INVESTIGATION AND INTEGRATION OF SPATIAL ANALYSES IN BENTHIC HABITAT MAPPING WITH APPLICATION TO NEARSHORE ARCTIC ENVIRONMENTS

By © Benjamin Misiuk

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Geography

Memorial University of Newfoundland

October 2019

St. John's, Newfoundland

Abstract

The field of benthic habitat mapping has entered an era of automated statistical methods that have increased the capacity to produce maps as marine management tools. Spurred by a confluence of advances in acoustic remote sensing, open-source statistical tools, GIS, and computing power, these methods facilitate quick and objective mapping of habitats and physical seabed characteristics. Their performance and accessibility have led to widespread uptake, yet key spatial issues associated with these methods have not fully translated into the benthic habitat mapping workflow. Towards establishing "best practices", this thesis explores the application of several spatial concepts to benthic habitat mapping using three Canadian Arctic case studies.

Relationships between seabed morphology and benthic habitats are well-established. Though recognized as a critical element in the field of geomorphometry, the scale dependence of these relationships is commonly neglected in habitat mapping. Chapter 2 provides evidence of the scale dependence of benthic terrain variables and demonstrates methods for testing and selecting from among many variables and scales for modelling the distribution of sediment grain size near Qikiqtarjuaq, Nunavut.

Given challenges associated with marine data collection that are pronounced in the Arctic, benthic habitat maps commonly utilize multi-year and multisource datasets. Despite apparent advantages, there can be substantial challenges associated with the compatibility and spatial properties of such data. Chapter 3 demonstrates that spatially autocorrelated samples are likely to inflate estimates of predictive performance and uses a spatial

resampling strategy to estimate and correct for inflation in a multi-model Arctic clam habitat map near Qikiqtarjuaq, Nunavut.

Classified seabed maps are a common requirement for marine management and one of two broad approaches are often selected to produce them. Chapter 4 examines differences between classification and continuous modelling approaches in a spatial context to produce classified seabed sediment maps for inner Frobisher Bay, Nunavut. Non-spatial methods failed to indicate whether models could extrapolate to unsampled areas, which was a requirement for this study. When evaluated in a spatial context, the qualities of the classification approach made it more suitable, which was a function of ground-truth dataset characteristics and the predictive goals of the model. Non-spatial techniques may be appropriate for interpolation, but the ability to extrapolate needs to be examined in a spatial context.

Acknowledgments

Francis Misiuk taught me the principles of hard work, which I have attempted to apply to my studies; Lynn Misiuk gave guidance, support, and literature that was instrumental to helping me through this degree. Nicholas Misiuk has always been a creative influence, and has affected the way I approach habitat mapping – it is partly an art. My grandparents, Cecile Misiuk, and Elaine and Charles McLellan, have supported me unconditionally, giving me the confidence to take risks. John Misiuk was a meticulous engineer and craftsman; I have tried to emulate his attention to detail to the best of my ability. This dissertation is dedicated to his memory.

My supervisors, Trevor Bell and Evan Edinger, provided me with more opportunities than were required during this degree and I am grateful. Their support was constant, even during my missteps, which required a great deal of patience. They prioritized my progress and experience as a student alongside the progress of our research, which speaks to their quality as supervisors and their character as people.

Craig Brown and Alec Aitken, my supervisory committee members, also provided more resources and mentorship than was required of them. They were attentive to my progress as a student and were highly involved in all stages of the research presented here. The quality of this dissertation and of my education have been greatly improved thanks to them.

The research conducted in Chapters 2 and 3 was only possible thanks to the support of Jonah Keyookta and the hamlet of Qikiqtarjuaq. Officially, Jonah was our boatman, but he

assumed many additional roles and advocated for our work. Qikiqtarjuaq is a beautiful community and they welcomed us warmly.

Thanks to the Government of Nunavut Department of Environment, Fisheries and Sealing Division, who provided funding and in-kind support for the research presented in this dissertation, and to ArcticNet who also provided funding along with ship time. Thanks to the captain and crew of the RV *Nuliajuk*, who assisted with data collection for each project.

Vincent Lecours encouraged me to pursue my interests and publish my research – I was lucky to have his mentorship. Thanks to Emilie Novaczek for feedback and discussions on research methods, and for coming to dive in the Arctic. I am indebted to Markus Diesing for making his expertise on seabed sediment mapping available to my interrogation. Natalya Dawe provided endless support and great cooking during the most challenging parts of this degree.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALR	Additive Log-Ratio
AUC	Area Under the Receiver Operating Characteristic Curve
BGS	British Geological Survey
BPI	Benthic Position Index
BRT	Boosted Regression Trees
BTM	Benthic Terrain Modeler
CCGS	Canadian Coast Guard Ship
CV	Cross-Validation
DFO	Department of Fisheries and Oceans
EUNIS	European Nature Information System
FMGT	Fledermaus Geocoder Toolbox
GEBCO	General Bathymetric Chart of the Oceans
GIS	Geographic Information System
GPS	Global Positioning System
LOO CV	Leave-One-Out Cross-Validation
MAE	Mean Absolute Error
Maxent	Maximum Entropy
MBES	Multibeam Echosounder
MSP	Marine Spatial Planning
RDMV	Relative Difference to the Mean Value
RV	Research Vessel

SA:PA	Surface Area: Planar Area
SCUBA	Self-Contained Underwater Breathing Apparatus
SD	Standard Deviation
SDM	Species Distribution Model
SLOO CV	Spatial Leave-One-Out Cross-Validation
SR-LOO CV	Spatially Resampled Leave-One-Out Cross-Validation
TASSE	Terrain Attribute Selection for Spatial Ecology
USGS	United States Geological Survey
VE	Variance Explained
VRM	Vector Ruggedness Measure

List of Appendices

1. Introduction and Overview

1.1 Background

Against the backdrop of changing climate and increased global stress on marine ecosystems (Halpern *et al*., 2008), the need for knowledge on the distribution of ocean life has become urgent. One of the greatest challenges in producing this information is the heterogeneity of marine ecosystems. Mapping distributions of marine habitats is critical to managing human impacts on marine ecosystems and maintaining their health, yet there is no "one size fits all" approach to this. It is therefore necessary to develop mapping approaches that are flexible yet robust.

Some common themes have emerged from the recent widespread adoption of marine habitat mapping methods, which have been recognized as fundamental concepts in several reviews of the field (Todd & Greene, 2007; Brown *et al*., 2011; Harris & Baker, 2012; Lecours *et al*., 2015). A suite of marine-specific guiding principles is developing around these concepts. Though a single mapping methodology will not be applicable in all circumstances, broad principles can be defined that provide a general framework around which marine habitat maps are produced.

One such set of concepts are spatial in nature. Marine habitat mapping is a fundamentally spatial endeavor and is becoming increasingly spatially explicit. For example, full coverage spatially continuous map predictions are now taken for granted as the primary products of marine habitat mapping (Brown *et al*., 2011), and are increasingly used as inputs for other

applications (Pickrill & Todd, 2003; Harris & Baker, 2012). These include marine protected area planning and management (e.g., Jordan *et al*., 2005; Copeland *et al*., 2013), conservation target assessment (Ross & Howell, 2013), regional classifications such as identification of seascapes (e.g., Roff *et al*., 2003) and recently, fisheries stock assessment (Smith *et al*., 2017).

The transition from spatially discontinuous samples to spatially continuous maps is rife with assumptions though – for instance, that gradational shifts in marine habitat can be accurately represented using hard-boundary discrete classes (Strong *et al*., 2018), that observed biological and ecological phenomena can be modelled and mapped using physical abiotic surrogates (Huang *et al*., 2011), and that benthic habitats can be mapped by considering them in only two dimensions (but see Duffy & Chown, 2017). Adopting these assumptions introduces some unknown amount of error to a habitat map, as it is a simplification of reality (Smits *et al*., 1999; Foody, 2002). Estimating the magnitude of this error is one of the most important jobs of the map producer (Barry & Elith, 2006), as it can indicate whether assumptions have held to a degree that we may rely on predictions to inform real-world decision-making (Mitchell *et al*., 2018). Spatial characteristics of habitat mapping data, however, can affect the evaluation of map quality (Segurado *et al*., 2006), yet these are often not considered. These issues can be especially insidious in that they potentially remain undetected unless explicitly investigated or guarded against.

This dissertation is built around two multi-year studies in the eastern Canadian Arctic. The motivation for these projects was a need for scientific knowledge on the distribution of

seabed habitats for local management near two communities on Baffin Island, Nunavut, Canada. Alongside each locally relevant primary mapping objective though, this dissertation explores the integration of specific spatial analyses to the habitat mapping workflow. Each analysis is generally linked to its corresponding study, for instance through data characteristics or limitations, yet the concepts are generalizable. This is discussed in the conclusion of each manuscript and is illustrated in the cumulative manner by which each concept is carried forth as components to following chapters.

1.2 Research Purpose and Gap

Access to marine resources is vital to northern lifestyles, but exploitation simultaneously impacts marine ecosystems. Historically, resource use in the Arctic has focused on subsistence hunting and fishing but now also includes recreational and commercial harvests, while additional marine activities such as shipping and coastal infrastructure development stimulate economic growth in remote communities. These activities may not access marine ecosystems directly, but still impact them (Harris, 2012). Coupled with rapidly changing climate in the eastern Canadian Arctic (Bell & Brown, 2018), anthropogenic activities have the potential to cumulatively impact the fine- and broad-scale distribution and health of marine ecosystems (Halpern *et al.*, 2008; Cheung *et al.*, 2009; Pinsky *et al.*, 2013). Because coastal resource use is central to life in the Arctic, it is necessary to balance the needs of resource users with conservation of marine life.

Foremost among approaches for achieving this balance in multi-use coastal systems is marine spatial planning (MSP). Using this framework, spatial information on the distribution of ecologically important areas is analysed in the context of human activity in order to allocate marine space (Ehler & Douvere, 2009). The goal of this ecosystem-based approach is to meet social and economic objectives without compromising marine biodiversity or ecosystem health. Though not the only approach for balancing competing interests in multi-use marine systems, MSP has been widely adopted as a model for how spatial information can be implemented effectively for ecosystem-based management (Foley *et al*., 2010).

Ecosystem-based management frameworks such as marine spatial planning require spatially explicit information from which to delimit ecologically important areas, and benthic habitat mapping has been widely adopted as a method for generating this knowledge (Cogan *et al*., 2009; Baker & Harris, 2012). Benthic habitat mapping is broadly defined as "the use of spatially continuous environmental data sets to represent and predict biological patterns on the seafloor" (Brown *et al*., 2011). "Habitat mapping" is therefore a general term that refers to the methodology used to predict distributions of species (e.g., Brown *et al*., 2012), communities (e.g., Pesch *et al*., 2007), biotopes (e.g., Buhl-Mortensen *et al*., 2009), and abiotic features as surrogates for ecosystem components (e.g., Huang *et al*., 2011; Siwabessy *et al*., 2018). These applications assume that the "habitat" response is a function of any number of explanatory environmental variables, and they rely on assumed or modelled relationships to predict the response at unsampled locations.

Modern benthic habitat maps commonly utilize seabed environmental variables that are derived from multibeam echosounder (MBES) data and accessed via geographic information systems (GIS). These include bathymetry, backscatter, and a suite of variables that describe seabed terrain, such as slope, curvature, roughness, and relative position (Lecours *et al*., 2017a; Walbridge *et al*., 2018). While not necessarily direct indicators of benthic habitat, these morphometric descriptors are often used as indirect surrogates to represent topographic characteristics such as bottom complexity and exposure to currents (St-Onge & Miron, 2007; Guinan *et al*., 2009; McArthur *et al*., 2010; Rengstorf *et al*., 2012), which can be important benthic habitat factors (Frederiksen *et al*., 1992; Davies *et al*., 2009). Terrain variables are calculated using geomorphometry – the science of quantitative land-surface analysis (Pike *et al*., 2009), which is only recently feasible in the marine realm thanks to full-coverage digital bathymetric models (Lecours *et al.*, 2016). These variables are increasingly applied using automated statistical methods – often "machine learning" statistical models (e.g., Maxent, Random Forest, Boosted Regression Trees, artificial neural networks; Reiss *et al*., 2015), or object-based approaches (e.g., Diesing *et al*., 2014; Lark *et al*., 2015; Lacharité *et al*., 2018).

The technology and methods used to generate benthic habitat maps have progressed rapidly in the last two decades and continue to do so. Advanced modelling tools have been integrated into popular GIS and statistical software, making them easily accessible to ecologists. These methods have been introduced in the habitat mapping literature (terrestrial and marine) as highly flexible, requiring little to no model tuning, variable selection, or parameterization (e.g., Elith *et al*., 2006; Olden *et al*., 2008). To ecologists,

this one-size-fits-all approach is extremely attractive, as ecological data commonly violate assumptions of parametric models, which generally require extensive tuning and variable selection (Olden *et al*., 2008). The performance and flexibility of modern predictive mapping tools can increase the capacity for producing highly accurate maps, yet the uncritical application of these tools as a result of their perceived flexibility is likely to ignore several important spatial concepts, especially in the marine realm.

1.2.1 Spatial Concepts

The scale-dependence of terrain variables – that the measurement of the variable is dependent on the scale at which is it observed – is well-accepted in terrestrial geomorphometry (Shary *et al*., 2002; Lechner *et al*., 2012). By default, terrain variables in most GIS software are calculated at a single local window of analysis (e.g., 3 x 3 raster cells) using "neighbourhood" statistics. This imposes an arbitrary scale onto the variable, which is dependent on the resolution of the input data and the algorithm by which it is calculated (MacMillan & Shary, 2009), and is therefore not selected by the user. This has been demonstrated in a marine context (e.g., Dolan, 2012; Dolan & Lucieer, 2014), prompting calls for a better integration of spatial scale concepts into the benthic habitat mapping workflow (Hopkins, 2009; Lecours *et al*., 2015).

A second issue is spatial autocorrelation. Because marine sampling is generally costly and inefficient compared to terrestrial sampling, data collection is commonly focused on maximizing sample size – and rightly so, sample size is one of the most important factors

dictating the performance of distribution models (Stockwell & Peterson, 2002; Hernandez *et al*., 2006; Guisan *et al*., 2007). Transect sampling, for example, is an efficient strategy for collecting many observations at a single location, or continuously over some distance (e.g., Foster *et al*., 2014). These samples are likely to contain spatial autocorrelation, which can violate assumptions of statistical models if unaccounted for, potentially leading to inflated estimates of model performance and spurious conclusions (Segurado *et al*., 2006). Again, this issue has been widely discussed and addressed in the ecology, landscape ecology, and geography literatures (e.g., Legendre, 1993; Diniz-Filho *et al*., 2003; Legendre *et al*., 2004; Segurado *et al*., 2006; Drew *et al*., 2011; Millard & Richardson, 2015; Roberts *et al*., 2017), but has been identified as an issue in marine habitat mapping that is commonly ignored (Vierod *et al*., 2014; Lecours *et al*., 2015). Furthermore, challenges associated with data collection in remote locations such as the Arctic encourage the use of opportunistic and multisource datasets, which may contain undesirable spatial characteristics when originally collected for other purposes. In such cases, it may be necessary to explore methods that are robust to the confounding effects of non-independence.

Finally, these spatial issues need to be considered when comparing and choosing between mapping and modelling approaches. There is an important distinction to be made between interpolative and extrapolative prediction, which can ultimately determine the suitability of modelling methods and evaluation criteria (Heikkinen *et al*., 2012; Roberts *et al*., 2017). For instance, spatial autocorrelation can be considered an asset when interpolating using geostatistics or spatially explicit modelling approaches (e.g., Li *et al*., 2010, 2017;

Buscombe & Grams, 2018), yet can be considered a liability when creating models that must extrapolate to new locations (Legendre, 1993; Dormann, 2007). The most common methods currently used to evaluate habitat maps are typically better suited to interpolative prediction (Bell & Schlaepfer, 2016; Roberts *et al*., 2017) and more attention needs to be paid to whether this is appropriate, or whether alternative approaches that evaluate extrapolative performance are required.

In summary, benthic habitat maps have become essential tools for informing management of multi-use coastal systems such as those in the Canadian Arctic, yet several important spatial issues regarding their production are commonly ignored. These issues have been raised, especially in reviews of the field (e.g., Brown *et al*., 2011; Vierod *et al*., 2014; Lecours *et al*., 2015), yet there is a disconnect between highlighting these spatial concepts and implementing corresponding spatial analyses as obligatory steps in the habitat mapping workflow.

1.3 Methods

This dissertation aims to address the disconnect in implementing the three following specific spatial concepts in the habitat mapping workflow by focusing on each, in turn, through case studies in the Canadian Arctic: 1) the importance of scale-specific variable selection, 2) the effects of spatially autocorrelated datasets, and 3) the need for spatially explicit model comparison. These have been recognized and demonstrated – an important step is to place them in an applied context showing how and why they need to be

incorporated in benthic habitat mapping and the potential consequences of neglecting to do so. Therefore, alongside applied results for two Arctic communities, this dissertation provides marine-specific recommendations for applying these spatial concepts in the benthic habitat mapping workflow.

Challenges associated with Arctic research make it a surprisingly suitable setting in which to demonstrate the importance of these concepts. Logistical and environmental limitations common to marine research are often amplified in the Arctic, and especially in challenging fjordic environments (Syvitski & Schafer, 1985). Icebergs and sea-ice cover, for instance, commonly inhibit ground-truth sampling, potentially resulting in unsampled areas that must be mapped via extrapolation. Sampling is further limited temporally by a short open-water period and the logistical and financial expense associated with operation in remote northern locations. These limitations require pragmatic approaches that may increase the likelihood of violating spatial assumptions. Multi-year surveys and opportunistic multisource datasets, for example, can be repurposed and leveraged to supplement limited data. When collected for purposes other than habitat mapping though, multisource datasets may contain undesirable properties, such as non-independent sample design, or a limited range of sampled environments.

1.3.1 Study Areas

The hamlet of Qikiqtarjuaq, and capital city of Iqaluit (Figure 1.1), are two locations in Nunavut where benthic mapping efforts are under way to inform the management of coastal

systems and resources. Case studies in each of these locations were based on locally relevant goals, and mapping projects were designed in response to specific community needs. Several multisource datasets were used to train supervised statistical models to predict the distributions of biotic and abiotic benthic features at each location. In all cases, MBES data were used to generate variables that explain the distributions of species and habitat characteristics observed in the ground truth. Modelling was conducted using machine learning algorithms that are highly relevant given their recent popularity among the benthic habitat mapping community.

Figure 1.1. Northern Canada and study locations on Baffin Island, Nunavut. Note distortion of landmasses with increasing latitude caused by the Mercator projection.

Qikiqtarjuaq is known for its unique diver-based clam harvest (Siferd, 2005). Seabed sediment (Chapter 2) and clam habitat (Chapter 3) maps were produced to support the assessment and potential commercialization of this small-scale fishery. Underwater

photographs were originally collected in 2003 to survey the local population of soft-shelled clams (*Mya* spp.) as part of a fishery stock assessment (Siferd, 2005). Additional video and benthic grab samples were collected between 2013-2015 from a 24-foot freighter canoe (Figure 1.2A) to characterize seabed habitats for the studies in this dissertation. SCUBA diving was also conducted by the field team in 2015 to observe clam habitat and to collect samples (Figure 1.2B, C). The research vessel (freighter canoe) was piloted by Jonah Keyookta – a local hunter and fisherman who is extremely knowledgeable about the study area. MBES data were collected opportunistically by two vessels, the RV *Nuliajuk* and CCGS *Amundsen*, over a four-year period.

Figure 1.2. (A) Freighter canoe used for field work in Qikiqtarjuaq, NU; (B) and (C) SCUBA diving for clam habitat observation and sampling.

Coastal infrastructure development near Iqaluit, including construction of a new deep-water port and potential fibre optic cable installation, is on-going and requires detailed information on the seabed geology of inner Frobisher Bay. Grab samples and underwater

video were collected to characterize surficial seabed sediments (Chapter 4) as part of the larger multidisciplinary project "Integrated Marine Geoscience to Guide Environmental Impact Assessment and Sustainable Development in Frobisher Bay, Nunavut" (Deering *et al*., 2018). Along with ground-truth sampling, targeted multibeam sonar surveys were conducted by the RV *Nuliajuk* (Figure 1.3) over a four-year period. These data were combined with opportunistic multibeam collected over five years by the CCGS *Amundsen* to generate harmonized bathymetry and backscatter mosaics for the research presented in this dissertation, and for other projects aimed at characterizing the seabed geology in this region.

Figure 1.3. RV *Nuliajuk* used for mapping near Qikiqtarjuaq, and sampling and mapping in Frobisher Bay.

1.4 Significance

There are important reasons for demonstrating the implementation of spatial analyses in a benthic habitat mapping context. The first is simply because they are commonly ignored – these case studies demonstrate that the issues are important to consider when producing habitat maps and are not reserved for simulation and theoretical exploration. Ideally, each of these concepts would always be considered when producing a benthic habitat map. The second reason for demonstrating these in an applied marine context is that there are components specific to benthic habitat mapping that are not present in the fields from which its methods were sourced, such as terrestrial habitat mapping, landscape ecology, and terrestrial remote sensing. These can require unique treatment; they include:

1) characteristics of remotely-sensed acoustic data, such as the influence of substantial data artefacts and noise on the selection of data resolution (and therefore scale; Lecours *et al*., 2017b), and the incorporation of multisource harmonized acoustic datasets (e.g., Lacharité *et al*., 2018), wherein the acoustic response of the seabed is highly dependent on MBES system parameters such as operating frequency (Lurton & Lamarche, 2015; Brown *et al*., 2019);

2) the characteristics of seabed terrain variables. Some of these are unique to the marine realm, such as benthic position index (BPI; Lundblad *et al*., 2006), are implemented in marine-specific ways (e.g., Walbridge *et al*., 2018), and all of which contain an implicit scale-dependent third dimension of interaction with the water column (Duffy & Chown, 2017). This has led to the use of terrain variables as surrogates for near-bottom

oceanographic conditions, such as currents, in benthic habitat mapping (McArthur *et al*., 2010; Lecours *et al*., 2015);

3) the limitations of marine ground-truth. These data often contain a non-negligible amount of location inaccuracy due to the difficulties associated with positioning at depth (Harris & Baker, 2012), which can affect the selection of acoustic data resolution, and therefore terrain variable scale (Lecours *et al*., 2016). Some marine sampling methods are also spatially imprecise, such as trawls, in which start and end locations are recorded, but the locations of specific samples captured along the trawl path are unknown. Furthermore, marine data are typically expensive to obtain, promoting the use of multisource and opportunistic datasets, or approaches that prioritize efficiency of data collection over spatial independence.

Because these marine-specific data characteristics all have spatial implications, it is important that they are addressed in a marine context. This increases the relevance of results to the field of marine habitat mapping and allows for specific recommendations – for example that address the use of transect data, marine-specific variables, or multisource acoustic data. Marine habitat map producers regularly encounter challenges associated with these issues. Therefore, throughout this dissertation, I advocate for addressing spatial concepts contextually.

1.5 Organization of Dissertation

In addition to this introductory chapter, this dissertation includes three manuscripts and a summary chapter. Each manuscript focuses on a deliverable habitat mapping product and on incorporating a specific spatial analysis into the methods used to produce it. Each manuscript is a stand-alone paper that has been peer-reviewed and published in the scientific literature, with minor changes in format and content based on the journal style and editorial process. Correspondingly, each chapter contains individual literature reviews, methods, results, and conclusions. A small amount of overlap in introductory material can be expected between the manuscripts as a function of this dissertation style.

Chapter 2 applies a methodology for selecting an appropriate spatial scale for terrain variables used in seabed sediment distribution models near Qikiqtarjuaq, Nunavut. This chapter shows that the arbitrary scale imposed by the data resolution is not necessarily optimal. It further demonstrates how the scale selection of terrain variables can be integrated into habitat mapping workflows that use morphological predictors.

Chapter 3 builds on the work at Qikiqtarjuaq from Chapter 2, using several multisource datasets and statistical models to predict soft-shell clam abundance in support of community-based fishery development. The spatial configuration of these data made them non-independent, and it was necessary to account for spatial autocorrelation to avoid a substantial inflation of apparent predictive accuracy when assessing the quality of model predictions.

Chapter 4 incorporates concepts from both preceding chapters to model seabed sediment classes throughout inner Frobisher Bay, Nunavut. This chapter compares the spatial qualities of different approaches for producing classified (i.e., thematic) sediment maps, and emphasizes the importance of differentiating between predictive goals – for instance, whether models are required to interpolate or extrapolate. This is seldom considered, yet can ultimately determine the appropriateness of modelling and evaluation methods.

The concluding chapter synthesizes the spatial concepts explored in the manuscripts. This chapter argues that adopting spatial analyses in the benthic habitat mapping workflow is critical to meeting assumptions inherent in common mapping methodologies. It offers several recommendations for implementing these that are broadly relevant to most habitat mapping applications and are reasonable to include in almost any workflow.

1.6 References

Baker, E. K., and Harris, P. T. 2012. Habitat mapping and marine management. *In* Seafloor Geomorphology as Benthic Habitat: Geohab Atlas of Seafloor Geomorphic Features and Benthic Habitats, pp. 23–38. Ed. by P. T. Harris and E. K. Baker. Elsevier, Amsterdam.

Barry, S., and Elith, J. 2006. Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43(3): 413–423.

Bell, D. M., and Schlaepfer, D. R. 2016. On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling*, 330: 50–59.

Bell, T., and Brown, T. M. (Eds). 2018. From Science to Policy in the Eastern Canadian Arctic: An Integrated Regional Impact Study (IRIS) of Climate Change and Modernization. ArcticNet, Quebec City. 560 pp.

Brown, C. J., Beaudoin, J., Brissette, M., and Gazzola, V. 2019. Multispectral multibeam echo sounder backscatter as a tool for improved seafloor characterization. *Geosciences*, 9(3): 126.

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(3): 502–520.

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.

Buhl-Mortensen, P., Dolan, M., and Buhl-Mortensen, L. 2009. Prediction of benthic biotopes on a Norwegian offshore bank using a combination of multivariate analysis and GIS classification. *ICES Journal of Marine Science*, 66(9): 2026–2032.

Buscombe, D., and Grams, P. E. 2018. Probabilistic substrate classification with multispectral acoustic backscatter: A comparison of discriminative and generative models. *Geosciences*, 8(11): 395.

Cheung, W. W. L., Lam, V. W. Y., Sarmiento, J. L., Kearney, K., Watson, R., and Pauly, D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries*, 10(3): 235–251.

Cogan, C. B., Todd, B. J., Lawton, P., and Noji, T. T. 2009. The role of marine habitat mapping in ecosystem-based management. *ICES Journal of Marine Science*, 66: 2033–2042.

Copeland, A., Edinger, E., Devillers, R., Bell, T., LeBlanc, P., and Wroblewski, J. 2013. Marine habitat mapping in support of Marine Protected Area management in a subarctic fjord: Gilbert Bay, Labrador, Canada. *Journal of Coastal Conservation*, 17(2): 225–237.

Davies, A. J., Duineveld, G. C. A., Lavaleye, M. S. S., Bergman, M. J. N., van Haren, H., and Roberts, J. M. 2009. Downwelling and deep-water bottom currents as food supply mechanisms to the cold-water coral *Lophelia pertusa* (Scleractinia) at the Mingulay Reef Complex. *Limnology and Oceanography*, 54(2): 620–629.

Deering, R., Misiuk, B., Bell, T., Forbes, D. L., Edinger, E., Tremblay, T., Telka, A., *et al.* 2018. Characterization of the seabed and postglacial sediments of inner Frobisher Bay, Baffin Island, Nunavut. Summary of Activities 2018. Canada-Nunavut Geoscience Office.

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D. 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, 84: 107–119.

Diniz-Filho, J. A. F., Bini, L. M., and Hawkins, B. A. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, 12(1): 53–64.

Dolan, M. F. J. 2012. Calculation of slope angle from bathymetry data using GIS – effects of computation algorithms, 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(2): 187–219.

Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16(2): 129–138.

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

Duffy, G., and Chown, S. 2017. Explicitly integrating a third dimension in marine species distribution modelling. *Marine Ecology Progress Series*, 564: 1–8.

Ehler, C., and Douvere, F. 2009. Marine Spatial Planning: A step-by-step approach toward ecosystem-based management. Intergovernmental Oceanographic Commission and Man and the Biosphere Programme, IOC Manual and Guides No. 5, ICAM Dossier No. 6. UNESCO, Paris.

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(2): 129–151.

Foley, M. M., Halpern, B. S., Micheli, F., Armsby, M. H., Caldwell, M. R., Crain, C. M., Prahler, E., *et al.* 2010. Guiding ecological principles for marine spatial planning. *Marine Policy*, 34(5): 955–966.

Foody, G. M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1): 185–201.

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(3): 287–297.

Frederiksen, R., Jensen, A., and Westerberg, H. 1992. The distribution of the scleractinian coral *Lophelia pertusa* around the Faroe Islands and the relation to internal tidal mixing. *Sarsia*, 77(2): 157–171.

Guinan, J., Brown, C., Dolan, M. F. J., and Grehan, A. J. 2009. Ecological niche modelling of the distribution of cold-water coral habitat using underwater remote sensing data. *Ecological Informatics*, 4(2): 83–92.

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., and Peterson, A. T. 2007. What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, 77(4): 615–630.

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., *et al.* 2008. A global map of human impact on marine ecosystems. *Science*, 319(5865): 948–952.

Harris, P. T. 2012. Anthropogenic threats to benthic habitats. *In* Seafloor Geomorphology as Benthic Habitat: Geohab Atlas of Seafloor Geomorphic Features and Benthic Habitats, pp. 39–60. Ed. by P. T. Harris and E. K. Baker. Elsevier, Amsterdam.

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.

Heikkinen, R. K., Marmion, M., and Luoto, M. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, 35(3): 276–288.

Hernandez, P. A., Graham, C. H., Master, L. L., and Albert, D. L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29(5): 773–785.

Hopkins, R. L. 2009. Use of landscape pattern metrics and multiscale data in aquatic species distribution models: A case study of a freshwater mussel. *Landscape Ecology*, 24(7): 943–955.

Huang, Z., Brooke, B. P., and Harris, P. T. 2011. A new approach to mapping marine benthic habitats using physical environmental data. *Continental Shelf Research*, 31(2): S4–S16.

Jordan, A., Lawler, M., Halley, V., and Barrett, N. 2005. Seabed habitat mapping in the Kent Group of islands and its role in marine protected area planning. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 15(1): 51–70.

Lacharité, M., Brown, C. J., and Gazzola, V. 2018. Multisource multibeam backscatter data: Developing a strategy for the production of benthic habitat maps using semi-automated seafloor classification methods. *Marine Geophysical Research*, 39(1–2): 307–322.

Lark, R. M., Marchant, B. P., Dove, D., Green, S. L., Stewart, H., and Diesing, M. 2015. Combining observations with acoustic swath bathymetry and backscatter to map seabed sediment texture classes: The empirical best linear unbiased predictor. *Sedimentary Geology*, 328: 17–32.

Lechner, A. M., Langford, W. T., Jones, S. D., Bekessy, S. A., and Gordon, A. 2012. Investigating species–environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. *Ecological Complexity*, 11: 91–102.

Lecours, V., Devillers, R., Schneider, D. C., Lucieer, V. L., Brown, C. J., and Edinger, E. N. 2015. Spatial scale and geographic context in benthic habitat mapping: Review and future directions. *Marine Ecology Progress Series*, 535: 259–284.

Lecours, V., Dolan, M. F. J., Micallef, A., and Lucieer, V. L. 2016. A review of marine geomorphometry, the quantitative study of the seafloor. *Hydrology and Earth System Sciences*, 20(8): 3207–3244.

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.

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(9): 5391–5406.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology*, 74(6): 1659–1673.

Legendre, P., Dale, M. R. T., Fortin, M.-J., Casgrain, P., and Gurevitch, J. 2004. Effects of spatial structures on the results of field experiments. *Ecology*, 85(12): 3202–3214.

Li, J., Potter, A., Huang, Z., Daniell, J. J., and Heap, A. D. 2010. Predicting seabed mud content across the Australian margin: Comparison of statistical and mathematical techniques using a simulation experiment. Geoscience Australia, Canberra.

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.

Liu, C., Berry, P. M., Dawson, T. P., and Pearson, R. G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3): 385–393.

Lundblad, E. R., Wright, D. J., Miller, J., Larkin, E. M., Rinehart, R., Naar, D. F., Donahue, B. T., *et al.* 2006. A benthic terrain classification scheme for American Samoa. *Marine Geodesy*, 29(2): 89–111.

Lurton, X., and Lamarche, G. (Eds). 2015. Backscatter measurements by seafloor-mapping sonars. Guidelines and recommendations. GeoHab Backscatter Working Group.

MacMillan, R. A., and Shary, P. A. 2009. Landforms and landform elements in geomorphometry. *Developments in Soil Science*, 33: 227–254.

McArthur, M. A., Brooke, B. P., Przeslawski, R., Ryan, D. A., Lucieer, V. L., Nichol, S., McCallum, A. W., *et al.* 2010. On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuarine, Coastal and Shelf Science*, 88: 21–32.

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(7): 8489–8515.

Mitchell, P. J., Downie, A.-L., and Diesing, M. 2018. How good is my map? A tool for semi-automated thematic mapping and spatially explicit confidence assessment. *Environmental Modelling & Software*, 108: 111–122.

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(2): 171–193.

Pesch, R., Pehlke, H., Jerosch, K., Schröder, W., and Schlüter, M. 2007. Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea. *Environmental Monitoring and Assessment*, 136(1–3): 313–325.

Pickrill, R. A., and Todd, B. J. 2003. The multiple roles of acoustic mapping in integrated ocean management, Canadian Atlantic continental margin. *Ocean & Coastal Management*, 46(6–7): 601–614.

Pike, R. J., Evans, I. S., and Hengl, T. 2009. Geomorphometry: A brief guide. *In* Geomorphometry: Concepts, Software, Applications, pp. 3–30. Elsevier, Amsterdam, Netherlands.

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. 2013. Marine taxa track local climate velocities. *Science*, 341(6151): 1239–1242.

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(2): 297–315.

Rengstorf, A. M., Grehan, A., Yesson, C., and Brown, C. 2012. Towards high-resolution habitat suitability modeling of vulnerable marine ecosystems in the deep-sea: Resolving terrain attribute dependencies. *Marine Geodesy*, 35(4): 343–361.

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(8): 913–929.

Roff, J. C., Taylor, M. E., and Laughren, J. 2003. Geophysical approaches to the classification, delineation and monitoring of marine habitats and their communities. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 13(1): 77–90.

Ross, R. E., and Howell, K. L. 2013. Use of predictive habitat modelling to assess the distribution and extent of the current protection of 'listed' deep-sea habitats. *Diversity and Distributions*, 19(4): 433–445.

Segurado, P., Araújo, M. B., and Kunin, W. E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43(3): 433–444.

Shary, P. A., Sharaya, L. S., and Mitusov, A. V. 2002. Fundamental quantitative methods of land surface analysis. *Geoderma*, 107(1–2): 1–32.

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.

Siwabessy, P. J. W., Tran, M., Picard, K., Brooke, B. P., Huang, Z., Smit, N., Williams, D. K., *et al.* 2018. Modelling the distribution of hard seabed using calibrated multibeam

acoustic backscatter data in a tropical, macrotidal embayment: Darwin Harbour, Australia. *Marine Geophysical Research*, 39(1–2): 249–269.

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(5): 650–667.

Smits, P. C., Dellepiane, S. G., and Schowengerdt, R. A. 1999. Quality assessment of image classification algorithms for land-cover mapping: A review and a proposal for a cost-based approach. *International Journal of Remote Sensing*, 20(8): 1461–1486.

Stockwell, D. R. B., and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148(1): 1–13.

St-Onge, P., and Miron, G. 2007. Effects of current speed, shell length and type of sediment on the erosion and transport of juvenile softshell clams (*Mya arenaria*). *Journal of Experimental Marine Biology and Ecology*, 349(1): 12–26.

Strong, J. A., Clements, A., Lillis, H., Galparsoro, I., Bildstein, T., and Pesch, R. 2018. A review of the influence of marine habitat classification schemes on mapping studies: Inherent assumptions, influence on end products, and suggestions for future developments. *ICES Journal of Marine Science*, 76(1): 10–22.

Syvitski, J. P. M., and Schafer, C. T. 1985. Sedimentology of Arctic Fjords Experiment (SAFE): Project introduction. *Arctic*, 38(4): 264–270.

Todd, B. J., and Greene, H. G. 2007. Mapping the seafloor for habitat characterization: Geological Association of Canada Special Paper 47. Geological Association of Canada, St. John's. 505 pp.

Vierod, A. D. T., Guinotte, J. M., and Davies, A. J. 2014. Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. *Deep Sea Research Part II: Topical Studies in Oceanography*, 99: 6–18.

Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. J. 2018. Unified geomorphological analysis workflows with Benthic Terrain Modeler. *Geosciences*, 8(3): 94.

Co-authorship Statement

This dissertation is presented in manuscript format. Each manuscript has also been published, and the author of this dissertation is the principal researcher in each case.

Chapter 2 was published in volume 13, issue 2 of the journal *PLoS ONE* (doi: https://doi.org/10.1371/journal.pone.0193647). The author of this dissertation was the primary contributor to study design, data collection and processing of field samples in 2014 and 2015, data analysis, and manuscript authorship. Dr. Vincent Lecours contributed to design of the multiscale analysis, provided expertise on the use of seabed terrain variables for statistical modelling, and assisted with writing the manuscript. Co-supervisor Dr. Trevor Bell initiated this project in partnership with the Hamlet of Qikiqtarjuaq, and the Government of Nunavut, Department of Environment, Fisheries and Sealing Division. Dr. Bell led the 2013 field survey, contributed to research design and methodology, and assisted with writing the manuscript.

Chapter 3 was published as an advance article in the *ICES Journal of Marine Science* (doi: https://doi.org/10.1093/icesjms/fsz099). The author of this dissertation was the primary contributor to study design and methodology, data collection and processing in 2014 and 2015, analysis, and manuscript authorship. Co-supervisor Dr. Trevor Bell initiated this project in the same capacity as in Chapter 2, led the 2013 field survey, contributed to research design and methodology, and assisted with writing the manuscript. Supervisory committee member Dr. Alec Aitken assisted with data collection in 2013 and 2014, sample processing, and manuscript preparation. Supervisory committee member Dr. Craig Brown

provided feedback on the methodology and manuscript. Co-supervisor Evan Edinger contributed to field preparation and study design and provided feedback on the manuscript.

Chapter 4 was published in the special issue on "Geological Seafloor Mapping" (2019, volume 9, issue 6) of the journal *Geosciences* (doi: https://doi.org/10.3390/geosciences9060254). The author of this dissertation was the primary contributor to study design and methodology, data collection and processing in 2016, and manuscript authorship. Co-author Dr. Markus Diesing contributed to the study design and methodology, provided expertise on the analysis of marine sediment data, and assisted with writing the manuscript. Supervisory committee member Dr. Alec Aitken contributed to data collection and processing, provided expertise on Quaternary history and geomorphology, and assisted with writing the manuscript. Co-supervisors Dr. Evan Edinger and Dr. Trevor Bell contributed to research design and provided feedback on the manuscript. Dr. Edinger is the co-leader of the larger ArcticNet project "Integrated Marine Geoscience to Guide Environmental Impact Assessment and Sustainable Development in Frobisher Bay, Nunavut", of which Chapter 4 is a part. Supervisory committee member Dr. Craig Brown provided feedback on the methodology and on preparation of the manuscript.

