



OPEN ACCESS

EDITED BY

Chris Taylor,
National Centers for Coastal Ocean Science
(NOAA), United States

REVIEWED BY

Giorgio Castellani,
National Research Council (CNR), Italy
Markus Diesing,
Geological Survey of Norway, Norway

*CORRESPONDENCE

Emily Sklar

✉ emily.sklar@dal.ca

RECEIVED 03 October 2023

ACCEPTED 01 February 2024

PUBLISHED 16 February 2024

CITATION

Sklar E, Bushuev E, Misiuk B,
Labbé-Morissette G and Brown CJ
(2024) Seafloor morphology and
substrate mapping in the Gulf of
St Lawrence, Canada, using machine
learning approaches.
Front. Mar. Sci. 11:1306396.
doi: 10.3389/fmars.2024.1306396

COPYRIGHT

© 2024 Sklar, Bushuev, Misiuk,
Labbé-Morissette and Brown. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Seafloor morphology and substrate mapping in the Gulf of St Lawrence, Canada, using machine learning approaches

Emily Sklar^{1*}, Esther Bushuev¹, Benjamin Misiuk^{1,2,3},
Guillaume Labbé-Morissette⁴ and Craig J. Brown¹

¹Seascape Ecology and Mapping Lab, Department of Oceanography, Dalhousie University, Halifax, NS, Canada, ²Department of Geography, Memorial University of Newfoundland, St. John's, NL, Canada, ³Department of Earth Sciences, Memorial University of Newfoundland, St. John's, NL, Canada, ⁴Research & Development, Interdisciplinary Centre for the Development of Ocean Mapping (CIDCO), Rimouski, QC, Canada

Detailed maps of seafloor substrata and morphology can act as valuable proxies for predicting and understanding the distributions of benthic communities and are important for guiding conservation initiatives. High resolution acoustic remote sensing data can facilitate the production of detailed seafloor maps, but are cost-prohibitive to collect and not widely available. In the absence of targeted high resolution data, global bathymetric data of a lower resolution, combined with legacy seafloor sampling data, can provide an alternative for generating maps of seafloor substrate and morphology. Here we apply regression random forest to legacy data in the Gulf of St Lawrence, Canada, to generate a map of seabed sediment distribution. We further apply k-means clustering to a principal component analysis output to identify seafloor morphology classes from the GEBCO bathymetric grid. The morphology classification identified most morphological features but could not discriminate valleys and canyons. The random forest results were in line with previous sediment mapping work done in the area, but a large proportion of zero values skewed the explained variance. In both models, improvements may be possible with the introduction of more predictor variables. These models prove useful for generating regional seafloor maps that may be used for future management and conservation applications.

KEYWORDS

random forest, principal component analysis, K-means (KM) clustering, seafloor morphology, seafloor mapping, sediment, bathymetry, legacy data

1 Introduction

Coastal marine ecosystems face significant anthropogenic pressures due to the goods and services they provide and the ease with which these may be accessed (Halpern et al., 2008). Globally, over 90% of international trade occurs via shipping (Mudryk et al., 2021), and these routes must pass through coastal ecosystems to get to port. Coastal fisheries also represent a significant impact, with fishing effort in many coastal regions increasing over the years and leading to habitat degradation caused by the gear deployed (Stewart et al., 2010). The Gulf of St. Lawrence (GSL) on the east coast of Canada is one of many such ecosystems facing these anthropogenic threats. It includes busy shipping routes that connect the Atlantic Ocean to the Great Lakes, and supports total fisheries landings valued at over 788 million dollars (DFO, 2021). Because of its prominent location, lucrative fishing grounds, and access it provides to inland North America, the GSL is considered one of the most important parts of the Canadian coast (Loring and Nota, 1973). Informed and sustainable management is critical to ensure ecological health of the GSL and continued use of these resources.

Seafloor sediment composition and morphology can act as effective surrogates for understanding biodiversity patterns at the seafloor which can be valuable for marine conservation planning (McArthur et al., 2010; Tecchiato et al., 2015; Wilson et al., 2018). Sessile filter feeders such as sponges often require hard substrate on which to anchor, while fine-grained sediments provide habitat for burrowers. Seafloor morphology may correlate strongly with hydrodynamics and sedimentation (Tecchiato et al., 2015; Miramontes et al., 2019), and can be a useful proxy for understanding spatial patterns of fauna and seafloor substrates. For instance, steep-sloped features such as seamounts and submarine canyons propagate internal tides, which act as efficient mechanisms for food transport (Mohn et al., 2014). Particulate organic matter may be transported along the faces of such morphological features due to the interactions between topography and internal tides, which enables settlement of suspension-feeding cold-water corals.

