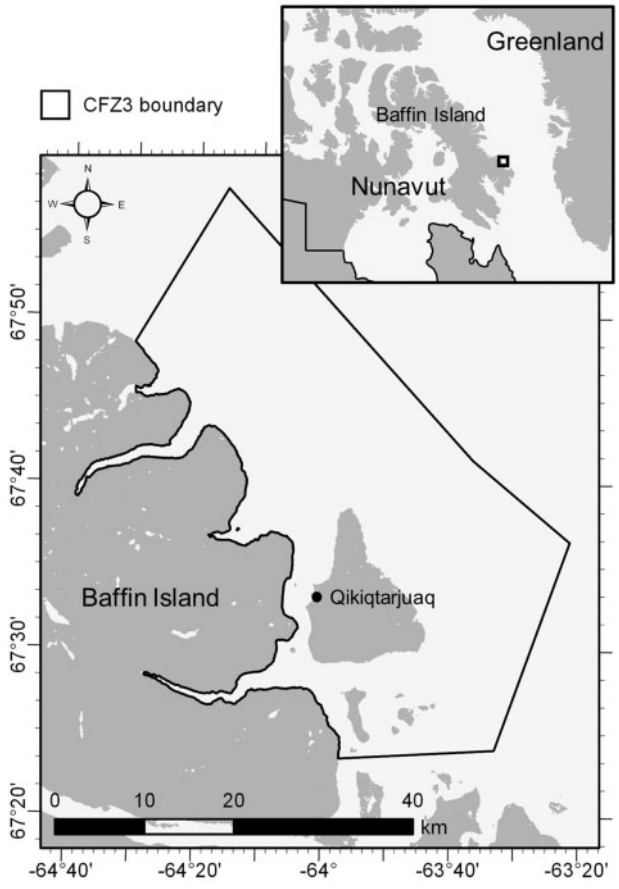

Figure 1. Location of Qikiqtarjuaq study area within fishing zone CFZ3 and eastern Nunavut, Canada (inset map). Basemaps obtained from the Canadian Land Cover GeoBase Series, containing information licenced under the Open Government Licence—Canada.

allow juveniles (spat) to settle and burrow, where they remain for their entire adult life. From their burrows, *Mya* filter feed by extending their siphon above the substrate surface to capture food, which settles through the water column, or is delivered via currents.

Data and methods

In SDM, spatially continuous environmental data explaining the habitat preferences of an organism are used to predict their distribution. Seabed morphology has been recognized as an integral component of benthic habitat and has been used to successfully predict the distribution of benthic taxa, including bivalves (Brown *et al.*, 2012). Benthic substrate properties were also expected to contribute to *Mya* habitat suitability, as they are infaunal organisms. Our modelling approach applied sonar-derived seabed morphological data and sediment grain size models to predict the abundance of *Mya* observed from the underwater image ground truth.

Environmental data

Multibeam echosounder (MBES) bathymetry and backscatter data (Figure 2) were collected near Qikiqtarjuaq to characterize *Mya* habitat. MBES collect depth soundings (bathymetry in m) and measurements of acoustic reflectivity (backscatter in dB) simultaneously, allowing for the approximation of fine-scale seabed morphology and substrate properties. MBES data were collected by two different vessels over a 4-year period and were used to derive explanatory environmental variables for the *Mya* presence–absence and abundance models. The *CCGS Amundsen* mapped ~20 km^2 in the deepest part of the study area in 2007 using a Kongsberg EM300 30 kHz echosounder, and the *RV Nuliajuk* mapped the remaining area in 2012 and 2013 using an EM3002 300 kHz echosounder, and in 2015 using an EM2040C 200–400 kHz echosounder.

Lecours *et al.* (2017a) suggested a combination of six terrain variables that capture most of the morphological information of a surface, which can be derived from a bathymetric model using the "Terrain Attribute Selection for Spatial Ecology" (TASSE) toolbox (Lecours, 2017; Table 1) in ESRI ArcGIS. In addition to these, we produced eight terrain variables commonly used to describe seabed morphology using the "Benthic Terrain Modeler" (BTM; Walbridge *et al.*, 2018) and spatial analyst toolboxes in ESRI ArcGIS (Table 1). These were calculated as "multiscale" terrain variables by averaging over a series of increasing neighbourhood sizes to incorporate information from a range of spatial scales between 5 and 275 m (Dolan, 2012; Dolan and Lucieer, 2014). While backscatter is commonly used as a proxy for seabed hardness (low backscatter corresponds to soft/fine sediments, high backscatter to hard/coarse; Harris and Baker, 2012), Diesing and Stephens (2015) suggested that the local variability in backscatter can inform on other substrate properties. We calculated the local variability in backscatter (i.e. the range of backscatter values in a 3 × 3-pixel neighbourhood; hereafter referred to as Δbackscatter) from the backscatter layer using the "Focal Statistics" and "Raster Calculator" tools in ESRI ArcGIS. We also included predicted proportions of mud, sand, and gravel, modelled for the study area by Misiuk *et al.* (2018).

Mya ground-truth data

Siferd (2005) collected and analysed drop-camera bottom photographs from sites randomly selected along the coast near Qikiqtarjuaq to quantify the abundance of *Mya* (individuals per m^2). Ten photographs were taken in transects parallel to shore at ~10, 20, 30, and 40 m water depths, resulting in four transects of 10 closely spaced sample points per site. Photographs were taken using a Nikon D1X digital camera in a Seacam underwater housing mounted on a frame with legs standing 70 cm above the seabed; the area photographed was ~0.5 m^2. *Mya* abundance was quantified by counting their siphons, which protrude above the substrate surface and are visible in underwater image frames. Photographs overlapping MBES coverage (n = 1827) were used for this analysis.

Underwater towed-video transects were collected in 2015 to supplement Siferd's (2005) dataset at a broader depth range (up to 200 m). Four-minute drifts were recorded from a 24 ft Nor-West freighter canoe using a GoPro Hero3 in a waterproof case, mounted to a housing with underwater lights and a live-feed Deep Blue Pro underwater camera. Two green lasers were attached to the camera housing, spaced 5 cm apart to provide scale in the underwater video. Positioning was obtained using a Garmin 18× PC GPS with a live feed to the underwater video, providing continuous locational information from the surface for the duration of the video. The surface GPS accuracy was rated at <3 m, but the