The author of this dissertation has contributed to other work related to these manuscripts that has been submitted for publication elsewhere and is not included here. The technical report, "Characterization of the seabed and postglacial sediments of inner Frobisher Bay, Baffin Island, Nunavut", was published in the 2018 Canada-Nunavut Geoscience Office

Summary of Activities, along with the primary author, Robert Deering, and co-authors Dr. Trevor Bell, Dr. Donald Forbes, Dr. Evan Edinger, Tommy Tremblay, Dr. Alice Telka, Dr. Alec Aitken, and Dr. Calvin Campbell. This publication is also a part of the larger ArcticNet project "Integrated Marine Geoscience to Guide Environmental Impact Assessment and Sustainable Development in Frobisher Bay, Nunavut". It details the surficial sediment characteristics of the bay in the context of local geomorphology and Quaternary geological history. The author of this dissertation assisted with writing the manuscript and was the primary cartographer and contributor to the sections describing the distribution of sediment grain size.

The author of this dissertation was the primary contributor to the essay titled: "Exploring the use of Inuit knowledge for mapping marine habitats", including writing, mapping, and GIS analysis. This essay is published the special issue of the *Journal of Ocean Technology*: "Tech-knowledge-y: Addressing Ocean Challenges Through Innovative Technology and Traditional Knowledge" (2019, volume 14, issue 1). It describes potential habitat mapping applications for an extensive Inuit knowledge database on the distributions of marine species, using an example from Frobisher Bay. Co-supervisors Dr. Trevor Bell and Dr. Evan Edinger, and supervisory committee member Dr. Alec Aitken, contributed writing and feedback to this manuscript. Teresa Tufts, the lead on the Nunavut Coastal Resource Inventory project for the Government of Nunavut, also provided feedback on the essay.

2. A Multiscale Approach to Mapping Seabed Sediments

2.1 Introduction

Marine ecosystems provide a broad range of services to humans, including food, extractive resources, and cultural identity (Galparsoro *et al*., 2014; Thurber *et al*., 2014). 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 (Myers & Worm, 2003; Halpern *et al*., 2008). 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 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 (Harris & Baker, 2012a; Harris & Baker, 2012b). Benthic habitat mapping is broadly defined as "the use of spatially continuous data sets to represent and predict biological patterns on the seafloor" (Brown *et al*., 2011). 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*. (2010) 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 (Ysebaert *et al*., 2002; Reiss *et al*., 2015). 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 (Coggan *et al*., 2012; Stephens & Diesing, 2015). 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 square kilometres) at a high spatial resolution (metres) when coupled with ground-truth substrate samples (e.g., Diesing *et al*., 2014; Stephens & Diesing, 2014; Stephens & Diesing, 2015). 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 (Diesing, 2015). This quantitative predictive approach represents a departure from subjective expert-based interpretation, towards more objective repeatable methods (Diesing, 2015; Lark *et al*., 2015).

Recent biological and geological modelling 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., Brown *et al*., 2012; Bučas *et al*., 2013), habitats (e.g., Dunn & Halpin, 2009; Copeland *et al*., 2013; Calvert *et al*., 2015), or sediment properties such as grain size or the presence of rock (e.g., Diesing *et al*., 2014; Downie *et al*., 2016). 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 (Lecours *et al*., 2016). While these processes are scale-dependent (e.g., Eidens *et al*., 2015), 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 operational environment of the 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 data 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 (Gambi & Danovaro, 2006; Dolan, 2012; Bradter *et al*., 2013). 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 (Lecours *et al*., 2015a). *Multiple scale* analyses are those that consider data at multiple successive scales, and *multiscale* analyses are those that integrate information from multiple scales simultaneously (Dolan, 2012). Several terrestrial (e.g.,

Dixon & Earls, 2009; Seo *et al*., 2009; Wolff *et al*., 2016) and marine (e.g., Dolan & Lucieer, 2014; Ross *et al*., 2015; Miyamoto *et al*., 2017; Porskamp *et al*., 2018) 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 (Lechner *et al*., 2012), the adoption of a multiscale approach ensures that the relevant scale-dependent patterns and processes are captured (Lecours *et al*., 2015).

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.

2.2 Data and Methods

2.2.1 Setting

This study was conducted in the coastal zone near the hamlet of Qikiqtarjuaq on the east-central coast of Baffin Island, Nunavut, Canada (Figure 2.1A). The surrounding terrain is mountainous, having been shaped by repeated glacial cycles during the Quaternary – it currently hosts upland ice caps and alpine glaciers (Margreth *et al*., 2017). For example,

deep valleys and fjords (e.g., > 300 m deep south of Qikiqtarjuaq; Figure 2.1B) 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 (Cowan, 2015). Surface currents of 0.5-0.8 m/s in the channel winnow the local seabed (Gilbert, 1980).

2.2.2 Primary Data

Bathymetry and backscatter data were collected using a MBES over the course of five years in the coastal zone near Qikiqtarjuaq (Figure 2.1). 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 (Figure 2.1B). The CCGS *Amundsen* collected data in the deepest area (> 600 m) using a Kongsberg EM300 30 kHz (variable beam width) echosounder in 2007 (Muggah, 2011). 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, which was operated at 200 or 300 kHz depending on water depth (Muggah, 2014, 2015). 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 Appendix A.

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.

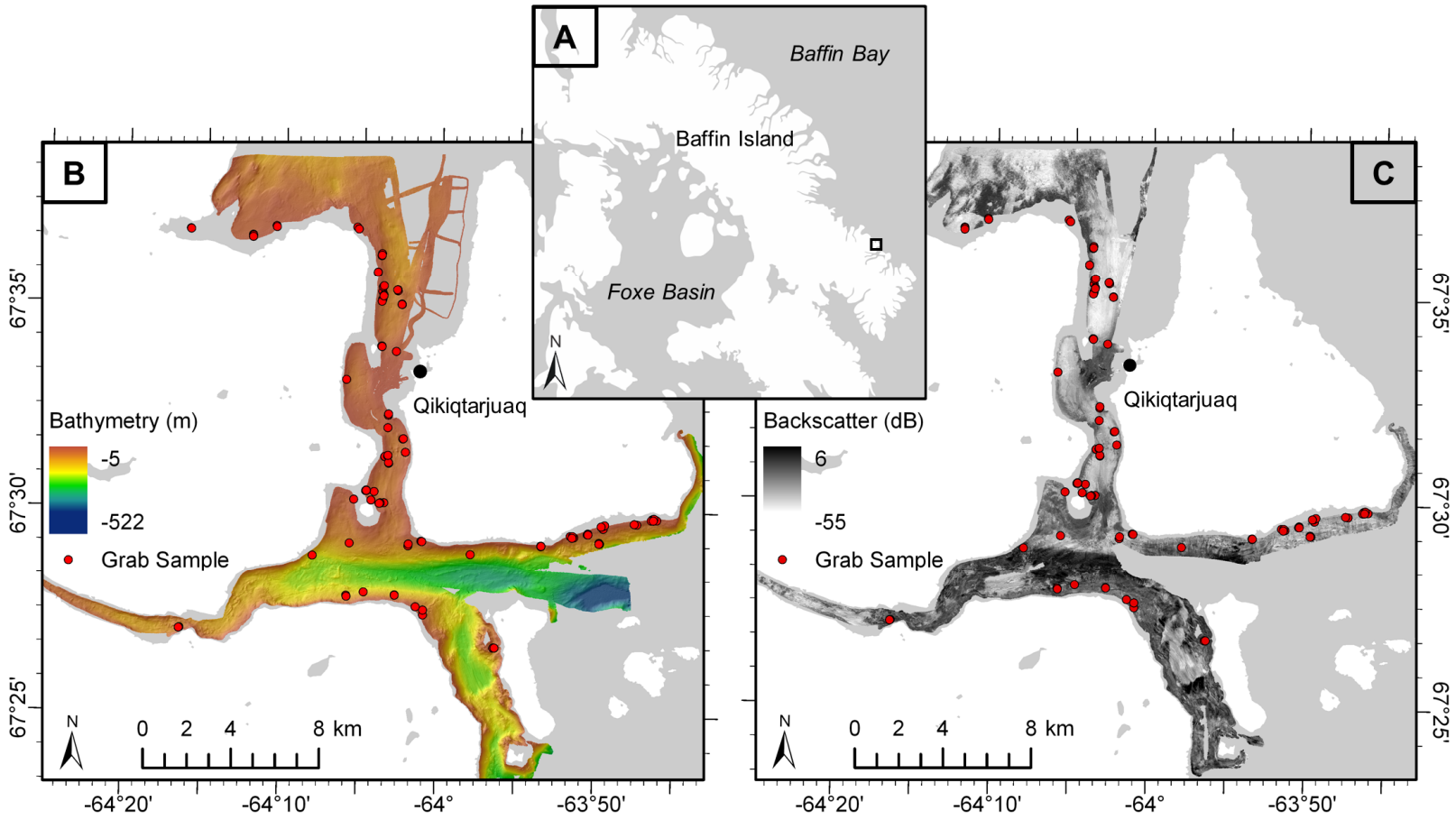


Figure 2.1. (A) Location of study site on east Baffin Island, NU, Canada. (B) Bathymetry data collected via MBES, with grab sample sites in red. (C) Backscatter data collected via MBES, with grab sample sites in red. (A) was modified from the USGS National Map, available under the public domain; basemap in (B) and (C) was obtained from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

In addition to MBES data, 109 sediment grab samples were collected between 2014 and 2015 to measure the grain size of sediment (Figures 2.1B, C). 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 but were 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.

2.2.3 Secondary Data

While many terrain attributes can be derived from bathymetric data to describe seabed morphology, Lecours *et al*. (2017a) 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 (Lecours, 2015) in ESRI ArcGIS v10.3.1 to compute values for those six terrain attributes using a default 3 x 3-pixel window of analysis at the native raster resolution of 5 m. The six terrain attributes include eastness and northness (unitless sine- 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*. (2017a), 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 (Wilson *et al*., 2007), 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 (Rengstorf *et al*., 2012; Tong *et al*., 2016). 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; Wilson *et al*., 2007). 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; Lundblad *et al*., 2006), rugosity, and ruggedness were derived using the Benthic Terrain Modeler (BTM) toolbox (Walbridge *et al*., 2018). Backscatter heterogeneity (hereafter Δbackscatter; dB), which is useful for differentiating coarse and fine substrates (Diesing & Stephens, 2015), was derived from the primary backscatter layer using the same function applied to calculate surface roughness (i.e., obtaining "minimum" and "maximum" 3 x 3-pixel neighborhood layers, then subtracting "maximum" - "minimum"; cf. "backscatter roughness"; Diesing & Stephens, 2015). ΔBackscatter was calculated using the "Focal Statistics" and "Raster Calculator" tools in ESRI ArcGIS v10.3.1. Distance from the coast (in metres), a potential driver of grain size distribution (Diesing, 2015), was calculated from a coastal polygon layer generated by Cowan (2015) 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 (Lecours *et al*., 2017b) these areas were excluded from the analysis, resulting in several narrow data gaps (see Figure A4 in Appendix A).

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 (Dolan, 2012). 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 neighborhoods followed the Fibonacci sequence (rounded up when even) - a convenient number series of increasing interval size (Wilson *et al*., 2007; Giusti *et al*., 2014). This resulted in 129 potential variables for predicting the response of sediment grain size, at eight different spatial scales (Table 2.1).

Table 2.1. Multiple scale explanatory variables selected for modelling 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	Lecours, 2015
	Northness	5,15,25,45,65,105,175,275	TASSE	Lecours, 2015
	RDMV	5,15,25,45,65,105,175,275	TASSE	Lecours, 2015
	Standard Deviation	5,15,25,45,65,105,175,275	TASSE	Lecours, 2015
	Slope	5,15,25,45,65,105,175,275	TASSE	Lecours, 2015
	Fine BPI*	5,15,25,45,65,105,175,275	BTM	Walbridge *et al.*, 2018
	Broad BPI*	5,15,25,45,65,105,175,275	BTM	Walbridge *et al.*, 2018
	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	Walbridge *et al.*, 2018
	Rugosity	5,15,25,45,65,105,175,275	BTM	Walbridge *et al.*, 2018
	Ruggedness	5,15,25,45,65,105,175,275	BTM	Walbridge *et al.*, 2018
Backscatter		5,15,25,45,65,105,175,275	-	-
	ΔBackscatter	5,15,25,45,65,105,175,275	Focal Statistics	Diesing & Stephens, 2015
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.

2.2.4 Grain Size Distribution Modelling

2.2.4.1 Response Variable

Ninety-eight grab samples were from locations within the MBES survey and were used to model the distribution of sediment grain size. Following recommendations by Aitchison (1982) and Stephens & Diesing (2015), 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 modelling (Equations 2.1, 2.2), 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 modelling (Stephens & Diesing, 2015):

$$ALR_{ms} = \log(\frac{mud}{sand}) \tag{2.1}$$

$$ALR_{gs} = \log(\frac{gravel}{sand}) \tag{2.2}$$

Some sediment samples had mud or gravel fractions equal to zero, which may be due to: 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 replacement" method (reviewed in Martín-Fernández & Thió-Henestrosa *et al*., 2006), 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*. (2012) and Diesing *et al*. (2014), 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.

2.2.4.2 Statistical Modelling

Boosted Regression Trees (BRT; Friedman *et al*., 2000; Breiman, 2001), a popular tree-based machine learning technique for ecological modelling (Olden *et al*., 2008), were used to model the responses of ALR_{ms} and ALR_{gs} to the different explanatory variables at different scales (see Table 2.1). This non-parametric technique can accommodate large numbers of non-linear categorical and numerical explanatory variables simultaneously, while automatically modelling interaction between predictors (Olden *et al*., 2008; Franklin, 2009). BRTs were chosen because they commonly outperform other quantitative modelling methods (Elith *et al*., 2006; Guisan *et al*., 2007; Reiss *et al*., 2015), ignore unimportant variables, are insensitive to outliers, and tend to avoid over-fitting (Friedman *et al*., 2000; Breiman, 2001; Olden *et al*., 2008). 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 (Elith *et al*., 2008). Details on how BRTs work have been detailed by Friedman *et al*. (2000), and with ecological examples in Elith *et al*. (2008).

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 (Evans *et al*., 2011; Diesing, 2015). 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 (Elith *et al*., 2008; Hopkins, 2009). Thus, we first fitted models of both ALR_{ms} and ALR_{gs} iteratively for each individual variable, but at all scales (Figure 2.2, 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 modelling was conducted in R v3.2.3, with code modified from that provided by Elith & Leathwick (2008) and Ridgeway (2015).

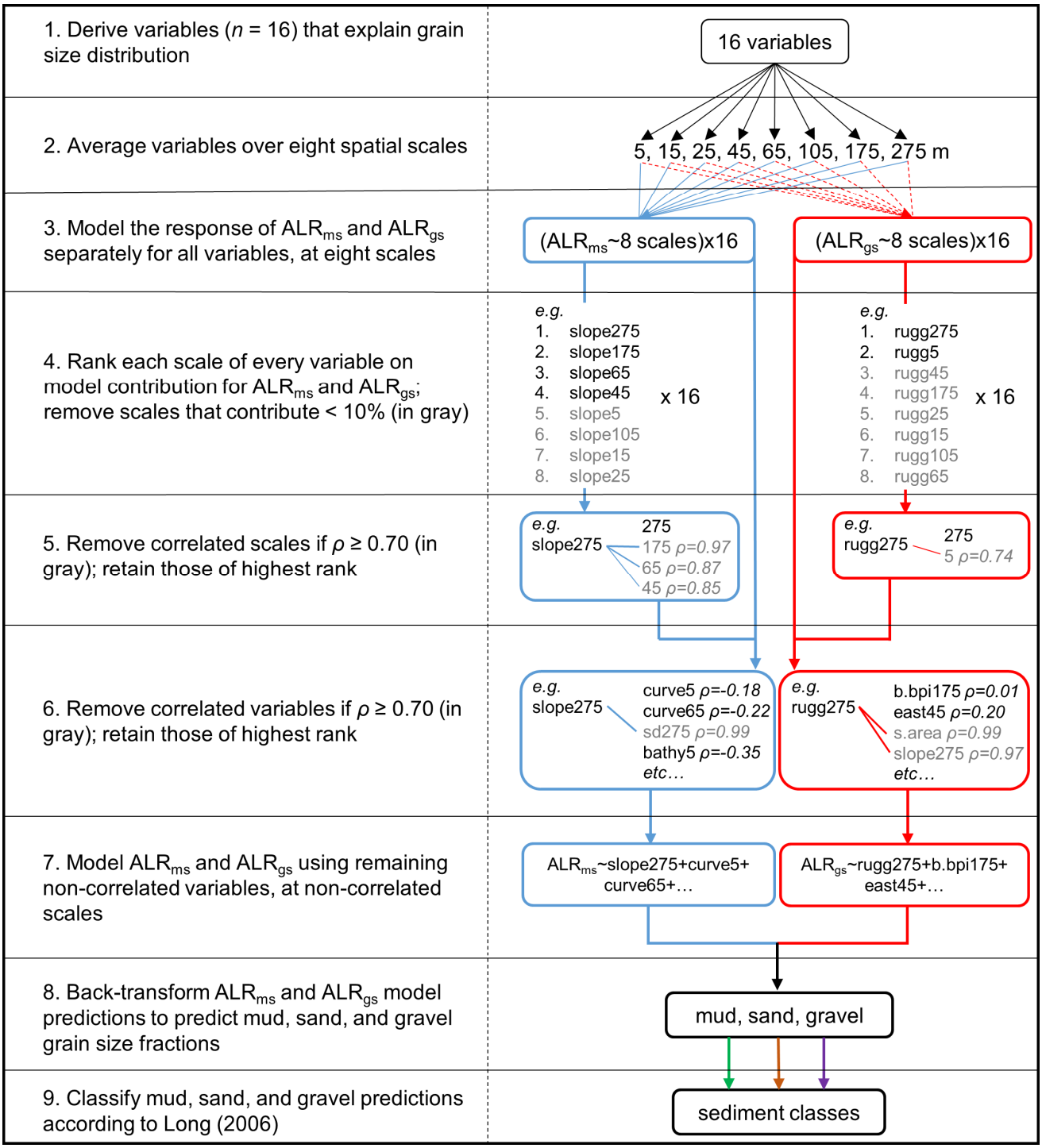


Figure 2.2. Procedure for selecting explanatory variables at multiple scales to model the response of ALR_{ms} and ALR_{gs} and predict the distribution of grain size classes.

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 ≥ 10% to a given model were subsequently considered. Correlation between scales of a given variable was measured using Spearman's rank correlation coefficient (Figure 2.2, 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 (Gottschalk *et al*., 2011; Downie *et al*., 2016). 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 (Figure 2.2, step 6). This methodology allowed for the selection of the most important non-correlated scales of each variable for inclusion in modelling ALR_{ms} and ALR_{gs}. Models were then fitted for both ALR_{ms} and ALR_{gs} using the remaining non-correlated variables (Figure 2.2, step 7).

ALR_{ms} and ALR_{gs} were back-transformed to produce individual predictions of mud, sand, and gravel fractions (Figure 2.2, step 8). Back-transformation was performed using Equations 2.3-2.5 with the "Raster Calculator" tool in ESRI ArcGIS v.10.3.1 (Diesing, 2015):

$$mud = \frac{\exp(ALR_{ms})}{\exp(ALR_{ms}) + \exp(ALR_{gs}) + 1} \tag{2.3}$$

$$gravel = \frac{\exp(ALR_{gs})}{\exp(ALR_{ms}) + \exp(ALR_{gs}) + 1} \tag{2.4}$$

$$sand = 1 - (mud + gravel) \tag{2.5}$$

These functions resulted in continuous predictions of mud, sand, and gravel fractions summing to 1 for all locations with MBES data. Following Stephens & Diesing (2015), mud, sand, and gravel

fractions were classified according to Long (2006) to produce a single map of sediment distribution (Figure 2.2, 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.

2.2.4.3 Model Evaluation

Elith & Leathwick's (2008) extension to the Generalized Boosted Regression Models ("*gbm*") package in R (Ridgeway, 2015) 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 & Leathwick's (2008) code calculates average percent deviance explained over 10 model folds by default, which is useful for evaluating model fit (Elith *et al*., 2008; Gottschalk *et al*., 2011). 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 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 (Guisan & Zimmermann, 2000; Potts & Elith, 2006).

2.3 Results

2.3.1 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 modelling 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 (Figure 2.3). Similarly, out of 55 scale-specific variables considered for modelling ALR_{gs}, 11 were at the default 5 m scale. All scales tested were selected for modelling at least once, yet 5, 175, and 275 m were most common for both ALR_{ms} and ALR_{gs} (Figure 2.3).

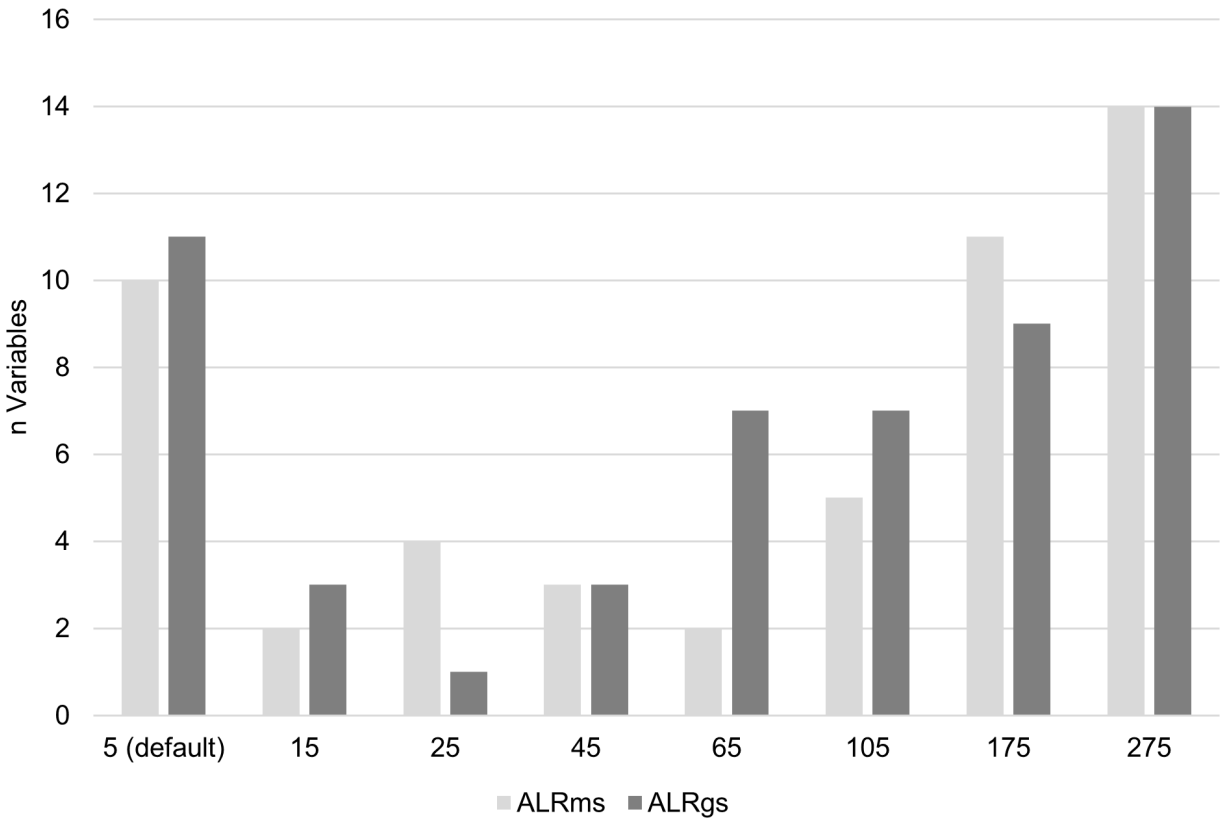


Figure 2.3. Number of times each scale contributed ≥ 10% to test models and was selected for modelling.

Eleven variables were ultimately selected to model the response of ALR_{ms} (mud and sand) at five different scales (Figure 2.4): 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.

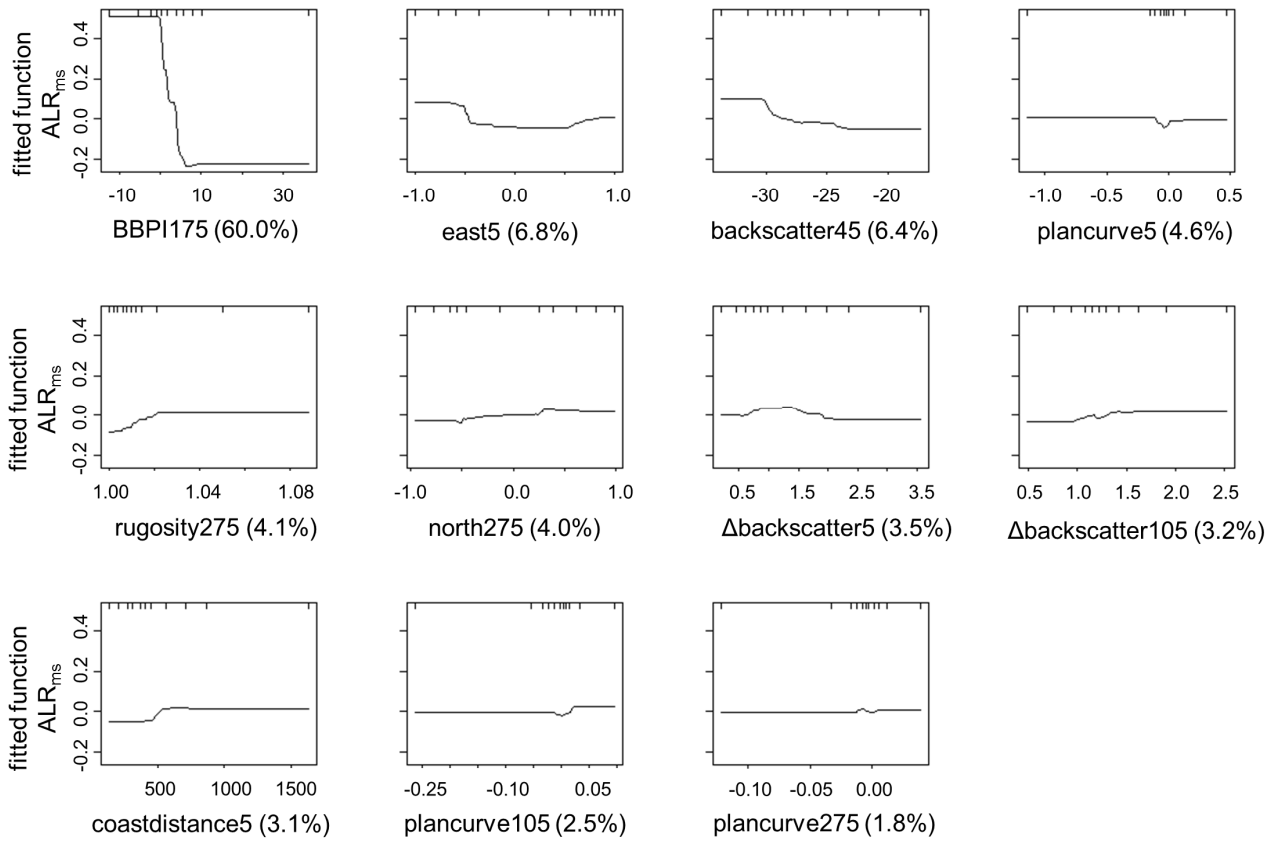


Figure 2.4. Partial dependence plots for multiple scale variables selected to model ALR_{ms} with percent contribution to the model and data deciles on the upper x-axis.

Fifteen variables were selected to model ALR_{gs} (gravel and sand) at five different scales (Figure 2.5): 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 (Figure 2.5).

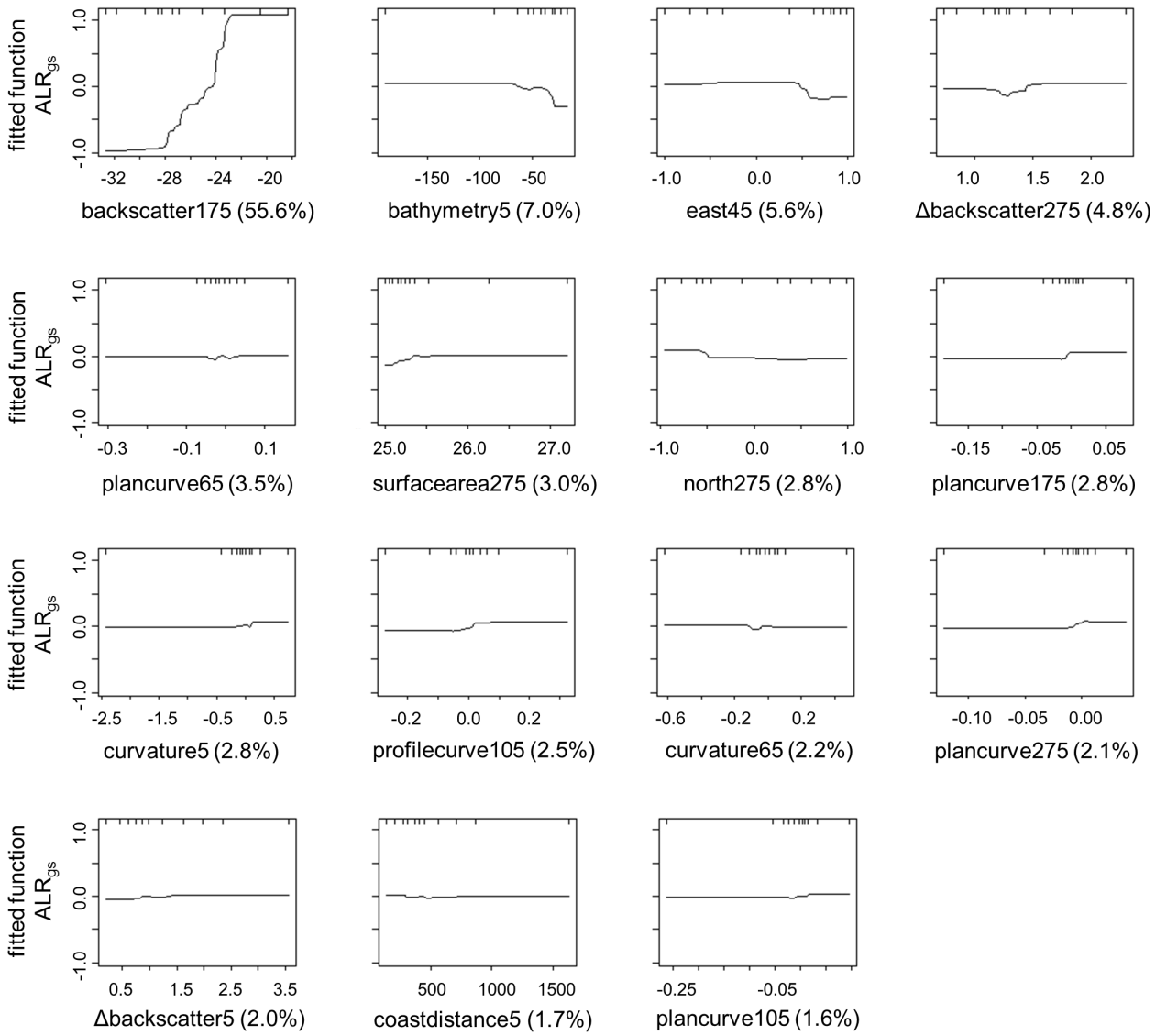


Figure 2.5. Partial dependence plots for multiple scale variables selected to model ALR_{gs}, with percent contribution to the model and data deciles on the upper x-axis.

2.3.2 Prediction

Back-transformed additive log-ratios produced continuous predictions of mud, sand, and gravel fractions over the area of environmental data coverage (Figure 2.6). 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 abundant in the deeper fjord

basin to the south (Figure 2.6B). 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 (Figure 2.6C). Mud was the least abundant size 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 (Figure 2.6A).

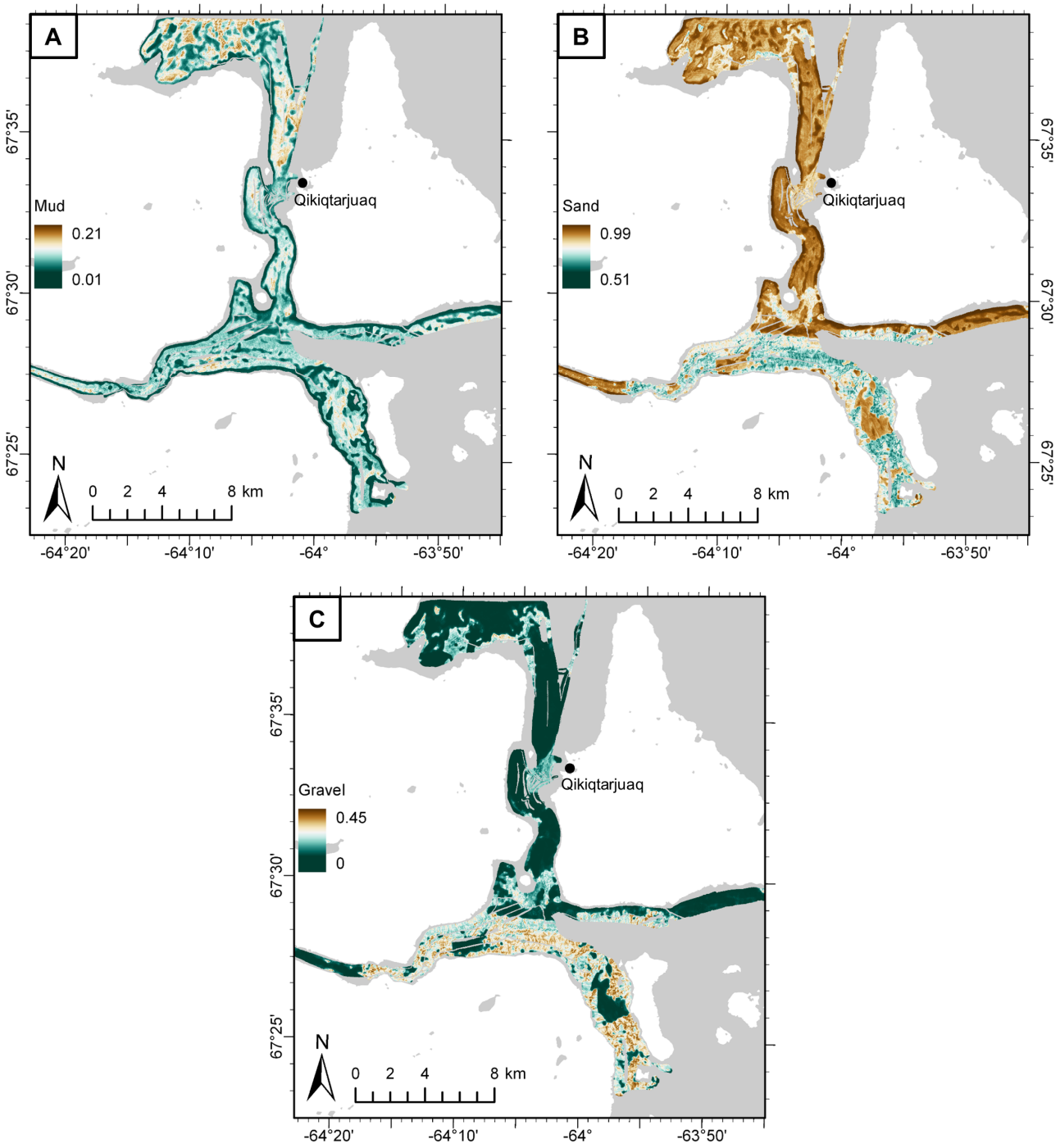


Figure 2.6. Predicted proportions of (A) mud, (B) sand, and (C) gravel fractions. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

2.3.3 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 ($\rho_{mud} = 0.772$), followed by sand ($\rho_{sand} = 0.712$) and gravel ($\rho_{gravel} = 0.578$). Ten-fold CV also produced a map of standard deviation for each grain size fraction, indicating areas of high and low model consensus (Figure 2.7). Mud predictions had the lowest mean standard deviation ($\sigma_{mud} = 0.007$), with the greatest model consensus in areas of low mud proportion, near the coasts (Figure 2.7A). Sand and gravel had similar mean standard deviations ($\sigma_{sand} = 0.021$, $\sigma_{gravel} = 0.019$) that were also distributed similarly in space (Figures 2.7B, C). Standard deviations for sand and gravel were highest in areas of high predicted gravel proportion and lowest in sandier areas.

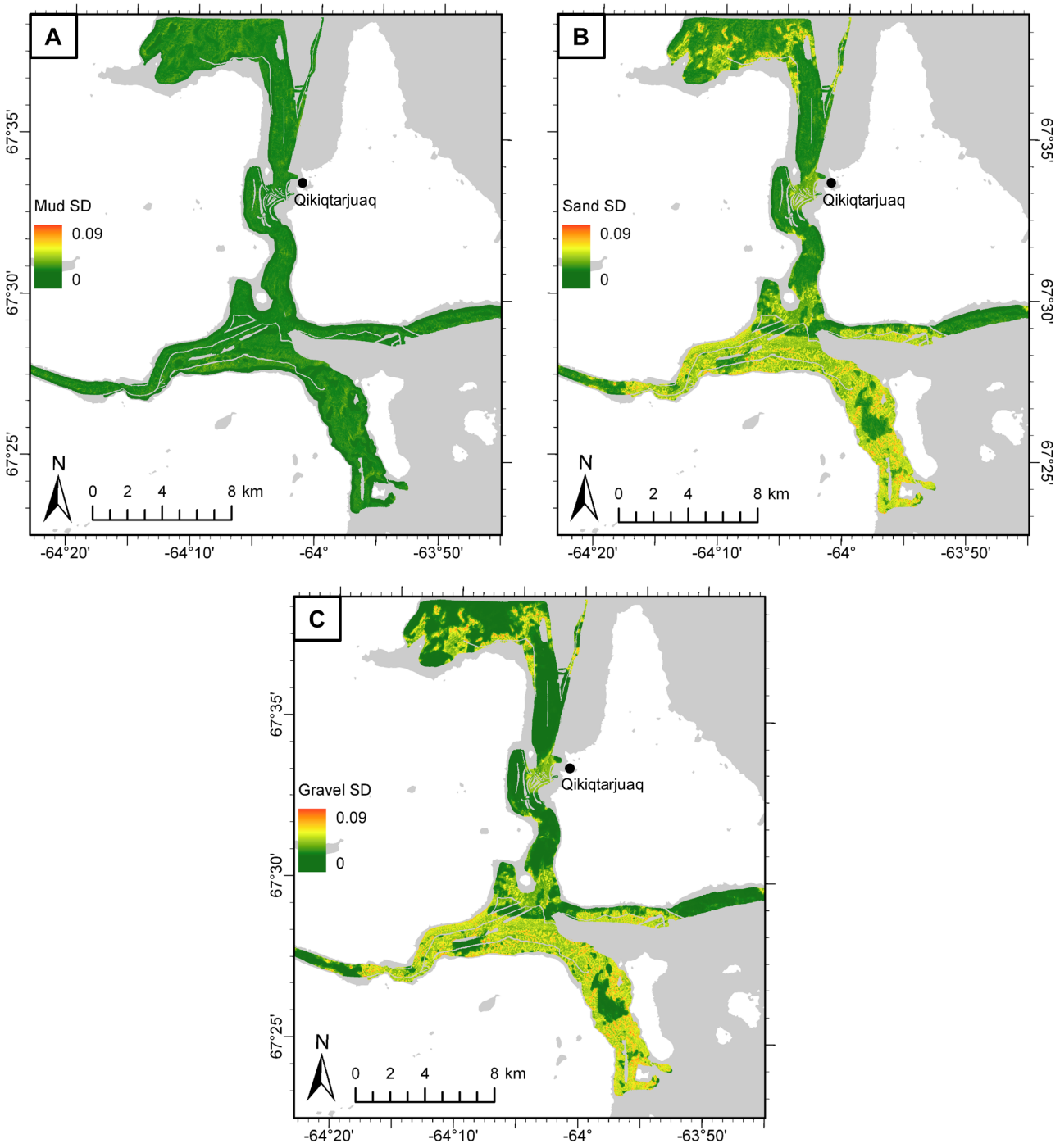


Figure 2.7. Ten-fold CV standard deviations (SD) for (A) mud, (B) sand, and (C) gravel predictions. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

2.3.4 Classification

Following Stephens & Diesing (2015), predictive maps of mud, sand, and gravel distribution were classified according to Long's (2006) modification of Folk's (1954) original scheme in order to facilitate interpretation and application as a management tool (Figure 2.8). 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 Figure A4 in Appendix A). 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.