Morphological features can be defined by the values of their bathymetric derivatives (e.g., slope degree, terrain ruggedness, bathymetric variance, etc.). These derivatives can be calculated from readily available digital elevation models (DEMs). The General Bathymetric Chart of the Oceans (GEBCO; GEBCO Compilation Group, 2021) has created a global bathymetric grid using a variety of datasets. The grid is primarily derived from satellite altimetry measurements, but also includes other modern datasets such as multibeam echosounder data. Legacy datasets are incorporated as well, and in the GSL these legacy bathymetric datasets have been collected for over a century and consist of lead line data and single beam echosounder data (CHS, 2022). As a result, the morphology of the GSL seafloor is generally understood, but a morphological classification scheme has yet to be applied.

Over a period of 10 years, Loring and Nota (1973) collected sediment samples and seafloor images to produce a map of the sediment distribution of the GSL. This interpretation required an expert depth of localised knowledge on the geological history and hydrodynamics of the region (Diesing et al., 2014). The Loring &

Nota interpretation considered local hydrodynamics and bathymetry, but the physical oceanographic models (e.g., Wang et al., 2018; Li et al., 2021) and DEMs (e.g., GEBCO Compilation Group, 2021) available today were not available to them at the time. Their sediment map is discretised, with transitions between classes presented as solid boundaries, manually interpreted from the discrete physical seafloor point sediment samples. In reality, sediment boundaries may be gradational rather than abrupt. Mapping sediment as a continuous variable instead of a discrete one may allow for more accurate estimates of species distributions when it is used as a predictor (Wilson et al., 2018). Modern quantitative modelling approaches can offer an alternative way to produce continuous coverage maps depicting gradational changes in substrate parameters, and may additionally be used to infer sediment composition in areas where ground truth validation is not available (Misiuk and Brown, 2024). This can be achieved by using geospatial models that treat substrate parameters as a response variable to be predicted using continuous coverage environmental data sets (e.g. bathymetry, seabed morphology, physical oceanographic parameters such as current speed and direction, etc.). Machine learning algorithms are increasingly applied to predict sediment parameters with high accuracy (e.g., Diesing et al., 2014; Stephens and Diesing, 2014; Misiuk et al., 2019). Such approaches also show promise for classification of seafloor morphology using bathymetric data and derivatives (e.g., Jasiewicz and Stepinski, 2013; Maschmeyer et al., 2019).

The goals of this paper is to 1) apply a machine learning methodology to predict sediment grain size fractions observed in the GSL legacy dataset using a modern suite of environmental predictors and generate continuous maps of grain size distributions, and 2) apply a morphological classification scheme to the seafloor of the GSL.

2 Materials and methods

2.1 Study area

The GSL (Figure 1) is bordered by the Canadian provinces of Quebec, New Brunswick, Nova Scotia, Prince Edward Island, and Newfoundland & Labrador. It connects the St. Lawrence Estuary to the Northwest Atlantic Ocean via the Cabot Strait and the Strait of Belle Isle on either side of Newfoundland. The GSL covers a total area of 240,000 km² and contains 3,553 km³ of water (Dufour and Ouellet, 2007). It has an average depth of 152 m, with ~25% of the area shallower than 75 m (Environment Canada, 2013). The deepest part of the gulf is the Laurentian Channel, which begins at the St. Lawrence Estuary and flows out into the Atlantic via the Cabot Strait. As the channel reaches the Cabot Strait, it attains a maximum depth of approximately 540 m (GEBCO Compilation Group, 2021).