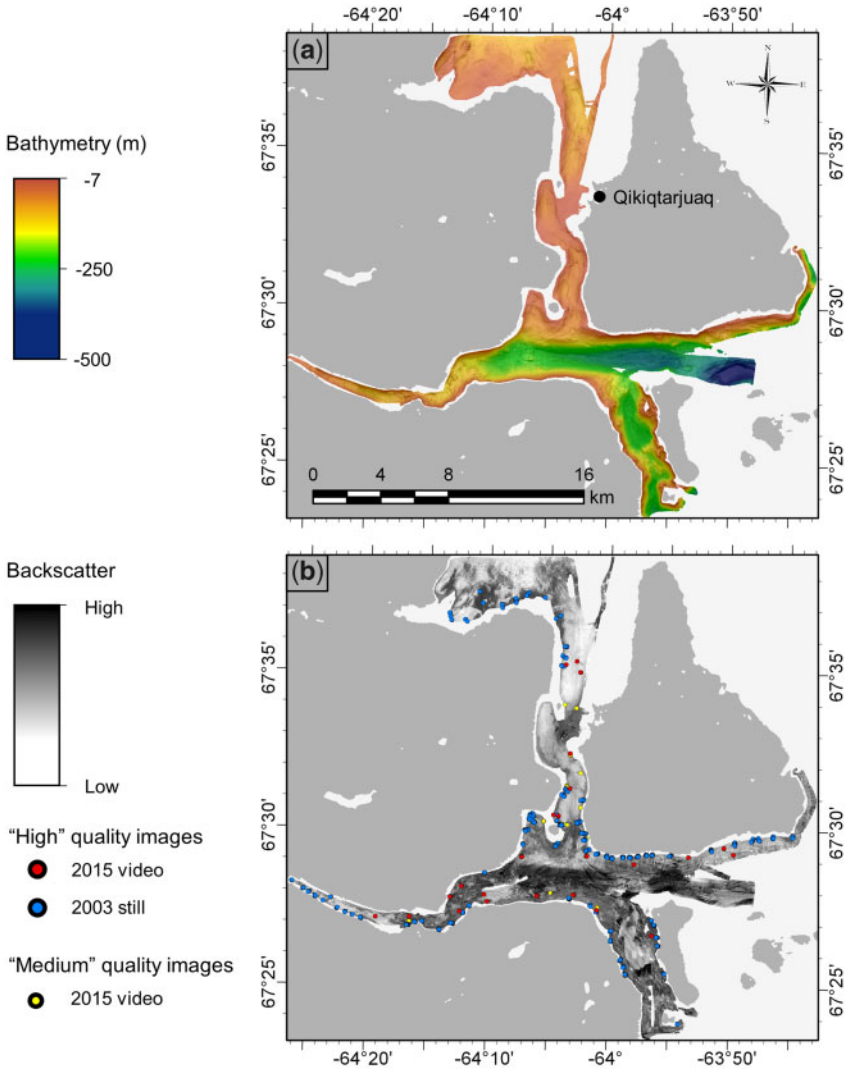

Figure 2. (a) Multibeam bathymetry near Qikiqtarjuaq. (b) Multibeam backscatter with ground truth image samples. Basemaps were obtained from the Canadian Land Cover GeoBase Series, containing information licenced under the Open Government Licence—Canada.

accuracy of the camera position underwater was likely >5 m depending on depth and current conditions. Sample sites were randomly selected over the MBES coverage, stratified by environmental variables expected to influence the abundance of *Mya* (water depth, seabed slope, and backscatter). Still frames were extracted from underwater video ($n = 938$) at ~2.5 m intervals.

Underwater video sampling in 2015 was designed to replicate Siferd's (2005) data, yet not all images were of comparable quality. Siferd's photos were taken from a stationary platform from which *Mya* abundance could be consistently quantified. Underwater video from 2015 was high resolution, but the drift speed, water clarity, and light availability at greater depths limited the quality of some frames. Video frames were therefore ranked for quality to determine their compatibility with still-frame data. "High" quality frames ($n = 250$) were of comparable quality to drop-camera stills (i.e. *Mya* abundance could be readily quantified). In "medium" quality frames ($n = 301$) the analyst was not confident that all syphons could be identified but was able to confidently determine presence or absence. In "low" quality frames ($n = 387$) presence or absence could not be confidently confirmed. Siferd's (2005) drop-camera still dataset and the "high" quality 2015 video frames therefore constituted the abundance modelling dataset, while these data plus the "medium" quality video frames were used to model presence–absence. "Low" quality video frames were not used.

Statistical modelling

Welsh *et al.* (1996) and Barry and Welsh (2002) recommended a combined modelling approach for zero-inflated species abundances, in which the presence or absence of a species is modelled first, then abundance conditional on presence. While "zero-inflated" generally implies the use of a parametric distribution, and many zeroes do not necessarily mean zero-inflation (see Warton, 2005), the *Mya* abundance dataset had absences that were not predicted by the abundance model, making a combined modelling approach useful. Furthermore, this approach acknowledges that the environmental variables influencing whether species are present and whether they are abundant may not be identical (Van Horne, 1983; Johnston *et al.*, 2015; Tingley *et al.*, 2016). Recall that

Table 1. Multiscale variables tested for inclusion in *Mya* presence–absence and abundance models.

Variable	Calculation method	Method source
Bathymetry	Primary data	–
Eastness	TASSE	Lecours *et al.* (2017a)
Northness	TASSE	Lecours *et al.* (2017a)
RDMV[a]	TASSE	Lecours *et al.* (2017a)
SD[b]	TASSE	Lecours *et al.* (2017a)
Slope	TASSE	Lecours *et al.* (2017a)
Broad BPI[c]	BTM	Walbridge *et al.* (2018)
Fine BPI[d]	BTM	Walbridge *et al.* (2018)
Surface area	BTM	Walbridge *et al.* (2018)
Rugosity	BTM	Walbridge *et al.* (2018)
Ruggedness	BTM	Walbridge *et al.* (2018)
Curvature	Spatial analyst toolbox	ESRI ArcGIS
Profile curvature	Spatial analyst toolbox	ESRI ArcGIS
Plan curvature	Spatial analyst toolbox	ESRI ArcGIS
Backscatter	Primary data	–
ΔBackscatter	Focal statistics	Diesing and Stephens (2015)
Mud proportion	BRT model	Misiuk *et al.* (2018)
Sand proportion	BRT model	Misiuk *et al.* (2018)
Gravel proportion	BRT model	Misiuk *et al.* (2018)

[a]Relative difference to the mean value; a unitless measure of local topographic position.
[b]Standard deviation of bathymetry values in a local neighbourhood.
[c]Broad benthic position index; inner radius of 15, outer radius of 50.
[d]Fine benthic position index; inner radius of 1 and outer radius of 20.

Table 2. Underwater image samples used for abundance and presence–absence modelling datasets.

		2015 survey		
	Siferd (2005)	"Medium" quality images	"High" quality images	Total
Abundance (*n*)	1 813	–	172	1 985
Presence–absence (*n*)	1 827	274	172	2 273

abundance was quantified for all "high" quality video data collected in 2015 and all of Siferd's (2005) data, yet "medium" quality data were only sufficient for presence–absence. Images that overlapped the MBES data coverage were therefore used to produce two separate modelling datasets: presence–absence observations from Siferd's (2005) images along with "medium" and "high" quality images from the 2015 survey (n = 2273), and abundance observations from Siferd's (2005) images along with only "high" quality images from the 2015 survey (n = 1985; Table 2). We thus created separate models of presence–absence and abundance using the two datasets, and ultimately combined them for a single ensemble map prediction.

The boosted regression trees (BRTs) machine learning algorithm has been shown to consistently perform well compared to other SDM techniques due to its flexibility for fitting non-parametric environmental relationships and its robustness to noisy data (Olden *et al.*, 2008; Franklin, 2009). Reiss *et al.* (2015) discussed how machine learning techniques can outperform regression-based models at predicting a quantitative response, such as abundance. BRTs were trained using the "gbm.step" function in the R package "dismo" (Hijmans *et al.*, 2017). A Bernoulli deviance loss function was used for the *Mya* presence–absence model and Poisson deviance was used for abundance. Ten stochastic models were initially trained for each dataset (presence–absence and abundance) using all multiscale variables to explore individual variable contributions to the models. BRTs can return information on the relative contribution of each variable to the model, and these were used to rank their importance for predicting the presence–absence and abundance of *Mya*. Spearman's rank correlation was then assessed between all variables, and when variables had correlation $\rho \geq 0.7$, the variable of lower rank was removed (Gottschalk *et al.*, 2011; Millard and Richardson, 2015; Jarnevich *et al.*, 2017). Retained variables for both datasets were used in the full presence–absence and abundance models.