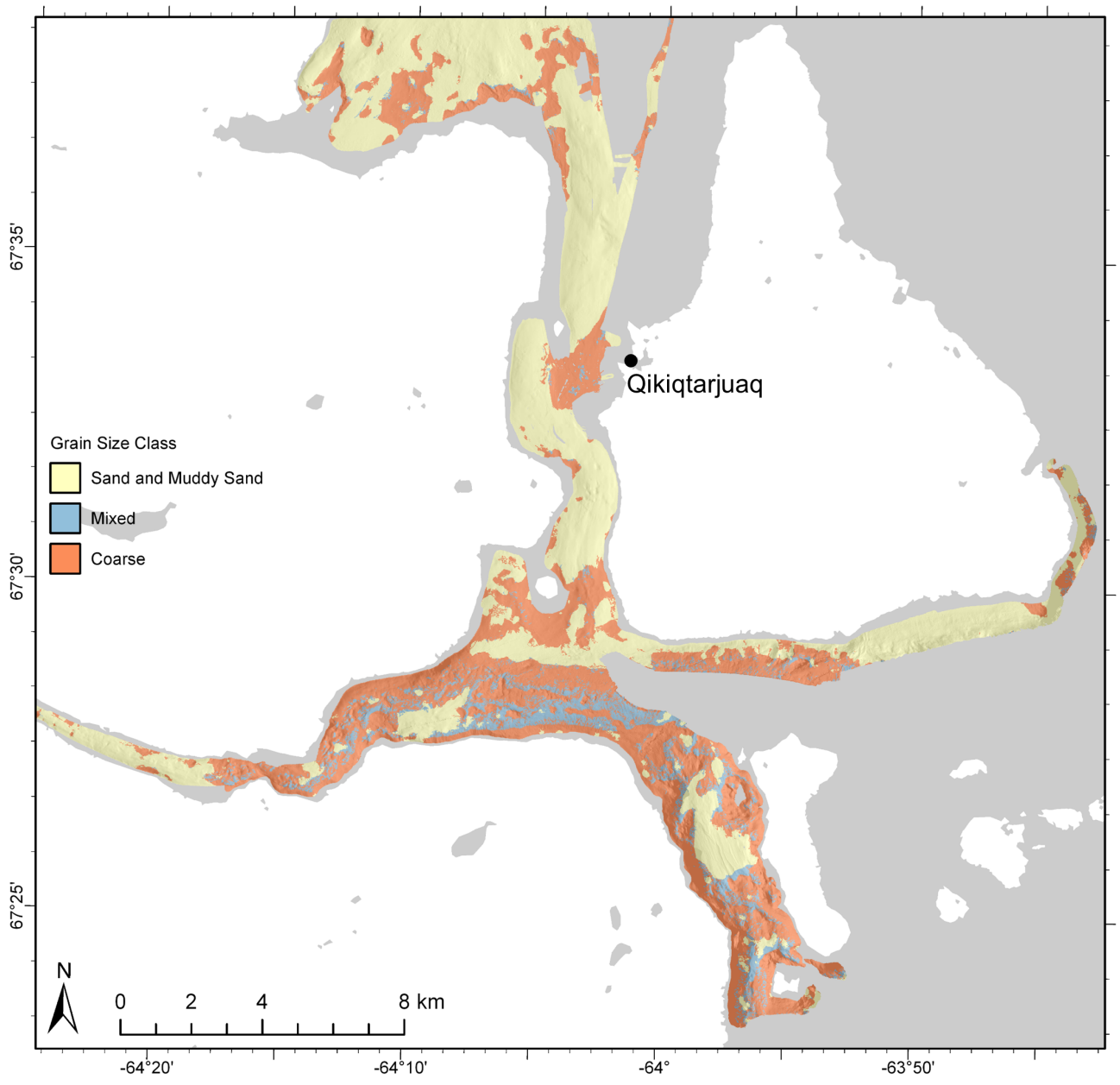


Figure 2.8. Predictions of mud, sand, and gravel classified according to Long's (2006) modification of Folk's (1954) original classification scheme. See text for discussion. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

2.4 Discussion

2.4.1 Scale Selection

The variable and scale selection process (Figure 2.2) demonstrated that the default scale of explanatory terrain variables was not necessarily the best option for modelling the distribution of sediment grain size, confirming previous observations that terrain variables are scale dependent (e.g., Dolan, 2012; Lechner *et al*., 2012; Dolan & Lucieer, 2014). 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 modelling predictions. Since the use of the same environmental variables at different scales will ultimately produce different modelling results (Albani *et al*., 2004; Wilson *et al*., 2007; Hopkins, 2009; Rengstorf *et al*., 2012; Dixon Hamil *et al*., 2016), the effects of scale selection will be propagated throughout the modelling process to 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 (Figure 2.5). 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 modelling process (removing correlated variables; Figure 2.2) more than half of the variables selected at the coarsest resolution (275 m) were correlated with other variables and were removed.

Other authors (e.g., Chow & Hodgson, 2009; Rengstorf *et al*., 2012; Lecours *et al*., 2017c) 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 (Figures 2.4, 2.5).

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 (Gottschalk *et al*., 2011). Noise present in the primary bathymetry and backscatter data layers can also be propagated, and even amplified in derivative layers (Lecours *et al*., 2017c), which can also affect model performance (Lecours et al., 2017b). A decrease in spatial resolution can reduce data noise in derivative layers (Gottschalk *et al*., 2011; Tong *et al*., 2016). By averaging derivative data layers over an increasing area, noise can be smoothed out and made less distinct (see Dolan & Lucieer, 2014, for 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 (Figure 2.3). 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.

2.4.2 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 (Figure 2.4) 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 (Figure 2.6). 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 (Figure 2.4), 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 (Tong *et al.*, 2012; Dolan & Lucieer, 2014; Tong *et al.*, 2016). Morphology influences the speed and orientation of currents and describes the exposure of the 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 5 m scale in our study is potentially a result of its surrogacy for current information (Tong *et al.*, 2013; Miyamoto *et al.*, 2017). The moderate response of ALR_{ms} to backscatter at an intermediate scale (45 m) suggests that sand was slightly more acoustically reflective than mud (Figure 2.4). 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 (Gottschalk *et al*., 2011). 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 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 (Figure 2.5). This corroborates findings by other authors (e.g., Goff *et al*., 2000) 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, in which the "salt and pepper" qualities of the backscatter data at fine scales was reduced by averaging values with neighboring pixels. 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 (Brown *et al*., 2011; Rengstorf *et al*., 2012; Rengstorf *et al*., 2013; Miyamoto *et al*., 2017). 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.

2.4.3 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 (Figure 2.6). 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 (Fulton, 1995; Wheeler *et al.*, 1996). 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 (Figure 2.8).

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; Eigenbrod *et al.*, 2008; Elith *et al.*, 2008; Elith & Graham, 2009), 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; Elith & Graham, 2009). Each modelling 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., Nielsen *et al*., 2005; Elith & Graham, 2009; Gottschalk *et al*., 2011). 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., Figure 2.2, 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 (Figure 2.6C) displaying the most variability between the 10 model folds, as measured by standard deviation (Figure 2.7). 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 > 4 mm, 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 (Figure 2.6C) represents the distribution of gravel ≤ 4 mm, and the "mixed" and "coarse" classes (Figure 2.8) 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" (Figure 2.9), 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., Diesing *et al*., 2015).

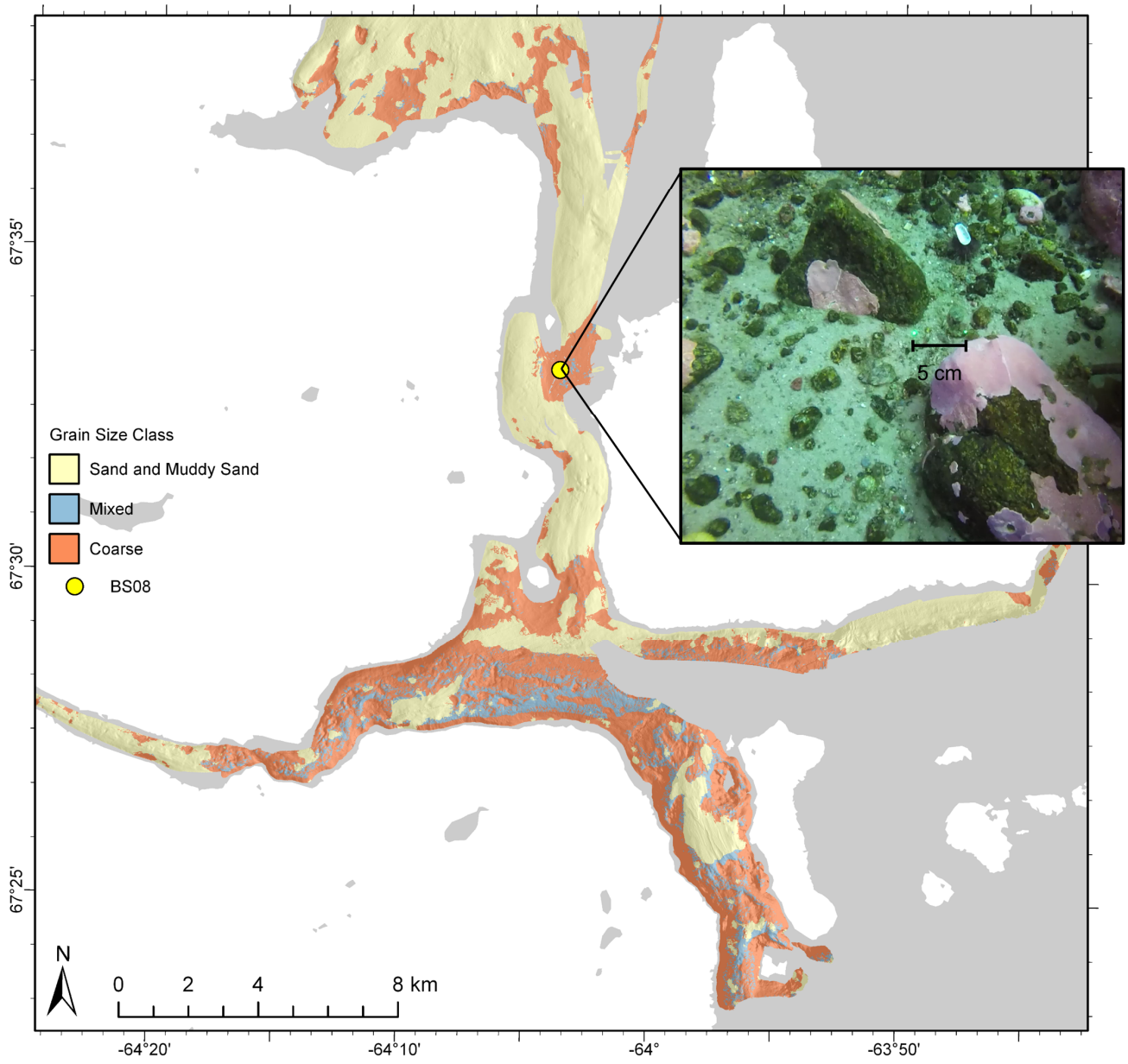


Figure 2.9. Clasts too large to sample in an area classified as "coarse", with 5 cm scale lasers. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

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 (Figure 2.1C) 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 (Figure 2.7). Despite this limitation, manual inspection suggests that predictions were largely based on high backscatter return (Figure 2.1C). Backscatter was the single most important variable in predicting ALR_{gs}, accounting for over 55% of explained deviance in the model (Figure 2.5), providing some confidence that these predictions are well-founded.

2.4.4 Classification

Long's (2006) simplification of Folk's (1954) 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 (Figure 2.8). 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 (Diesing, 2015).

2.5 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 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 Dolan, 2012) 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.

2.6 References

Aitchison, J. 1982. The statistical analysis of compositional data. *Journal of the Royal Statistical Society*, 44(2): 139–177.

Albani, M., Klinkenberg, B., Andison, D. W., and Kimmins, J. P. 2004. The choice of window size in approximating topographic surfaces from Digital Elevation Models. *International Journal of Geographical Information Science*, 18(6): 577–593.

Bradter, U., Kunin, W. E., Altringham, J. D., Thom, T. J., and Benton, T. G. 2013. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. *Methods in Ecology and Evolution*, 4(2): 167–174.

Breiman, L. 2001. Statistical modeling: The two cultures. *Statistical Science*, 16(3): 199–231.

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(3): 502–520.

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.

Bučas, M., Bergström, U., Downie, A.-L., Sundblad, G., Gullström, M., von Numers, M., Šiaulys, A., *et al.* 2013. Empirical modelling of benthic species distribution, abundance, and diversity in the Baltic Sea: Evaluating the scope for predictive mapping using different modelling approaches. *ICES Journal of Marine Science*, 70(6): 1233–1243.

Calvert, J., Strong, J. A., Service, M., McGonigle, C., and Quinn, R. 2015. An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data. *ICES Journal of Marine Science*, 72(5): 1498–1513.

Chow, T. E., and Hodgson, M. E. 2009. Effects of lidar post-spacing and DEM resolution to mean slope estimation. *International Journal of Geographical Information Science*, 23(10): 1277–1295.

Coggan, R., Barrio Froján, C. R. S., Diesing, M., and Aldridge, J. 2012. Spatial patterns in gravel habitats and communities in the central and eastern English Channel. *Estuarine, Coastal and Shelf Science*, 111: 118–128.

Copeland, A., Edinger, E., Devillers, R., Bell, T., LeBlanc, P., and Wroblewski, J. 2013. Marine habitat mapping in support of Marine Protected Area management in a subarctic fjord: Gilbert Bay, Labrador, Canada. *Journal of Coastal Conservation*, 17(2): 225–237.

Cowan, B. 2015. Shorelines beneath the sea: Geomorphology and characterization of the postglacial sea-level lowstand, Cumberland Peninsula, Baffin Island, Nunavut. Memorial University of Newfoundland, St. John's. 282 pp.

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D. 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, 84: 107–119.

Diesing, M. 2015. Quantitative spatial prediction of seabed sediment composition. EMODnet-Geology Phase II, C5818. Centre for Environment Fisheries and Aquaculture.

Diesing, M., Green, S. L., Stephens, D., Cooper, R., and Mellett, C. L. 2015. Semi-automated mapping of rock in the English Channel and Celtic Sea. JNCC Report, 569. JNCC, Peterborough.

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

Dixon, B., and Earls, J. 2009. Resample or not?! Effects of resolution of DEMs in watershed modeling. *Hydrological Processes*, 23(12): 1714–1724.

Dixon Hamil, K.-A., Iannone III, B. V., Huang, W. K., Fei, S., and Zhang, H. 2016. Cross-scale contradictions in ecological relationships. *Landscape Ecology*, 31(1): 7–18.

Dolan, M. F. J. 2012. Calculation of slope angle from bathymetry data using GIS – effects of computation algorithms, 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(2): 187–219.

Downie, A.-L., Dove, D., Westhead, K., Diesing, M., Green, S. L., and Cooper, R. 2016. Semi-automated mapping of rock in the North Sea. JNCC Report, 592. JNCC, Peterborough.

Dunn, D. C., and Halpin, P. N. 2009. Rugosity-based regional modeling of hard-bottom habitat. *Marine Ecology Progress Series*, 377: 1–11.

Eidens, C., Hauffe, T., Bayraktarov, E., Wild, C., and Wilke, T. 2015. Multi-scale processes drive benthic community structure in upwelling-affected coral reefs. *Frontiers in Marine Science*, 2.

Eigenbrod, F., Hecnar, S. J., and Fahrig, L. 2008. The relative effects of road traffic and forest cover on anuran populations. *Biological Conservation*, 141(1): 35–46.

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(2): 129–151.

Elith, J., and Leathwick, J. R. 2008. Tutorial for running boosted regression trees [Appendix S3], *in* A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 802–813.

Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 802–813.

Elith, J., and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32(1): 66–77.

Evans, J. S., Murphy, M. A., Holden, Z. A., and Cushman, S. A. 2011. Modeling species distribution and change using random forest. *In* Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications, pp. 139–159. Ed. by C. A. Drew, Y. F. Wiersma, and F. Huettmann. Springer, New York.

Folk, R. L. 1954. The distinction between grain size and mineral composition in sedimentary-rock nomenclature. *The Journal of Geology*, 62(4): 344–359.

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

Friedman, J., Hastie, T., and Tibshirani, R. 2000. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 28(2): 337–407.

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

Galparsoro, I., Borja, A., and Uyarra, M. C. 2014. Mapping ecosystem services provided by benthic habitats in the European North Atlantic Ocean. *Frontiers in Marine Science*, 1.

Gambi, C., and Danovaro, R. 2006. A multiple-scale analysis of metazoan meiofaunal distribution in the deep Mediterranean Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, 53(7): 1117–1134.

Gilbert, R. 1980. Environmental studies in Maktak, Coronation, and North Pangnirtung fiords, Baffin Island, N.W.T. Petro-Canada Exploration Inc. Final Report. Queen's University, Kingston, ON.

Giusti, M., Innocenti, C., and Canese, S. 2014. Predicting suitable habitat for the gold coral *Savalia savaglia* (Bertoloni, 1819) (Cnidaria, Zoantharia) in the South Tyrrhenian Sea. *Continental Shelf Research*, 81: 19–28.

Goff, J. A., Olson, H. C., and Duncan, C. S. 2000. Correlation of side-scan backscatter intensity with grain-size distribution of shelf sediments, New Jersey margin. *Geo-Marine Letters*, 20(1): 43–49.

Gottschalk, T. K., Aue, B., Hotes, S., and Ekschmitt, K. 2011. Influence of grain size on species–habitat models. *Ecological Modelling*, 222(18): 3403–3412.

Guisan, A., and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3): 147–186.

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., and Peterson, A. T. 2007. What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, 77(4): 615–630.

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., *et al.* 2008. A global map of human impact on marine ecosystems. *Science*, 319(5865): 948–952.

Harris, P. T., and Baker, E. K. 2012a. 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.

Harris, P. T., and Baker, E. K. 2012b. GeoHab atlas of seafloor geomorphic features and benthic habitats: Synthesis and lessons learned. *In* Seafloor Geomorphology as Benthic Habitat: Geohab Atlas of Seafloor Geomorphic Features and Benthic Habitats, pp. 871-890. Ed. by P. T. Harris and E. K. Baker. Elsevier, Amsterdam.

Hopkins, R. L. 2009. Use of landscape pattern metrics and multiscale data in aquatic species distribution models: A case study of a freshwater mussel. *Landscape Ecology*, 24(7): 943–955.

Lark, R. M., Dove, D., Green, S. L., Richardson, A. E., Stewart, H., and Stevenson, A. 2012. Spatial prediction of seabed sediment texture classes by cokriging from a legacy database of point observations. *Sedimentary Geology*, 281: 35–49.

Lark, R. M., Marchant, B. P., Dove, D., Green, S. L., Stewart, H., and Diesing, M. 2015. Combining observations with acoustic swath bathymetry and backscatter to map seabed sediment texture classes: The empirical best linear unbiased predictor. *Sedimentary Geology*, 328: 17–32.

Lechner, A. M., Langford, W. T., Jones, S. D., Bekessy, S. A., and Gordon, A. 2012. Investigating species–environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. *Ecological Complexity*, 11: 91–102.

Lecours, V. 2015. Terrain attribute selection for spatial ecology (TASSE). ArcGIS toolbox version 1.0. doi: 10.13140/RG.2.2.15014.52800

Lecours, V., Devillers, R., Schneider, D. C., Lucieer, V. L., Brown, C. J., and Edinger, E. N. 2015. Spatial scale and geographic context in benthic habitat mapping: Review and future directions. *Marine Ecology Progress Series*, 535: 259–284.

Lecours, V., Dolan, M. F. J., Micallef, A., and Lucieer, V. L. 2016. A review of marine geomorphometry, the quantitative study of the seafloor. *Hydrology and Earth System Sciences*, 20(8): 3207–3244.

Lecours, V., Devillers, R., Edinger, E. N., Brown, C. J., and Lucieer, V. L. 2017b. Influence of artefacts in marine digital terrain models on habitat maps and species distribution models: A multiscale assessment. *Remote Sensing in Ecology and Conservation*, 3(4): 232–246.

Lecours, V., Devillers, R., Lucieer, V. L., and Brown, C. J. 2017c. 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(9): 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.

Long, D. 2006. BGS detailed explanation of seabed sediment modified folk classification. MESH Report. British Geological Survey.

Lundblad, E. R., Wright, D. J., Miller, J., Larkin, E. M., Rinehart, R., Naar, D. F., Donahue, B. T., *et al.* 2006. A benthic terrain classification scheme for American Samoa. *Marine Geodesy*, 29(2): 89–111.

Margreth, A., Gosse, J. C., and Dyke, A. S. 2017. Wisconsinan and early Holocene glacial dynamics of Cumberland Peninsula, Baffin Island, Arctic Canada. *Quaternary Science Reviews*, 168: 79–100.

Martín-Fernández, J. A., and Thió-Henestrosa, S. 2006. Rounded zeros: Some practical aspects for compositional data. *In* Compositional Data Analysis in the Geosciences: From Theory to Practice, pp. 191–201. Ed. by A. Buccianti, G. Mateu-Figueras, and V. Pawlowsky-Glahn. Geological Society, London.

McArthur, M. A., Brooke, B. P., Przeslawski, R., Ryan, D. A., Lucieer, V. L., Nichol, S., McCallum, A. W., *et al.* 2010. On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuarine, Coastal and Shelf Science*, 88: 21–32.

Miyamoto, M., Kiyota, M., Murase, H., Nakamura, T., and Hayashibara, T. 2017. Effects of bathymetric grid-cell sizes on habitat suitability analysis of cold-water gorgonian corals on seamounts. *Marine Geodesy*, 40(4): 205–223.

Muggah, J. 2011. 2007 multibeam sonar data collected from the CCGS *Amundsen*. OMG Arctic Metadata. Available from: http://www.omg.unb.ca/Projects/Arctic/metadata/Amundsen_2007_MB_General.html.

Muggah, J. 2014. 2013 multibeam sonar data collected from the M/V *Nuliajuk*. OMG Arctic Metadata. Available from: http://www.omg.unb.ca/Projects/Arctic/metadata/Nuliajuk_2013_MB_General.html.

Muggah, J. 2015. 2014 multibeam sonar data collected from the M/V *Nuliajuk*. OMG Arctic Metadata. Available from: http://www.omg.unb.ca/Projects/Arctic/metadata/Nuliajuk_2014_MB_General.html.

Myers, R. A., and Worm, B. 2003. Rapid worldwide depletion of predatory fish communities. *Nature*, 423(6937): 280–283.

Nielsen, S. E., Johnson, C. J., Heard, D. C., and Boyce, M. S. 2005. Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. *Ecography*, 28(2): 197–208.

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(2): 171–193.

Porskamp, P., Rattray, A., Young, M., and Ierodiaconou, D. 2018. Multiscale and hierarchical classification for benthic habitat mapping. *Geosciences*, 8(4): 119.

Potts, J. M., and Elith, J. 2006. Comparing species abundance models. *Ecological Modelling*, 199(2): 153–163.

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(2): 297–315.

Rengstorf, A. M., Grehan, A., Yesson, C., and Brown, C. 2012. Towards high-resolution habitat suitability modeling of vulnerable marine ecosystems in the deep-sea: Resolving terrain attribute dependencies. *Marine Geodesy*, 35(4): 343–361.

Rengstorf, A. M., Yesson, C., Brown, C., and Grehan, A. J. 2013. High-resolution habitat suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea. *Journal of Biogeography*, 40(9): 1702–1714.

Ridgeway, G. 2015. Generalized boosted regression models. R package version 2.1.1. Available from: https://CRAN.R-project.org/package=gbm

Ross, L. K., Ross, R. E., Stewart, H. A., and Howell, K. L. 2015. The influence of data resolution on predicted distribution and estimates of extent of current protection of three 'listed' deep-sea habitats. *PLOS ONE*, 10(10): e0140061.

Seo, C., Thorne, J. H., Hannah, L., and Thuiller, W. 2009. Scale effects in species distribution models: Implications for conservation planning under climate change. *Biology Letters*, 5(1): 39–43.

Stephens, D., and Diesing, M. 2014. A comparison of supervised classification methods for the prediction of substrate type using acoustic and legacy grain-size data. *PLOS ONE*, 9(4): e93950.

Stephens, D., and Diesing, M. 2015. Towards quantitative spatial models of seabed sediment composition. *PLOS ONE*, 10(11): e0142502.

Thurber, A. R., Sweetman, A. K., Narayanaswamy, B. E., Jones, D. O. B., Ingels, J., and Hansman, R. L. 2014. Ecosystem function and services provided by the deep sea. *Biogeosciences*, 11(14): 3941–3963.

Tong, R., Purser, A., Unnithan, V., and Guinan, J. 2012. Multivariate statistical analysis of distribution of deep-water gorgonian corals in relation to seabed topography on the Norwegian Margin. *PLoS ONE*, 7(8): e43534.

Tong, R., Purser, A., Guinan, J., and Unnithan, V. 2013. Modeling the habitat suitability for deep-water gorgonian corals based on terrain variables. *Ecological Informatics*, 13: 123–132.

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.

Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. J. 2018. Unified geomorphological analysis workflows with Benthic Terrain Modeler. *Geosciences*, 8(3): 94.

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.

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., and Grehan, A. J. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, 30(1–2): 3–35.

Wolff, C., Vafeidis, A. T., Lincke, D., Marasmi, C., and Hinkel, J. 2016. Effects of scale and input data on assessing the future impacts of coastal flooding: An application of DIVA for the Emilia-Romagna Coast. *Frontiers in Marine Science*, 3.

Ysebaert, T., Meire, P., Herman, P., and Verbeek, H. 2002. Macrobenthic species response surfaces along estuarine gradients: Prediction by logistic regression. *Marine Ecology Progress Series*, 225: 79–95.

3. Mapping Arctic Clam Abundance using Multiple Datasets, Models, and a Spatially Explicit Accuracy Assessment

3.1 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 (e.g., Guisan & Zimmermann, 2000; Elith *et al*., 2006; Franklin, 2009; Miller, 2010; Drew *et al*., 2011).

In this study we built an SDM to investigate the environmental drivers of soft-shell clam (*Mya* spp.; hereafter referred to as "*Mya*") distribution and 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 predicts *Mya* abundance continuously 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. This assumes that mapped habitats are static and are fully representative of a species' habitat; it is important to note that a species' realized habitat is also affected by other external temporal factors such as changing climate or fishing stress.

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 conditions that 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 & 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. Additionally, 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 over 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 (e.g., 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 (e.g., Brown *et al*., 2012; Porskamp *et al*., 2018), spatial blocking (Roberts *et al*., 2017), spatial sub-sampling (e.g., Kendall *et al*., 2005; Segurado *et al*., 2006; Veloz, 2009), and geostatistical models (e.g., 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 1) supplement Siferd's (2005) survey data by sampling a broader range of environmental variables to determine the extent of *Mya* habitat, 2) model *Mya* abundance using this combined dataset, including predictions of

absence where habitat is unsuitable, and 3) estimate the magnitude of inflation caused by spatial autocorrelation in the sample data to provide accurate estimates of model performance.

3.2 Study Area and Species

Qikiqtarjuaq is located on the west coast of Broughton Island, off eastern Baffin Island, Nunavut, Canada (Figure 3.1). 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 & 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 3.1), 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 approximately 10 years to mature and can live up to 60 years (cf. 40 years; Hewitt & Dale, 1984). Previous surveys (Petersen, 1978; Abraham & Dillon, 1986; Siferd, 2005) have also suggested that

Mya prefer shallow depths, and unconsolidated substrates that 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.

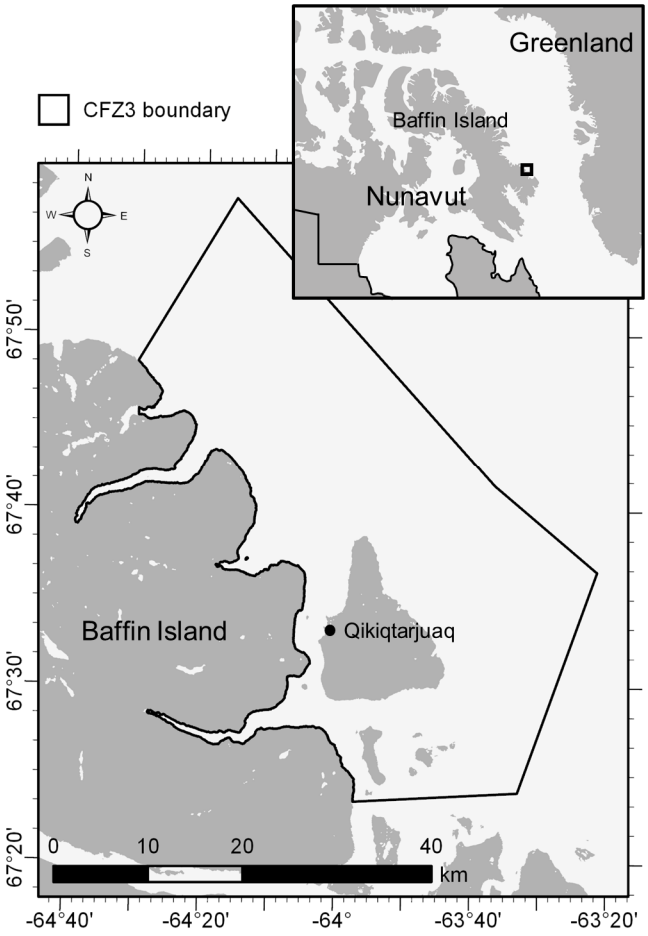


Figure 3.1. Location of Qikiqtarjuaq study area within fishing zone CFZ3 and eastern Nunvaut, Canada (inset map). Basemaps obtained from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

3.3 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 (e.g., 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.

3.3.1 Environmental Data

Multibeam echosounder (MBES) bathymetry and backscatter data were collected near Qikiqtarjuaq to characterize *Mya* habitat (Figure 3.2). 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 four-year period and were used to derive explanatory environmental variables for the *Mya* presence-absence and abundance models at a 5 m resolution. 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 3.1) 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 3.1). These were calculated as "multiscale" terrain variables by averaging over a series of increasing neighborhood sizes to incorporate

information from a range of spatial scales between 5-275 m (Dolan, 2012; Dolan & 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 & Baker, 2012), Diesing & Stephens (2015) suggested that the local variability in backscatter can also aid in differentiating coarse from fine sediments, as the former are characterized not only by an increase in backscatter, but also by an increase in backscatter variability. We calculated the local variability in backscatter (i.e., the range of backscatter values in a 3 x 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).

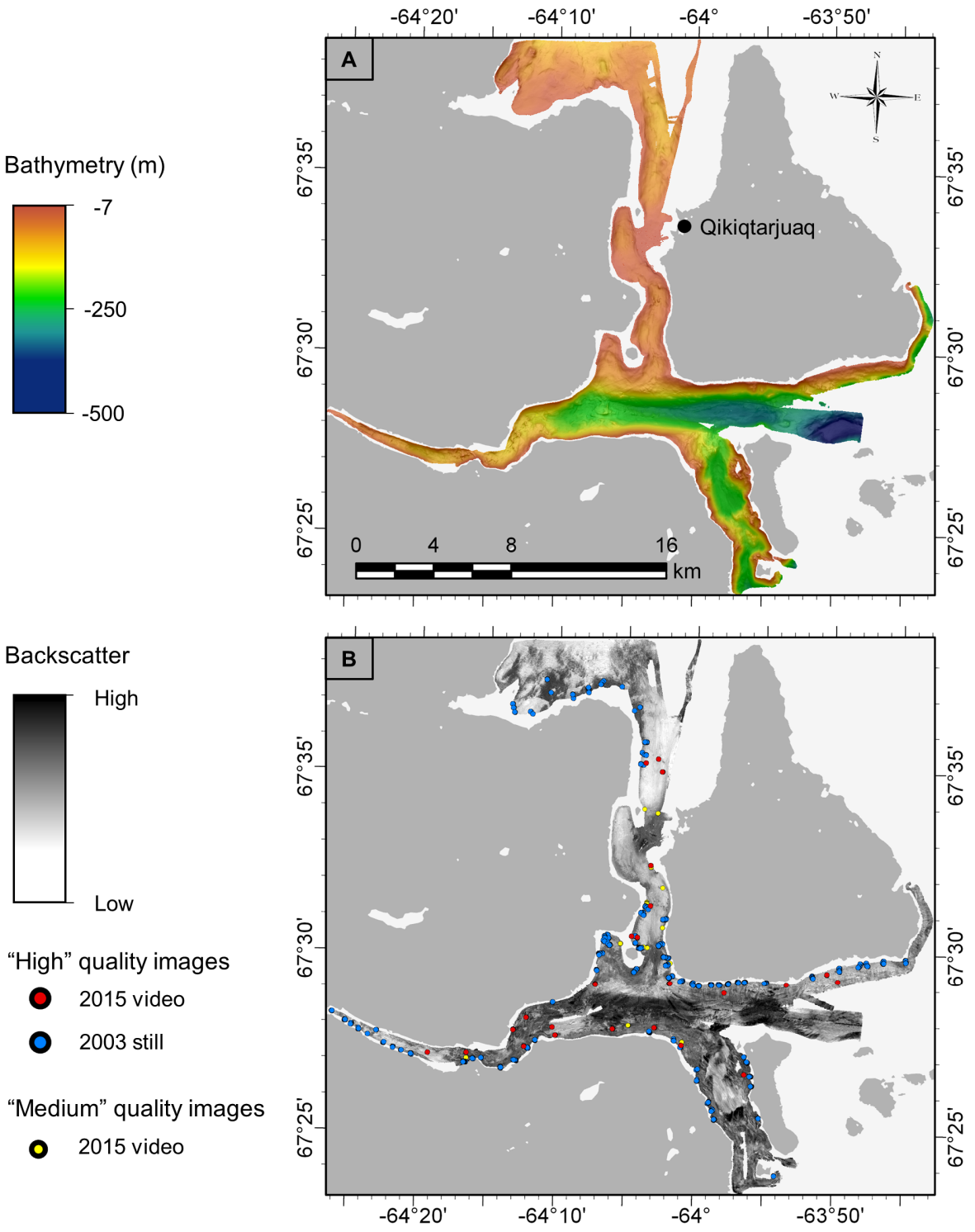


Figure 3.2. (A) Multibeam bathymetry near Qikiqtarjuaq. (B) Multibeam backscatter with ground truth image samples. Basemaps were obtained from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

Table 3.1. Multiscale variables tested for inclusion in *Mya* presence-absence and abundance models.

Variable	Calculation method	Method source
Bathymetry	Primary data	-
Eastness	TASSE	Lecours (2017)
Northness	TASSE	Lecours (2017)
RDMV[1]	TASSE	Lecours (2017)
SD[2]	TASSE	Lecours (2017)
Slope	TASSE	Lecours (2017)
Broad BPI[3]	BTM	Walbridge *et al.* (2018)
Fine BPI[4]	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 & 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)

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

3.3.2 *Mya* Ground-truth Data

Siferd (2005) collected and analyzed drop-camera bottom photographs from sites randomly selected along the coast near Qikiqtarjuaq to quantify the abundance of *Mya* (individuals per m^2). Ten down-facing 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 (e.g., Figure 3.3).

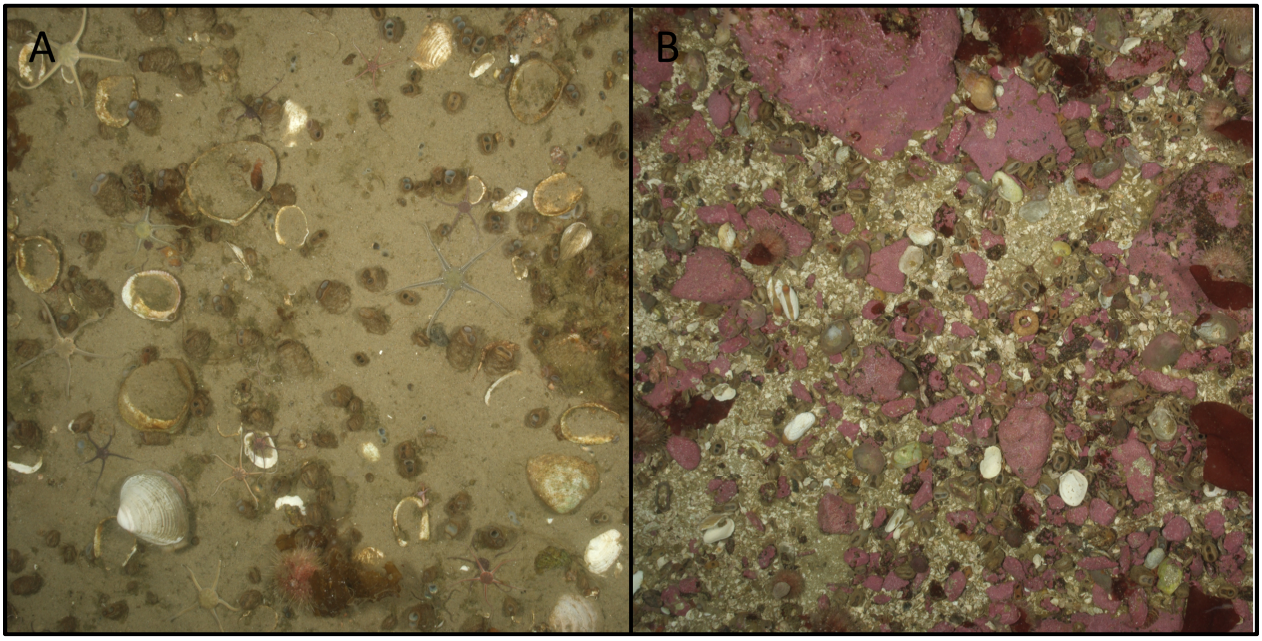


Figure 3.3. Underwater images from Siferd (2005) of (A) sandy substrate and (B) gravelly substrate, both showing abundant clam populations.