The Laurentian Channel divides the GSL into northern and southern regions. To the south lies a plateau with an average depth of 80 m (Dufour and Ouellet, 2007). On this plateau is Prince Edward Island and Les îles-De-La-Madeleine – a small island chain under Quebec jurisdiction. To the northwest is the St. Lawrence Estuary, divided into upper and lower sections, with the lower section considered as part of the gulf. Eastward from the estuary,

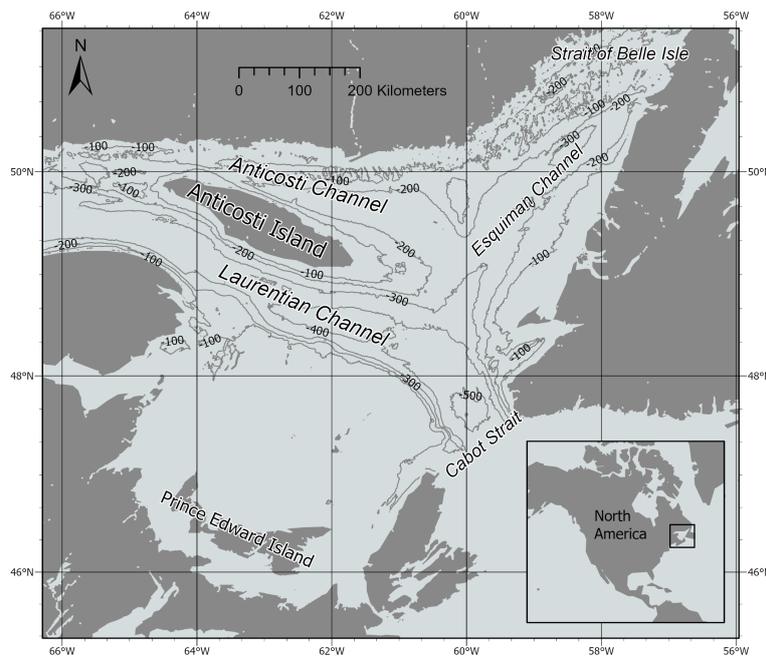


FIGURE 1
Study site - Gulf of St. Lawrence, Canada. Contour lines are drawn at 100 m intervals.

Anticosti Island splits the channel into the Laurentian Channel and the Anticosti Channel. The Anticosti Channel connects to the Esquimaux Channel to the southeast of the island. The Esquimaux Channel enters the gulf from the Strait of Belle Isle between Newfoundland and Labrador.

The GSL was covered by the Laurentide Ice Sheet (LIS) until approximately 11,500 years ago (Casse et al., 2017). The rapid retreat of the LIS at this time heavily influenced changes in sediment deposition due to increased meltwater input into the GSL. In the Laurentian Channel, a >450 m thick Quaternary sedimentary succession has developed primarily due to high sedimentation rates brought on by the LIS retreat and its associated meltwater (Casse et al., 2017). The channel itself developed along a faulted contact zone before being modified by glacial erosion during the Quaternary period (Loring and Nota, 1973; Casse et al., 2017). The deepening of the channel at the Cabot Strait is likely due to forced narrowing by the surrounding terrestrial landforms increasing and deepening the glacial erosion process (Loring and Nota, 1973).

The north shore of the Gulf, from the lower St. Lawrence Estuary to the Strait of Belle Isle, is lined with submarine valleys and canyons (Loring and Nota, 1973; Normandeau et al., 2015). Many of these are pre-Paleozoic in origin, but were further carved by ice while the Esquimaux and Anticosti channels were undergoing a transition from fluvial valleys to glacial troughs (Loring and Nota, 1973). The predominance of canyons and valleys along the north shore, especially compared to their near absence on the other shores of the Gulf, can be attributed to a steep slope gradient from shore to seafloor as well as the high volume of sediment that was transported southward during the deglaciation that occurred 11,500 years ago (Normandeau et al., 2015).

2.2 Predictor variables

Seventeen predictor variables were used in the random forest and are provided in Table 1. These predictors were selected based on previous sediment modelling work (Diesing et al., 2014; Stephens and Diesing, 2014; Misiuk et al., 2018; 2019; Bushuev et al., 2023).