The results of the presence–absence model were probabilities of occurrence for *Mya* at a given location from 0 to 1; these were converted to presences and absences using a threshold probability, above which *Mya* were predicted as "present" and below which were predicted as "absent." We selected the threshold that maximized the cross-validated accuracy of abundance predictions (see Model evaluation section). The results of the abundance model were predicted densities of *Mya* individuals per m^2. Abundance predictions were multiplied by predicted occurrence of *Mya* (0 or 1), resulting in abundance predictions conditional on presence. Both models were predicted over the full extent of the environmental data.

Model evaluation

An important step in SDM is evaluating model performance (Franklin, 2009), which is commonly based on assessing a model's ability to generalize to test data that were not used to train it.

There are several methods for obtaining independent test data. The most obvious, and arguably most robust (cf. Hijmans, 2012), is to collect an independent test sample dataset (Araújo and Guisan, 2006; Elith *et al.*, 2006); but this is often not feasible in the marine realm. A common approach is thus to withhold a proportion of the sample data from model training (e.g. 25%) to test predictive performance. A more robust approach is cross-validation, in which the sample data are randomly partitioned into k sets (e.g. 10), $k-1$ of which are used to train a given model fold, with the excluded set withheld for testing. This is repeated over k folds that are subsequently averaged for prediction and model evaluation. Using this method, all data are used to both train and test a model. When $k = n$ (the total number of samples), each sample in the dataset is withheld in turn to test model predictions—known as "leave-one-out cross-validation" (LOO CV; Hastie *et al.*, 2009).

To conduct an unbiased assessment of accuracy we used a spatial (buffered) LOO CV (SLOO CV; Le Rest *et al.*, 2014; Valavi *et al.*, 2018). Using this approach, the first sample point in the dataset is withheld for testing and a spatial buffer is placed around it, up to a distance beyond which the effects of spatial autocorrelation are negligible. Any sample points falling within the buffer are also omitted from the model training fold. The model is trained using all remaining sample points, and the value at the withheld site is predicted. The process is then repeated for each point in the dataset, and the performance metrics from all sites are averaged. SLOO CV is an effective method for evaluating model performance when samples are not spatially independent, and is flexible with regards to clustered or irregular sampling compared to other methods (e.g. blocking; Roberts *et al.*, 2017).

The distance of the SLOO CV buffer was determined by variogram analysis. We investigated spatial structure in ESRI ArcGIS Pro v.2.3 by calculating the average length of all sample transects and the average distance between transects, using their mean centre. The average distance between samples within transects was estimated given the number of samples and transect length. Isotropic and directional variograms were generated for *Mya* abundance model residuals using the "*gstat*" package in R (Pebesma, 2004; Gräler *et al.*, 2016), and variogram models were fit using the ESRI ArcGIS Geostatistical Wizard to estimate spatial autocorrelation. The Geostatistical Wizard was also used to test for anisotropy and fit directional variograms. The variogram model major range was used as a buffer distance for SLOO CV, beyond which the effects of autocorrelation on the model results were assumed to be negligible (Wagner and Fortin, 2005; Roberts *et al.*, 2017).

The area under the receiver operating characteristic curve (AUC) was calculated to measure the threshold-independent accuracy of the presence–absence model, and the correct classification rate and Cohen's kappa were used to measure the threshold-dependent accuracy. The linear correlation between observed and predicted *Mya* abundance, conditional on presence, was assessed using Pearson's correlation coefficient, and non-linear correlation was assessed using Spearman's coefficient. The mean absolute error (MAE) was calculated between observed and predicted values to estimate the magnitude of error in modelled predictions, and the percentage of variance explained (VE) was calculated to estimate the error in predictions relative to the variance in the dataset. Predictive accuracy estimates from the SLOO CV were compared to estimates from the internal 10-fold CV (cross-validation) within the "gbm.step" function to determine if apparent accuracy was inflated by spatial autocorrelation, and if so, the magnitude of inflation for each statistic.

Results

Response to environmental variables

Mya were prevalent in the dataset, occurring in 84% of images analysed. Eleven variables were selected to model their presence and absence after correlation reduction (Figure 3). Bathymetry was the most important variable, accounting for over 66% of the explanatory power in the presence–absence model. Image observations and the partial response of *Mya* to bathymetry suggested that they generally did not occur deeper than 70 m. The Δbackscatter variable suggested by Diesing and Stephens (2015) was the second most important, and the partial response plot showed an inverse relationship with *Mya* presence—probability of presence was predicted to decrease with increase in local backscatter variability. All other variables provided only minor contribution ($\leq$5%) to the model.

Where present, *Mya* where observed at densities of 2–472 individuals per m^2, and a different suite of 12 non-correlated variables were selected to model their abundance (Figure 4). The northness component of aspect was the most important variable, suggesting *Mya* were abundant on north- and south-facing slopes. The partial response plot of *Mya* to ruggedness suggests that abundance decreases with an increase in terrain variability. The response plot to bathymetry showed a decrease in abundance with increasing water depth greater than ~40 m. Many of the remaining variables displayed complex relationships with *Mya* abundance, but it was generally highest at intermediate levels of backscatter and a low mud proportion.

Statistical modelling and prediction

High probabilities of presence were predicted along the coast throughout much of the study area, while absences were generally predicted further from the coast, in deeper water (Figure 5). The most extensive patches of suitable *Mya* habitat were predicted in the southern half of Broughton Channel. Moderate probabilities near Qikiqtarjuaq and in the northernmost part of the study area indicate greater uncertainty in predicted habitat suitability. Extensive areas of low probability further from the coasts, and in the east–west-oriented Kingnelling Fjord indicate unsuitable habitat with low uncertainty.

A 0.61 probability threshold was selected based on maximizing the accuracy of the combined presence–absence and abundance predictions. Therefore, abundance was predicted for locations where the probability of presence exceeded 0.60, while absence was predicted for all other locations. The highest abundances (>200 individuals per m^2) were predicted in <50 m water depth in southwest and southeast Broughton Channel (Figure 6a), on the southern shore at the mouth of Kingnelling Fjord (Figure 6b), and directly south of Broughton Island. Substantial populations were also predicted near the southern shore of Kingnelling Fjord, southeast of Broughton Island, throughout most of the nearshore area of Broughton Channel, and in patches northwest of Broughton Channel. Moderate abundances generally surrounded areas of higher abundance south of Broughton Island and in Broughton Channel, and abundances were predicted low outside of these areas, near the limits of suitable habitat. *Mya* were predicted to be absent in >70 m water depth.