Underwater video was 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 down-facing 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 18x 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 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.

3.3.3 Statistical Modelling

Welsh *et al*. (1996) and Barry & 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 the 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 (e.g., Van Horne, 1983; Johnston *et al*., 2015; Tingley *et al*., 2016). Recall that 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. This resulted in two separate modelling datasets (Table 3.2): 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). We thus created separate models of presence-absence and abundance using the two datasets, and ultimately combined them for a single ensemble map prediction.

Table 3.2. Underwater image samples used for abundance and presence-absence modelling datasets.

	Siferd (2005)	2015 survey		Total
		"Medium" quality images	"High" quality images	
Abundance (n)	1813	-	172	1985
Presence-absence (n)	1827	274	172	2273

The Boosted Regression Trees (BRT) machine learning algorithm has been shown to consistently perform well compared to other SDM techniques due to its flexibility in fitting complex environmental relationships and its robustness to noisy data (Olden *et al*., 2008; Franklin, 2009). Elith *et al*., (2008) demonstrated the flexibility and robustness of BRTs in an ecological context, and Reiss *et al*. (2015) discussed how such techniques can outperform regression-based models at predicting a quantitative response, such as species abundance. BRTs were trained using the "gbm.step" function in the R package "*dismo*" (Hijmans *et al*., 2017). This package uses a deviance loss function, and we specified a Bernoulli error distribution for the *Mya* presence-absence model and Poisson 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 (e.g., Gottschalk *et al*., 2011; Millard & 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 next section on model evaluation). 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 (e.g., Clark *et al*., 2014). Both models were predicted over the full extent of the environmental data.

3.3.4 Model Evaluation

An important step in SDM is evaluating model performance (Franklin, 2009), which is commonly based on assessing a model's ability to predict data points that are withheld from model training. 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 & 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

success. 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) leave-one-out cross-validation (SLOO CV; e.g., 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 with other methods (e.g., blocking; Roberts *et al*., 2017).

The distance of the SLOO CV buffer was determined using empirical variogram analysis. We investigated spatial structure in ESRI ArcGIS by calculating the average length of each sample transect and the average distance between transects, using their mean center. The average distance between samples within transects were estimated given the number of samples and transect length. An empirical variogram of *Mya* abundance was generated to observe the similarity in observations over a range of distances, and to determine a suitable buffer distance at which the effects of

autocorrelation were negligible, based on the major range of the variogram model (Wagner & 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 with estimates from the internal 10-fold CV 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.

3.4 Results

3.4.1 Response to Environmental Variables

Mya were prevalent in the dataset, occurring in 84% of images analyzed. Eleven variables were selected to model their presence and absence after correlation reduction (Figure 3.4). 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 & 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 contributions (≤ 5%) to the model.

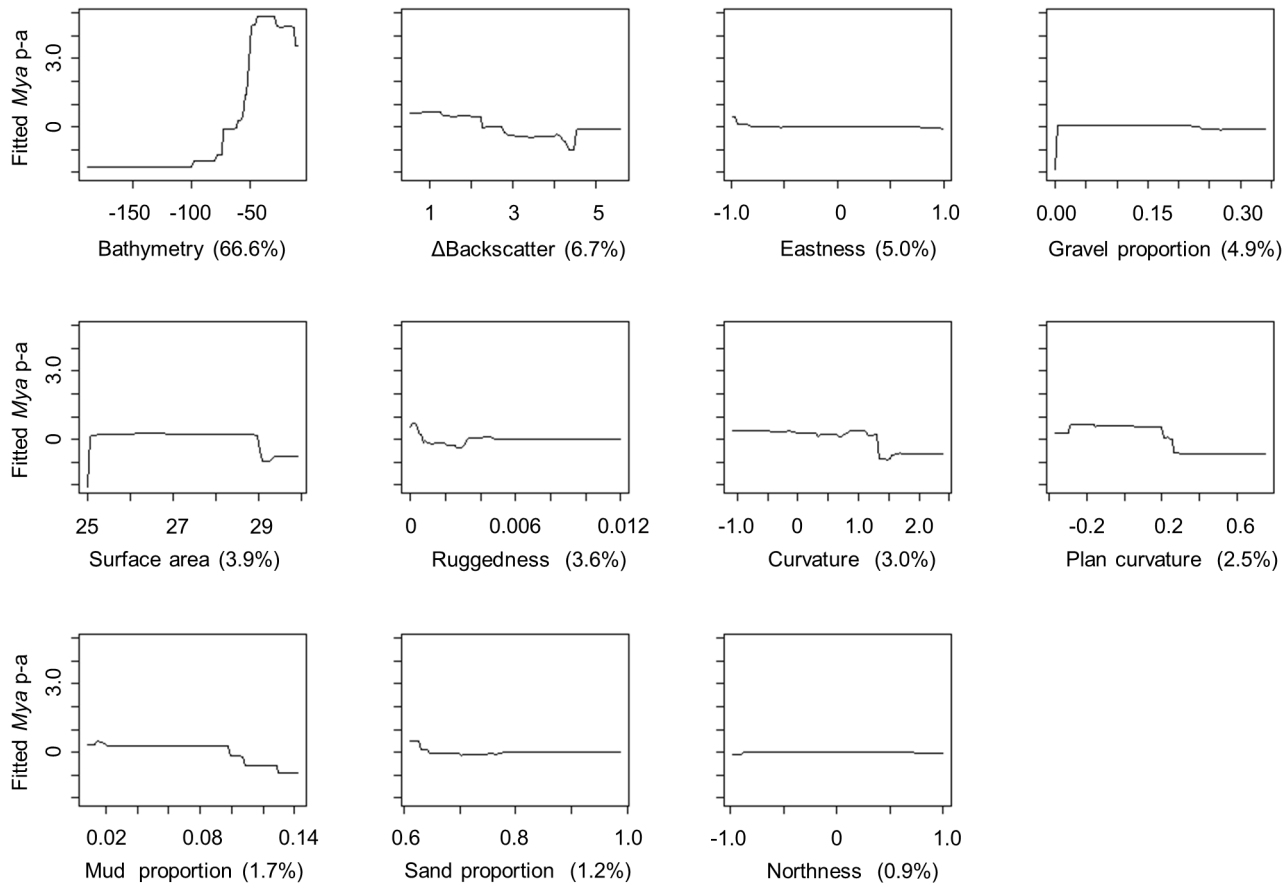


Figure 3.4. Partial response plots and percent contribution of explanatory variables (Table 3.1) to *Mya* presence-absence (p-a) model.

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 3.5). 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.

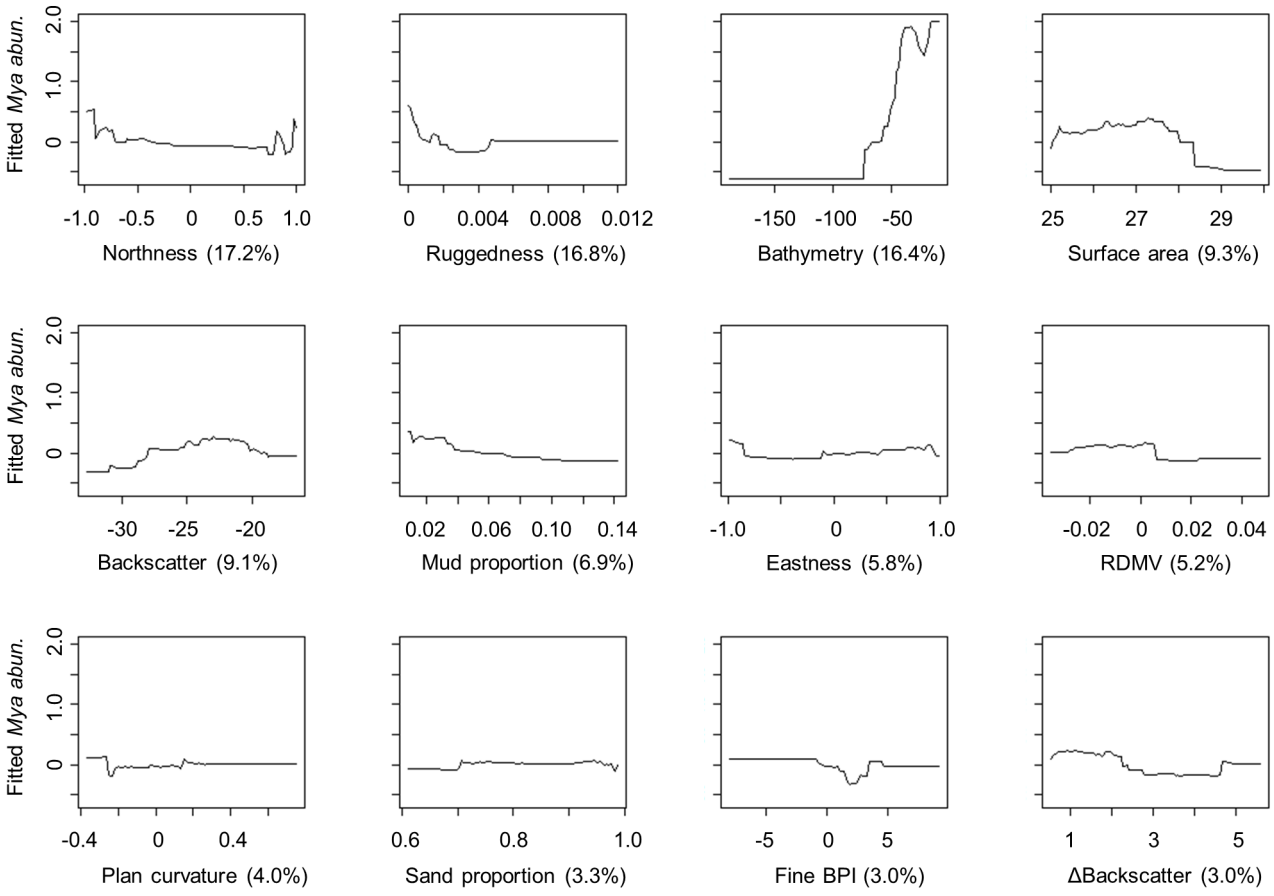


Figure 3.5. Partial response plots and percent contribution of explanatory variables (Table 3.1) to *Mya* abundance (abun.) model.

3.4.2 Statistical Modelling and Prediction

A probability threshold of 0.61 was selected based on maximizing the spatially independent accuracy of the combined presence-absence and abundance predictions. Therefore, locations where probability of presence was higher than 0.61 were considered as presences, while all other

areas were considered as absences. *Mya* were predicted present in much of the coastal area, but were most prevalent in Broughton Channel, where they were also predicted to occur further from the coast (Figure 3.6).

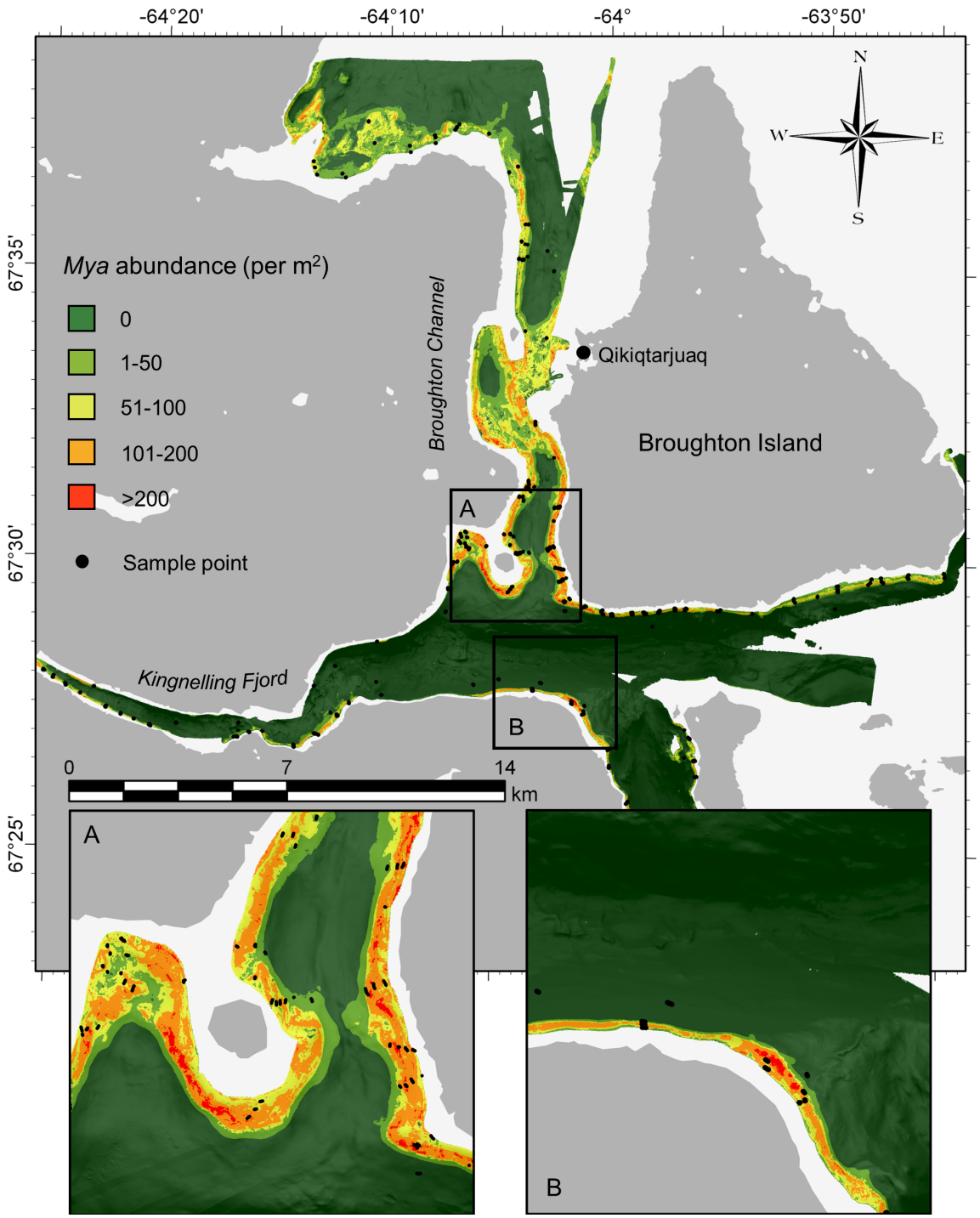


Figure 3.6. Combined prediction of *Mya* abundance, conditional on presence, near Qikiqtarjuaq, with insets (A) at the southern part of Broughton Channel, and (B) on the southern shore at the mouth of Kingnelling Fjord.

The highest abundances (> 200 individuals per m^2) were predicted in < 50 m water depth in southwest and southeast Broughton Channel (Figure 3.6A), on the southern shore at the mouth of

Kingnelling Fjord (Figure 3.6B), 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.

3.4.3 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. Based on these metrics, we calculated an empirical variogram for *Mya* abundance using 10 m lags and fit a circular model to observe the effects that clustered transect sampling had on the spatial structure of the data (Figure 3.7). The variogram demonstrated that, at the scale of interest, samples were spatially autocorrelated up to ~40 m. Therefore, nearly all samples within a given transect were expected to contain some degree of spatial autocorrelation.

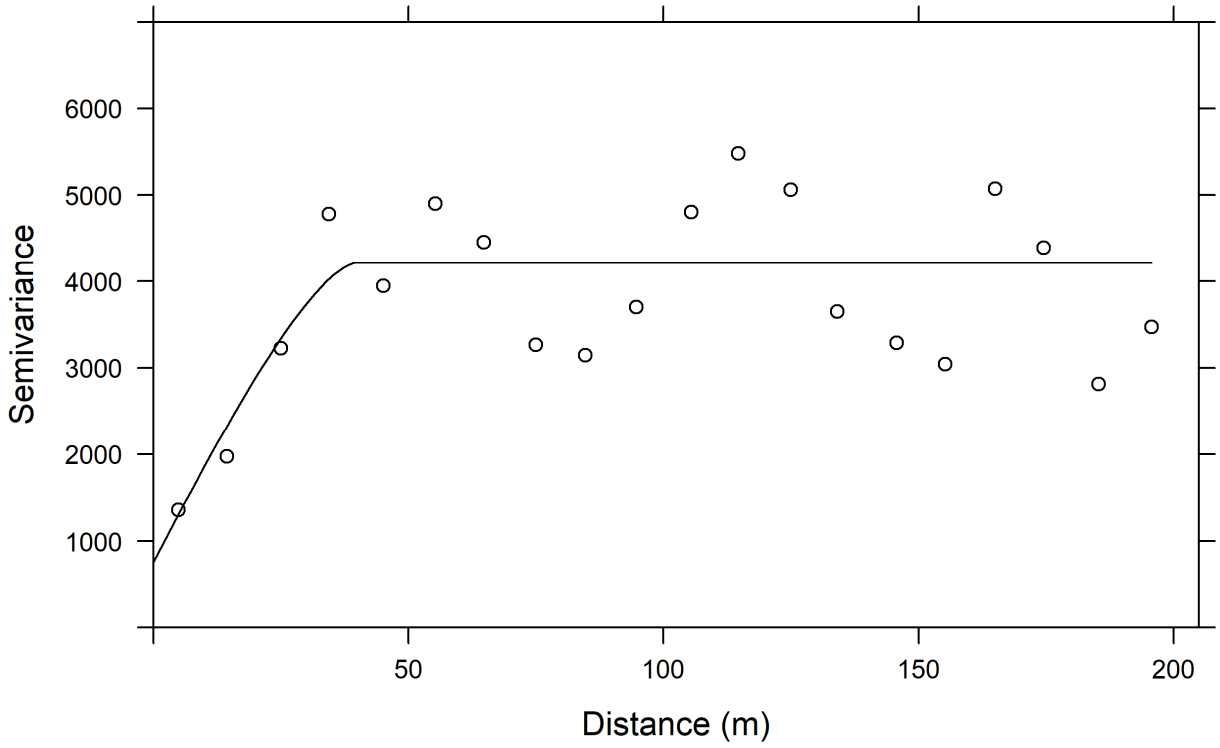


Figure 3.7. Circular variogram model of *Mya* abundance with 10 m lags. Partial sill = 3453, nugget = 760, range = 39 m.

SLOO CV was used with a 40 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 0.88 and 0.60, respectively. Pearson's and Spearman's correlation coefficients between observed and predicted abundances were $r = 0.56$ and $\rho = 0.65$, respectively; MAE and VE values were 43.59 and 0.28. 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 3.3). The combined abundance-conditional-on-presence predictions evaluated using SLOO CV were slightly more accurate than abundance alone, with $r = 0.56$, $\rho = 0.66$, MAE = 43.07, and VE = 0.28, yet they included "zero" predictions, unlike the abundance model in isolation (Figure 3.6).

Table 3.3. Performance of combined abundance model estimated using SLOO CV with a 40 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	0.56	0.78	0.22
Spearman ρ	0.65	0.81	0.15
MAE	43.59	31.57	11.50
VE	0.28	0.61	0.34

3.5 Discussion

The presence and absence of *Mya* was predicted primarily by bathymetry (Figure 3.4). 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, or potentially finer-grained substrates, as coarse substrates can be characterized by an increase in backscatter variability (Diesing & Stephens, 2015). 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.

On the other hand, *Mya* abundance predictions were influenced considerably by several topographic and substrate variables (Figure 3.5). 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 influence larval dispersal and control the flow of food to benthic filter feeders (Tong *et al*., 2016; Lacharité &

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 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 (Figure 3.5). *Mya* had a specific response to seabed substrate – favoring areas of moderate backscatter intensity (i.e., seabed hardness) and low mud proportion, which was also suggested by Pfitzenmeyer (1972) and Abraham & Dillon (1986). Anecdotal observation generally supports these predictions, with sandy or mixed sandy/gravelly substrates appearing to contain the highest abundances.

There were suspected sources of spatial and temporal modelling error that likely influenced the accuracy of predicted abundance. Though the positional accuracy of the GPS was < 3 m, it is likely that positional uncertainty exceeded 5 m at the greatest depths (near 200 m; Rattray *et al*., 2014). Therefore, it is possible that the positional uncertainty of some deep samples was greater than the 5 m raster resolution of the modelling data layers. The effect of this uncertainty on model accuracy was likely mitigated by the use of multiscale predictor variables, which incorporate information from broader scales. Ideally though, locational inaccuracy would be accounted for to limit the amount of error propagated to model predictions.

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 of 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.

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 abundances at depths > 70 m, but not necessarily absence. By combining the model of abundance with binary presence-absence predictions, we were able to predict unsuitable *Mya* habitat where the species is absent, while leaving abundance predictions of abundance intact where

habitat is suitable. This approach also acknowledges that the environmental factors determining if *Mya* are present and whether they are abundant may not be the same. 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 3.4), yet seabed topography and substrate properties ultimately influenced predictions of how abundantly *Mya* colonize these areas (Figure 3.5). This approach is similar to the use of parametric conditional or hurdle models for zero-inflated count data (Welsh *et al*., 1996; Martin *et al*., 2005).

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

The empirical variogram of *Mya* abundance suggested that samples within 40 m of one another generally contained some amount of spatial autocorrelation, which includes 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 3.3). Using 10-fold CV for evaluation, the abundance model seems highly accurate at both linear and non-linear monotonic prediction (r = 0.79, ρ = 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 = 0.56, ρ = 0.65), and an average error that was 72% of the variance in the observed data (VE = 0.28).

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 only the abundance model. This also utilized a greater proportion of the sample dataset, which were of sufficient quality to determine presence or absence of *Mya* but not abundance. Using a spatial leave-one-out cross-validation demonstrates how failing to account for spatial autocorrelation when evaluating species distribution models can substantially inflate estimates of predictive performance. Transect sampling is common in marine science, yet its effects on species distribution models are often not considered. 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.

3.6 Conclusions

Results suggest that although bathymetry is the primary limiting factor to *Mya* habitat, seabed topography, morphology, and substrate properties jointly predicted 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 the relationship 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. When designing discrete transect surveys (i.e., discrete series of continuous observations), modelers may consider placing greater emphasis on selecting a greater number of short transects rather than fewer long transects in order to maximize the amount of information available for modelling, while minimizing the number of autocorrelated observations. Regardless, we recommend that exploring the spatial dependence of ground-truth data should be considered a compulsory step in marine SDM. This can only help to improve the transparency of model quality and limitations for map users.

3.7 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. Biological Report, US Fish and Wildlife Service Series, TR EL-82-4. US Fish and Wildlife Service.

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(6): 969–978.

Araújo, M. B., and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33(10): 1677–1688.

Bahn, V., and McGill, B. J. 2013. Testing the predictive performance of distribution models. *Oikos*, 122(3): 321–331.

Barry, S. C., and Welsh, A. H. 2002. Generalized additive modelling and zero inflated count data. *Ecological Modelling*, 157(2–3): 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(4): 577–598.

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(3): 502–520.

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.

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(5): 990–999.

Dale, M. R. T., and Fortin, M.-J. 2002. Spatial autocorrelation and statistical tests in ecology. *Écoscience*, 9(2): 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 algorithms, 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(2): 187–219.

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

Dyke, A. S., Andrews, J. T., and Miller, G. H. 1982. Quaternary geology of Cumberland Peninsula, Baffin Island, District of Franklin. Memoir 403. Geological Survey of Canada, 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(2): 129–151.

Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 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(1): 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(3): 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(18): 3403–3412.

Guisan, A., and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3): 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 Series in Statistics. 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(10): 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, 84–1B. Geological Survey of Canada.

Hijmans, R. J. 2012. Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology*, 93(3): 679–688.

Hijmans, R. J., Phillips, S. J., and Elith, J. 2017. dismo: Species Distribution Modeling. R package version 1.1-4. Available from: https://CRAN.R-project.org/package=dismo

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(7): 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.

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(7): 811–820.

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., 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.

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(9): 5391–5406.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology*, 74(6): 1659–1673.

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(11): 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(7): 8489–8515.

Miller, J. 2010. Species distribution modeling. *Geography Compass*, 4(6): 490–509.

Misiuk, B., Lecours, V., and Bell, T. 2018. A multiscale approach to mapping seabed sediments. *PLoS ONE*, 13(2): e0193647.

Mullahy, J. 1986. Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3): 341–365.

Nunavut Department of Environment - Fisheries and Sealing Division. 2012. Nunavut coastal resource inventory - Iqaluit. Accessed from: https://www.gov.nu.ca/sites/default/files/ncri_iqaluit_en.pdf

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(2): 171–193.

Petersen, G. H. 1978. Life cycles and population dynamics of marine benthic bivalves from the Disko Bugt area of West Greenland. *Ophelia*, 17(1): 95–120.

Pfitzenmeyer, H. T. 1972. Tentative outline for inventory of molluscs: *Mya arenaria* (soft-shell clam). *Chesapeake Science*, 13: 182–184.

Porskamp, P., Rattray, A., Young, M., and Ierodiaconou, D. 2018. Multiscale and hierarchical classification for benthic habitat mapping. *Geosciences*, 8(4): 119.

Rattray, A., Ierodiaconou, D., Monk, J., Laurenson, L. J. B., and Kennedy, P. 2014. Quantification of spatial and thematic uncertainty in the application of underwater video for benthic habitat mapping. *Marine Geodesy*, 37(3): 315–336.

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(2): 297–315.

Ridout, M., Demétrio, C. G., 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(8): 913–929.

Segurado, P., Araújo, M. B., and Kunin, W. E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43(3): 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.

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(5): 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(5): 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(2): 225–232.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. *The Journal of Wildlife Management*, 47(4): 893–901.

Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36(12): 2290–2299.

Wagner, H. H., and Fortin, M.-J. 2005. Spatial analysis of landscapes: Concepts and statistics. *Ecology*, 86(8): 1975–1987.

Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. J. 2018. Unified geomorphological analysis workflows with Benthic Terrain Modeler. *Geosciences*, 8(3): 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(3): 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(1–3): 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.

4. A Spatially Explicit Comparison of Quantitative and Categorical Modelling Approaches for Mapping Seabed Sediments

4.1 Introduction

There is growing pressure on marine ecosystems due to human use, especially near coasts where interactions between terrestrial and marine drivers have the potential to generate large cumulative impacts (Halpern *et al*., 2008). Coastal ecosystems provide many important goods and services to both coastal and inland inhabitants (Costanza *et al*., 1997; Ghermandi *et al*., 2010; Galparsoro *et al*., 2014). Therefore, it is often necessary to balance competing demands from stakeholders with the sustainable management of marine resources and ecology (Baker & Harris, 2012). Marine spatial planning (MSP) is a framework by which this can be accomplished (Ehler & Douvere, 2009). Using MSP, local maps of ecology are analyzed alongside those of human use to identify overlaps and conflicts. This spatial information is used to implement management plans for the current and future use of the marine system (Ehler & Douvere, 2009). As the primary means of conveying spatial information, maps are, therefore, key components to such management initiatives.

Seafloor substrate maps are particularly useful for determining the distribution of coastal marine biota. Substrate composition can be a strong predictor of benthic biodiversity (McArthur *et al*., 2010). The presence of hard substrata, for example, can provide attachment surfaces for sessile animals, marine algae, and the grazers that feed on them, while soft sediments provide habitat for many infaunal invertebrates (McArthur *et al*., 2010). Substrate composition also modifies seabed morphological complexity by providing structure and shelter for marine fauna – factors that

correlate with biodiversity (Beaman *et al*., 2005). Substrate maps can therefore provide information on the distributions of single species or biodiversity, both of which may be important components of a given management framework.

A variety of methods for producing seabed sediment maps have been explored (e.g., Che Hasan *et al*., 2012; Diesing *et al*., 2014; Stephens & Diesing, 2014). Surficial sediment maps were traditionally produced by manual interpretation of ground-truth data in the context of local geomorphology, and often, sonar data (e.g., Todd *et al*., 1999), but modern methods increasingly rely on automated objective approaches (Brown *et al*., 2011). These have recently become feasible thanks to the widespread accessibility of digital data, powerful GIS tools, and high-performance computing – they allow for mapping a range of substrate characteristics. Grab sample and sediment core data, for example, have been used to model sediment grain size (Stephens & Diesing, 2014; Galparsoro *et al*., 2015) and particulate organic carbon content in unconsolidated sediments (Diesing *et al*., 2017) based on MBES data to produce continuous map predictions, while the presence of rock or hard substrates has been modelled and mapped from underwater video (Li *et al*., 2016; Siwabessy *et al*., 2018). There are now a variety of approaches to choose from for a given mapping application and it is important to select those that fit the given geographic and dataset characteristics.

Coupled with high resolution acoustic mapping, automated statistical methods are among the most promising recent approaches to mapping seabed sediments. They perform well compared to other methods (Li *et al*., 2010; Diesing *et al*., 2014) and are objective, providing several advantages over manual or subjective approaches (Ierodiaconou *et al*., 2011). In supervised modelling, ground-truth sediment samples (e.g., grabs, cores, video observations) are used to train a statistical model based on environmental data (e.g., depth, seabed morphology, acoustic seabed properties).

Statistical relationships between sediment samples (response variable) and environmental data at the sample location (explanatory variables) are used to predict sediment characteristics at unsampled locations. With spatially continuous remotely sensed environmental data, it is therefore possible to produce full-coverage seabed sediment maps from relatively sparse sediment samples.

For producing classified maps of sediment grain size, several common textural classification schemes such as Folk (1954) place grain size samples on a ternary diagram according to the ratio of sand:mud and the percentage of gravel (Figure 4.1A). Similar textural schemes coarsen the thematic resolution of Folk's by aggregating to fewer classes, such as the British Geological Survey (BGS) modification for small-scale (1:1,000,000) maps, which eliminates the "slightly gravelly" classes (Figure 4.1B; Long, 2006). To account for substrate types used in the European Nature Information System (EUNIS) habitat classification, a further simplified version of Folk's classification with only four classes has been suggested, and is widely used (Figure 4.1C; Connor *et al*., 2006; Long, 2006). Among other criteria, the selection of a classification scheme may be for compatibility with regional management systems (e.g., EUNIS; Davies *et al*., 2004), for alignment with existing literature (Strong *et al*., 2018), or for matching with ground-truth data.

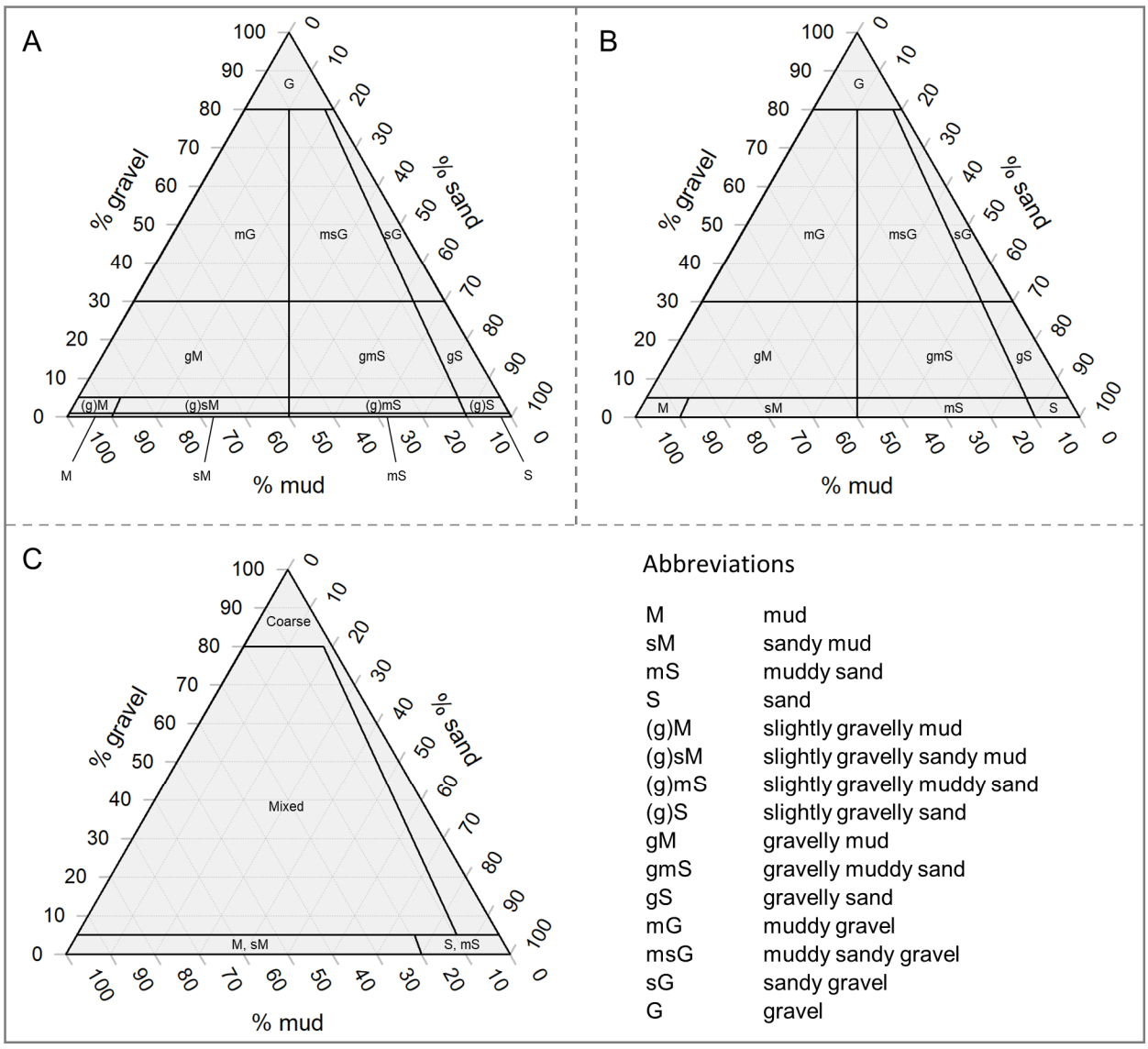


Figure 4.1. Ternary diagrams with (A) Folk, (B) simplified Folk, and (C) EUNIS classes.

Using a supervised modelling approach, ground-truth sediment data are commonly treated in two ways to produce classified maps of seabed sediment according to schemes such as those described in Figure 4.1:

1. Quantitative measures of a substrate property, such as grain size fraction (e.g., percent mud, sand, and gravel), are used to predict quantitative gradational values across the full environmental data coverage (e.g., Stephens & Diesing, 2015). These predictions are useful for management or

further modelling, but some applications require classified or thematic maps, which can be produced by classifying the quantitative predictions according to some scheme (e.g., Figure 4.1). This is useful for summarizing sediment composition in a single map, or for ensuring compatibility with regional management plans or similar research (see Strong *et al.*, 2018 for discussion of classification and compatibility).

2. Ground-truth data are aggregated according to a classification scheme prior to modelling, thereby treating them as categorical variables (e.g., Diesing & Stephens, 2015). It may also be the case that inherited data (e.g., from the literature, online databases, legacy data) are already classified, and the quantitative data are unavailable, or that datasets consist of sediment classes derived from visual assessment. In these cases, the options available to the modeller are limited, and the categorical classification approach may be the logical choice. Using this approach, a model can predict the occurrence of the observed classes over the full extent of the environmental data.

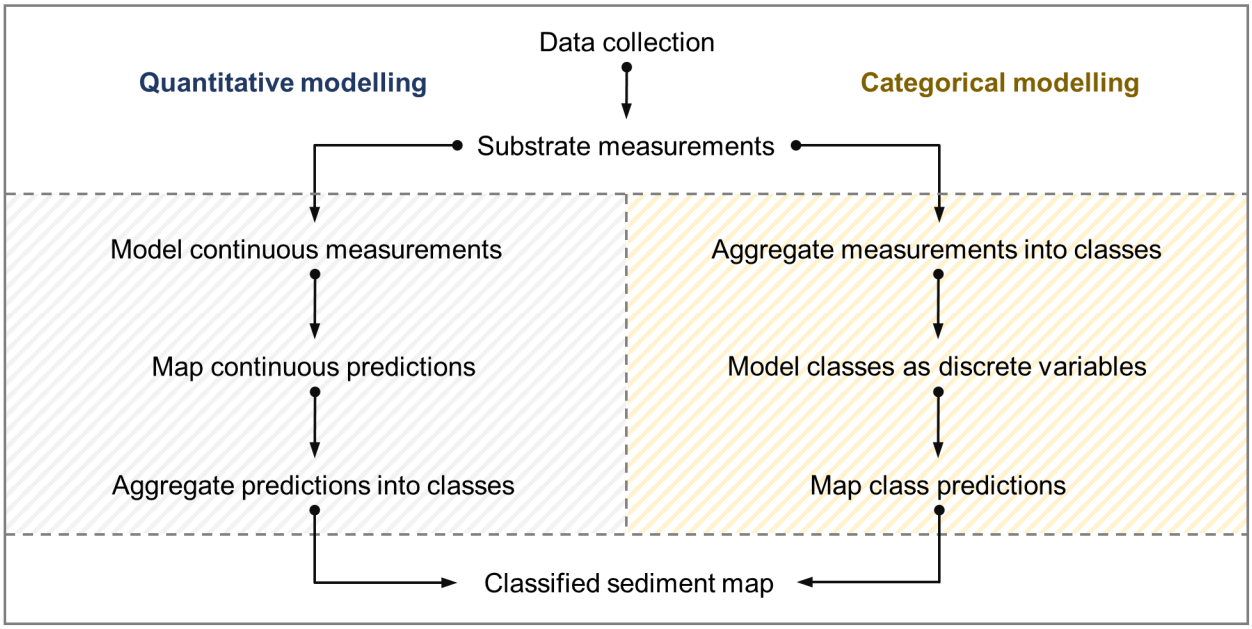


Figure 4.2. Two common supervised workflows for producing objective classified seabed sediment maps.

Here we will refer to these as "quantitative" and "categorical" modelling approaches. While the quantitative approach is also known as "continuous" or "regression" modelling, and categorical is commonly referred to as "classification", we use the terms "quantitative" and "categorical" to reduce confusion, since the other terms also have other meanings that are relevant here (e.g., *classified maps* can be produced from either approach, and all predictions are *spatially continuous*).

Each of these broad approaches contain numerous individual modelling techniques with their own intricacies, many of which have been compared in the ecological and conservation management literature (e.g., presence-absence models: Elith *et al*., 2006; regression models: Hernandez *et al*., 2006; machine learning: Olden *et al*., 2008; geostatistical and hybrid methods: Li *et al*., 2010, 2017). Modern open-source software has given users potentially unlimited access to most modelling methods. For seabed mapping, if the end goal is a classified seabed sediment map and if there are no pre-existing classification constraints (e.g., pre-classified data, regional compatibility), then understanding the suitability of the two approaches outlined in Figure 4.2 will help considerably in narrowing modelling decisions.