Bathymetric data were obtained from GEBCO. GEBCO is a global repository of bathymetric data compiled as part of the Nippon Foundation-GEBCO Seabed 2030 Project, which has the goal of mapping the entire seafloor by 2030 (GEBCO Compilation Group, 2021). The GEBCO 2021 data are gridded at 15 arc-second resolution, equivalent to approximately 450 m at the equator. The grid was downloaded for the extent of the GSL and projected to a custom Lambert Conformal Conic projection with a central meridian longitude of 61°W and standard parallel latitudes of 46°N and 50°N. Eight morphometric derivatives were calculated from the bathymetry data using the Benthic Terrain Modeller (BTM) toolbox in ArcGIS Pro 2.7.3 (Walbridge et al., 2018; Goes et al., 2019). These derivatives are bathymetric mean, bathymetric variance, standardised broadscale and finescale bathymetric position indices (BPI), eastness, northness, ruggedness, and slope (Table 1). BPI provides information on relative vertical position of a focal cell (Walbridge et al., 2018). The BPI radius values were selected based on previous work done with the BTM toolbox (Walbridge et al., 2018). Bathymetric mean and variance required a neighbourhood size for calculation. The neighbourhood size is the maximum number of cells used in the calculation of a terrain attribute (Misiuk et al., 2021). The neighbourhood size for variance and mean were selected for consistency with the spatial scale of the BPI measurements. Radius values and neighbourhood sizes used to calculate predictors are provided in Table 1.

TABLE 1 Predictor variables used in random forest sediment models.

Variable	Calculation Method	Data Source
Bathymetry (Supplementary Figure 4)	N/A	GEBCO Compilation Group, 2021
Bathymetric mean (Supplementary Figure 5)	Benthic Terrain Modeler (neighbourhood size = 25)	GEBCO Compilation Group, 2021
Bathymetric variance (Supplementary Figure 6)	Benthic Terrain Modeler (neighbourhood size = 25)	GEBCO Compilation Group, 2021
Standardised broadscale BPI (Supplementary Figure 7)	Benthic Terrain Modeler (inner radius=25, outer radius=250)	GEBCO Compilation Group, 2021
Standardised finescale BPI (Supplementary Figure 8)	Benthic Terrain Modeler (inner radius=3, outer radius=25)	GEBCO Compilation Group, 2021
Statistical aspect (eastness; Supplementary Figure 9)	Benthic Terrain Modeler	GEBCO Compilation Group, 2021
Statistical aspect (northness; Supplementary Figure 10)	Benthic Terrain Modeler	GEBCO Compilation Group, 2021
Terrain ruggedness (Vector ruggedness measure; Supplementary Figure 11)	Benthic Terrain Modeler (neighbourhood size = 25)	GEBCO Compilation Group, 2021
Slope degree (Supplementary Figure 12)	Benthic Terrain Modeler	GEBCO Compilation Group, 2021
Mean bottom current direction 1990-2015 (Supplementary Figure 13)	Inverse Distance Weighting	Wang et al., 2018
Mean bottom current magnitude 1990-2015 (Supplementary Figure 13)	Inverse Distance Weighting	Wang et al., 2018
Mean seafloor shear velocity 2002-2005 (Supplementary Figure 14)	Inverse Distance Weighting	Li et al., 2021
Maximum seafloor shear velocity 2002-2005 (Supplementary Figure 15)	Inverse Distance Weighting	Li et al., 2021
Maximum wave power 2002-2005 (Supplementary Figure 16)	Inverse Distance Weighting	Li et al., 2021
Mean wave power 2002-2005 (Supplementary Figure 17)	Inverse Distance Weighting	Li et al., 2021
Euclidean distance to mainland and islands > 5000 km ² (Supplementary Figure 18)	Spatial Analyst toolbox	Runfola et al., 2020
Euclidean distance to islands < 5000 km ² (Supplementary Figure 19)	Spatial Analyst toolbox	Runfola et al., 2020

N/A, Not Applicable.

Physical oceanographic predictor variables were interpolated using inverse distance weighting to match the resolution of the GEBCO grid. For benthic current magnitude and direction, data were obtained from the Bedford Institute of Oceanography North Atlantic Model (BNAM; Wang et al., 2018). BNAM is used by the Department of Fisheries and Oceans Canada (DFO) to model oceanographic conditions through space and time in the Northwest Atlantic. BNAM predictions were provided at a nominal resolution of 1/12° (approximately 6500 m). For seafloor shear velocity and wave power, model predictions were provided at a nominal resolution of 1/10° (approximately 7800 m; Li et al., 2021).