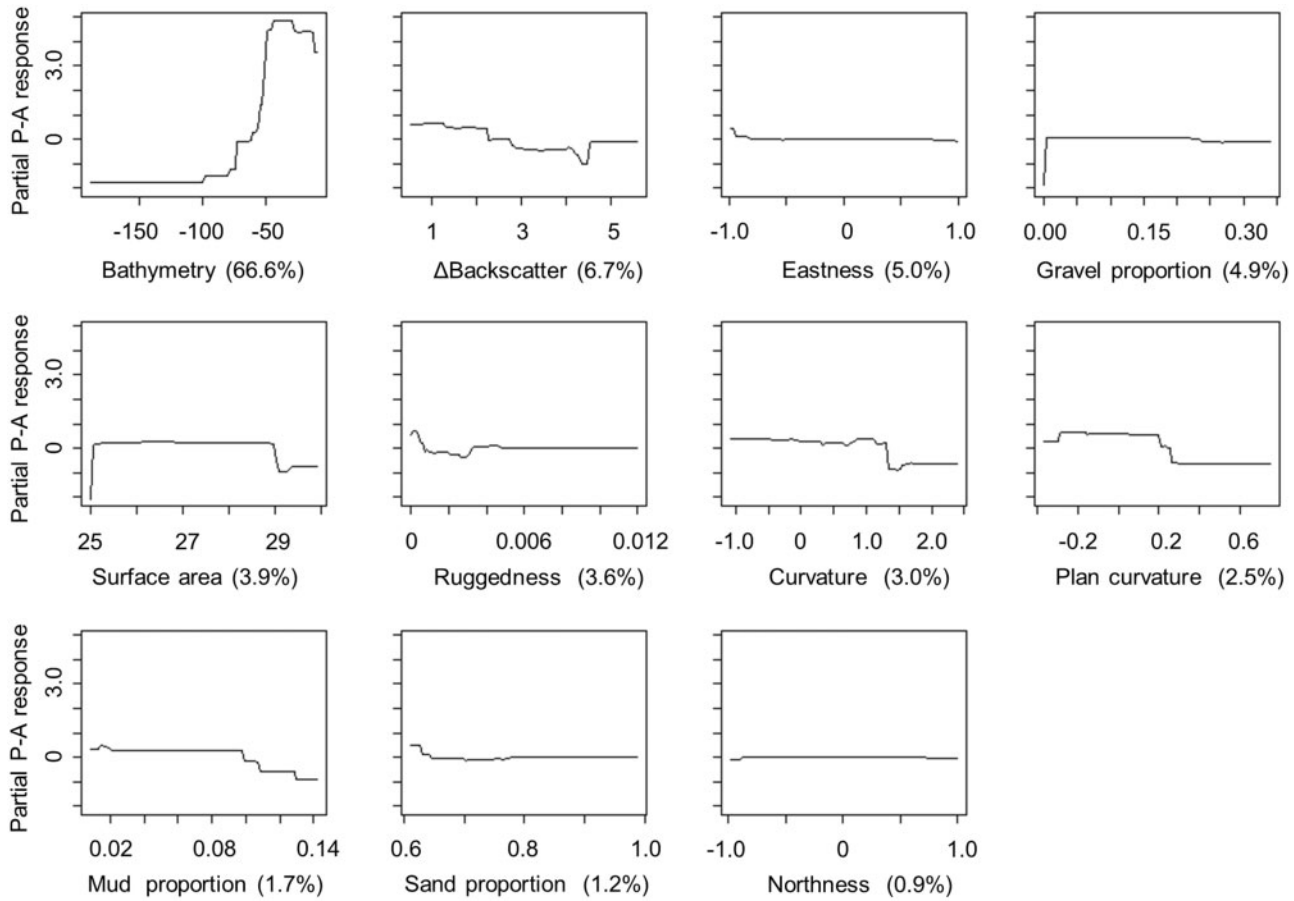

Figure 3. *Mya* log-odds scale presence–absence partial response plots with percent contribution of explanatory variables (Table 1).

Model evaluation

Image sample transects were ∼30 m long on average, and the average distance between samples within a transect was estimated at ∼3 m. On average, transects were spaced ∼128 m apart. We calculated the empirical variogram for the *Mya* abundance model residuals using 15 m lags and also tested for anisotropy using directional variograms. The abundance residuals exhibited anisotropy oriented at 69°, and a circular variogram model was fit with a major range of 60 m (Figure 7), suggesting that the residuals were autocorrelated up to that distance at this scale. Therefore, nearly all samples within a given transect were expected to contain residual spatial autocorrelation.

Based on the directional variogram major range, SLOO CV was used with a 60 m buffer to evaluate presence–absence and abundance models to estimate non-biased predictive performance. The presence–absence model had a correct classification rate of 92%, and AUC and kappa values of 0.87 and 0.59, respectively. Pearson's and Spearman's correlation coefficients between observed and predicted abundances were $r = 0.55$ and $\rho = 0.64$, respectively; MAE and VE values were 44.07 and 0.27. The 10-fold CV estimates of correlation for the abundance model were $r = 0.78$ and $\rho = 0.81$, and MAE and VE were 31.57 and 0.61 (Table 3). The combined abundance-conditional-on-presence predictions evaluated using SLOO CV were slightly more accurate than abundance alone, with $r = 0.55$, $\rho = 0.64$, MAE = 43.54, and VE = 0.27, yet they included "zero" predictions, unlike the abundance model in isolation (Figure 6).

Discussion

The presence and absence of *Mya* was predicted primarily by bathymetry (Figure 3). The partial response plot showed a strong decrease in likelihood of presence at depths >50 m, confirming findings by Ellis (1960) and Siferd (2005). It is likely that bathymetry is a proxy for several variables that define the ecological niche for *Mya* at this depth, possibly such as light, temperature, food availability, or water chemistry. Probability of presence also had a negative relationship with Δbackscatter suggesting that *Mya* are more likely to inhabit areas of relatively homogenous seabed hardness, but this only contributed marginally to the presence–absence model. The remaining topographic and substrate variables made only minor contributions to the model, suggesting that they may exercise subtle influence on the suitability of *Mya* habitat, yet may not form distinct environmental boundaries.

However, *Mya* abundance predictions were influenced considerably by several topographic and substrate variables (Figure 4). Surprisingly, the northness component of seabed aspect was the most important variable in predicting abundance. This may be caused by correlation with other important environmental factors such as bottom currents, which control the flow of food to benthic filter feeders (Tong *et al.*, 2016; Lacharité and Metaxas, 2018), or with local geomorphic features where *Mya* were observed, such as portions of east–west-oriented Kingnelling Fjord. Benthic SDM commonly rely on such surrogates to represent oceanographic information, as primary variables are seldom available. Down-scaled oceanographic models have the potential