There are apparent advantages and disadvantages to both quantitative and categorical sediment modelling approaches for producing classified maps. Unclassified quantitative predictions on their own constitute a useful result for further modelling and mapping and are flexible once produced – it is easy to classify and reclassify quantitative values as necessary. The modelling process can be complex though, potentially involving data transformations such as additive log-ratios for compositional data (Aitchison, 1982), multiple models for different log-ratios (Stephens & Diesing, 2015), and multiple corresponding tuning and variable selection procedures. On the other hand, the categorical modelling procedure can be more straightforward, requiring little data

manipulation once ground-truth measurements have been aggregated into classes. Categorical models may only require tuning parameters to a single set of variables to predict the entire range of sampled classes. Once produced though, classes are more static when compared to quantitative predictions. Though it may be possible to simply aggregate mapped classes to a more general scheme (e.g., Folk to simplified Folk; Figure 4.1), it may also be necessary to re-classify the ground-truth, select new variables, and re-tune model parameters for a new classification. This may be necessary if the original scheme is a poor match for the data or if there is a desire to test classifications at different hierarchical levels of detail.

The characteristics of the ground-truth data and the type of prediction required of the models may also be important qualities for determining their suitability for producing classified maps. For instance, sample size, distribution and bias, class prevalence, and spatial dependence are known to have profound effects on the performance of distribution models (Hirzel & Guisan, 2002; Araújo & Guisan, 2006; Hernandez *et al*., 2006), and particularly Random Forest (Millard & Richardson, 2015), which was used here. These and other dataset characteristics might influence the appropriateness of the approach selected for producing classified maps. For instance, rare classes may be difficult to model using a categorical approach when they have been sampled few times, but may cause less of an issue when modelled as quantitative variables. In some cases, clustered or uneven sampling may create spatial dependence in the response data (Hammond & Verbyla, 1996), violating assumptions of independence (Legendre, 1993). This could have unintended consequences for prediction and apparent model accuracy when extrapolating to new locations (Hammond & Verbyla, 1996; Segurado *et al*., 2006), and the magnitude of these consequences could depend partly on the modelling approach. Here we refer to extrapolation in a spatial sense as predictions outside of the sampled area, whereas interpolation is between sample locations. An

implicit assumption then, is that interpolation operates within the sampled environmental conditions, while extrapolation may predict outside of them.

The primary goal of this study was to create a classified seabed sediment map for inner Frobisher Bay, Nunavut, Canada from grab samples and underwater video using the Random Forest statistical modelling algorithm. Ground-truth characteristics, however, suggested that spatial dependence might be an issue when extrapolating seabed sediment characteristics to unsampled locations and evaluating these predictions. We therefore undertook a spatially explicit investigation of the qualities of the two broad modelling approaches – quantitative and categorical (Figure 4.2) – for predicting sediment grain size classes from grab samples. Coarse substrates that were not adequately represented in grab samples were modelled separately using underwater video data, and the two predictions were subsequently combined to produce a single map of surficial sediment distribution.

Specifically, when evaluating the quantitative and categorical modelling approaches for producing classified maps, we investigated: 1) their performance when extrapolating grain size predictions to new locations and if performance was affected by spatial autocorrelation, 2) the appropriateness of three levels of classification based on the relative proportions of grain size measurements, and 3) if the two approaches produced similar maps. Because the observations of coarse sediment from video transects were likely to be spatially autocorrelated, we investigated if the proximity of these samples 1) inflated the apparent accuracy of coarse substrate predictions, and 2) caused overfitting in model training. The results of these investigations informed the selection of modelling approach, while also providing spatially explicit accuracy estimates. Based on the results, we provide recommendations on the utility and potential pitfalls of these approaches in a spatial context.

4.2 Data and Methods

4.2.1 Study Area

Frobisher Bay is a long (~265 km), northwest-southeast-oriented macrotidal fjord located in southeastern Baffin Island, Nunavut, Canada (Figure 4.3). It can be conveniently partitioned into two morphologically distinct sections. The inner part is fjärd-like, spanning from the northwest head of the bay to the mid-bay islands, with a maximum depth of approximately 350 m. Much of this section is shallow (< 100 m) with extreme tides (> 10 m) resulting in extensive tidal flats. The mid-bay islands separate the predominantly muddy shallow inner bay from the coarse and bedrock-dominated outer bay, which deepens to over 800 m and opens to the North Atlantic. The southwest coast of the outer bay is the fault boundary of a half-graben and is characterized by steep rock cliffs.

This study focuses on inner Frobisher Bay. The morphology and orientation of submarine features here are a product of repeated Quaternary glaciations, the most recent of which receded between 9-7 ka (Hodgson, 2005; Tremblay *et al*., 2015). These have produced a complex, heterogeneous seabed, with erosional and depositional glacial features such as scour troughs and moraines indicative of southeast ice flow. Currently, seabed sediments are re-mobilized by several non-glacial processes, including tidal currents, submarine slope failures, and iceberg and sea-ice scour (Deering *et al*., 2018).

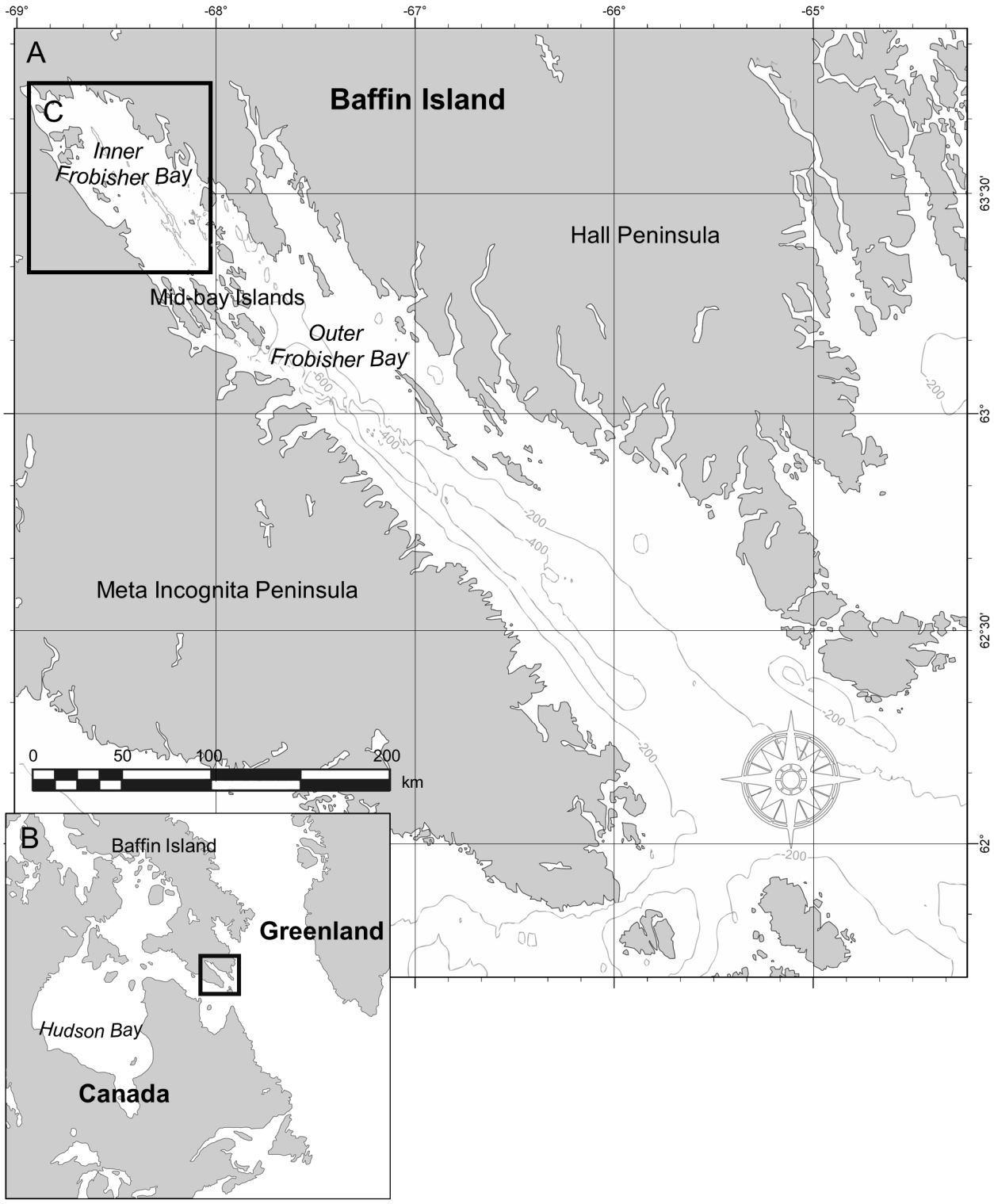


Figure 4.3. (A) Frobisher Bay, Nunavut, Canada with 200 m bathymetric contours from the GEBCO_2014 grid (Weatherall *et al*., 2015) and coastline reproduced from ESRI (2011), with (B) location on southeastern Baffin Island, and (C) the study area – inner Frobisher Bay.

4.2.2 Environmental Data

Multibeam echosounder (MBES) bathymetry and backscatter data (Figures 4.4, 4.5) were collected between 2006 and 2017 to characterize the seabed as part of the ArcticNet project "Integrated Marine Geoscience to Guide Environmental Impact Assessment and Sustainable Development in Frobisher Bay, Nunavut" (Deering *et al*., 2018). The CCGS *Amundsen* collected opportunistic MBES data during transit to and from Iqaluit between the years 2006-2008 with a Kongsberg EM300 30 kHz echosounder, and between 2009-2017 with an EM302 (30 kHz). The RV *Nuliajuk* completed targeted surveys in the bay between 2012-2013 with an EM3002 (300 kHz), and between 2014-2016 with an EM2040C (200-400 kHz). Details on MBES data processing are included in Appendix B.1.

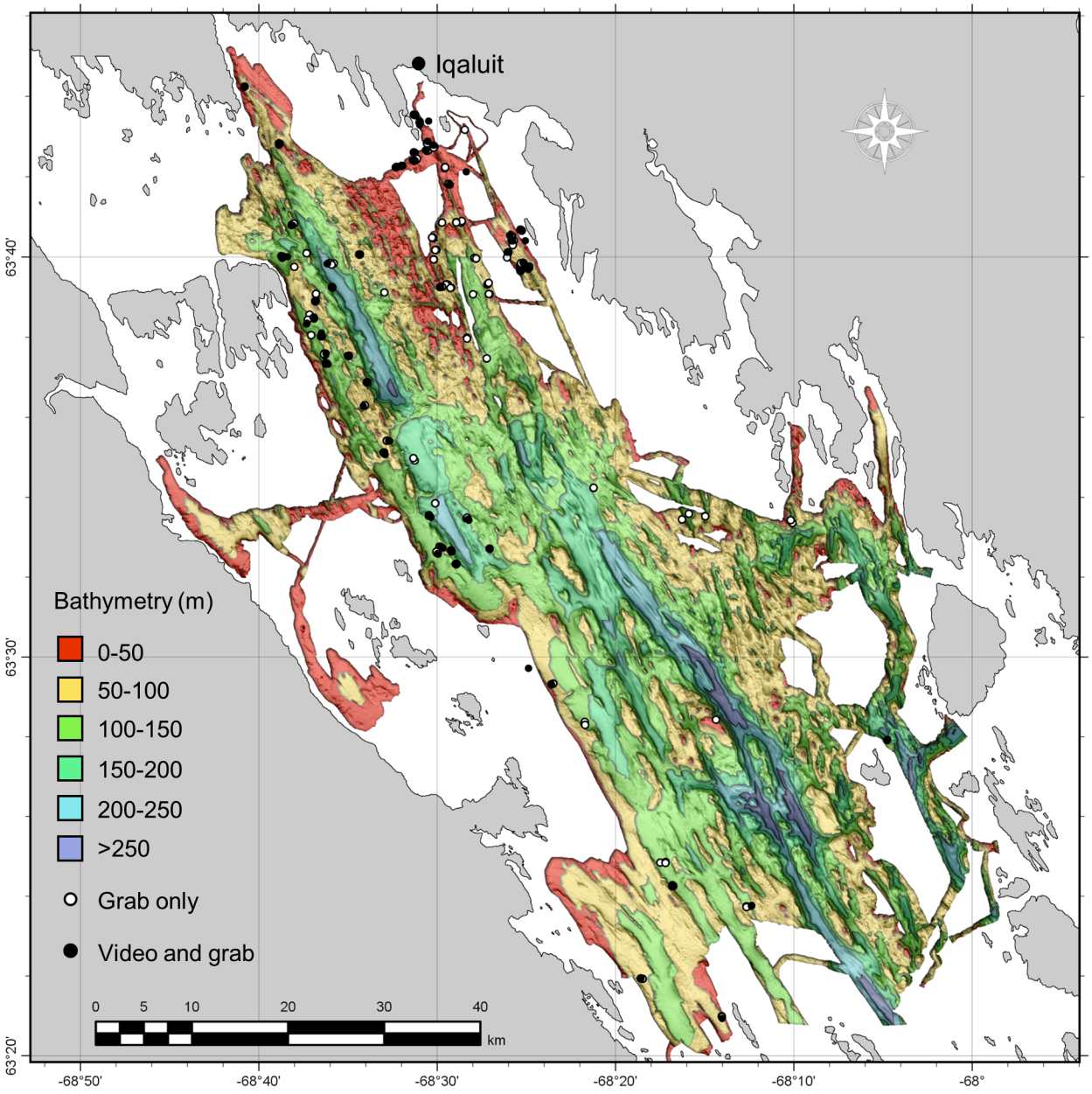


Figure 4.4. Inner Frobisher Bay MBES bathymetry contoured at 50 m with shaded terrain and sample sites.

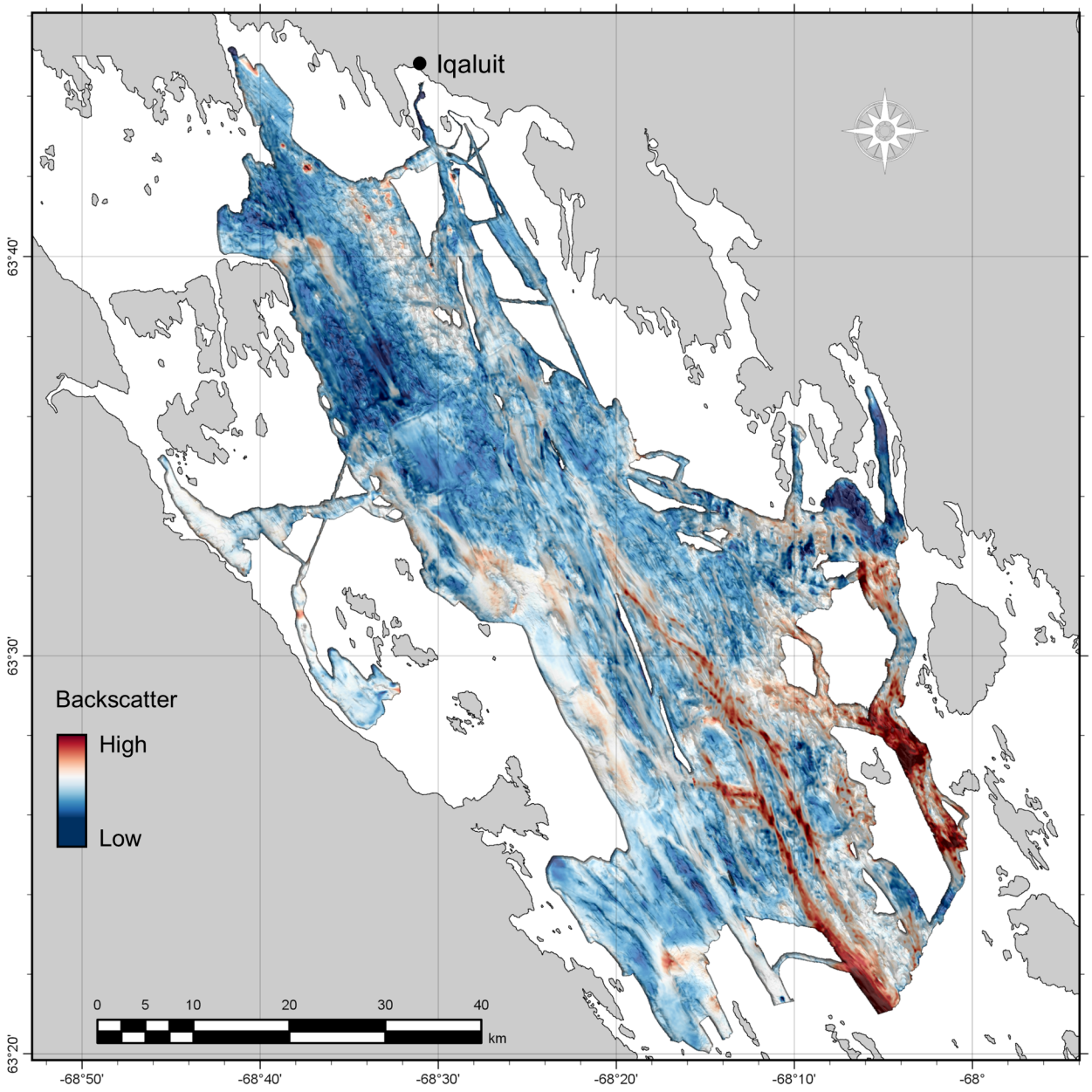


Figure 4.5. Inner Frobisher Bay relative MBES backscatter.

4.2.3 Ground-truth

Ideally, a sediment model would rely on a single consistent source of ground-truth data, yet grab samples commonly fail to accurately represent coarser sediments (Rees, 2009; Eleftheriou & Moore, 2013), while it can be difficult to consistently distinguish mud from sand in underwater video (e.g., Whitmire *et al*., 2007). To overcome these limitations, we modelled fine grain sizes (<

4 mm) using grab sample data, and coarse substrates (≥ 4 mm; i.e., pebble, cobble, boulder) from video observations.

Grab samples (n = 239) and underwater video (n = 78) were collected in 2015 and 2016 to provide substrate ground-truth for the MBES data (Figures 4.4, 4.5). Ground-truth sites were selected from the area of MBES coverage prior to each field season using a random approach, stratified by water depth up to 200 m and seabed slope. Because sampling and mapping occurred simultaneously, only part of the final mapped area was available for sample site selection each year, resulting in unsampled areas.

A live-feed Deep Blue Pro underwater camera with a GoPro Hero4 was deployed at each selected site to collect high-definition video for a four-minute drift. Two lights were attached to the camera mount to illuminate the seabed, and two green lasers, spaced 5 cm apart, were attached for scale. Positioning was obtained using a Garmin 18x PC GPS with video overlay, providing coordinates at the surface for the duration of the recording. GPS accuracy was rated at < 3 m, and though efforts were made to keep the location of the GPS above the camera during drifts, it is likely that the locational error was greater under windy or high current conditions due to horizontal drift of the camera from the vessel, especially at the greatest depths. Positional error potentially exceeded 10 m under high drift conditions; if this was suspected, the drift was cut short. Still frames were extracted for analysis every 10 s for the duration of the video. If coarse substrates (pebble, cobble, boulder) were visible in a frame, they were labelled as “present”. All observations were aggregated so that coarse substrates were labelled as “present” or “absent” for each 10 m raster cell that was sampled.

Up to three sediment samples were collected from the area near each site using either a 24 l Van Veen, or a 2.4 l Petite Ponar grab and individually georeferenced from the surface using the ship GPS location. Each grab sample was sub-sampled for ~100 g of sediment, which was considered sufficient for measuring grain size composition up to 4 mm. These were stored in a sample jar and frozen for transport. In the lab, samples were thawed and dried at low heat in an oven. Samples were dry-sieved for 5 minutes in a mechanical sieve shaker to separate mud (< .063 mm), sand (.063-2 mm), and gravel (2-4 mm) fractions. Many samples had a high proportion of flocculant mud that failed to disperse during dry sieving. To obtain an accurate measure of the mud fraction, samples were gently spray-rinsed through a .063 mm sieve and agitated by hand, washing away the mud fraction. The remaining sediments coarser than .063 mm were re-dried and weighed to estimate the proportion of mud that was lost. The weights of each fraction were divided by the total weight to obtain percent mud, sand, and gravel composition.

4.2.4 Statistical Modelling

The Random Forest machine learning algorithm was used to model both sediment grain size and presence of coarse substrates. Random Forest is a stochastic ensemble modelling method that uses bagging (Breiman, 1996) to combine the results of many individual classification or regression trees. This is an effective method for stabilizing predictions amidst noisy data. Because Random Forest is an ensemble of trees, it can handle complex non-linear responses with interaction. The algorithm also selects a subset of predictors to test at each tree node rather than selecting from the entire set of predictors, which can be effective at reducing overfitting, even amidst noisy data (Breiman, 2001). Furthermore, Random Forest can handle missing explanatory data values (e.g., at edges of certain terrain variable rasters), and large numbers of quantitative or categorical

variables, ostensibly ignoring unimportant ones (Liaw & Wiener, 2002). These qualities have made Random Forest popular for environmental modelling. It can perform both regression for quantitative response data and classification for categorical data, making it an ideal choice for this study.

For the categorical grain size modelling approach, samples were assigned class labels according to three ternary schemes prior to modelling to test different levels of data aggregation. The Folk (1954) and simplified Folk schemes were used according to Long (2006). Note that here, the "slightly gravelly" boundary for the Folk classification is at 1% rather than "trace", as in Folk (1954). The third classification was simply "muddy" or "sandy" if there was a majority of either size fraction. This was used instead of the EUNIS simplification to the Folk classification (Figure 4.1C), which was not appropriate given the data – most samples were muddy and < 5% gravel. The EUNIS simplification would aggregate nearly all the samples into the class "mud and sandy mud".

For the quantitative approach, percent mud, sand, and gravel measurements were transformed to an unbound additive log-ratio (ALR) scale (Aitchison, 1982) with respect to the mud fraction (Equations 4.1, 4.2). Note that the results are unaffected by the choice of size fraction to serve as the denominator (Pawlowsky-Glahn & Olea, 2004).

$$ALR_{sm} = \log(\frac{\%sand}{\%\, mud}) \tag{4.1}$$

$$ALR_{gm} = \log(\frac{\%gravel}{\%\, mud}) \tag{4.2}$$

Model predictions were then back transformed to a compositional scale bound between 0 and 1, corresponding to the relative percentage of each size fraction, and summing to 1 for each sample (Equations 4.3-4.5; Diesing, 2015).

$$sand = \frac{\exp(ALR_{sm})}{\exp(ALR_{sm}) + \exp(ALR_{gm}) + 1} \tag{4.3}$$

$$gravel = \frac{\exp(ALR_{gm})}{\exp(ALR_{sm}) + \exp(ALR_{gm}) + 1} \tag{4.4}$$

$$mud = 1 - (sand + gravel) \tag{4.5}$$

To produce a classified map from the quantitative output, predictions were classified according to the three schemes above.

The presence or absence of coarse substrates was recorded for each underwater video still frame to produce binary presence-absence data. This was used to train a categorical Random Forest model, essentially treating the presence or absence of coarse substrates as a two-class categorical response. Random Forest can output class probabilities rather than the class of majority vote, which allows the presence threshold to be tuned according to the classification goal, rather than the arbitrary default value of 0.5. It is important to tune the threshold for binary classifiers because model prediction and performance are sensitive to class prevalence (Liu *et al*., 2005).

4.2.5 Explanatory Variables

In addition to the primary MBES data (bathymetry and backscatter) and distance to the nearest coast, 11 secondary variables were tested for inclusion in each of quantitative and categorical grain

size models and the coarse substrate model using a multiscale approach (Appendix B.2). Five bathymetric derivatives suggested by Lecours *et al*. (2017) were calculated that describe most of the topographic structure of a surface (eastness and northness of slope, relative difference to the mean bathymetric value, slope, local bathymetric standard deviation). Local bathymetric standard deviation was omitted because it was highly correlated with seabed slope but did not perform as well. Three measures of curvature (total, plan, profile) and two measures of surface complexity (surface area:planar area [SA:PA], vector ruggedness measure [VRM]) were derived for their potential as topographic surrogates that influence bottom currents, and potentially sediment transport. The range of backscatter values in a circular neighborhood and the standard deviation in a 3 x 3-pixel neighborhood were derived for each spatial scale as potential surrogates for local substrate variability.

Random Forest is generally considered robust to noise, ignoring unimportant predictors, but there are benefits to variable reduction such as decreased variability between model runs and more accurate estimates of error (Millard & Richardson, 2015). We simplified the predictor set by removing variables that had Spearman's $\rho \geq 0.70$ – a common threshold for reducing correlated variables (e.g., Gottschalk *et al*., 2011; Dormann *et al*., 2013; Downie *et al*., 2016; Appendix B.2).

4.2.6 Evaluating Model Performance

All predictions were first tested using a leave-one-out cross validation (LOO CV). Using this approach, a single sample is removed from the dataset and all other samples (N-1) are used to train the model. The class of the withheld sample is then predicted using the N-1 model. This is repeated for every sample in the dataset, producing observed and predicted classes at every sample location, which are used to estimate predictive performance. Error matrices were computed to observe the

success in predicting the observed classes. From this we derived the percent correctly classified and the kappa coefficient, which reflects whether the model achieved better results than to be expected at random given the prevalence of each class. Because the performance of the coarse presence-absence model depends on the probability threshold, we used the threshold-independent area under the receiver operating characteristic curve (or, "area under the curve"; AUC) and maximum kappa values to compare candidate models.

Spatial autocorrelation is known to inflate estimates of predictive performance (Segurado *et al*., 2006; Bahn & McGill, 2013). To determine its effects on the modelling approaches tested here, and whether models were able to extrapolate to unsampled locations, we also conducted a spatial leave-one-out cross-validation (SLOO CV; Le Rest *et al*., 2014). This procedure is identical to LOO CV, except that a spatial buffer is placed around the withheld test point, and training data from within this buffer are omitted from both model training and testing so that there are no training data proximal to the test. This aims to eliminate spatial bias in accuracy assessment by removing points that are spatially autocorrelated with the test site up to the specified buffer distance.

We calculated empirical variograms for the observed grain size values and coarse substrate observations to determine a suitable buffer distance. We used the variogram model range to estimate the distance beyond which the effects of autocorrelation are negligible. This distance has been suggested as adequate for SLOO CV (Roberts *et al*., 2017). Variograms were calculated up to 5000 m and multiple models were tested for characterizing the major range using both the automated fitting within the "fit.variogram" function in the R package "*gstat*" and the Geostatistical Wizard in ArcGIS Pro v.2.2.3.

The coarse substrate model was trained using image frames from video transects, and these were expected to be highly autocorrelated due to their proximity. Therefore, in addition to the above two assessment procedures, we conducted a spatially resampled leave-one-out cross-validation (SR-LOO CV) to determine whether this spatial dependence affected model fitting in addition to performance estimation. Because Random Forest is an "embarrassingly parallel" algorithm (Liaw & Wiener, 2002), separate "forests" can be combined and treated as a single model to make predictions. The SR-LOO CV builds on SLOO CV by using the same spatial buffering procedure (i.e., the withheld test sample is spatially buffered; points within the buffer are excluded from training and testing), except that each training point for each "leave-one-out" iteration is also spatially buffered, so that no adjacent points are used for model fitting or testing (similar to the algorithm in Holland *et al*., 2004). Because this severely limits the number of samples available for training each iteration of Random Forest, we can randomly subsample each "leave-one-out" training dataset many times (100 here) to acquire different subsets of spatially independent training data. By producing only a small number of trees (*ntree*) from each of the 100 spatially independent data subsets, it is possible to achieve the same number of trees in a forest that are desirable in a full "leave-one-out" iteration (e.g., *ntree* = 2000 here). This effectively creates the same number of trees as in a normal Random Forest run but produces the "forest" using the combination of many small spatially independent subsets, rather than the full dataset.

The results of these cross-validations were analyzed to address the goals of this study. For grain size predictions, we first investigated the performance of the quantitative and categorical approaches to determine if these models could successfully extrapolate sediment grain sizes at unsampled locations, and whether spatial autocorrelation interfered with assessing map accuracy. Performance was assessed using multiple classification schemes at different levels of grain size

detail (i.e., aggregation) to select one that fit the data. We then compared maps produced using both approaches to determine if and how any differences in model performance manifested in the mapped predictions, and if the maps agreed. For predicting the presence of coarse substrates, we tested whether the proximity of sample points within video transects inflated estimates of predictive performance, and also whether their proximity caused overfitting of Random Forest models.

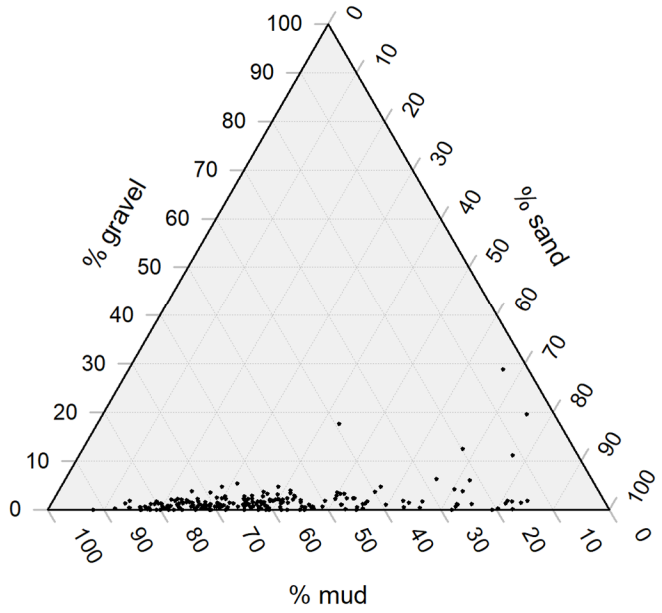
4.2.7 Map Prediction

Predictions of sediment grain size and the presence of coarse substrates were combined to produce a single map of seabed sediment distribution. The results from the accuracy assessments and map comparison, and also the qualities inherent in the two modelling approaches, were used to select a suitable model for predicting sediment grain size classes. To test for differences in candidate model performance, we used Monte Carlo permutation tests with 999 iterations at the 5 percent level of significance following McKenzie *et al*. (1996) and Foody (2004). The probability of coarse substrate presence was predicted for the entire study area. To combine these predictions with those of grain size, an occurrence threshold was set to maximize the sum of sensitivity and specificity, which aims to balance the class accuracy of predictions (Downie *et al*., 2016), and has been shown to perform well compared to other threshold selection criteria (Liu *et al*., 2005). Thus, the combined map predicts the sediment grain size class and whether coarse substrates are present for each pixel throughout the study area (e.g., “muddy with coarse substrate”).

4.3 Results

4.3.1 Grain Size Data

The sampled substrates were primarily muddy, with some sandy sediments (Figure 4.6). When classified according to Folk (1954), the most common class was (g)sM (40.72%), followed by sM (38.14%), and (g)mS (13.40%). The simplified Folk scheme eliminates the "slightly gravelly" modifier, aggregating the classes (g)sM and (g)mS with sM and mS, increasing these class proportions to 78.87% and 18.04%, respectively. In the "muddy/sandy" classification, 79.38% of samples fall into the "muddy" class, with the remaining 20.62% in "sandy". Coarse substrates were observed in 20.06% of raster cells containing underwater video observations (Table 4.1).



Folk

Class	Occurrences
M	1
sM	74
mS	9
S	0
(g)M	0
(g)sM	79
(g)mS	26
(g)S	0
gM	0
gmS	3
gS	2
mG	0
msG	0
sG	0
G	0

Simplified Folk

Class	Occurrences
M	1
sM	153
mS	35
S	0
gM	0
gmS	3
gS	2
mG	0
msG	0
sG	0
G	0

Muddy/sandy

Class	Occurrences
Muddy	154
Sandy	40

Figure 4.6. Grain size composition of sediment samples with class membership for the three schemes.

Table 4.1. Coarse substrate observations from underwater video frames.

Coarse substrates	Raster cells
Present	65
Absent	259

4.3.2 Spatial Autocorrelation

In general, the variogram model fitting in the ArcGIS Pro Geostatistical Wizard estimated a major range that was greater than in the "*gstat*" package but both variogram modelling tools were sensitive to input parameters (e.g., lag size, maximum distance, model type). The greater estimates of major range from the Geostatistical Wizard models were used to determine a buffer distance for the spatial leave-one-out cross-validations. Circular variogram models provided a distinct major range compared to the more asymptotic models (e.g., exponential, Gaussian, Bessel), were relatively stable with varying input parameters, and fit the data comparatively well. The major ranges of the circular mud and sand variogram models were 1497 m and 1210 m, respectively, when calculated at a maximum distance of 5000 m (see Appendix B.3). There appeared to be little change in gravel measurement variance with increasing distance, and they did not yield a useable variogram model. We selected a buffer distance of 1500 m for the SLOO CV based on the mud and sand major range estimates. The major range of the circular coarse substrate model was 192 m; we selected a buffer distance of 200 m for SLOO CV and SR-LOO CV methods based on this model.

4.3.3 Variable Selection

Different sets of scale dependent variables were selected for modelling ALR_{sm}, ALR_{gm}, grain size classes, and the presence of coarse substrates (Tables 4.2, 4.3). Backscatter range was commonly

correlated with the backscatter SD ($\rho \geq 0.70$), and only one of these two variables was generally selected, except in the classification model where the correlation between backscatter range at 250 m scale and backscatter SD at 100 m scale was below this threshold. The different curvature measures were often correlated at similar scales, and also to RDMV. Total curvature was correlated with one of these variables at every scale tested, and consistently had a weaker relationship with the response – it was therefore removed from all models. The two measures of complexity, SA:PA and VRM, were also correlated at similar scales.

Table 4.2. Variables selected for sediment grain size models.

ALR_{sm}		**ALR_{gm}**		**Classification**	
Variable	Scale (m)	Variable	Scale (m)	Variable	Scale (m)
Backscatter	-	Bathymetry	-	Bathymetry	-
Bathymetry	-	Backscatter	-	Backscatter	-
Distance from coast	-	Distance from coast	-	Distance from coast	-
Backscatter range	200	Backscatter SD	100	Backscatter range	250
Eastness	50	Eastness	100	Backscatter SD	100
Eastness	500	Eastness	400	Eastness	10
Northness	10	Northness	250	Eastness	450
Plan curvature	50	Plan curvature	50	Northness	10
Plan curvature	350	Plan curvature	300	Plan curvature	150
Profile curvature	450	Profile curvature	300	Plan curvature	300
RDMV	300	Profile curvature	450	Profile curvature	300
SA:PA	10	Slope	10	RDMV	200
Slope	10	Slope	450	RDMV	350
Slope	500	VRM	200	Slope	10
VRM	400	VRM	400	Slope	450
				VRM	10

Table 4.3. Variables selected for coarse substrate model.

Coarse substrates

Variable	Scale (m)
Bathymetry	-
Backscatter	-
Distance from coast	-
Backscatter SD	100
Eastness	100
Eastness	500
Northness	250
Plan curvature	100
Plan curvature	350
Profile curvature	10
RDMV	100
RDMV	300
Slope	200
VRM	100
VRM	350

4.3.4 Grain Size Model Evaluation and Comparison

The LOO CV suggested that all predictions were significantly more accurate using the categorical than the quantitative approach (Table 4.4). Of these, the simplified Folk and muddy/sandy classes were predicted accurately, but the Folk classes were not. The error matrix shows that this poorer performance was the result of several uncommon Folk classes that were not successfully predicted (Appendix B.5). This resulted in a percent correctly classified of only 54.64%, and predictions that were only marginally better than random (kappa = 0.25). The higher accuracy with the simplified Folk and muddy/sandy schemes is a product of rare classes being aggregated into broader ones, reducing their misclassification. Though the quantitative approach produced less accurate predictions than the categorical using LOO CV using all schemes, it was similar in that the

simplified schemes were predicted more successfully than the Folk. Again, this was a result of aggregating rare Folk classes into broader ones, making them easier to predict.

In contrast to the LOO CV results, there was generally no difference in performance between the two approaches when evaluated using SLOO CV. Predictions were generally less accurate using the SLOO CV (except for the quantitative Folk predictions; Table 4.4). The percent correctly classified was reduced only marginally – by 7.92% to not at all – yet kappa values all indicate that the performance of these models is hardly better than by random chance based on class prevalence. The disparity between percent correctly classified and kappa scores in the SLOO CV assessments are the result of an increased inability to predict the rarer classes (see error matrices in Appendix B.5).

Table 4.4. Performance of quantitative and categorical predictions using three schemes with spatial and non-spatial cross-validations.

		LOO CV		**SLOO CV (1500 m)**	
		Categorical	Quantitative	Categorical	Quantitative
Folk	% correctly classified	54.64*†	48.46*	46.72†	48.58
	Kappa	0.25*†	0.14*	0.12†	0.10
Simplified Folk	% correctly classified	85.50*†	82.25*‡	78.79†	78.11‡
	Kappa	0.52*†	0.34*‡	0.05†	0.04‡
muddy/ sandy	% correctly classified	86.29*†	84.62*‡	78.36†	79.88‡
	Kappa	0.53*†	0.41*‡	0.06**†	0.11**‡

*Significant difference between categorical and quantitative approaches using LOO CV.
**Significant difference between categorical and quantitative approaches using SLOO CV.
†Significant difference between LOO CV and SLOO CV using the categorical approach.
‡Significant difference between LOO CV and SLOO CV using the quantitative approach.

Despite similarities in predictive performance when evaluated using SLOO CV, there were obvious differences in map predictions between categorical and quantitative modelling approaches. The quantitative approach predicted the occurrence of classes that were not observed

in ground-truth samples, such as the Folk class gM in the deep channels in the southeast part of the bay (Figure 4.7B), which was the third most commonly predicted class, and was only predicted in unsampled areas. Conversely, the categorical approach generally predicted the most common class (sM) in these areas. (g)sM was the most commonly predicted Folk class in the classified quantitative map, occurring in 65.36% of raster cells, while it was only the second most common for the categorical map, occurring in 44.04% of cells. sM was the most common for the categorical map, occurring in 52.83% of cells, while it was the second most common in the classified quantitative map, occurring in only 27.79% of cells. The prediction of unsampled classes using the quantitative approach accounted for much of the disagreement between maps, but they also disagreed on the extent of the most common classes throughout the study area.