Euclidean distance from the coast was calculated as a potential proxy for terrestrial sediment input. Distance layers were calculated for both the mainland coast and from islands smaller than 5,000 km² based on the assumption sediment input from larger islands may differ substantially from smaller islands. Prince Edward Island and Anticosti are larger than 5,000 km² and were therefore considered “mainland”. The two Euclidean distance variables were calculated at the same resolution as the GEBCO grid from a

polygon shapefile of shorelines obtained from Runfola et al. (2020) using the Spatial Analyst toolbox in ArcGIS Pro 2.7.3.

2.3 Surficial sediment data

The original dataset used by Loring and Nota (1973) consisted of approximately 1500 sediment samples that were collected throughout the GSL using a 0.1 m² Van Veen grab (Loring and Nota, 1973). Of the original dataset, records containing grain size composition for 223 samples were recovered at the Bedford Institute of Oceanography (Figure 2). Data from the remaining of the original 1500 samples could not be located. Of the data recovered, 200 points contained non-zero values for mud, 214 non-zero values for sand, and 50 non-zero values for gravel.

Spatial autocorrelation for each of the three grain size classes was assessed using Global Moran's I (Moran, 1950). For a set of locations and an associated attribute, this statistic tests the null hypothesis that the attribute in question is randomly distributed by calculating Moran's Index with the model residuals. Moran's Index ranges

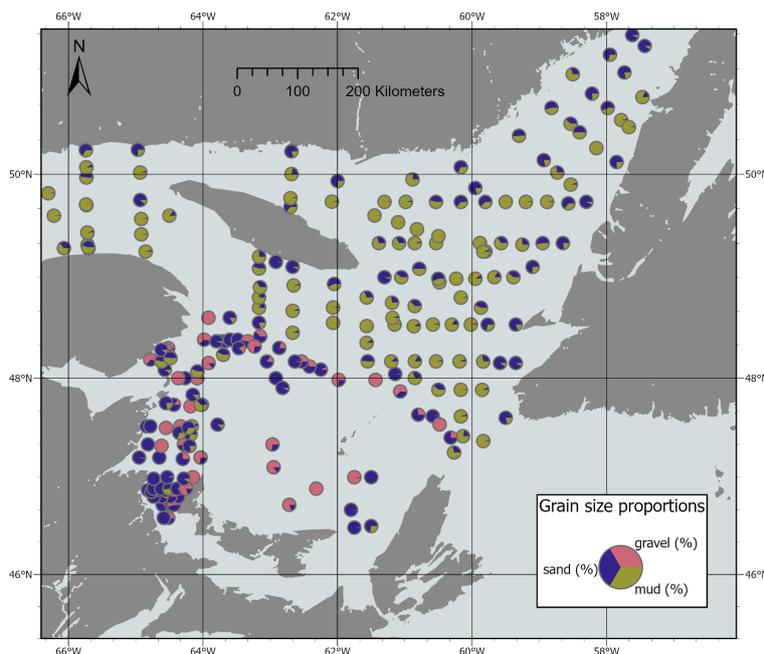


FIGURE 2

Distribution of sediment grain size samples from Loring and Nota (1973) that were recovered from the Bedford Institute of Oceanography. Location points are presented as pie charts that indicate the grain size fractions of the given sample.

from -1 to 1. If the value is close to -1, then the spatial distribution of the data is dispersed. If the value is close to 1, then the spatial distribution of the data is clustered. If the value is close to 0, then the spatial distribution of the data is random. The significance of the index value is determined by a z score and p value. The Global Moran's I test was carried out using the Spatial Autocorrelation tool in the Spatial Analyst toolbox of ArcGIS Pro 2.7.3.

2.4 Sediment modelling using random forest

Legacy sediment data were used to model each of the three grain size fractions using regression random forest to produce a broadscale map of sediment distribution in the gulf. Random forest is a machine learning algorithm that generates multiple classification or regression trees with a randomly selected subset of the provided predictor variables at each node in the tree (Breiman, 2001, 2002). Individual trees are additionally grown using bootstrapped samples of the training data to reduce the variance of the aggregated predictions, and the data not drawn for a given tree (the "out-of-bag" [OOB] observations) may be used to validate the model predictions. This is accomplished by aggregating predictions over all the OOB samples once the full model has been trained. A regression random forest was chosen to model sediment as it is suitable for interpolating large datasets and is robust against issues caused by noisy data and multicollinear or unimportant predictor variables. To generate the random forest model, the randomForest package in R was used (Liaw and Wiener, 2002). Five hundred trees (ntree) and six predictor variables (mtry) at each tree node were used for all three grain size fractions. The ntree value was