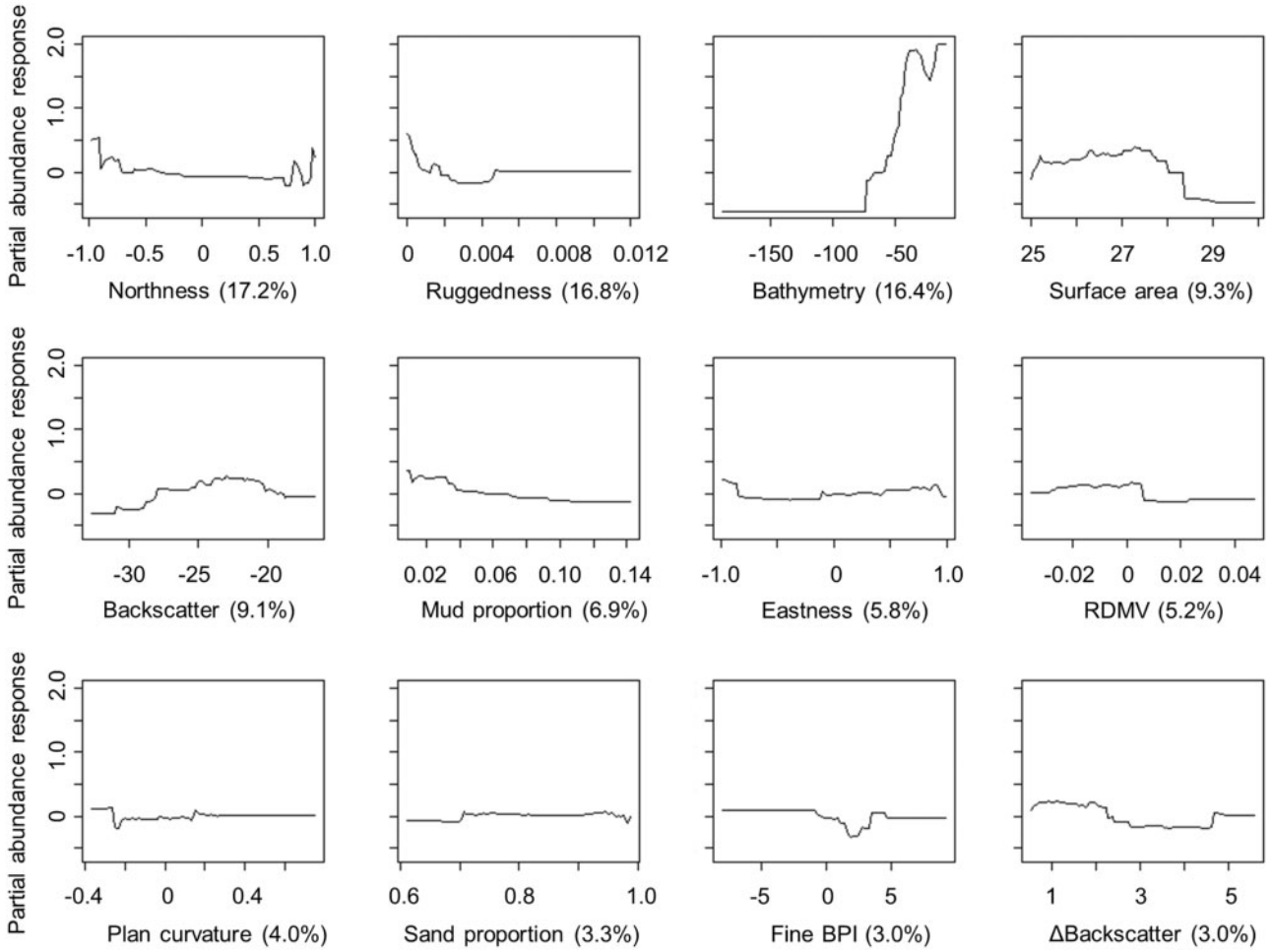

Figure 4. *Mya* log scale abundance partial response plots and percent contribution of explanatory variables (Table 1).

to provide important habitat information for benthic filter feeders, yet their joint use with terrain variables has not been thoroughly explored.

Bathymetric ruggedness was the second most important variable for predicting abundance; it is derived from the three-dimensional variability in terrain orientation. The partial response plot showed that *Mya* were more abundant in low ruggedness areas, but interpretation of this variable requires caution. Though comparable for most of the mapped area, ruggedness measurements appeared to differ between portions of the dataset derived from different MBES systems in the deepest part of the survey, near the mouth of Kingnelling Fjord. It is unlikely that this seriously impacted the statistical analysis because this deep area was not ground-truthed, and because BRTs are effective at ignoring noisy or unimportant data (Elith *et al.*, 2008), yet it demonstrates how combining MBES datasets from different sources could potentially lead to error. Possible sources of discrepancy between datasets include noise in the bathymetry data from acquisition that was amplified in the ruggedness measure (Lecours *et al.*, 2017b), error in the data caused by mapping near the depth limits of the echosounder, and differences in MBES system parameters such as beam width and operating frequency. The response of *Mya* abundance to bathymetry was similar to that of presence–absence, showing a decline in abundance with increasing depth. *Mya* had a specific response to seabed substrate—favouring areas of moderate backscatter intensity (i.e. seabed hardness) and low mud proportion, which was also suggested by Pfitzenmeyer (1972) and Abraham and Dillon (1986). Anecdotal observation generally supports these predictions, with sandy or mixed sandy/gravelly substrates appearing to contain the highest abundances.

The use of data collected over multiple years introduced the potential for temporal error in predicted abundance. Though major changes to the broad seabed morphology or current regimes are unexpected over a 10-year period, other external factors could affect the *Mya* population. Icebergs regularly scour the seabed near Qikiqtarjuaq and become grounded in shallow parts of the north–south-oriented channel—sometimes for multiple years. This could locally impact clam populations and is difficult to determine. Predation and harvest may also exert fine-scale influence on the abundance of the species. Walrus are a common predator of clams, and the small but active subsistence clam fishery in Qikiqtarjuaq operates in locations up to 20 m water depth throughout the area. Though these temporal components have the potential to introduce error to a multi-year study, the value of the extensive combined dataset likely outweighs the associated temporal error.

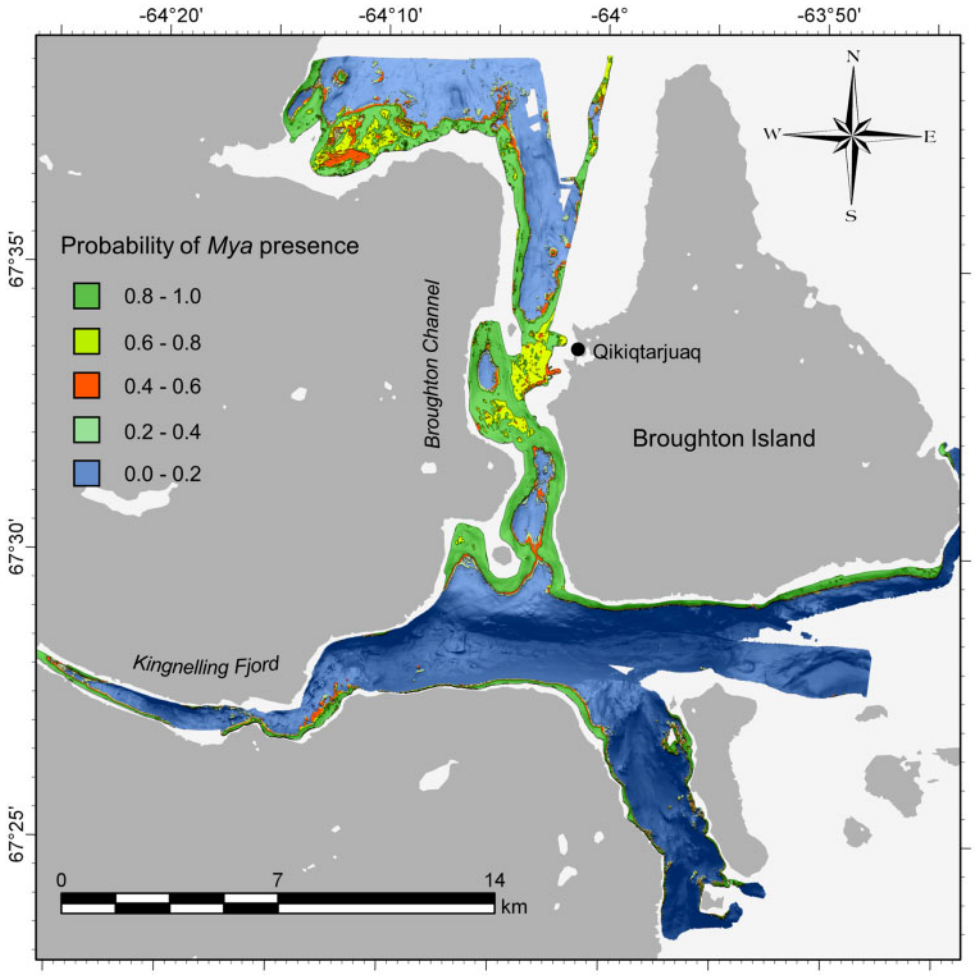

Figure 5. Predicted probability of *Mya* presence near Qikiqtarjuaq. Basemaps were obtained from the Canadian Land Cover GeoBase Series, containing information licenced under the Open Government Licence—Canada.