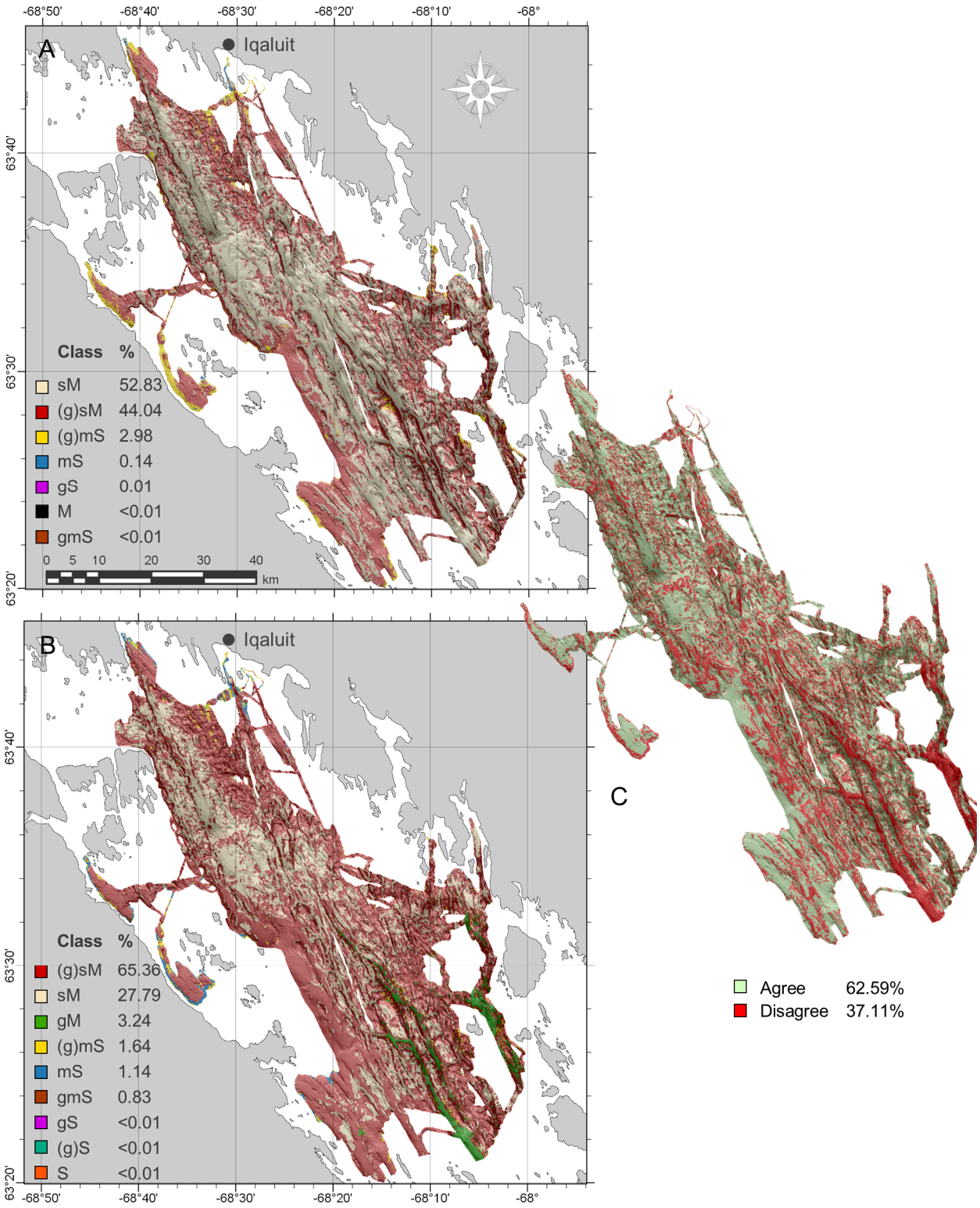


Figure 4.7. Predicted Folk grain size classes for (A) categorical, and (B) quantitative models, with (C) agreement between predictions.

The primary differences between categorical and quantitative approaches for the simplified Folk classified maps (Figure 4.8) was the prediction of unsampled classes. Again, the quantitative approach predicted extensive areas of gM in the deep southeast channel that were not predicted using the categorical approach. In all other areas the two map predictions were similar, with 93.15% and 96.37% of quantitative and categorical maps classified as sM, respectively. Both approaches predicted similar distributions of mS, primarily near Iqaluit, in the northernmost mapped area, and near the southwestern coast. “Muddy/sandy” maps (Figure 4.9) were highly similar between the two approaches with nearly 97% agreement, having eliminated all unsampled classes. “Sandy” sediment was primarily predicted where mS occurred in the simplified Folk maps. Predicted class proportions between the approaches were similar, with “sandy” sediment predicted in 3.61% of cells in the classified quantitative map and 4.60% in the categorical map – the remaining being “muddy”.

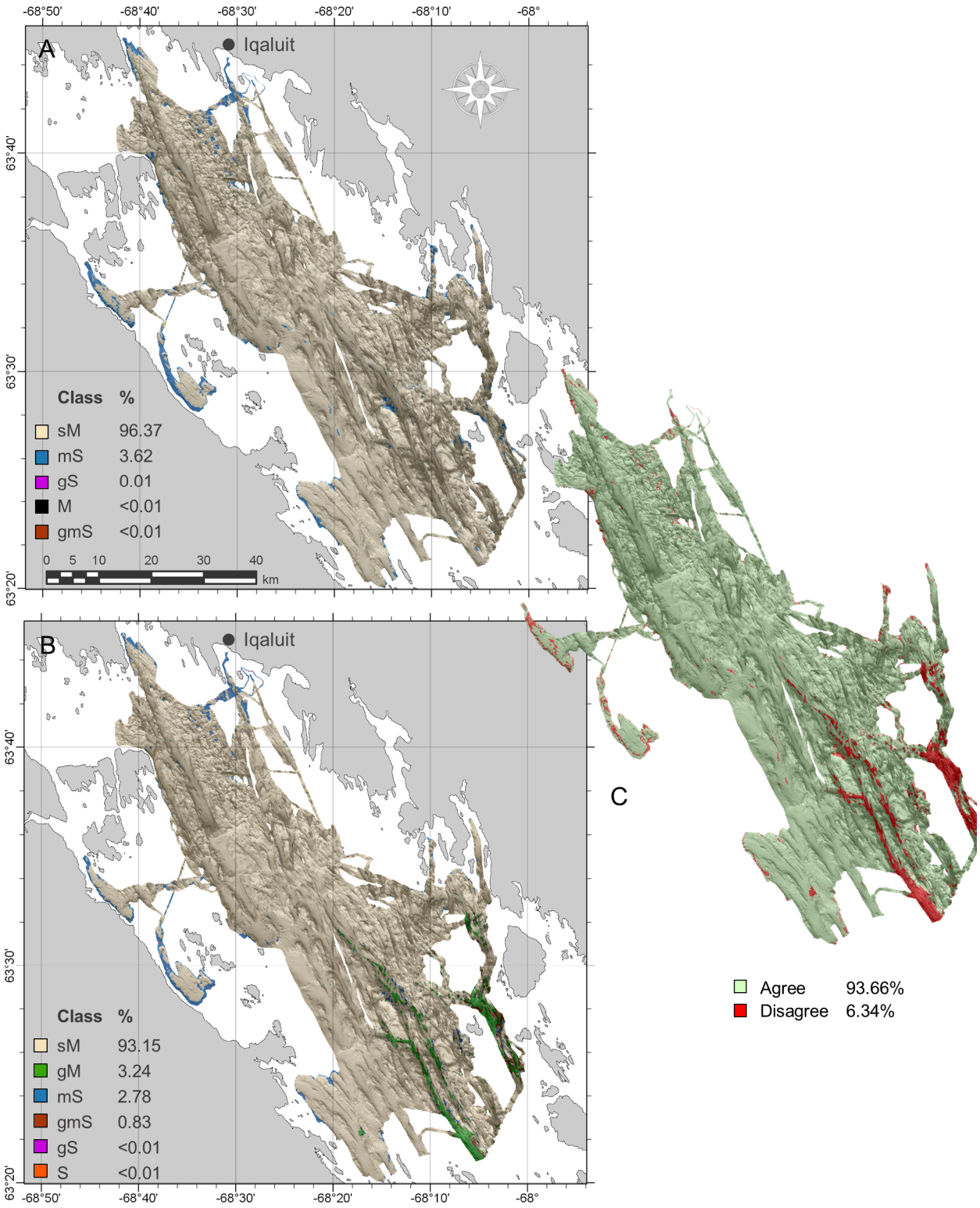


Figure 4.8. Predicted simplified Folk grain size classes for (A) categorical, and (B) quantitative models, with (C) agreement between predictions.

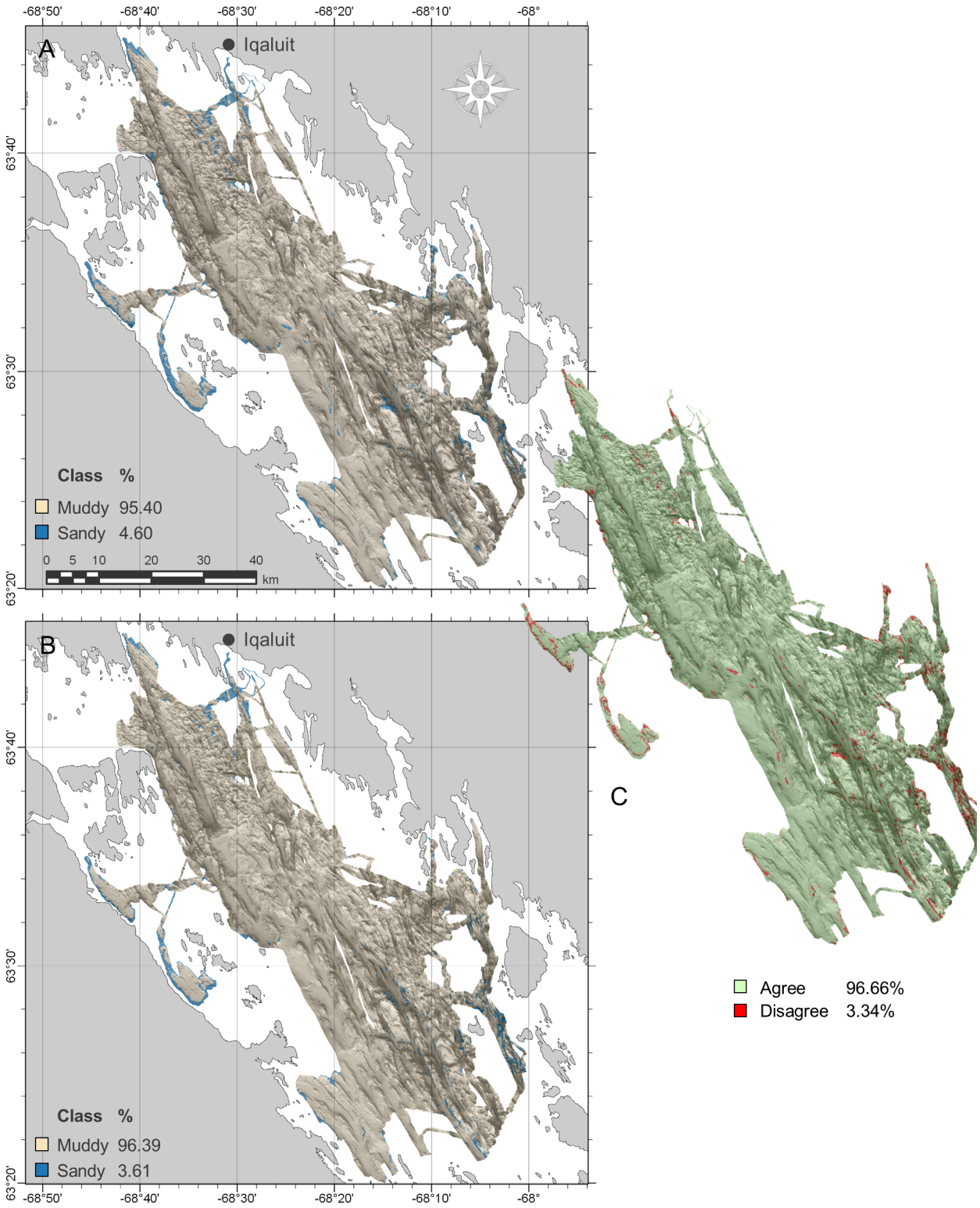


Figure 4.9. Predicted "muddy/sandy" grain size classes for (A) categorical, and (B) quantitative models, with (C) agreement between predictions.

4.3.5 Coarse Model Assessment

The LOO CV suggested that the presence of coarse substrates was predicted accurately, but the SLOO CV accuracy was significantly lower (Table 4.5). The maximum kappa value obtained using LOO CV (0.62; Table 4.5) suggested that the model had potential to predict much better than expected by chance (depending on the threshold selected). The threshold-independent AUC value (0.86) also suggested strong predictive performance. Conversely, SLOO CV yielded significantly lower maximum kappa and AUC values. Accuracy of the SR-LOO CV, however, was significantly higher than the SLOO CV with the same spatial constraints (i.e., 200 m buffer), suggesting potential model overfitting caused by the proximity of the training data that may have been alleviated when training data were forced to be independent.

Table 4.5. Threshold-independent accuracy of coarse substrate model using spatial and non-spatial CV approaches, and with spatially independent training data.

	LOO CV	SLOO CV (200 m)	SR-LOO CV (200 m)
AUC	0.86*	0.67*†	0.76†
Kappa	0.62*	0.24*†	0.40†

*Significant difference between LOO and SLOO CV.
†Significant difference between SLOO and SR-LOO CV.

The map of coarse substrates shows the probability of occurrence for each pixel (Figure 4.10). Coarse substrates were predicted to occur throughout the bay, but primarily on the flanks of topographic highs and on several coasts (see the northern and westernmost mapped areas; Figure 4.10). Coarse substrates were also predicted in the southeast section of the bay on the flanks of deep channels, and near the islands to the east, where backscatter return was high (Figure 4.5).

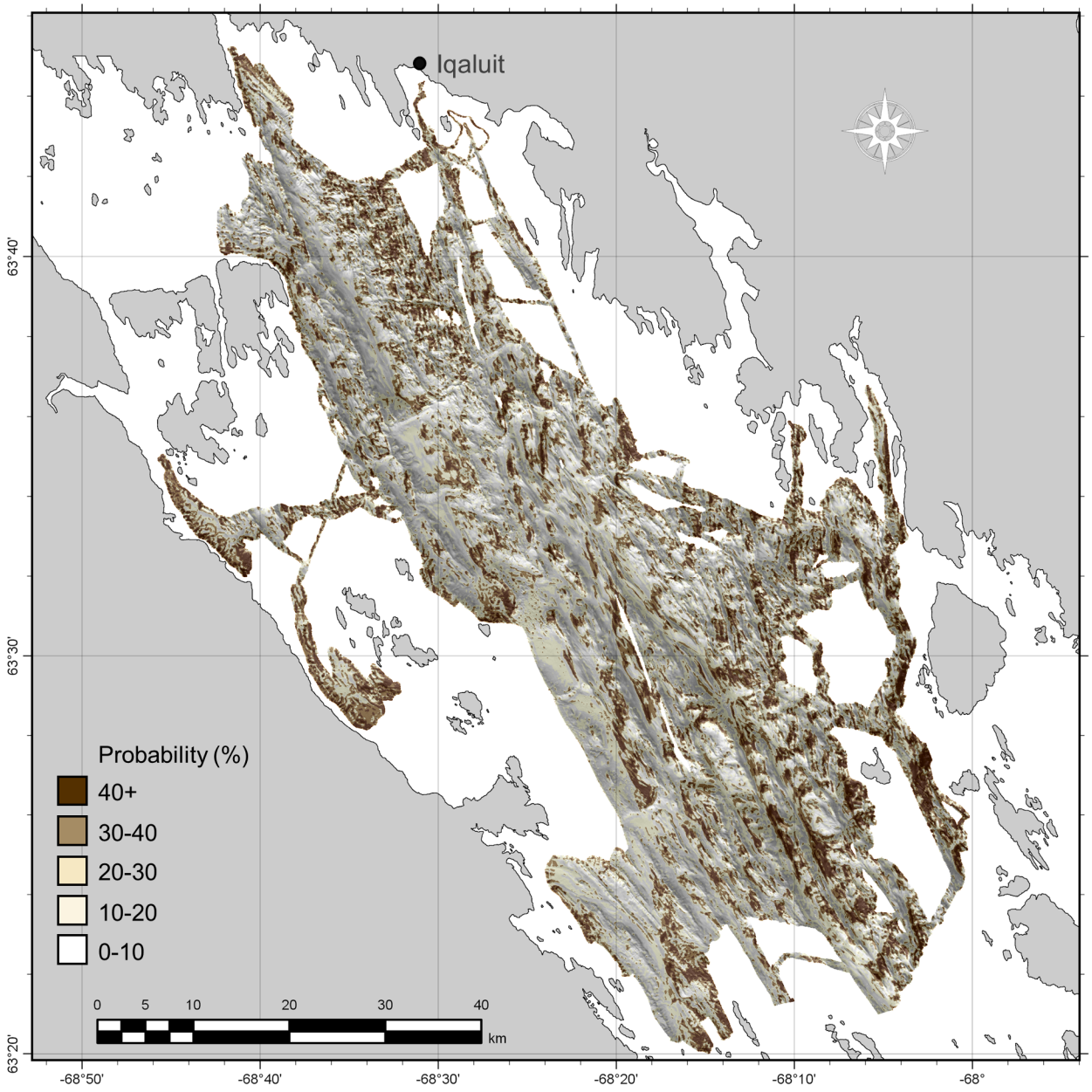


Figure 4.10. Predicted probability of coarse substrates using the SR-LOO CV (200 m) model.

4.3.6 Combined Map and Model Tuning

The two-class categorical muddy/sandy predictions were selected to combine with coarse substrate predictions to produce a single map of surficial sediments (Figure 4.11). Because the grain size scheme is binary, the categorical approach has the distinct advantage of a flexible threshold of occurrence, which can be readily optimized. Setting the threshold to maximize the sum of

sensitivity and specificity (0.18 here; Liu *et al*., 2005) for predicting "sandy" sediments using SLOO CV yielded kappa = 0.27, which is significantly higher than the classified quantitative predictions (0.11; p = 0.009). Having selected this model, the performance was tested after dropping further unimportant predictors identified from estimates of variable importance. Maintaining only the top six variables (bathymetry, backscatter, 300 m profile curvature, 10 m and 450 m slopes, 10 m VRM) resulted in more accurate and more stable predictions using SLOO CV (Table 4.6). At the 0.18 threshold, there was no significant difference in the percent correctly classified (p = 0.190) or kappa (p = 0.387) between LOO CV and SLOO CV.

The presence or absence of coarse substrates was dichotomized using a 0.27 threshold of occurrence to maximize the sum of sensitivity and specificity (Liu *et al*., 2005). These predictions were 75.34% accurate with kappa = 0.40 using the SR-LOO CV method (Table 4.6). The final seabed sediment class was determined by specifying the predicted grain size class ("muddy" or "sandy") and whether coarse substrates are present (Figure 4.11).

Table 4.6. Accuracies of grain size and coarse substrate components of combined seabed sediment predictions.

	Muddy/sandy (variables reduced)	Coarse presence-absence
% correctly classified	70.37	75.64
Kappa	0.34	0.40

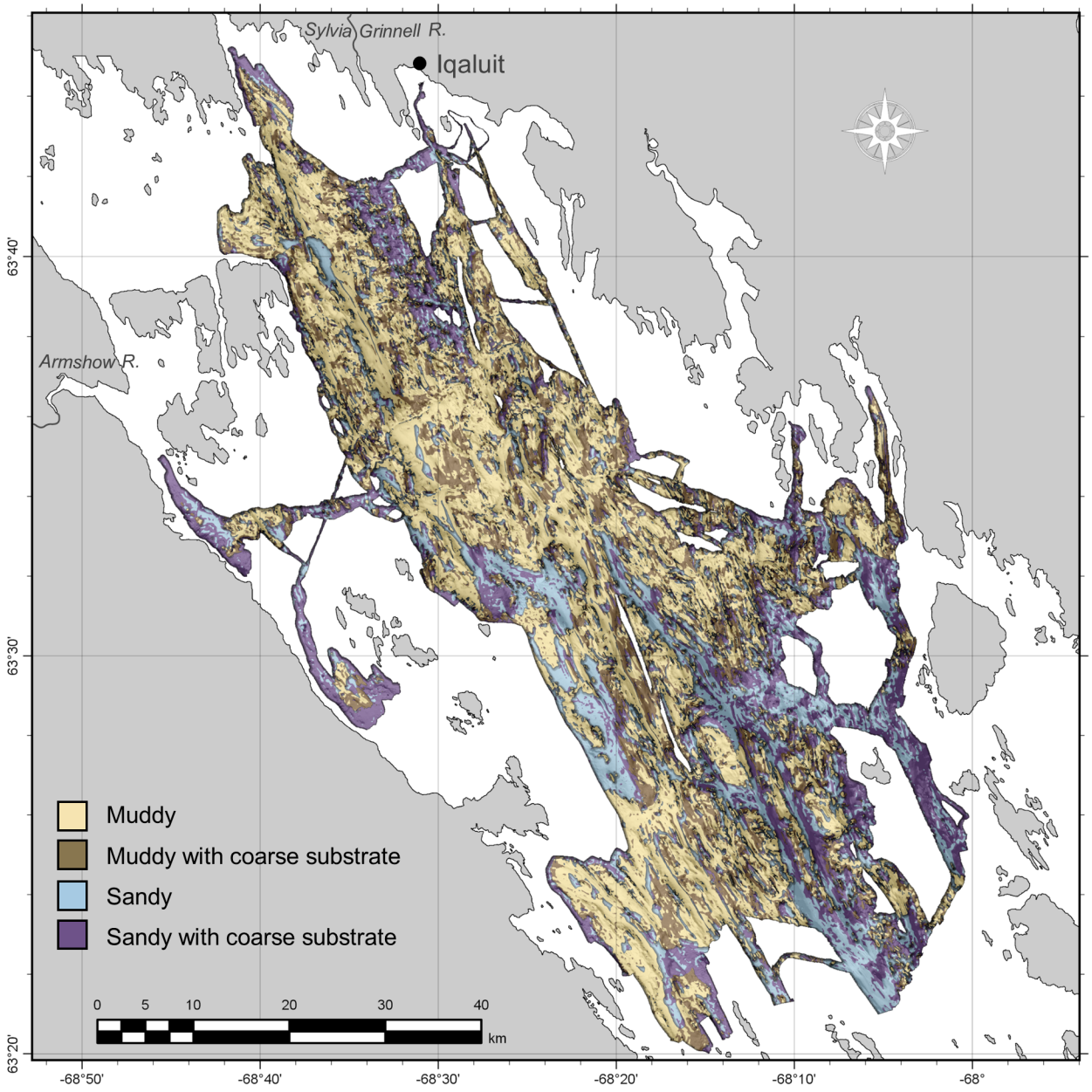


Figure 4.11. Combined map of grain size classification and coarse substrate predictions.

4.4 Discussion

The predicted seabed sediment classes generally agreed with expectation given the geomorphology of the bay, yet particular locations without ground-truth data require further investigation. The majority of the low-relief seabed was classified as "muddy", which is not surprising given what was observed in grab samples and underwater video (e.g., Figure 4.12A). Sandy sediments (e.g.,

Figures 4.12C, D) predicted south and southwest of Iqaluit may be partially attributable to sediment input from the Sylvia Grinnell River, directly west of the city. This is also an area of distinct sea-ice scouring (Deering *et al*., 2018), with higher acoustic backscatter than the surrounding seabed (Figure 4.5), and several distinctly reflective features that were classified as "sandy with coarse substrate". This class was also predicted at several locations along the coast, fining to muddier grain sizes with increasing distance and depth. Otherwise, exposed coarse substrates predicted along the flanks of steep topographic features may be attributable to current winnowing of unstable fine sediments (Syvitski, 1989). This is likely the case in the high-relief, deep southeastern channels, where coarse substrates were predicted extensively. Further investigation is necessary in these deep channels though – this was an unsampled area of high disagreement between the categorical and quantitative models (Figures 4.7, 4.8). One might expect a muddier composition at the bottoms of these deep channels, yet sandier grain sizes were predicted, likely as a product of the high backscatter response (Figure 4.5).

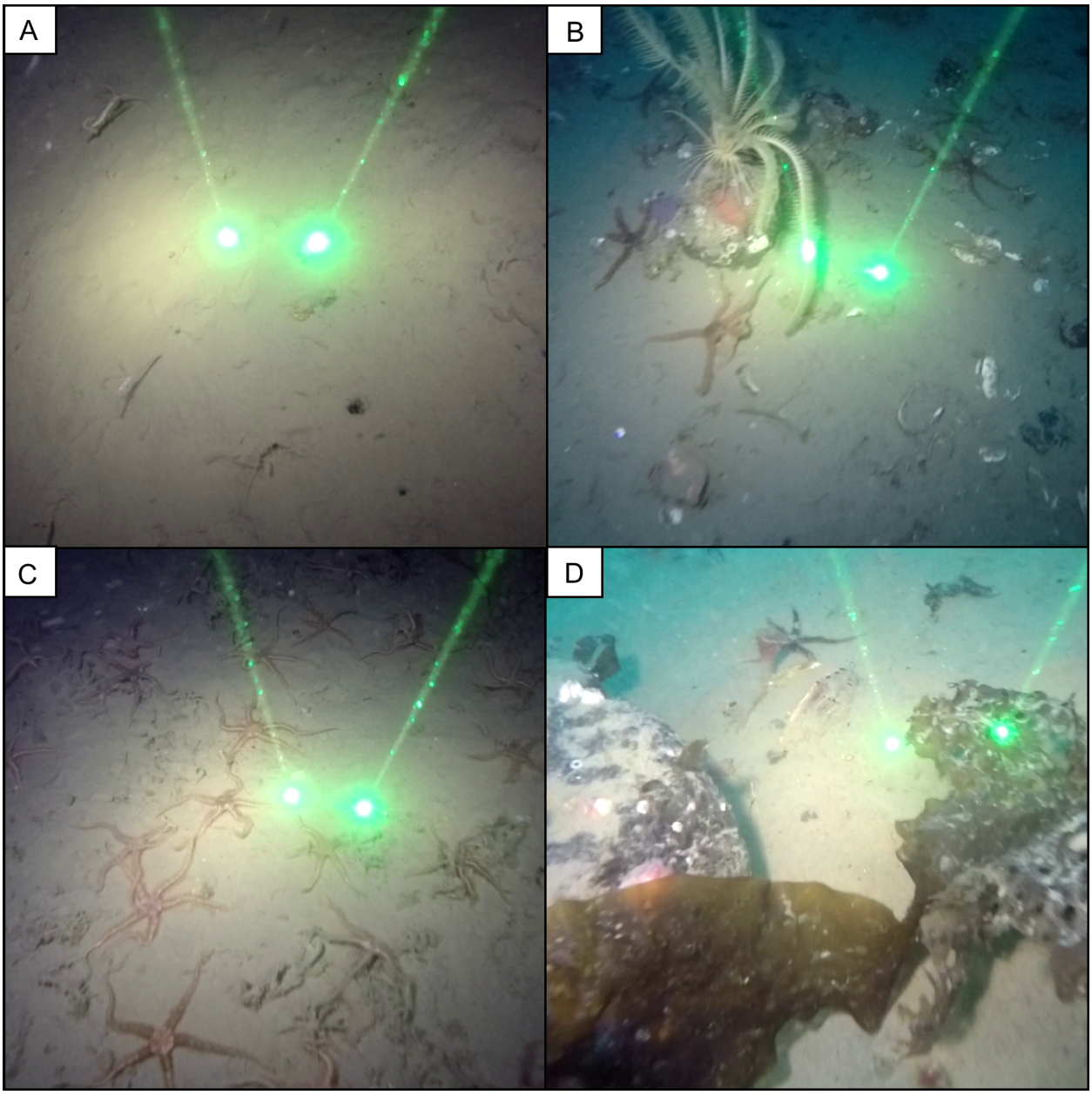


Figure 4.12. Examples of (A) muddy, (B) muddy with coarse substrate, (C) sandy, and (D) sandy with coarse substrate classes observed in Frobisher Bay video samples.

4.4.1 Model Comparison

There was little difference in accuracy between quantitative and categorical approaches when using spatially explicit cross-validation methods, but their maps differed substantially. Using a two-class scheme, it was possible to tune the threshold of occurrence for the probabilistic output

of the categorical model to obtain a higher accuracy than the quantitative model, and this was selected for the final map. The most noticeable difference between maps was that the quantitative approach predicted extensive patches of sediment classes that were not observed in the ground-truth data, where the categorical approach predicted the most commonly observed classes. The predicted proportions and distributions of the classes also differed between approaches.

Although the quantitative approach failed at extrapolation in this study, it has several characteristics that may be desirable in other situations. Because classification of quantitative predictions is done *post hoc*, this method might avoid some of the difficulty associated with predicting unbalanced classes – one of the major shortcomings of the categorical approach. Furthermore, as demonstrated here, predictions are not constrained to the classes that were sampled. Thus, if the model were fit well, it may be possible to predict rare and unsampled classes at new locations, while this is not feasible with the categorical approach. This may be a particularly useful quality if unsampled areas are expected to contain different sediment characteristics than the sample sites, yet it requires a high degree of confidence in the modelled relationships between grain size composition and the explanatory variables – ideally, a second ground truthing effort should be undertaken to determine the usefulness of such predictions. The spatial leave-one-out CV error matrices for classified quantitative predictions (Appendix B.5) failed to indicate that the model could successfully predict rare classes in this study, and we did not have the confidence to adopt predictions of unobserved classes in unsampled areas. It is quite possible that these areas do actually contain different sedimentary characteristics though, as their morphology and backscatter characteristics were unique, but there is no way to confirm this. Sampling these areas would be a priority in future work.

It is also worth considering characteristics of the unclassified quantitative predictions of mud, sand, and gravel, which may offer some advantages over classification. These predictions represent gradational changes in sediment composition, which are more realistic, and potentially more desirable, than discrete classes. If classes are required, quantitative predictions are completely flexible with regards to classification scheme. Because the quantitative values remain the same, it is not necessary to run through the model fitting procedure to test different classifications, the class boundaries simply need to be adjusted. Other methods could also be used to optimize the classification of quantitative predictions to produce relevant and distinct classes, such as multivariate clustering. This way, it is possible to define an appropriate number of classes with boundaries that are most relevant to a given study area.

The qualities of the categorical approach ultimately made it more suitable for extrapolation in this study. The main difficulty was that rare classes were seldom predicted correctly, and it was therefore important to select a classification scheme that fit the data, thereby reducing the number of classes with few observations. For instance, four of the Folk and three of the simplified Folk classes contained < 10 observations, each of which had 0% correctly classified using SLOO CV. Furthermore, because the samples are spatially autocorrelated, it is likely that the samples of rare classes occur close to one another. In this case, using any sort of spatial CV approach in which proximal samples are omitted (e.g., SLOO, SR-LOO, blocking, etc.) is likely to omit even more, if not all, samples of a rare class. This further reduces the sampled prevalence and hinders the predictive accuracy of the model.

Designating grab samples as “muddy” or “sandy” created meaningful classes given these data (Figure 4.6), yet the class prevalence was still unbalanced. The second solution afforded by the categorical approach is that the threshold of occurrence for the rarer class (sandy) can be optimized.

This produced significantly higher extrapolative (i.e., SLOO CV) accuracy than the quantitative approach and was similar to the predictive accuracy of the LOO CV after removing superfluous variables.

4.4.2 Spatial Assessment

Spatial autocorrelation inflated estimates of predictive accuracy regardless of the modelling approach or classification scheme for both grain size and coarse substrate models, hindering the ability to determine whether the models could successfully extrapolate grain size classes at unsampled locations. Similar to LOO CV, many common model validation techniques (e.g., sample partitioning, *k*-fold CV) have no spatial component. For this study, non-spatial techniques failed to correctly estimate the model's ability to extrapolate. If LOO CV were used in isolation to evaluate the categorical simplified Folk predictions, for example, the percent correctly classified and kappa values (85.50%; 0.52) would suggest the model is highly accurate and reliable. In reality though, it fails to extrapolate beyond the sphere of spatial autocorrelation influence, with predictions no better than random.

The SR-LOO CV for the coarse substrate model suggested not only that spatial autocorrelation inflated estimates of accuracy, but also that Random Forest was spatially overfitting, hindering extrapolation. This is an important issue for severely autocorrelated sample distributions that is not necessarily solved by other spatial validation approaches such as SLOO CV or spatial blocking, which allow adjacent samples for model fitting. Though our method (SR-LOO CV) shows promise for reducing overfitting and providing non-biased estimates of accuracy, we note that the computational effort is not realistic for many studies. One hundred random samples of each spatially buffered leave-one-out sample set (n = 324) yielded 32,400 sub-samples and

corresponding Random Forest models for one SR-LOO CV run. Furthermore, it required the "embarrassingly parallel" qualities of Random Forest, complicating its implementation with most other modelling methods. Simpler alternatives could involve aggregating sample transects to a single point and adjusting the raster resolution, but this may be less attractive if a high resolution is desired, and how such methods compare with the SR-LOO CV remains to be explored.

4.4.3 Spatial Prediction

It is important to distinguish between interpolating a well-sampled area and extrapolating to unsampled locations (Roberts *et al*., 2017). Though it is becoming standard practice to report predictive accuracy, it is less common to differentiate between these predictive spatial qualities, which are largely determined by sampling effort and distribution. Again, if interpolation is the goal, with somewhat uniform and well-distributed sampling, then standard non-spatial model evaluation methods may be appropriate (e.g., LOO CV, *k*-fold CV, partitioning). If samples are clustered, with parts of the study area unsampled, then it is necessary to evaluate for extrapolation, which may require a spatially explicit approach, as was the case here. This also may affect the appropriateness of categorical and quantitative approaches – if extrapolating to a potentially new sedimentary environment is the goal, and if there is confidence in the modelled relationships between sediment and explanatory variables, then the quantitative approach may be useful for identifying unsampled or rare sediment classes. Here, we found that the flexibility of the threshold of occurrence using the categorical approach resulted in superior extrapolative performance compared to the quantitative approach for a binary classification scheme. Otherwise, there was little statistical difference in performance between categorical and classified quantitative approaches for the Folk and simplified Folk schemes (Table 4.4). Given a set of classification

requirements (e.g., regional compatibility), and a desire to maximize predictive accuracy of predetermined classes, it may be necessary to test both approaches – our results do not suggest the consistent superiority of one method over the other.

Recently there have been calls for greater transparency in reporting map quality, including uncertainty and error, to determine whether thematic maps are fit for purpose (e.g., Barry & Elith, 2006; Mitchell *et al*., 2018). This becomes especially important when providing maps as tools for management, where end-users may lack the technical understanding to critically evaluate a map (Lecours, 2017). The spatial component of distribution modelling is a potential source of data error (Barry & Elith, 2006) that is commonly neglected (Bell & Schlaepfer, 2016), yet which can be exacerbated due to marine sampling constraints. Here we have demonstrated the necessity of spatially explicit analysis for comparing the error and predictions between two seabed sediment mapping approaches, and the potential pitfalls of neglecting to do so. These qualities of the categorical and quantitative approaches have not yet been thoroughly compared in the seabed mapping literature to our knowledge. The SR-LOO CV approach used here to model the presence of coarse substrates is similar to the variable scale selection procedure in Holland *et al*. (2004) but uses "embarrassingly parallel" Random Forests so that no samples are fully omitted. This is the first application of the approach in this context to our knowledge. Though the SR-LOO CV method was well suited to modelling video transect data in this study, we acknowledge several other useful strategies (e.g., spatial blocking; Roberts *et al*., 2017) and tools (e.g., the "*blockCV*" R package; Valavi *et al*., 2018) that are worth considering to address spatial sampling bias. We have applied these methods to modelling seabed sediments, but they are potentially applicable to other similar benthic distribution models including those of species and biotopes.

4.5 Conclusions

Neither categorical nor quantitative methods performed consistently better between the classification schemes tested, but the differences in the mapped predictions and the qualities of the models ultimately determined their suitability. The ability of the quantitative approach to predict rare and unsampled classes may be an important quality depending on sample distribution and mapping goals, yet we found the probabilistic threshold qualities of the categorical approach with a binary scheme (i.e., "muddy/sandy") were more accommodating for extrapolating predictions in areas far from sampled sites. Extrapolation was a necessary quality for these models because of sample site clustering and expansive unsampled areas.

There was evidence that the proximity of transect video observations caused both inflated estimates of accuracy and overfitting in the Random Forest models of coarse substrates. These results strongly suggest that the spatial dependence of transect data should be examined if the model will be extrapolating. It should not be taken for granted that Random Forest will not overfit amidst severely autocorrelated data, and one of these or other methods may be necessary to ensure spatial independence, regardless of the modelling algorithm used.

From the results of this study we recommend that seabed map producers be specific about their predictive goals – especially whether the models are required to extrapolate to new environments and locations, or whether they will "fill in the gaps" between sample sites (i.e., interpolate). This distinction can determine the appropriateness of model evaluation methods and potentially the suitability of categorical or quantitative approaches. We found it necessary to use spatially explicit strategies to evaluate whether the models in this study were able to extrapolate, and that modelling

highly autocorrelated data required both model fitting and testing in a spatially independent context.

4.6 References

Aitchison, J. 1982. The statistical analysis of compositional data. *Journal of the Royal Statistical Society*, 44(2): 139–177.

Araújo, M. B., and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33(10): 1677–1688.

Bahn, V., and McGill, B. J. 2013. Testing the predictive performance of distribution models. *Oikos*, 122(3): 321–331.

Baker, E. K., and Harris, P. T. 2012. Habitat mapping and marine management. *In* Seafloor Geomorphology as Benthic Habitat: Geohab Atlas of Seafloor Geomorphic Features and Benthic Habitats, pp. 23–38. Ed. by P. T. Harris and E. K. Baker. Elsevier, Amsterdam.

Barry, S., and Elith, J. 2006. Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43(3): 413–423.

Beaman, R. J., Daniell, J. J., and Harris, P. T. 2005. Geology-benthos relationships on a temperate rocky bank, eastern Bass Strait, Australia. *Marine and Freshwater Research*, 56(7): 943–958.

Bell, D. M., and Schlaepfer, D. R. 2016. On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling*, 330: 50–59.

Breiman, L. 1996. Bagging predictors. *Machine Learning*, 24(2): 123–140.

Breiman, L. 2001. Random forests. *Machine Learning*, 45(1): 5–32.

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(3): 502–520.

Che Hasan, R., Ierodiaconou, D., and Monk, J. 2012. Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. *Remote Sensing*, 4(11): 3427–3443.

Connor, D. W., Gilliland, P. M., Golding, N., Robinson, P., Todd, D., and Verling, E. 2006. UKSeaMap: The mapping of seabed and water column features of UK seas. Joint Nature Conservation Committee, Peterborough. 104 pp.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., *et al.* 1997. The value of the world's ecosystem services and natural capital. *Nature*, 387(6630): 253–260.

Davies, C. E., Moss, D., and Hill, M. O. 2004. EUNIS habitat classification revised. European Environment Agency.

Deering, R., Misiuk, B., Bell, T., Forbes, D. L., Edinger, E., Tremblay, T., Telka, A., *et al.* 2018. Characterization of the seabed and postglacial sediments of inner Frobisher Bay, Baffin Island, Nunavut. Summary of Activities 2018. Canada-Nunavut Geoscience Office.

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D. 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, 84: 107–119.

Diesing, M. 2015. Quantitative spatial prediction of seabed sediment composition. EMODnet-Geology Phase II, C5818. Centre for Environment Fisheries and Aquaculture.

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

Diesing, M., Kröger, S., Parker, R., Jenkins, C., Mason, C., and Weston, K. 2017. Predicting the standing stock of organic carbon in surface sediments of the North–West European continental shelf. *Biogeochemistry*, 135(1–2): 183–200.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., *et al.* 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1): 27–46.