selected by plotting the OOB error rate against number of trees used and selecting a value of ntree that corresponded to stabilised OOB error values. The mtry value was selected based on a trial-and-error procedure laid out by Breiman (2002), where multiple values are attempted, beginning with the square root of the total number of predictors, and the testing set error is checked for each attempt. The code used to run the random forest models is provided online (https://github.com/emilysklar/sediment_rf).

Model performance was evaluated using root mean squared error (RMSE) and variance explained of the OOB observations. RMSE calculates the root of the average squared difference between predicted and observed values:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

where y_i and \hat{y}_i are observed and predicted values of the response, respectively. The variance explained is calculated using the ratio of the mean squared error to the variance of the response observations:

$$\% VE = \left(1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}\right) 100$$

Predictor variable importance was evaluated using the mean decrease in residual sum of squares (RSS). The more important a predictor variable is to the model, the more the RSS decreases when it is used at a tree node (Breiman, 2002).

Mud, sand, and gravel percentages of seafloor substrate are compositional, and predicted values at each data point must sum to unity. The additive log-ratio (ALR) transformation was initially

applied to enforce a compositional output by modelling two variables that are the log-ratios between percentages of sand and gravel, and mud and gravel (Stephens and Diesing, 2015). A preponderance of zero observations within the gravel class necessitated imputation of small non-zero values to enable logarithmic transformation for a large proportion (~78%) of data points (Lark et al., 2012). We therefore additionally trialled a separate approach wherein the raw data values are modelled separately for each class and the outputs from these models are optimised to a compositional scale after prediction. The ALR models were outperformed by the optimised model outputs, which were selected for all models presented hereafter. Additional details and comparison are provided in [Supplementary Material S2](#). Code for performing the optimisation is provided online (<https://github.com/benjaminmisiuk/sNet>). After modelling, predicted proportions of mud, sand, and gravel were additionally classified into grain size classes according to Folk (1954).

2.5 Seafloor morphology classification

An automated data-driven approach was used to distinguish morphological features of the GSL. Morphometric features of the bathymetric surface were initially classified using the *r.geomorphon* tool in GrassGIS (Jasiewicz and Stepinski, 2013), yet the classified output contained a large number of data artefacts relict from the GEBCO bathymetry input raster. To obtain a more interpretable and useable morphological classification, unsupervised classification was used to identify objective morphometric features (Bushuev et al., 2023). Principal components analysis (PCA) is an ordination technique used to obtain a lower-dimensional linearly independent set of features from a high-dimensional collinear input (Ismail et al., 2015; Joliffe and Cadima, 2016; Lever et al., 2017). PCA was applied to bathymetry, bathymetric mean, bathymetric variance, broad- and fine-scale BPIs, and slope raster layers using the *RStoolbox* package in R (Leutner et al., 2022; Table 1). The first four principal components

were retained, which accounted for 94.3% of the variance of the input variables. K-means clustering was then performed on the four principal components to yield 10 clusters ($k = 10$). K-means is an unsupervised learning algorithm that partitions a pre-defined number of clusters in such a way that within-cluster variance is minimised to the greatest extent possible (Lloyd, 1982; Malik and Tuckfield, 2019). The elbow method (Thorndike, 1953) was initially attempted to determine what the optimal value of k was, but the results were inconclusive. Trial-and-error was then carried out with the k -means clustering being run multiple times, each time with a different value for k , to determine what the optimal number k value was. Each iteration of the model was assessed by qualitatively comparing the output to the GEBCO bathymetric grid for the area. Code for the PCA k -means clustering procedure is available online (https://github.com/esther-bushuev/morphology_clustering).

The 10 k -means clusters were used to identify eight morphological classes in the GSL (Table 2), based on previous work reclassifying morphological clustering outputs (Iwahashi et al., 2018). Classes were determined based on the definitions provided in the literature, box plots of the distribution of values for each predictor variable at each cluster, and by comparing the model output to an output from the *r.geomorphon* tool. The PCA k -means approach failed to correctly classify canyons and valleys, instead identifying elongated slope features between ridges. The valley class from the *r.geomorphon* output was therefore supplanted into the model output wherever it occurred. The PCA k -means output was retained at all other locations. The output was compared qualitatively to expert interpretation of the bathymetry raster to evaluate the quality of the classification.