The *Mya* sample dataset contained "zero" values where no individuals were observed, and the abundance model alone was unable to reliably predict these. For instance, the model predicted decreased abundance at depths >70 m, but not necessarily absence. By combining the model of abundance with binary presence–absence predictions, *Mya* were predicted absent in areas of unsuitable habitat while leaving predictions of abundance intact where habitat is suitable. This approach is similar to the use of a parametric conditional or hurdle model for zero-inflated count data (Welsh *et al.*, 1996; Martin *et al.*, 2005), yet differs at the step at which maps of presence–absence and abundance are combined by using the threshold of occurrence to determine where abundance will be predicted (Rooper *et al.*, 2016), rather than using the product of the abundance estimate and the probability of presence (Barry and Welsh, 2002). This produces exact zero values where habitat is predicted to be unsuitable, and leaves the abundance estimates unaffected where the probability of presence is greater than the threshold but <1. This approach acknowledges that the environmental factors determining if *Mya* are present and whether they are abundant may not be the same, and therefore treats these as distinct phenomena. This is supported by the plots of partial response, which show that bathymetry was the main determining factor in predicting where habitat is potentially suitable (Figure 3), yet seabed topography and substrate properties ultimately influenced predictions of how abundantly *Mya* colonize these areas (Figure 4). The final map (Figure 6) can therefore be conceptualized as a combination of two separate and distinct predictions, rather than as a hurdle model.

In addition to providing more realistic predictions of *Mya* distribution, the integration of presence–absence and abundance models maximized use of the data. This approach allowed for the use of all "moderate" and "high" quality images for modelling. Furthermore, because SLOO CV tests only one sample point at a time, nearly the full dataset is used for each model evaluation fold. This produces model folds that are expected to be very close to the full model, which was used for the final prediction. It does not require that samples are excluded from analysis, which is a common approach to dealing with spatially autocorrelated data (Dale and Fortin, 2002; Segurado *et al.*, 2006). Though it is important to use information within the modelling dataset as efficiently as possible, large amounts of autocorrelation can potentially produce a pseudo-replication effect (Segurado *et al.*, 2006), meaning that multiple proximal samples add little or no new information to the model. It is worth considering whether the potential for inflation due to autocorrelation outweighs the loss of information caused by sample aggregation or omission.

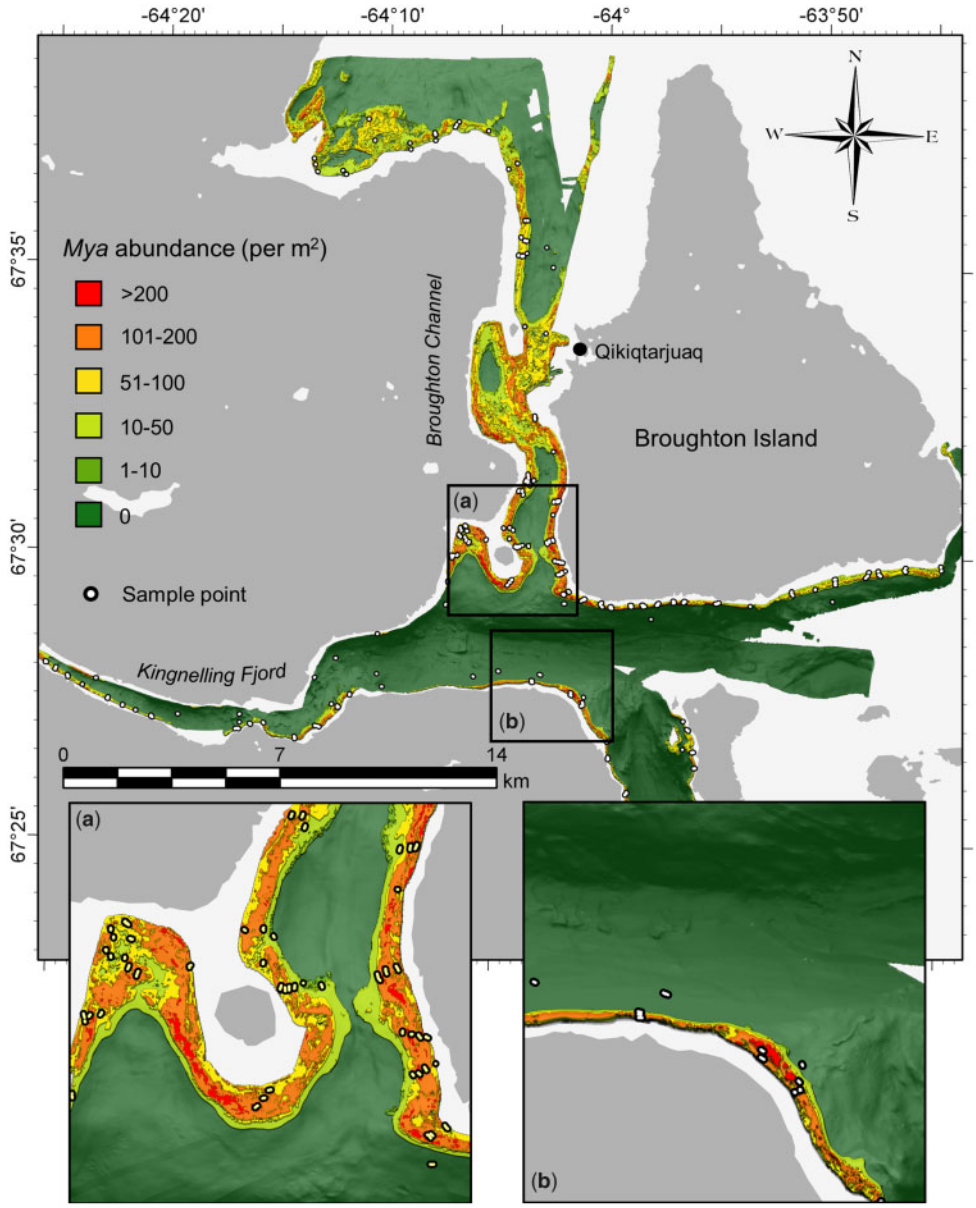

Figure 6. Combined prediction of *Mya* abundance, conditional on presence, near Qikiqtarjuaq, with insets (a) at the southern part of Broughton Channel, and (b) on the southern shore at the mouth of Kingnelling Fjord. Basemaps were obtained from the Canadian Land Cover GeoBase Series, containing information licenced under the Open Government Licence—Canada.

Spatial leave-one-out CV is a flexible, albeit computationally intensive compromise.

The directional variogram of *Mya* abundance residuals suggested that spatial autocorrelation between samples within ~60 m of on another might be influencing apparent model performance, which would include nearly all samples within a given transect. These spatial properties largely represent those of the 2005 sample data, which comprised the majority of the combined dataset, though the 2015 survey was designed to be similar. The difference in performance between spatially dependent (internal 10-fold CV) and independent (SLOO CV) model evaluations confirmed that this bias inflated all measures of apparent predictive performance substantially (Table 3). Using 10-fold CV for evaluation, the abundance model seems highly accurate at both linear and non-linear monotonic prediction ($r = 0.79$, $\rho = 0.81$), with an average error that was 39% of the variance in the observed data (VE = 0.61). Spatially independent evaluation (SLOO CV), however, suggested substantially weaker linear and non-linear monotonic correlation ($r = 0.55$, $\rho = 0.64$), and an average error that was 73% of the variance in the observed data (VE = 0.27).

These results have applied relevance for managing the clam fishery in Qikiqtarjuaq, but the methods highlight several important concepts for marine SDM. Combining abundance with presence–absence predictions increased the ability to distinguish between suitable and non-suitable habitat by incorporating predicted absences that were not available using the abundance model. This also utilized a greater proportion of the sample

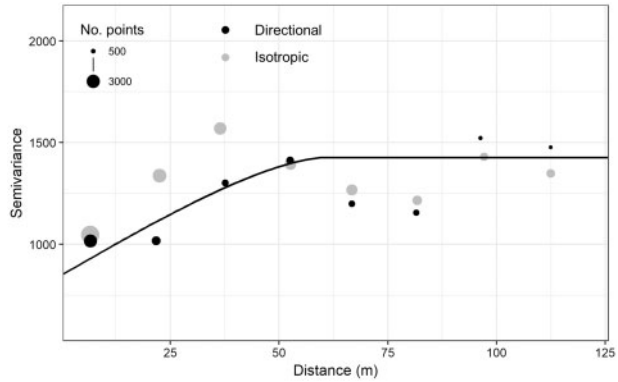

Figure 7. Isotropic and directional (69°) variograms of *Mya* abundance residuals with directional circular variogram model (nugget = 850, sill = 576, range = 60).