Downie, A.-L., Dove, D., Westhead, K., Diesing, M., Green, S. L., and Cooper, R. 2016. Semi-automated mapping of rock in the North Sea. JNCC Report, 592. JNCC, Peterborough.

Ehler, C., and Douvere, F. 2009. Marine Spatial Planning: A step-by-step approach toward ecosystem-based management. Intergovernmental Oceanographic Commission and Man and the Biosphere Programme, IOC Manual and Guides No. 5, ICAM Dossier No. 6. UNESCO, Paris.

Eleftheriou, A., and Moore, D. C. 2013. Macrofauna techniques. *In* Methods for the Study of Marine Benthos, Fourth edition, pp. 175–251. Ed. by A. Eleftheriou. Wiley-Blackwell, Chichester.

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(2): 129–151.

ESRI. World light gray canvas base [basemap]. 2011. Available from: Available from: https://www.arcgis.com/home/item.html?id=8b3d38c0819547faa83f7b7aca80bd76

Folk, R. L. 1954. The distinction between grain size and mineral composition in sedimentary-rock nomenclature. *The Journal of Geology*, 62(4): 344–359.

Foody, G. M. 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing*, 70(5): 627–633.

Galparsoro, I., Borja, A., and Uyarra, M. C. 2014. Mapping ecosystem services provided by benthic habitats in the European North Atlantic Ocean. *Frontiers in Marine Science*, 1.

Galparsoro, I., Agrafojo, X., Roche, M., and Degrendele, K. 2015. Comparison of supervised and unsupervised automatic classification methods for sediment types mapping using multibeam echosounder and grab sampling. *Italian Journal of Geosciences*, 134(1): 41–49.

Ghermandi, A., Nunes, P. A. L. D., Portela, R., Nalini, R., and Teelucksingh, S. S. 2010. Recreational, cultural and aesthetic services from estuarine and coastal ecosystems. *SSRN Electronic Journal*.

Gottschalk, T. K., Aue, B., Hotes, S., and Ekschmitt, K. 2011. Influence of grain size on species–habitat models. *Ecological Modelling*, 222(18): 3403–3412.

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., *et al.* 2008. A global map of human impact on marine ecosystems. *Science*, 319(5865): 948–952.

Hammond, T. O., and Verbyla, D. L. 1996. Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, 17(6): 1261–1266.

Hernandez, P. A., Graham, C. H., Master, L. L., and Albert, D. L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29(5): 773–785.

Hirzel, A., and Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, 157(2–3): 331–341.

Hodgson, D. A. 2005. Quaternary geology of western Meta Incognita Peninsula and Iqaluit area, Baffin Island, Nunavut. Bulletin, 582. Geological Survey of Canada.

Holland, J. D., Bert, D. G., and Fahrig, L. 2004. Determining the spatial scale of species' response to habitat. *BioScience*, 54(3): 227–233.

Ierodiaconou, D., Monk, J., Rattray, A., Laurenson, L., and Versace, V. L. 2011. Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. *Continental Shelf Research*, 31(2): S28–S38.

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(7): 811–820.

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

Lecours, V. 2017. On the use of maps and models in conservation and resource management (warning: results may vary). *Frontiers in Marine Science*, 4.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology*, 74(6): 1659–1673.

Li, J., Potter, A., Huang, Z., Daniell, J. J., and Heap, A. D. 2010. Predicting seabed mud content across the Australian margin: Comparison of statistical and mathematical techniques using a simulation experiment. Geoscience Australia, Canberra.

Li, J., Tran, M., and Siwabessy, J. 2016. Selecting optimal random Forest predictive models: A case study on predicting the spatial distribution of seabed hardness. *PLOS ONE*, 11(2): e0149089.

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.

Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. *R News*, 2: 18–22.

Liu, C., Berry, P. M., Dawson, T. P., and Pearson, R. G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3): 385–393.

Long, D. 2006. BGS detailed explanation of seabed sediment modified folk classification. MESH Report. British Geological Survey.

McArthur, M. A., Brooke, B. P., Przeslawski, R., Ryan, D. A., Lucieer, V. L., Nichol, S., McCallum, A. W., *et al.* 2010. On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuarine, Coastal and Shelf Science*, 88: 21–32.

McKenzie, D. P., Mackinnon, A. J., Péladeau, N., Onghena, P., Bruce, P. C., Clarke, D. M., Harrigan, S., *et al.* 1996. Comparing correlated kappas by resampling: Is one level of agreement significantly different from another? *Journal of Psychiatric Research*, 30(6): 483–492.

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(7): 8489–8515.

Misiuk, B., Lecours, V., and Bell, T. 2018. A multiscale approach to mapping seabed sediments. *PLoS ONE*, 13(2): e0193647.

Mitchell, P. J., Downie, A.-L., and Diesing, M. 2018. How good is my map? A tool for semi-automated thematic mapping and spatially explicit confidence assessment. *Environmental Modelling & Software*, 108: 111–122.

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(2): 171–193.

Pawlowsky-Glahn, V., and Olea, R. A. 2004. Geostatistical Analysis of Compositional Data. Studies in Mathematical Geology. Oxford University Press, New York. 304 pp.

Rees, H. L. (Ed). 2009. Guidelines for the study of the epibenthos of subtidal environments. ICES Techniques in Marine Environmental Sciences. International Council for the Exploration of the Sea, Copenhagen. 88 pp.

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(8): 913–929.

Segurado, P., Araújo, M. B., and Kunin, W. E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43(3): 433–444.

Siwabessy, P. J. W., Tran, M., Picard, K., Brooke, B. P., Huang, Z., Smit, N., Williams, D. K., *et al.* 2018. Modelling the distribution of hard seabed using calibrated multibeam acoustic backscatter data in a tropical, macrotidal embayment: Darwin Harbour, Australia. *Marine Geophysical Research*, 39(1–2): 249–269.

Stephens, D., and Diesing, M. 2014. A comparison of supervised classification methods for the prediction of substrate type using acoustic and legacy grain-size data. *PLOS ONE*, 9(4): e93950.

Stephens, D., and Diesing, M. 2015. Towards quantitative spatial models of seabed sediment composition. *PLOS ONE*, 10(11): e0142502.

Strong, J. A., Clements, A., Lillis, H., Galparsoro, I., Bildstein, T., and Pesch, R. 2018. A review of the influence of marine habitat classification schemes on mapping studies: Inherent assumptions, influence on end products, and suggestions for future developments. *ICES Journal of Marine Science*, 76(1): 10–22.

Syvitski, J. P. M. 1989. On the deposition of sediment within glacier-influenced fjords: Oceanographic controls. *Marine Geology*, 85(2–4): 301–329.

Todd, B. J., Fader, G. B. J., Courtney, R. C., and Pickrill, R. A. 1999. Quaternary geology and surficial sediment processes, Browns Bank, Scotian Shelf, based on multibeam bathymetry. *Marine Geology*, 162(1): 165–214.

Tremblay, T., Day, S., McNeil, R., Smith, K., Richardson, M., and Shirley, J. 2015. Overview of surficial geology mapping and geochemistry in the Sylvia Grinnell Lake area, Baffin Island, Nunavut. Summary of Activities 2015. Canada-Nunavut Geoscience Office.

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(2): 225–232.

Weatherall, P., Marks, K. M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., Rovere, M., *et al.* 2015. A new digital bathymetric model of the world's oceans. *Earth and Space Science*, 2(8): 331–345.

Whitmire, C. E., Embley, R. W., Wakefield, W. W., Merle, S. G., and Tissot, B. N. 2007. A quantitative approach for using multibeam sonar data to map benthic habitats. *In* Mapping the Seafloor for Habitat Characterization: Geological Association of Canada, Special Paper 47, pp. 111–126. Geological Association of Canada, St. John's.

5. Summary

Research and reviews in the terrestrial (Dormann, 2007; Bahn & McGill, 2013) and marine (Wilson *et al*., 2007; Dolan & Lucieer, 2014; Vierod *et al*., 2014; Lecours *et al*., 2015) literature have suggested that closer attention should be paid to spatial concepts and analyses in habitat mapping. This dissertation demonstrates why these concepts are particularly important in marine-specific contexts and addresses the disconnect between recommendations and the integration of spatial analyses as essential steps in marine habitat mapping. Demonstrating the incorporation of these analyses in applied studies is a logical step in establishing them as best practices. Manuscripts within this dissertation use a case study approach to illustrate the incorporation of several spatial analyses into the habitat mapping workflow, while simultaneously producing habitat maps for a variety of practical applications that are tailored to the needs of local resource managers.

These case studies are well suited to demonstrating the benefits and importance of spatial analyses, partially as a function of the limitations associated with habitat mapping in challenging remote locations such as the Canadian Arctic. The multisource data qualities and sampling constraints in these Arctic case studies are extreme examples of circumstances encountered in marine habitat mapping that affect the spatial qualities of marine data. The manuscripts demonstrate the potential magnitude of effects that these spatial issues have on habitat map results.

5.1 Findings

Chapter 2 demonstrates that the default scale of terrain variables imposed by the data resolution is not necessarily optimal and applies a methodology for empirically selecting relevant scales. Terrain variables that represent seabed morphology were informative predictors for differentiating

muddy and sandy substrates and were implemented at both broad and fine scales (between 5-275 m). The acoustic backscatter of the seabed and fine-scale bathymetry were important for differentiating gravelly from sandy substrates. These results support the argument that the relationship between response and environment is scale-dependent (Wilson *et al*., 2007; Lechner *et al*., 2012; Lecours *et al*., 2015), and that it is important to match the scale of environmental variables to the response. Terrain attributes calculated at one scale can even be conceptualized as unique variables compared to when calculated at other scales, and can be treated as such during the variable selection process.

Chapter 3 combined models of presence-absence and abundance in order to predict the distribution of the soft-shell clam *Mya* spp. using transect seabed imagery data. This was the first application of a "hurdle"-like model in a benthic habitat mapping context to the knowledge of the author, and results suggested this improved predictive accuracy, while simultaneously producing more realistic predictions of clam habitat distribution by incorporating predicted absences. Results of a spatially explicit accuracy assessment showed that the spatial configuration of image samples inflated the apparent predictive performance of these models by between 23-121%, depending on the accuracy statistic. These results were used to estimate the uninflated predictive accuracy. Transect data are common in marine science – if the spatial dependence associated with these data are addressed at all, current methods typically aggregate the data spatially or omit samples from the analysis in order to ensure independence. The protocol applied in this manuscript allows for the use of all available data, while providing a spatially independent accuracy assessment. This methodology is suitable to most habitat mapping applications, including those at broader thematic scales (e.g., biotope, seascape).

The comparison between categorical and quantitative modelling approaches for producing classified sediment maps in Chapter 4 demonstrates the importance of considering spatial data characteristics for selecting a modelling approach. For example, sample distribution and the distinction between predictive goals of the model can influence the appropriateness of certain methods. Though this comparison was inconclusive as to whether one modelling approach can consistently outperform the other, there were important differences in the mapped predictions that may partially determine their suitability. The quantitative approach, for instance, predicted rare and unsampled sediment classes, which may be an asset under certain circumstances. These predictions cannot be validated though, requiring an untested confidence in the statistical relationships between the response and environment established in the models. Some confidence can be obtained through an understanding of local geological and oceanographic context, but this can become increasingly uncertain at high spatial resolutions (i.e., fine detail), especially where predictions defy expectation. Samples in this study were clustered and uneven, and there was evidence of both apparent performance inflation and model overfitting caused by spatial autocorrelation, making the prediction of rare and unsampled classes tenuous at best. In this case, the flexibility of probabilistic predictions from the categorical approach made it more suitable and more accurate.

5.2 Recommendations

5.2.1 Recommendations for Map Producers

There are several recommendations based on the thesis findings that are reasonable for map producers to implement in almost any supervised habitat mapping workflow. It is not always

necessary to explore each of these issues exhaustively in every study, as even a cursory investigation may indicate important spatial data qualities. The extent to which these topics are explored should be a function of the study goals and data characteristics.

Chapter 2 suggests that ideally the scale of environmental variables should be optimized to model a given benthic habitat characteristic as efficiently as possible. The approach in Chapter 2 relied on testing many different scales to select a predictor set based on the relationship with the response, either as surrogates for other environmental factors (e.g., slope aspect) or as a more direct indicator of seabed characteristics (e.g., backscatter variability). Realistically, it may not always be feasible to test many different scales for many different variables as in Chapter 2; a more mechanistic approach might attempt to conceptually link the scale or resolution of environmental data to the phenomenon being modelled, for instance by attempting to match the scale of terrain variables to the scale of morphological features. In either case, this means taking a critical approach to selecting data resolution, rather than accepting a "default" scale that is a function of data processing or the variable derivation algorithm. The selection of variable scale might ideally occur after data collection and processing, but before variable selection/reduction and model fitting.

Results from Chapters 3 and 4 suggest that all studies should at least explore spatial dependencies in the data being modelled. The empirical variogram was used throughout this dissertation and is readily (and easily) calculated in popular GIS (e.g., ArcGIS Geostatistical Wizard) and open source statistical software packages (e.g., the "*gstat*" package in R). Other methods for estimating spatial dependence may also be appropriate. Regardless, these analyses can indicate the magnitude of spatial dependence in ground-truth data, which may be useful for selecting appropriate modelling and validation approaches. Spatial dependence can be investigated in the raw data prior to model fitting (Roberts *et al*., 2017), or in model residuals afterwards (Veloz, 2009; Valavi *et*

al., 2018), depending on the application. At the very least, exploring these data qualities will give the user a better sense of the potential limitations associated with their analysis.

One such potential limitation is the ability of a supervised model to extrapolate predictions to new locations. Results from Chapter 4 suggest that a greater distinction between interpolative and extrapolative prediction would be beneficial to understanding the suitability and limitations of distribution models. This also suggests that the spatial data configuration, such as their distribution and evenness, should be considered when selecting modelling and validation approaches. Presently, most validation approaches are well suited to estimating interpolative performance but may be inadequate for extrapolation. This distinction is rarely considered. Measures of spatial dependence in the raw data, such as the variogram, may help to determine this prior to modelling. Mapping spatial autocorrelation decay may be useful for assisting the map producer in visualizing spatial dependence among the data (e.g., Figure 5.1), which can be accomplished using the variogram model parameters (e.g., Appendix B.3). This can aid in identifying if and where sample clustering might cause predictive accuracy inflation and in determining whether spatial interpolation or extrapolation techniques are appropriate for modelling and accuracy assessment.

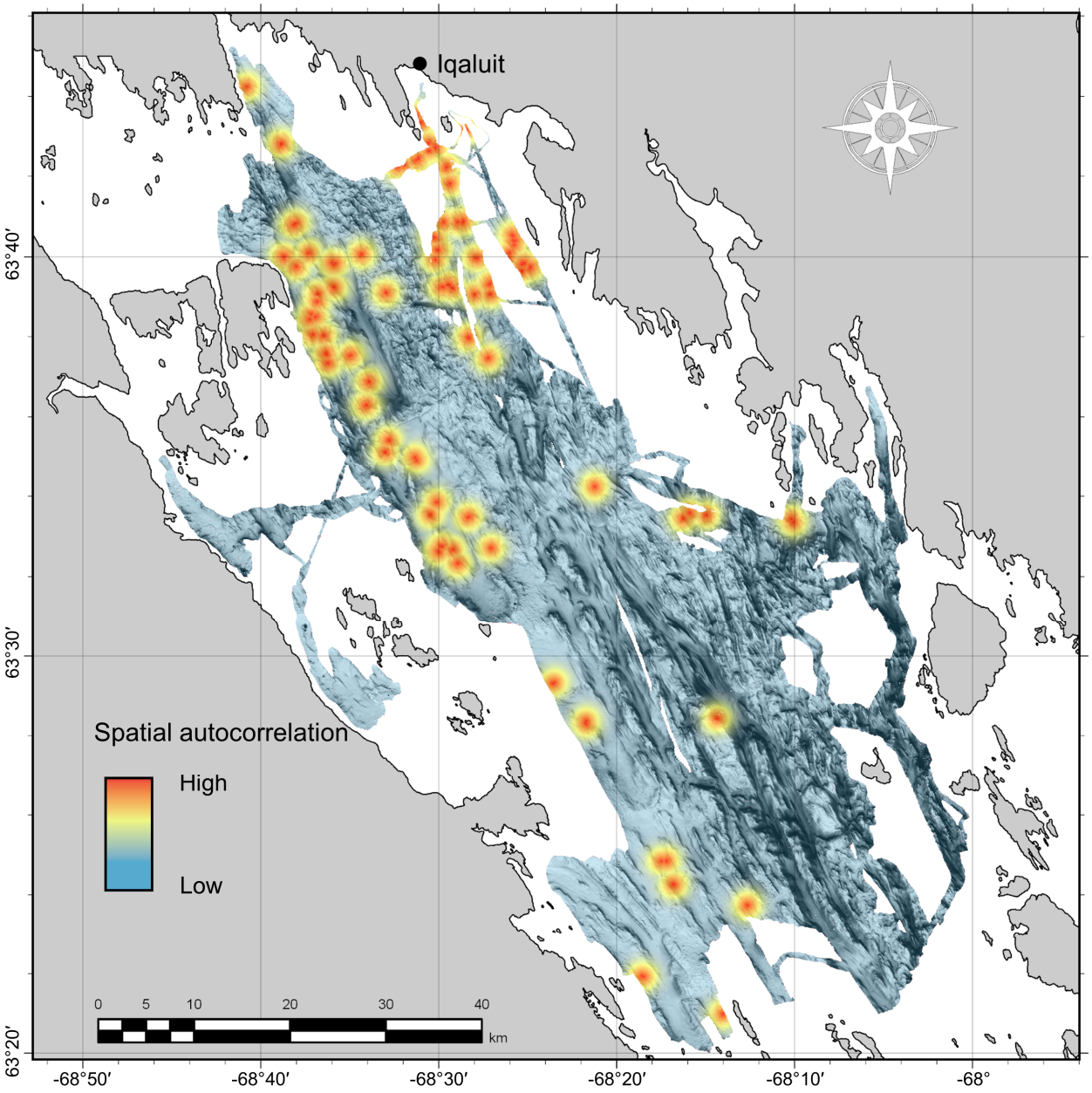


Figure 5.1. Spatial autocorrelation decay of mud measurements mapped from circular variogram model parameters (Appendix B.3) in Frobisher Bay (Chapter 4).

5.2.2 Recommendations for Map Users

There are also several recommendations based on the findings of this thesis that end-users of habitat maps should consider. A disconnect between map producers and users can hinder understanding of the quality or limitations of habitat maps, while understanding these

characteristics may be critical to determining how a map is used. Because the spatial issues considered in this thesis have implications for the perceived quality and appropriateness of benthic habitat maps and mapping methods, they are also relevant to map users.

The scale of analysis for producing benthic habitat maps should match the mapping or management objectives, and the interplay between data resolution and analysis scale is therefore important to understand. Remotely sensed habitat mapping data are often pre-processed and provided to map producers at a native resolution that is not informed by management goals. Map users should think critically about the scale of their objectives, and therefore the scale of data that is appropriate. This should inform the data resolution that is selected by the map producer. For example, MBES technology is constantly improving and acoustic data are being produced at increasingly higher resolutions (e.g., < 1 m). Though these data are highly useful for some applications, and are readily useable for habitat mapping, predictions at sub-metre scales may be inefficient for informing management of marine habitats, depending on the application. The presence of a commercially harvested benthic species such as clams, for example, might be irrelevant at sub-metre scales. Predictions may be more useful over metres, tens, or even hundreds of metres. Habitat map end-users may have a better understanding of a suitable analysis scale than the map producer, and it is important to convey this information. Lacking *a priori* information, it may also be possible to test for scale-dependence in the data, as in Chapter 2. Regardless, the scale of analysis should not be dictated solely by the native or default data resolution; it should be informed by consideration of criteria that indicate scale suitability, such as mapping purpose or empirical analysis, and this understanding should be shared between the map producer and user.

It is important to recognize that the accuracy reported by the map producer depends on predictive goals and study design, which may not necessarily align with management goals. Chapter 4

demonstrates how the distribution of samples is fundamentally linked to accuracy assessment. If predictions are derived in areas that have not been adequately sampled, for example, then the model must extrapolate to these areas, which may contain new environments. Most common map evaluation techniques that rely on withholding portions of the dataset for testing predictions will likely fail to represent the quality of predictions in extrapolated areas, which may be important to managers. The misalignment of predictive goals and management goals could potentially hinder the success of spatial management designs.

Relatedly, it is important for habitat map end-users to understand that reported map accuracy is not uniform across a study area. If spatial autocorrelation is present in the data, but is not accounted for during accuracy assessment, then reported accuracies are likely to represent predictions that are close to sample sites, but not those that are farther away. These accuracy metrics are therefore overoptimistic in some parts of the study area. If spatial autocorrelation is present and is accounted for during the accuracy assessment using methods similar to Chapters 3 and 4, then reported accuracies represent predictions that are farther away from samples (i.e., beyond the range of spatial autocorrelation), while predictions close to samples may be more accurate than the reported values. In such cases the reported accuracy is conservative, which is often preferable when samples are distributed unevenly or when predicting to under-sampled or unsampled environments. Such hierarchies of confidence in the reported map accuracies can be represented spatially, which is helpful for conveying the spatially explicit nature of performance metrics (e.g., Figure 5.2). Maps of accuracy confidence do not represent the predictive confudence, they describe to what degree the reported accuracy statistics represent different parts of the study area (e.g., from Section 4.3.4).

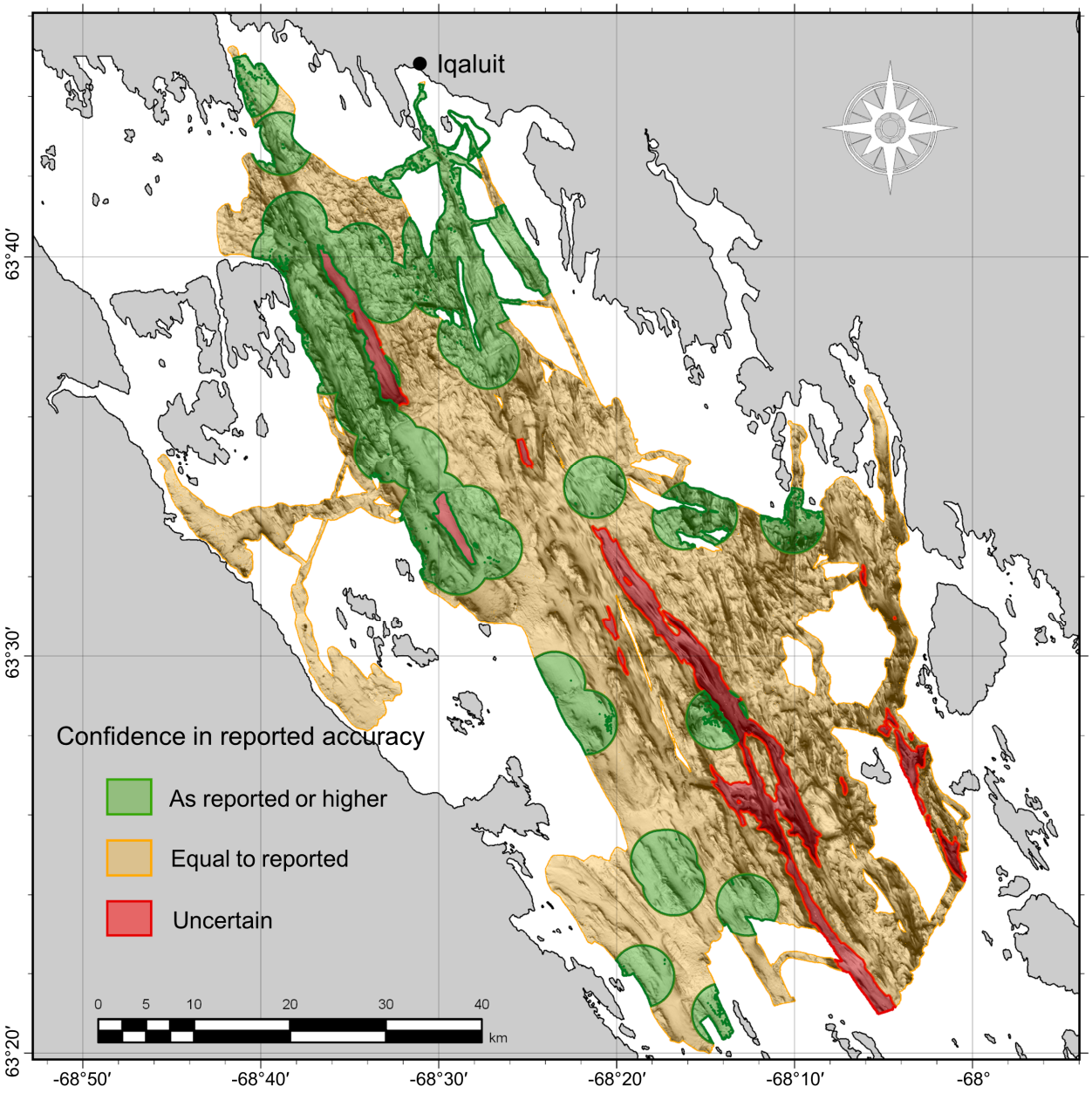


Figure 5.2. Example of mapped confidence of spatially explicit accuracy statistics for grain size predictions in Frobisher Bay (Chapter 4). Green zones within the range of spatial autocorrelation are at least as accurate as reported, but potentially more accurate; yellow zones outside the range of autocorrelation are expected to be as accurate as reported; red zones are beyond the range of sampled environments, where reported accuracies may not be representative.

Above all, it is important to understand that map predictions are not truth and are never fully correct. All seabed map predictions attempt to represent reality using incomplete information, and all maps therefore contain some associated amount of error. Minimizing and quantifying this error

are jobs of the map producer, but not all accuracy assessments are universally appropriate, and the method of assessment is often at the discretion of the map producer. End-users of benthic habitat maps can benefit greatly from an understanding of this process when interpreting the suitability of habitat maps for management applications.

5.3 Future Work and Outstanding Challenges

Chapter 2 demonstrated a method for selecting benthic terrain variables at multiple spatial scales, yet there are several potential approaches for deriving these variables. Multiple scale functionality is available in some older GIS software (e.g., Landserf, GRASS; Wood, 2009), but is not explicitly incorporated into widely used modern software packages (e.g., ESRI ArcGIS). In the latter, some variables that are derived via neighbourhood raster analyses can be optionally calculated at specific scales, but this is not the case for many variables that are commonly used in marine habitat mapping, such as slope, curvature, and surface roughness. Calculating these at multiple scales requires alternative approaches or workarounds, which have been summarized by Dolan (2012) and Dolan & Lucieer (2014). Given these potential alternative approaches, it is not currently clear whether some are more appropriate under certain circumstances. There are several reasons why it may be desirable to change the scale of a variable, for instance to capture specific terrain features, to match the scale of environmental data to ground truth data, or to reduce artefacts or errors in acoustic data (note that these are marine-specific applications). Future work should investigate the theoretical and practical implications of these different approaches for calculating variables at multiple scales and offer recommendations on their use.

Relatedly, the concept of “true” multiscale (hereafter, simply “multiscale”) has yet to be investigated in a marine context. This concept considers the selection of spatial scale a “basic

problem in geomorphometry" (Shary *et al*., 2002), and therefore attempts to derive variables whose scale is not fixed, but varies by location. Again, this functionality has been implemented in older GIS software but is generally unavailable in modern packages; thus, GIS users seek alternative approaches. Some of these may average multiple scales of a variable to produce a single multiscale layer (e.g., Dolan, 2012), while others aim to calculate each cell in a raster at the optimal scale (e.g., Lindsay & Newman, 2018). There may also be additional approaches that are specific to each variable, or that involve the use of principal components, yet these have yet to be explored. An entirely separate suite of approaches might examine the calculation of multiscale variables in the context of marine habitat mapping, for instance, linking the scales of response variable to species. This research avenue is almost completely unexplored and offers exciting possibilities for marine habitat mapping.

The approach adopted for predicting distributions of seabed sediments in Chapter 2 is highly data-driven – attempting to uncover statistical relationships between ground-truth samples and many candidate explanatory variables at specific scales. Previous research on fjordic sedimentation in the Arctic has explored the mechanisms behind these processes in detail (e.g., Gilbert, 1983; Syvitski & Schafer, 1985; Syvitski, 1989; Gilbert *et al*., 1993; Syvitski & Shaw, 1995). A separate approach might attempt to link the processes uncovered in these geomorphological investigations at specific scales to the ground-truth data – taking a mechanistic rather than data-driven modelling approach. A comparison between these approaches would be highly informative as to the potential utility of each in fjordic settings. Such comparisons have been undertaken elsewhere (e.g., North Sea, UK; Diesing *et al*., 2014), but not in a fjordic environment at the time of writing, to the author's knowledge. The morphological heterogeneity of these estuarine environments makes

them complex to map – how data-driven and mechanistic approaches compare at predicting sediment and bottom type distributions under these condition remains to be seen.

Chapters 3 and 4 of this dissertation applied spatially explicit cross-validation approaches to model transect data to reduce the effects of spatial autocorrelation on predictive accuracy. There are other approaches for ensuring the independence of transect ground-truth data that involve aggregating observations, adjusting the environmental data resolution, or ensemble modelling, which may be simpler to implement, and similarly effective. Future work should compare these approaches to determine whether some are preferable universally, or under certain circumstances. Another solution is to handle this issue *a priori* by selecting a ground-truth sample design that ensures the spatial independence of sample points, yet this is not always feasible when using inherited or legacy data. Furthermore, with transect data, the decision ultimately must be made to either aggregate samples within a transect or treat them as discrete points. Results from Chapter 4 suggest that this decision might be affected by the predictive goals of the model (i.e., interpolation versus extrapolation).

The use of multibeam backscatter has become widespread in marine habitat mapping, yet substantial difficulties arise when attempting to use data acquired from different surveys, or by different systems. The backscatter measurements acquired by multibeam echosounders are most often uncalibrated, meaning that these data are relative to a single survey and not directly comparable with others (Lucieer *et al*., 2018). Furthermore, the acoustic response of the seabed is affected by other acquisition parameters, such as the operating frequency (Brown *et al*., 2019), potentially resulting in disparate measurements for the same area. Nonetheless, the use of data collected opportunistically by multiple vessels is often a necessity in marine habitat mapping (Lacharité *et al*., 2018), as it was in each manuscript of this dissertation. Harmonizing disparate

backscatter datasets has been attempted (Hughes Clarke *et al*., 2008); each manuscript in this thesis relied on similar "bulk shift" approaches to produce a single backscatter layer for use as an explanatory variable in predictive models. A limited analysis included in the Chapter 2 Appendix (A.1) suggested that, as a function of the sample distribution in this context, a bulk shift harmonization produced more accurate results than analyzing each dataset in isolation and combining the results *post-hoc* (e.g., Lacharité *et al*., 2018). Otherwise, there has been little research on the effectiveness and appropriateness of bulk shift and other backscatter harmonization methods. Future work should compare different methods for harmonizing multisource backscatter datasets and investigate the advantages and disadvantages of such approaches compared to the *post-hoc* combination of results produced using disparate backscatter datasets.

There is a clear need to address the mismatch in spatial scales of marine data types for benthic habitat mapping to move towards an integration of benthic and oceanographic habitat variables. Although the use of MBES data to describe seabed terrain and substrate at fine scales has become commonplace, it is still relatively uncommon to include oceanographic predictors alongside MBES data to characterise benthic habitats at fine scales (Brown *et al*., 2011). We often consider benthic habitats in two dimensions, but it is critical to remember that these are three-dimensional systems and neglecting to characterise water column properties that overlay the seabed is failing to consider a fundamental component of seabed habitat. Oceanographic variables are readily available at broad scales, but it is important to develop robust approaches for obtaining this information at scales relevant for high resolution habitat mapping. Approaches such as downscaling oceanographic models, interpolating discrete temporally variable oceanographic measurements, and collecting continuous oceanographic data while simultaneously mapping the seabed need to be further developed to fill this critical data gap.

5.4 Conclusions

As marine habitat mapping has developed as an independent field of research it has acquired recommendations, commonly from other disciplines, that have evolved into "best practices". The testing of predictive accuracy on independent data, for instance, has become an integral component to habitat modelling and mapping – analyses may be considered incomplete without it. The spatial analyses discussed here have yet to be adopted in this manner, yet doing so could benefit habitat map producers substantially.

To progress towards a better integration of these concepts into the habitat mapping workflow, this dissertation demonstrates the application and importance of spatial analyses contextually, building directly on recommendations from this field. Habitat map producers face unique challenges in the marine realm that require specific solutions. This dissertation is therefore part of a larger effort to define best practices for marine habitat mapping in order to increase the quality and transparency of these methods. Above all else, it is important to consider that the effectiveness of spatial management strategies, which are urgently needed amidst increasing anthropogenic impacts and changing climate, are limited by the quality of information used to inform them. Thus, the mandate of habitat map producers is to create the best maps possible and to accurately report results. To achieve both goals we need to be active in investigating marine-specific habitat mapping methods, and in establishing best practices within this field.

5.5 References

Bahn, V., and McGill, B. J. 2013. Testing the predictive performance of distribution models. *Oikos*, 122(3): 321–331.

Brown, C. J., Beaudoin, J., Brissette, M., and Gazzola, V. 2019. Multispectral multibeam echo sounder backscatter as a tool for improved seafloor characterization. *Geosciences*, 9(3): 126.

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(3): 502–520.

Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D. 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, 84: 107–119.

Dolan, M. F. J. 2012. Calculation of slope angle from bathymetry data using GIS – effects of computation algorithms, 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(2): 187–219.

Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16(2): 129–138.

Gilbert, R. 1983. Sedimentary processes of Canadian arctic fjords. *Sedimentary Geology*, 36(2–4): 147–175.

Gilbert, R., Aitken, A. E., and Lemmen, D. S. 1993. The glacimarine sedimentary environment of Expedition Fiord, Canadian High Arctic. *Marine Geology*, 110(3–4): 257–273.

Hughes Clarke, J. E., Iwanowska, K. K., Parrott, R., Duffy, G., Lamplugh, M., and Griffin, J. 2008. Inter-calibrating multi-source, multi-platform backscatter data sets to assist in compiling regional sediment type maps: Bay of Fundy. *In* Proceedings of the Canadian Hydrographic Conference and National Surveyors Conference 2008. Victoria.

Lacharité, M., Brown, C. J., and Gazzola, V. 2018. Multisource multibeam backscatter data: Developing a strategy for the production of benthic habitat maps using semi-automated seafloor classification methods. *Marine Geophysical Research*, 39(1–2): 307–322.

Lechner, A. M., Langford, W. T., Jones, S. D., Bekessy, S. A., and Gordon, A. 2012. Investigating species–environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. *Ecological Complexity*, 11: 91–102.

Lecours, V., Devillers, R., Schneider, D. C., Lucieer, V. L., Brown, C. J., and Edinger, E. N. 2015. Spatial scale and geographic context in benthic habitat mapping: Review and future directions. *Marine Ecology Progress Series*, 535: 259–284.

Lindsay, J. B., and Newman, D. R. 2018. Hyper-scale analysis of surface roughness. *PeerJ Preprints*.

Lucieer, V., Roche, M., Degrendele, K., Malik, M., Dolan, M. F. J., and Lamarche, G. 2018. User expectations for multibeam echo sounders backscatter strength data-looking back into the future. *Marine Geophysical Research*, 39(1–2): 23–40.

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(8): 913–929.

Shary, P. A., Sharaya, L. S., and Mitusov, A. V. 2002. Fundamental quantitative methods of land surface analysis. *Geoderma*, 107(1–2): 1–32.

Syvitski, J. P. M., and Schafer, C. T. 1985. Sedimentology of Arctic Fjords Experiment (SAFE): Project introduction. *Arctic*, 38(4): 264–270.

Syvitski, J. P. M. 1989. On the deposition of sediment within glacier-influenced fjords: Oceanographic controls. *Marine Geology*, 85(2–4): 301–329.

Syvitski, J. P. M., and Shaw, J. 1995. Chapter 5: Sedimentology and geomorphology of fjords. *In* Geomorphology and Sedimentology of Estuaries, pp. 113–178. Ed. by G. M. E. Perillo. Elsevier, Amsterdam.

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(2): 225–232.

Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36(12): 2290–2299.

Vierod, A. D. T., Guinotte, J. M., and Davies, A. J. 2014. Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. *Deep Sea Research Part II: Topical Studies in Oceanography*, 99: 6–18.

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., and Grehan, A. J. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, 30(1–2): 3–35.

Wood, J. 2009. Chapter 14 Geomorphometry in LandSerf. *In* Developments in Soil Science, pp. 333–349. Ed. by H. Tomislav and I. R. Hannes. Elsevier.

Appendix A. Chapter 2

A.1 Backscatter Data Harmonization

The use of multiple non-calibrated backscatter datasets presents several difficulties, yet due to the high cost of data collection and the importance of these data for seabed mapping, methods have been developed to facilitate their combination. Seabed acoustic reflectivity has been used as a proxy for bottom hardness and substrate properties, yet acoustic measurements are dependent on several water column conditions (e.g., temperature, salinity), MBES system-specific parameters (e.g., operating frequency, pulse length), and survey conditions (e.g., vessel speed, survey overlap) (Lacharité *et al*., 2018). For measurements to represent seabed characteristics, radiometric and geometric corrections must be applied to raw backscatter intensity to correct for these factors (Lurton & Lamarche, 2015). Corrections to raw backscatter intensity produce relative measurements that are specific to a single MBES system, and often to a single survey, unless calibrations have been performed to the systems prior to the survey. Furthermore, even after confounding factors have been controlled for, relative backscatter measurements are a function of operating frequency as well as substrate properties (Hughes Clarke *et al*., 2008; Hillman *et al*., 2018).

There are two broad methodologies for combining non-calibrated backscatter datasets: analyzing datasets separately and combining results *post-hoc* (e.g., Lacharité *et al*., 2018), or combining them into one harmonized dataset prior to analysis (e.g., Hughes Clarke *et al*., 2008). Both methodologies were tested here. Modelling the response of grain size to the single harmonized backscatter dataset produced maps with less-noticeable boundaries and edge effects caused by dataset combination (Figure A1). The harmonized approach also produced more accurate results

than creating independent models of each backscatter dataset and combining the results (Table A1).