Table 3. Performance of abundance model estimated using SLOO CV with a 60 m buffer, internal CV from the "gbm.step" function, and estimate of inflation caused by spatial autocorrelation bias.

	Abundance SLOO CV	Abundance 10-fold CV	Apparent inflation
Pearson *r*	0.55	0.78	0.23
Spearman ρ	0.64	0.81	0.17
MAE	44.07	31.57	12.69
VE	0.27	0.61	0.34

dataset, which were of sufficient quality to determine presence or absence of *Mya* but not abundance. Using a spatial LOO CV demonstrates that failing to account for spatial autocorrelation when evaluating SDMs can substantially inflate estimates of predictive performance. Transect sampling is common in marine science, yet the spatial characteristics of these data and the effects they have on benthic habitat maps are often not considered (but some notable exceptions include Kendall *et al.*, 2005; Foster *et al.*, 2014; Perkins *et al.*, 2019). Furthermore, SLOO CV may be useful for limited datasets where subsampling would constrain the predictive ability of the model, as it does not require large subsampling.

Conclusions

Results suggest that although bathymetry is the primary limiting factor to *Mya* habitat, seabed topography, morphology, and substrate properties jointly predict how abundantly they occur within the appropriate depth range. Different environmental variables appear to influence whether *Mya* are present, and whether they are abundant, reinforcing that relationships between habitat suitability and species abundance can be non-linear. Therefore, when abundance models fail to predict species absence, combined approaches such as mixture and hurdle models that are more flexible at modelling zero values may be useful (Mullahy, 1986; Welsh *et al.*, 1996).

We found that nearly all samples within a sample transect were spatially autocorrelated, which inflated estimates of model accuracy substantially. These results demonstrate that the spatial dependence of sample points can impact the interpretation of model quality, and this reinforces the importance of a non-biased evaluation. This is especially relevant in the marine realm, where transect data are common, yet the spatial qualities of the data are often not considered. At the very least we recommend exploring the spatial structure of ground-truth data as a compulsory step in marine SDM. This can only help to improve the transparency of model quality and limitations for map users.

Acknowledgements

Thanks to Tim Siferd (DFO) for sharing data that was used for this analysis, to Beth Cowan, Manasie Kendall, and Emilie Novaczek for assistance with data collection and lab work, and to Jonah Keyookta for boat operation and expertise, guiding, interpretation, and project support. We also appreciate the effort of two reviewers who provided especially thorough comments—we feel their feedback has greatly improved the quality of this work. Finally, our deepest thanks to the community of Qikiqtarjuaq for their support and for the opportunity to conduct this research.

Funding

This work was supported by the Government of Nunavut, Department of Environment, Fisheries and Sealing Division and ArcticNet (268150-2011).

References

Abraham, B. J., and Dillon, P. L. 1986. Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (mid-Atlantic)—softshell clam. U.S. Fish and Wildlife Service Biological Report 82(11.68). U.S. Army Corps of Engineers, TR EL-82-4.

Aitken, A. E., Risk, M. J., and Howard, J. D. 1988. Animal–sediment relationships on a subarctic intertidal flat, Pangnirtung Fiord, Baffin Island, Canada. Journal of Sedimentary Research, 58: 969–978.

Araújo, M. B., and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33: 1677–1688.

Bahn, V., and McGill, B. J. 2013. Testing the predictive performance of distribution models. Oikos, 122: 321–331.

Barry, S. C., and Welsh, A. H. 2002. Generalized additive modelling and zero inflated count data. Ecological Modelling, 157: 179–188.

Brigham, J. K. 1983. Stratigraphy, amino acid geochronology, and correlation of Quaternary sea-level and glacial events, Broughton Island, Arctic Canada. Canadian Journal of Earth Sciences, 20: 577–598.

Brown, C. J., Sameoto, J. A., and Smith, S. J. 2012. Multiple methods, maps, and management applications: purpose made seafloor maps in support of ocean management. Journal of Sea Research, 72: 1–13.

Brown, C. J., Smith, S. J., Lawton, P., and Anderson, J. T. 2011. Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuarine, Coastal and Shelf Science, 92: 502–520.

Clark, J. S., Gelfand, A. E., Woodall, C. W., and Zhu, K. 2014. More than the sum of the parts: forest climate response from joint species distribution models. Ecological Applications, 24: 990–999.

Dale, M. R. T., and Fortin, M.-J. 2002. Spatial autocorrelation and statistical tests in ecology. Écoscience, 9: 162–167.

Diesing, M., and Stephens, D. 2015. A multi-model ensemble approach to seabed mapping. Journal of Sea Research, 100: 62–69.

Dolan, M. F. J. 2012. Calculation of slope angle from bathymetry data using GIS—effects of computation algorithm, data resolution and analysis scale. NGU Report, 2012.041. Geological Survey of Norway, Trondheim, Norway.

Dolan, M. F. J., and Lucieer, V. L. 2014. Variation and uncertainty in bathymetric slope calculations using geographic information systems. Marine Geodesy, 37: 187–219.

Drew, C. A., Wiersma, Y. F., and Huettmann, F. (Eds). 2011. Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications. Springer, New York.

Dyke, A. S., Andrews, J. T., and Miller, G. H. 1982. Quaternary geology of Cumberland Peninsula, Baffin Island, District of Franklin. Geological Survey of Canada, Memoir 403. Ottawa.

Elith, J., Graham, C. H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J. *et al.* 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29: 129–151.

Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology, 77: 802–813.

Ellis, D. V. 1960. Marine infaunal benthos in Arctic North America. Technical Paper, 5. Arctic Institute of North America.

Forbes, D. L., and Taylor, R. B. 1994. Ice in the shore zone and the geomorphology of cold coasts. Progress in Physical Geography: Earth and Environment, 18: 59–89.

Foster, S. D., Hosack, G. R., Hill, N. A., Barrett, N. S., and Lucieer, V. L. 2014. Choosing between strategies for designing surveys: autonomous underwater vehicles. Methods in Ecology and Evolution, 5: 287–297.

Franklin, J. 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge.

Fulton, R. J. 1995. Surficial materials of Canada [map]. Geological Survey of Canada, "A" Series Map 1880A. Natural Resources Canada.

Gottschalk, T. K., Aue, B., Hotes, S., and Ekschmitt, K. 2011. Influence of grain size on species–habitat models. Ecological Modelling, 222: 3403–3412.

Gräler, B., Pebesma, E., and Heuvelink, G. 2016. Spatio-temporal interpolation using gstat. The R Journal, 8: 204.

Guisan, A., and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135: 147–186.

Harris, P. T., and Baker, E. K. 2012. Why map benthic habitats? In Seafloor Geomorphology as Benthic Habitat: Geohab Atlas of Seafloor Geomorphic Features and Benthic Habitats, pp. 3–22. Ed. by P. T. Harris and E. K. Baker. Elsevier, Amsterdam.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.

Hattab, T., Lasram, F. B. R., Albouy, C., Sammari, C., Romdhane, M. S., Cury, P., Leprieur, F. *et al.* 2013. The use of a predictive habitat model and a fuzzy logic approach for marine management and planning. PLoS One, 8: e76430.