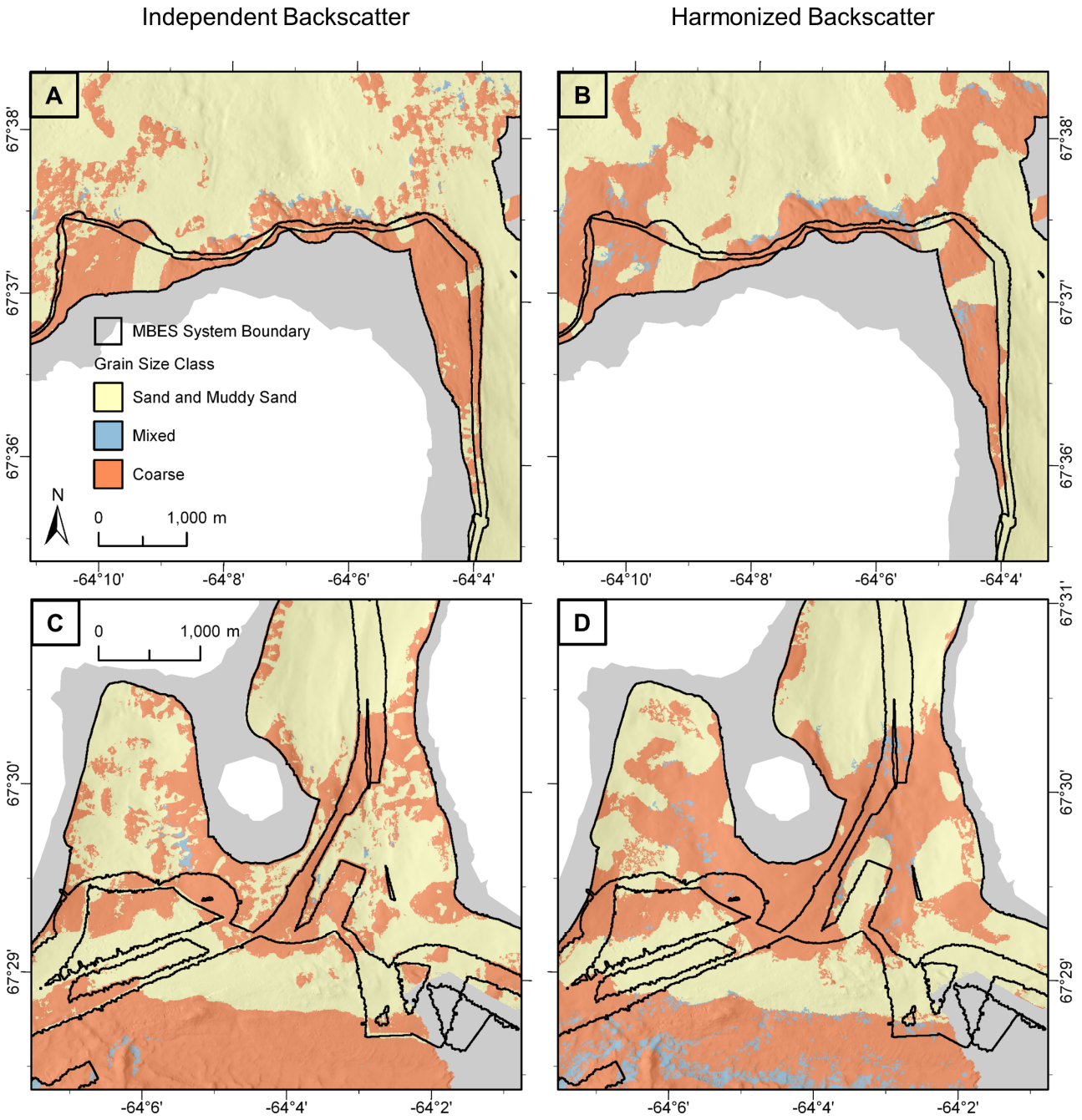


Figure A1. Differing grain size class predictions from modelling independent (A, C), and harmonized (B, D) backscatter datasets. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

Table A1. Accuracies of independently modelled backscatter mosaics and harmonized backscatter mosaic.

	ρ_{mud}	ρ_{sand}	ρ_{gravel}	ALR_{ms} Deviance Explained	ALR_{gs} Deviance Explained
EM3002	0.744	0.524	0.439	44.8%	17.4%
EM2040C	0.331	0.677	0.500	44.1%	59.2%
Average of Independent Models	0.538	0.600	0.469	44.4%	38.3%
Harmonized	0.772	0.712	0.578	56.3%	46.4%

Similar to Hughes Clarke *et al*. (2008) we applied bulk shifts to align backscatter datasets from different surveys to create a single harmonized backscatter dataset. Using the most extensive survey as reference, we observed how the surveys differed in areas where they overlapped. Hughes Clarke *et al*. (2008) applied bulk shifts to backscatter datasets from five different MBES systems ranging from 93 to 300 kHz by adding or subtracting the relative differences (in dB) of overlapping survey areas with respect to a reference backscatter dataset. We attempted both this method and a bulk shift that multiplied the backscatter datasets from different surveys based on the median factor by which datasets differed with respect to the reference (EM3002 in 2012). The multiplication method affected the range of the data as well as the median and produced backscatter layers that were closer to the reference on average, with a lower standard deviation (Table A2). Visual analysis suggested that this method was effective at harmonizing backscatter datasets from different surveys and MBES systems (Figure A2), with few noticeable boundaries between the different datasets in the resulting layer.

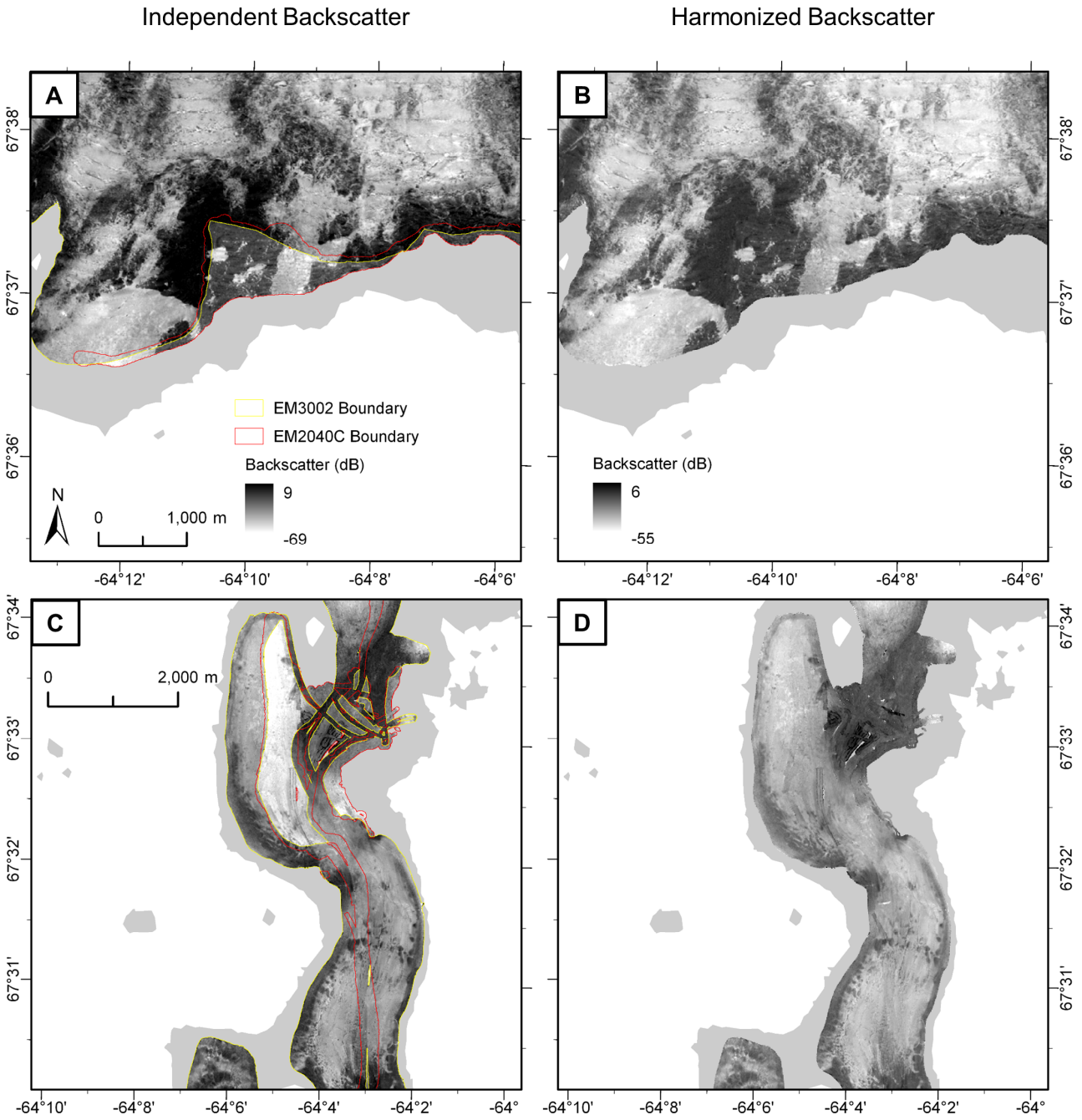


Figure A2. (A, C) Non-corrected backscatter mosaics, and (B, D) bulk shift harmonized backscatter mosaic. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

Table A2. Mean absolute error between independent backscatter mosaics with respect to reference survey (2012; EM3002) after bulk shift.

	2013 (EM3002)		2014 (EM2040C)		2015 (EM2040C)	
	Additive Shift	**Factor Shift**	**Additive Shift**	**Factor Shift**	**Additive Shift**	**Factor Shift**
Mean (dB)	2.17	2.16	2.67	2.59	3.03	2.75
SD (dB)	1.99	1.96	2.41	2.32	2.84	2.56

The decision to use a single harmonized backscatter dataset that integrates surveys from four different years was based on the performance of this dataset compared to using them individually, and the quality of the map products. The harmonized backscatter layer is a relative and imperfect proxy for seabed hardness, yet results suggest this is the most efficient way to utilize such valuable data. Furthermore, the harmonized backscatter layer was the single most important variable for differentiating gravel from sand; the Boosted Regression Trees algorithm that we used for modelling ignores noisy or non-important variables (Elith *et al*., 2008), suggesting that the harmonized backscatter data produced using these methods was useful as a proxy for substrate properties.

A.2 Bathymetry and Terrain Variable Harmonization

Bathymetric raster layers from each survey were mosaicked to a 5 m grid to form a single harmonized layer prior to deriving terrain variables and averaging to create multiple scales. The beam width of the EM3002 echosounder was 1.5° x 1.5°, the EM2040C was 1° x 1°, and the EM300 (backscatter not used) was variable between 1° x 1°, 1° x 2°, 2° x 2°, and 2° x 4°, meaning that the acoustic footprint and sounding density differed between the

systems. The differences in inherent data resolution of the MBES systems may have resulted in small discrepancies in bathymetric resolution when gridded as a 5 m raster, which were not initially apparent, but which manifest in derivative variables (Hughes Clarke, 2003). A low-pass smoothing filter was applied to higher-order variables that seemed potentially affected by the differences in inherent resolution (e.g., Figure A3), including measures of curvature and relative difference to the mean value (RDMV; a measure of topographic variability). This issue can also be resolved when raster layers are averaged to create coarser scale variables, which is a positive byproduct of a multiple scale approach. The performance of these variables and others at fine scales (Figures 2.3-2.5 in Chapter 2) suggested that the low-pass filter was effective at reconciling differences in inherent resolution, and that discrepancies between data from the different echosounders did not seriously inhibit the use of these variables as predictors.

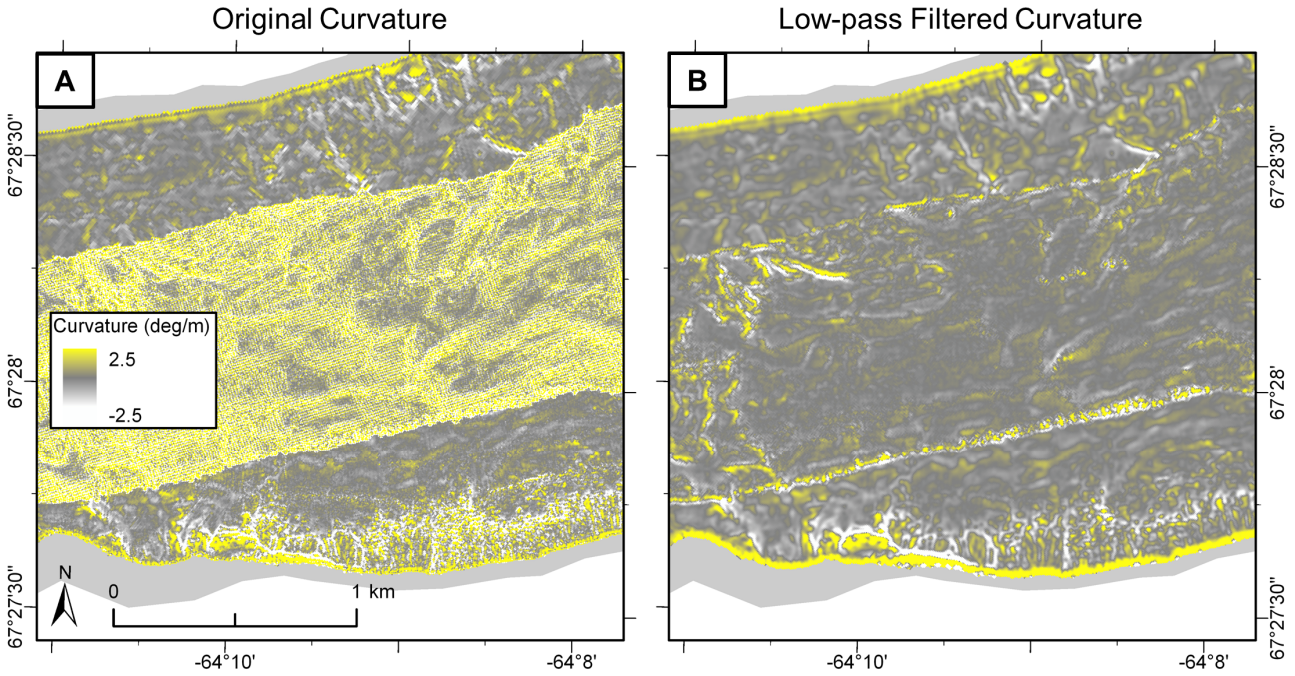


Figure A3. (A) Curvature layer derived from 5 m gridded bathymetry data. (B) Curvature layer after low-pass filtering. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

Even after applying a low-pass filter to the variables listed above, there were clear artefacts present in some areas where surveys from different years overlapped (e.g., Figure A3). Rather than risk impacting the analysis with these incorrect measurements, these areas were removed prior to modelling (e.g., Figure A4). These areas were omitted from all variables in order to maintain equal geographic extent.

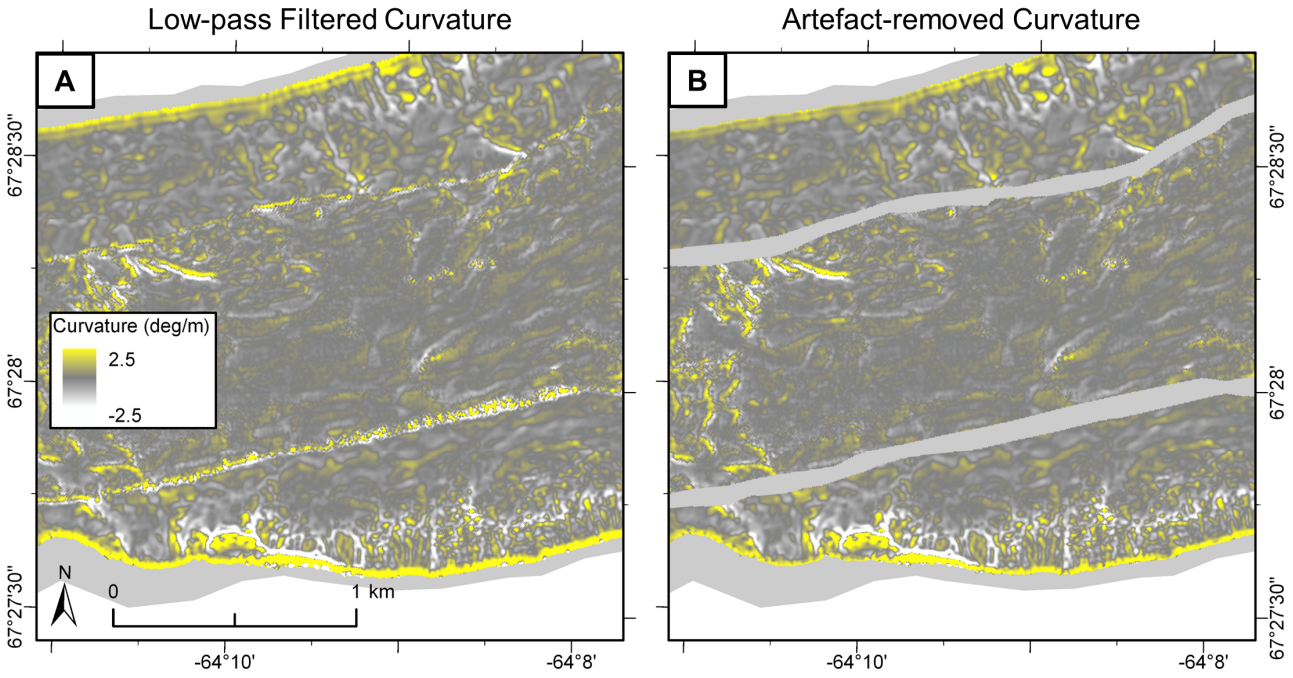


Figure A4. (A) Curvature layer after low-pass filtering. (B) Curvature layer after removing artefacts at MBES survey boundaries. Basemap from the Canadian Land Cover GeoBase Series, containing information licensed under the Open Government Licence – Canada.

A.3 References

Elith, J., Leathwick, J. R., and Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 802–813.

Hillman, J. I. T., Lamarche, G., Pallentin, A., Pecher, I. A., Gorman, A. R., and Schneider von Deimling, J. 2018. Validation of automated supervised segmentation of multibeam backscatter data from the Chatham Rise, New Zealand. *Marine Geophysical Research*, 39(1–2): 205–227.

Hughes Clarke, J. E. 2003. Dynamic motion residuals in swatch sonar data: Ironing out the creases. *International Hydrographic Review*, 4(1): 6–23.

Hughes Clarke, J. E., Iwanowska, K. K., Parrott, R., Duffy, G., Lamplugh, M., and Griffin, J. 2008. Inter-calibrating multi-source, multi-platform backscatter data sets to assist in

compiling regional sediment type maps: Bay of Fundy. *In* Proceedings of the Canadian Hydrographic Conference and National Surveyors Conference 2008. Victoria.

Lacharité, M., Brown, C. J., and Gazzola, V. 2018. Multisource multibeam backscatter data: Developing a strategy for the production of benthic habitat maps using semi-automated seafloor classification methods. *Marine Geophysical Research*, 39(1–2): 307–322.

Lurton, X., and Lamarche, G. (Eds). 2015. Backscatter measurements by seafloor-mapping sonars. Guidelines and recommendations. GeoHab Backscatter Working Group.

Appendix B. Chapter 4

B.1 Multibeam Echosounder Data Processing

Bathymetry data were imported to Qimera version 1.7; erroneous values were removed manually or using conservative spline filters. The data were corrected for tides using the Arctic9 tide model (Collins *et al*., 2011). Data acquired in each survey year and system were processed separately, exported as a 10 m floating point geoTIFF grid, and mosaicked in ESRI ArcGIS Pro v.2.1 to a single raster (Figure 4.4 in Chapter 4).

Uncalibrated MBES backscatter data from each survey year and system were processed using the Fledermaus Geocoder Toolbox (FMGT) and exported separately as floating point geoTIFF grid files. Focal statistics were used in ESRI ArcGIS Pro to smooth the data over a 5 x 5-cell neighborhood to reduce noise. The use of multisource backscatter datasets presents several difficulties as relative dB values partially depend on the acquisition parameters of individual MBES systems (e.g., operating frequency; Lurton & Lamarche, 2015). If each survey has been adequately ground-truthed, disparate datasets can be analyzed separately and their results combined (e.g., Lacharité *et al*., 2018). Here, some of the datasets had few or no ground truth samples. We therefore adopted a normalization approach by which separate datasets were harmonized using a "bulk shift" methodology (e.g., Hughes Clarke *et al*., 2008; Misiuk *et al*., 2018). This standardizes each survey using the most extensive as reference, operating under the assumption that relative backscatter strength is a function of substrate properties and is relatively stable throughout a given

survey. All surveys were thus mosaicked to a single raster at 10 m resolution, and a low pass filter was applied to smooth out remaining data noise (Figure 4.5 in Chapter 4).

B.2 Variable Scale Selection

For the quantitative models, we calculated Spearman's correlation coefficient for each scale of each predictor variable with grain size composition (ALR_{sm} and ALR_{gm}) to test for non-parametric monotonic relationships. We attempted to determine up to two appropriate scales (i.e., "intrinsic scales"; Lechner *et al*., 2012) for each predictor by identifying local peaks in the plot of correlation vs. variable scale. Because calculating correlation coefficients between a multi-level categorical response (viz., grain size classes) and quantitative predictors is not as straightforward, we used univariate multinomial logistic regressions to test the ability of each predictor at each scale to explain the sediment grain size class. The residual deviance of the univariate models was plotted against variable scale and up to two local minima were identified in each graph as intrinsic scales. All correlation scores and multinomial logistic regressions were calculated in R using the "cor" and "multinom" functions within the "*stats*" and "*nnet*" packages.

We tested whether selected scales of a given variable were correlated with each other and removed the weakest variable (based on relationship with the response) if Spearman's $\rho \geq 0.70$ between predictors (e.g., Gottschalk *et al*., 2011; Dormann *et al*., 2013; Downie *et al*., 2016). We then tested the correlation between all remaining scales of all variables and removed weaker variables in cases where Spearman's $\rho \geq 0.70$.

B.3 Variogram Analysis

Variograms were calculated for measurements of each grain size fraction and the presence of coarse sediments. The following model fits were obtained from the ArcGIS Geostatistical Wizard, but variograms and models were produced using the "*gstat*" package in R.

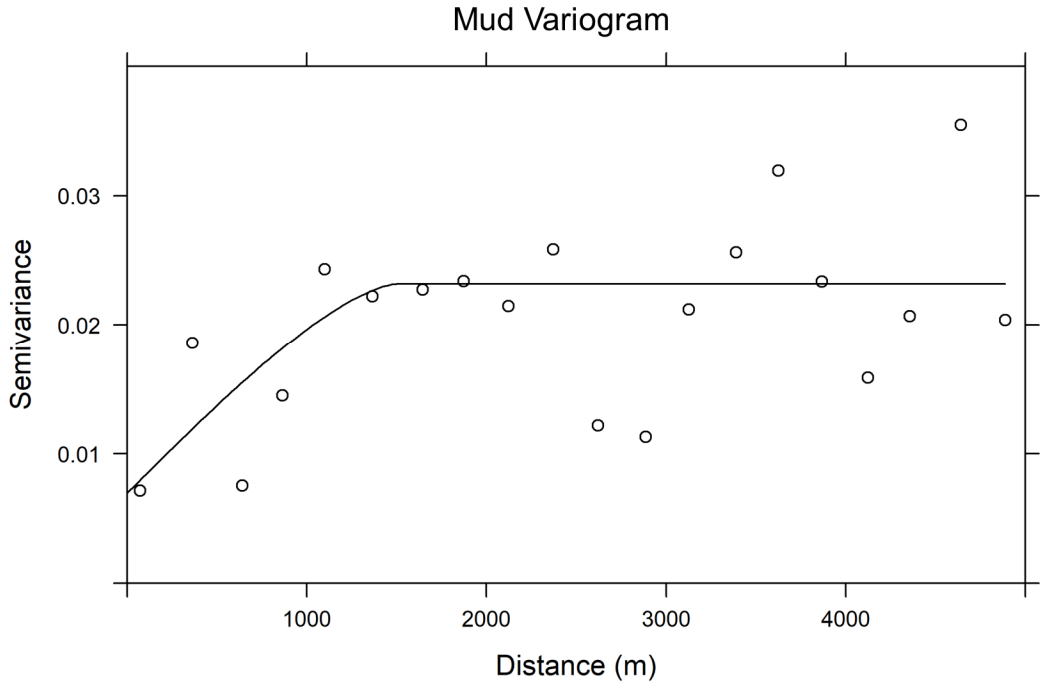


Figure B1. Mud fraction variogram circular model with 250 m lags; nugget = 0.0069, partial sill = 0.0162, major range = 1496.63 m.

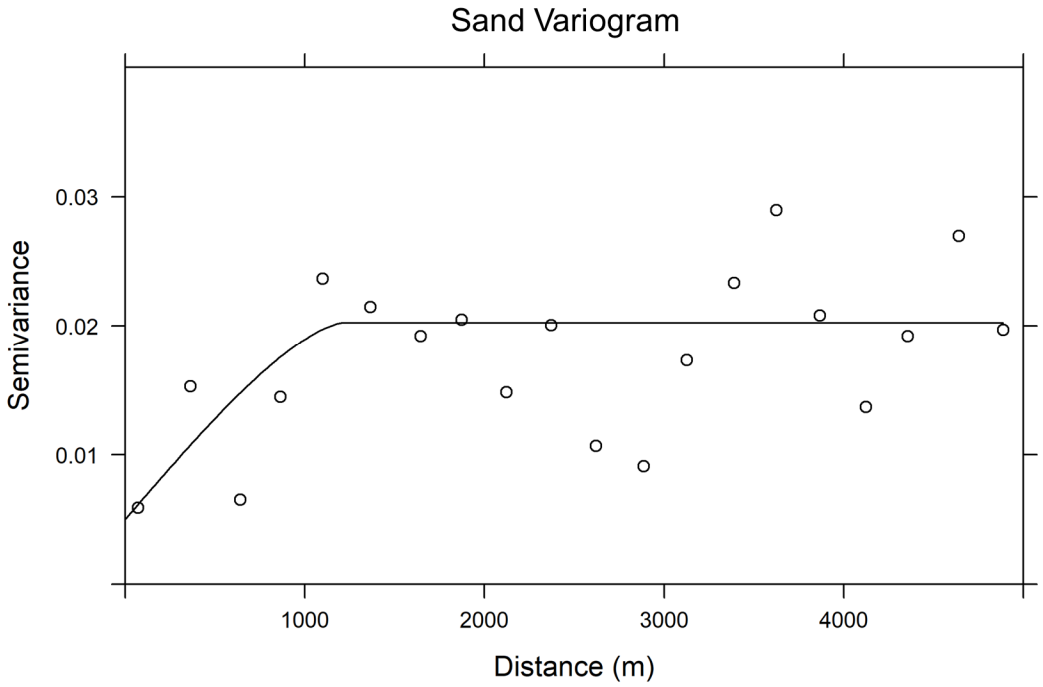


Figure B2. Sand fraction variogram circular model with 250 m lags; nugget = 0.0050, partial sill = 0.0152, major range = 1210.05 m.

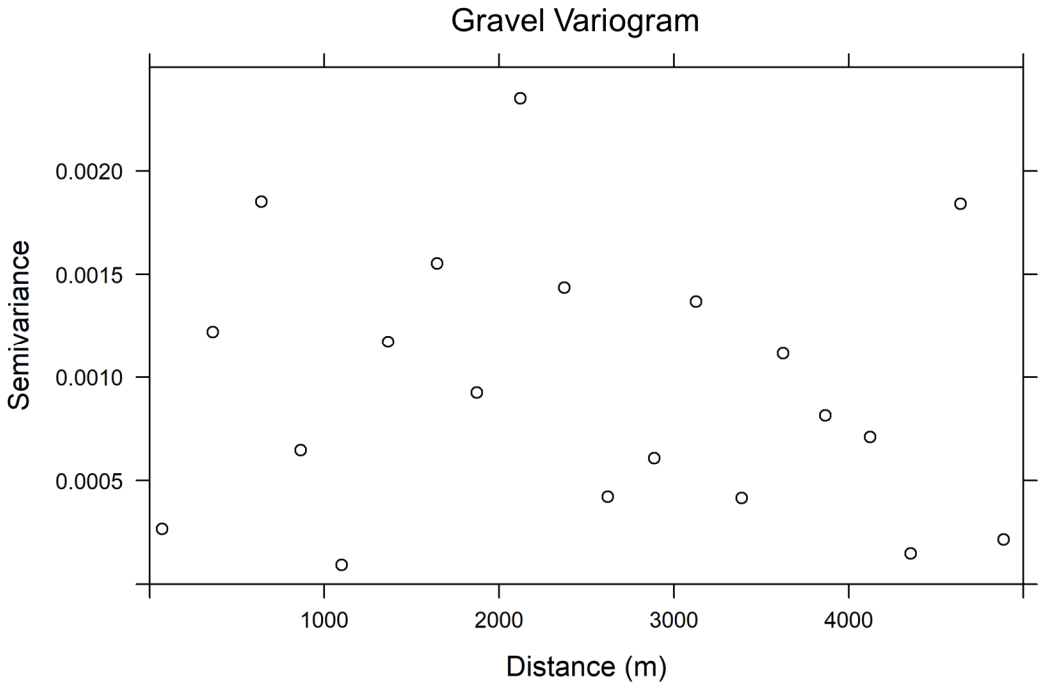


Figure B3. Gravel fraction variogram with 250 m lags.

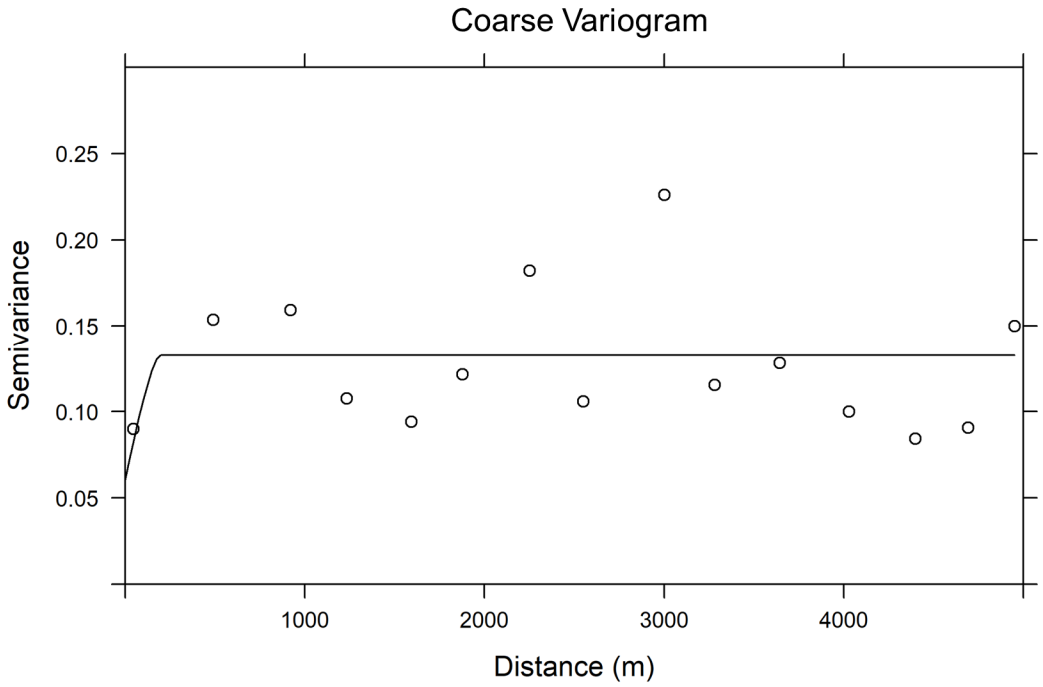


Figure B4. Coarse substrate variogram circular model with 250 m lags; nugget = 0.0605, partial sill = 0.0721, major range = 191.87 m.

B.4 Continuous Quantitative Model Performance

The predictive performance of quantitative model continuous predictions was estimated using: 1) Pearson's and Spearman's coefficients to determine the strength of linear and non-linear correlation between predicted the observed values, 2) mean absolute error (MAE) to determine, on average, the error in predicted percentages of mud, sand, and gravel, and 3) the percent variance explained (%VE), to quantify the accuracy of the model standardized to the variance of the observed grain size fraction values.

Nearly all continuous quantitative predictions of sediment grain size seemed accurate using leave-one-out cross-validation (LOO CV) yet were significantly less accurate using spatial (buffered) leave-one-out cross-validation (SLOO CV; Table B1). The LOO CV %VE

suggested that the relative error of these predictions was less than the variance in the observed values, but SLOO CV %VE values were significantly lower and negative, indicating a high amount of error relative to the measurements of each size class. Similarly, the non-relative error between predicted and observed values (i.e., the MAE) was significantly higher in SLOO CV predictions than LOO. With the exception of gravel predictions, all Pearson and Spearman correlation scores were lower in SLOO predictions than LOO CV.

Table B1. Accuracies of unclassified quantitative grain size predictions using spatial and non-spatial cross-validations.

		LOO CV	SLOO CV (1500 m)
Mud	%VE	60.96*	-6.42*
	MAE (%)	8.28*	13.51*
	Pearson	0.78*	0.13*
	Spearman	0.68*	0.06*
Sand	%VE	60.05*	-10.99*
	MAE (%)	8.01*	12.91*
	Pearson	0.78*	0.06*
	Spearman	0.67*	0.03*
Gravel	%VE	29.66*	-2.06*
	MAE (%)	1.13*	1.58*
	Pearson	0.61*	0.13*
	Spearman	0.40	0.32

*Significant difference between LOO and SLOO.

B.5 Error Matrices

Because the leave-one-out cross-validations (including spatial leave-one-out) produce a prediction for each sample point in the dataset, error matrices can be calculated between observed and predicted sediment classes for each sample. Note a slight difference in

observed grain size classes caused by the complete absence of gravel from some samples; quantitative samples were transformed to additive log-ratios, which do not allow zero values. The following tables (B2-B13) correspond to those discussed and compared in results section 4.3.4 ("Grain Size Model Evaluation and Comparison"). Error matrices were also calculated to estimate the predictive performance of the categorical grain size and coarse substrate models after optimizing the threshold of occurrence to maximize the sum of sensitivity + specificity. Tables B14 and B15 correspond to results section 4.3.6 ("Combined Map and Model Tuning").

Table B2. Categorical Folk LOO CV error matrix.

		Observed						
		(g)mS	(g)sM	gmS	gS	M	mS	sM
Predicted	(g)mS	10	4	0	0	0	2	2
	(g)sM	9	43	1	0	1	0	29
	gmS	0	0	0	0	0	0	0
	gS	0	0	0	2	0	0	0
	M	0	0	0	0	0	0	0
	mS	1	0	0	0	0	4	2
	sM	5	30	2	0	0	3	39

Table B3. Categorical Folk SLOO CV error matrix.

		Observed						
		(g)mS	(g)sM	gmS	gS	M	mS	sM
Predicted	(g)mS	0	1	0	2	0	0	2
	(g)sM	18	46	0	0	0	7	23
	gmS	0	0	0	0	0	0	0
	gS	0	0	0	0	0	0	0
	M	0	0	0	0	0	0	0
	mS	0	1	0	0	0	0	0
	sM	7	29	3	0	1	2	46

Table B4. Quantitative Folk LOO CV error matrix.

		Observed					
		(g)mS	(g)sM	gmS	gS	mS	sM
Predicted	(g)mS	4	3	0	0	0	0
	(g)sM	16	56	2	0	4	37
	gmS	2	0	0	2	0	0
	gS	0	0	0	0	0	0
	mS	2	0	0	0	2	0
	sM	1	19	1	0	0	18

Table B5. Quantitative Folk SLOO CV error matrix.

		Observed					
		(g)mS	(g)sM	gmS	gS	mS	sM
Predicted	(g)mS	0	1	0	2	0	0
	(g)sM	21	66	3	0	6	39
	gmS	1	0	0	0	0	0
	gS	0	0	0	0	0	0
	mS	0	0	0	0	0	0
	sM	3	11	0	0	0	16

Table B6. Categorical simplified Folk LOO CV error matrix.

		Observed				
		gmS	gS	M	mS	sM
Predicted	gmS	0	0	0	0	0
	gS	0	2	0	0	0
	M	0	0	0	0	0
	mS	0	0	0	18	7
	sM	3	0	1	16	142

Table B7. Categorical simplified Folk SLOO CV error matrix.

Predicted \ Observed	gmS	gS	M	mS	sM
gmS	0	0	0	0	0
gS	0	0	0	0	0
M	0	0	0	0	0
mS	0	2	0	1	2
sM	3	0	1	33	147

Table B8. Quantitative simplified Folk LOO CV error matrix.

Predicted \ Observed	gmS	gS	mS	sM
gmS	0	2	2	0
gS	0	0	0	0
mS	0	0	8	3
sM	3	0	21	130

Table B9. Quantitative simplified Folk SLOO CV error matrix.

Predicted \ Observed	gmS	gS	mS	sM
gmS	0	0	1	0
gS	0	0	0	0
mS	0	2	0	1
sM	3	0	30	132

Table B10. Categorical muddy/sandy LOO CV error matrix.

Predicted \ Observed	Muddy	Sandy
Muddy	142	19
Sandy	8	20

Table B11. Categorical muddy/sandy SLOO CV error matrix.

		Observed	
		Muddy	Sandy
Predicted	Muddy	145	36
	Sandy	5	3

Table B12. Quantitative muddy/sandy LOO CV error matrix.

		Observed	
		Muddy	Sandy
Predicted	Muddy	131	24
	Sandy	2	12

Table B13. Quantitative muddy/sandy SLOO CV error matrix.

		Observed	
		Muddy	Sandy
Predicted	Muddy	132	33
	Sandy	1	3

Table B14. Categorical muddy/sandy SLOO CV error matrix with threshold of occurrence set to 0.18.

		Observed	
		Present	Absent
Predicted	Present	30	47
	Absent	9	103

Table B15. Coarse substrate SR-LOO CV error matrix with threshold of occurrence set to 0.27.

		Observed	
		Present	Absent
Predicted	Present	44	57
	Absent	15	176

B.6 References

Collins, A. K., Hannah, C. G., and Greenberg, D. 2011. Validation of a high resolution modelling system for tides in the Canadian Arctic Archipelago. Canadian Technical Report of Hydrographic and Ocean Sciences, 273. Bedford Institute of Oceanography, Dartmouth.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., *et al.* 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1): 27–46.

Downie, A.-L., Dove, D., Westhead, K., Diesing, M., Green, S. L., and Cooper, R. 2016. Semi-automated mapping of rock in the North Sea. JNCC Report, 592. JNCC, Peterborough.

Gottschalk, T. K., Aue, B., Hotes, S., and Ekschmitt, K. 2011. Influence of grain size on species–habitat models. *Ecological Modelling*, 222(18): 3403–3412.

Hughes Clarke, J. E., Iwanowska, K. K., Parrott, R., Duffy, G., Lamplugh, M., and Griffin, J. 2008. Inter-calibrating multi-source, multi-platform backscatter data sets to assist in compiling regional sediment type maps: Bay of Fundy. *In* Proceedings of the Canadian Hydrographic Conference and National Surveyors Conference 2008. Victoria.

Lacharité, M., Brown, C. J., and Gazzola, V. 2018. Multisource multibeam backscatter data: Developing a strategy for the production of benthic habitat maps using semi-

automated seafloor classification methods. *Marine Geophysical Research*, 39(1–2): 307–322.

Lechner, A. M., Langford, W. T., Jones, S. D., Bekessy, S. A., and Gordon, A. 2012. Investigating species–environment relationships at multiple scales: Differentiating between intrinsic scale and the modifiable areal unit problem. *Ecological Complexity*, 11: 91–102.

Lurton, X., and Lamarche, G. (Eds). 2015. Backscatter measurements by seafloor-mapping sonars. Guidelines and recommendations. GeoHab Backscatter Working Group.

Misiuk, B., Lecours, V., and Bell, T. 2018. A multiscale approach to mapping seabed sediments. *PLoS ONE*, 13(2): e0193647.