Hewitt, R. A., and Dale, J. E. 1984. Growth increments of modern *Mya truncata* L. from the Canadian Arctic, Greenland, and Scotland. Current research: part B, Geological Survey of Canada, paper no. 84–1B: 179–186.

Hijmans, R. J. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology, 93: 679–688.

Hijmans, R. J., Phillips, S. J., and Elith, J. 2017. dismo: Species Distribution Modeling. R package version 1.1–4. https://CRAN.R-project.org/package=dismo (last accessed 17 February 2019).

Jarnevich, C. S., Talbert, M., Morisette, J., Aldridge, C., Brown, C. S., Kumar, S., Manier, D. *et al.* 2017. Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: an example with background selection. Ecological Modelling, 363: 48–56.

Johnston, A., Fink, D., Reynolds, M. D., Hochachka, W. M., Sullivan, B. L., Bruns, N. E., Hallstein, E. *et al.* 2015. Abundance models improve spatial and temporal prioritization of conservation resources. Ecological Applications, 25: 1749–1756.

Kendall, M. S., Jensen, O. P., Alexander, C., Field, D., McFall, G., Bohne, R., and Monaco, M. E. 2005. Benthic mapping using sonar, video transects, and an innovative approach to accuracy assessment: a characterization of bottom features in the Georgia Bight. Journal of Coastal Research, 216: 1154–1165.

Lacharité, M., and Metaxas, A. 2018. Environmental drivers of epibenthic megafauna on a deep temperate continental shelf: a multiscale approach. Progress in Oceanography, 162: 171–186.

Lecours, V. 2017. Terrain attribute selection for spatial ecology (TASSE). ArcGIS toolbox version 1.1. doi: 10.13140/RG.2.2.15014.52800.

Lecours, V., Devillers, R., Lucieer, V. L., and Brown, C. J. 2017b. Artefacts in marine digital terrain models: a multiscale analysis of their impact on the derivation of terrain attributes. IEEE Transactions on Geoscience and Remote Sensing, 55: 5391–5406.

Lecours, V., Devillers, R., Simms, A. E., Lucieer, V. L., and Brown, C. J. 2017a. Towards a framework for terrain attribute selection in environmental studies. Environmental Modelling & Software, 89: 19–30.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology, 74: 1659–1673.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., and Bretagnolle, V. 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Global Ecology and Biogeography, 23: 811–820.

Li, J., Alvarez, B., Siwabessy, J., Tran, M., Huang, Z., Przeslawski, R., Radke, L. *et al.* 2017. Application of random forest, generalised linear model and their hybrid methods with geostatistical techniques to count data: predicting sponge species richness. Environmental Modelling & Software, 97: 112–129.

Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J. *et al.* 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters, 8: 1235–1246.

Millard, K., and Richardson, M. 2015. On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. Remote Sensing, 7: 8489–8515.

Miller, J. 2010. Species distribution modeling. Geography Compass, 4: 490–509.

Misiuk, B., Lecours, V., and Bell, T. 2018. A multiscale approach to mapping seabed sediments. PLoS One, 13: e0193647.

Mullahy, J. 1986. Specification and testing of some modified count data models. Journal of Econometrics, 33: 341–365.

Nunavut Department of Environment—Fisheries and Sealing Division. 2012. Nunavut Coastal Resource Inventory—Iqaluit. https://www.gov.nu.ca/sites/default/files/ncri_iqaluit_en.pdf (last accessed 7 December 2018). 10 and 124 pp.

Olden, J. D., Lawler, J. J., and Poff, N. L. 2008. Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology, 83: 171–193.

Pebesma, E. J. 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30: 683–691.

Perkins, N. R., Hosack, G. R., Foster, S. D., Hill, N. A., and Barrett, N. S. 2019. Spatial properties of sessile benthic organisms and the design of repeat visual survey transects: the influence of spatial properties of sessile benthic organisms, transect relocation, and sampling effort on monitoring outcomes for visual surveys. Aquatic Conservation: Marine and Freshwater Ecosystems, 29: 59–71.

Petersen, G. H. 1978. Life cycles and population dynamics of marine benthic bivalves from the Disko Bugt area of West Greenland. Ophelia, 17: 95–120.

Pfitzenmeyer, H. T. 1972. Tentative outline for inventory of molluscs: *Mya arenaria* (soft-shell clam). Chesapeake Science, 13: s182–184.

Porskamp, P., Rattray, A., Young, M., and Ierodiaconou, D. 2018. Multiscale and hierarchical classification for benthic habitat mapping. Geosciences, 8: 119.

Reiss, H., Birchenough, S., Borja, A., Buhl-Mortensen, L., Craeymeersch, J., Dannheim, J., Darr, A. *et al.* 2015. Benthos distribution modelling and its relevance for marine ecosystem management. ICES Journal of Marine Science, 72: 297–315.

Ridout, M., Demétrio, C. G. B., and Hinde, J. 1998. Models for count data with many zeros. *In* Proceedings of the XIXth International Biometric Conference. International Biometric Society, Cape Town, South Africa.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S. *et al.* 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40: 913–929.

Rooper, C., Sigler, M., Goddard, P., Malecha, P., Towler, R., Williams, K., Wilborn, R. *et al.* 2016. Validation and improvement of species distribution models for structure-forming invertebrates in the eastern Bering Sea with an independent survey. Marine Ecology Progress Series, 551: 117–130.

Segurado, P., Araújo, M. B., and Kunin, W. E. 2006. Consequences of spatial autocorrelation for niche-based models. Journal of Applied Ecology, 43: 433–444.

Siferd, T. 2005. Assessment of a clam fishery near Qikiqtarjuaq, Nunavut. Canadian Technical Report of Fisheries and Aquatic Sciences. Department of Fisheries and Oceans Canada, Winnipeg, MB.

Smith, S. J., Sameoto, J. A., and Brown, C. J. 2017. Setting biological reference points for sea scallops (*Placopecten magellanicus*) allowing for the spatial distribution of productivity and fishing effort. Canadian Journal of Fisheries and Aquatic Sciences, 74: 650–667.

Tingley, M. W., Wilkerson, R. L., Howell, C. A., and Siegel, R. B. 2016. An integrated occupancy and space-use model to predict abundance of imperfectly detected, territorial vertebrates. Methods in Ecology and Evolution, 7: 508–517.

Tong, R., Purser, A., Guinan, J., Unnithan, V., Yu, J., and Zhang, C. 2016. Quantifying relationships between abundances of cold-water coral *Lophelia pertusa* and terrain features: a case study on the Norwegian margin. Continental Shelf Research, 116: 13–26.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G. 2018. BLOCKCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution, 10: 225–232.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. The Journal of Wildlife Management, 47: 893–901.

Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. Journal of Biogeography, 36: 2290–2299.

Wagner, H. H., and Fortin, M.-J. 2005. Spatial analysis of landscapes: concepts and statistics. Ecology, 86: 1975–1987.

Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. J. 2018. Unified geomorphological analysis workflows with Benthic Terrain Modeler. Geosciences, 8: 94.

Warton, D. I. 2005. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics, 16: 275–289.

Welsh, A. H., Cunningham, R. B., Donnelly, C. F., and Lindenmayer, D. B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling, 88: 297–308.

Wheeler, J. O., Hoffman, P. F., Card, K. D., Davidson, A., Sanford, B. V., Okulitch, A. V., and Roest, W. R. 1996. Geological map of Canada [map]. Geological Survey of Canada, "A" Series Map 1860A. Natural Resources Canada.

Handling editor: Joanna Norkko