3 Results

3.1 Substrate modelling

For all three grain size classes, the Global Moran's I test failed to reject the null hypothesis that the data was randomly distributed. For

TABLE 2 Morphological classifications assigned to the GSL.

Class	Definition	Adapted from
Shallow channel floor	A broad, low-gradient feature within a bathymetric low, flanked by higher and steeper gradient features. Shallower than 375 m.	Dove et al., 2020
Deep channel floor	A broad, low-gradient feature within a bathymetric low, flanked by higher and steeper gradient features. Deeper than 375 m.	Dove et al., 2020
Escarpment	A steep feature of relatively constant slope that separates horizontal or gently sloping features.	IHO, 2019
Slope	A sloped feature, not as steep as an escarpment, that deepens to a point where the gradient decreases.	IHO, 2019
Plane	A flat surface.	Dove et al., 2020
Ridge	A feature of varying complexity that is longer than it is wide, with one steep side.	IHO, 2019; Dove et al., 2020
Shoulder	A transitional feature between horizontal and sloped features, always at the upper edge of the sloped feature.	Jasiewicz and Stepinski, 2013
Footslope	A transitional feature between sloped and horizontal features, always at the lower edge of the sloped feature.	Jasiewicz and Stepinski, 2013
Valley/canyon	An elongated bathymetric low that slopes upward on both sides. Canyons have steeper slope gradients than valleys.	Jasiewicz and Stepinski, 2013; Dove et al., 2020

mud, the Moran's Index value was -0.035 with a p-value of 0.571. For sand, the Moran's Index value was -0.048 with a p-value of 0.422. For gravel, the Moran's Index value was -0.059 with a p-value of 0.311.

The random forest model for mud had the strongest performance, explaining 79.4% of variance in the mud observations (Table 3). Gravel, which contained the lowest number of non-zero values in the dataset, had the weakest performance, with 19.5% of variance explained by the model. Observed and predicted values for each model are provided in Figure 3. The line of best fit for the mud predictions was closest to the $x=y$ line, while gravel was furthest. This indicates that the gravel model residuals were mostly positive for observed values close to 0, and mostly negative for observed values close to 1.

For all three sediment classes modelled, bathymetry, bathymetric mean, and maximum shear velocity were three of the top four most important predictor variables (Figure 4). Broad-scale BPI was in the top four for sand and mud, but for gravel the fourth variable was maximum wave power. Mud percentage was highest when bathymetry values were deeper than approximately 300 m, while gravel and sand percentages were lowest in these areas and highest when bathymetry was shallower than approximately 80 m.

Modelled grain size fraction distributions and Folk classifications are presented in Figure 5. Gravelly sand was the most common Folk class, comprising approximately 31% of the total modelled area (Figure 5D, Table 4). Muddy sandy gravel was the rarest Folk class, covering <0.01% of the total modelled area.

3.2 Morphology classification

Each k-means cluster of the PCA outputs was assigned to a single class except for the plane and escarpment classes, which each comprise two k-means clusters. The two plane clusters plot close together in multidimensional space according to the first three principal components (Figure 6), as do the two escarpment clusters. Interquartile ranges (IQR) of 4 out of the 7 predictor variables additionally overlapped for the two plane clusters (Figure 7). Shallow and deep channel floors also had overlapping IQRs for 4 of 7 predictors, but were retained as separate classes due to their multivariate distance (Figure 6) and the clear division of the clusters at a depth of 375 m. The two escarpment clusters were similar in that their boxplot maxima were higher than any other clusters for all 7 predictor variables, and they often had wider value ranges than any other class (Figure 7).

Shallow channel floor, deep channel floor, and plane classes were defined by low IQRs and a low median value for slope, bathymetric variance, bathymetric standard deviation, finescale BPI, and broad-scale BPI compared to the other classes in the model (Figure 7). These low values imply relatively flat, level features.

TABLE 3 Model validation statistics for each sediment class.

Sediment class	% Variance explained	RMSE
Mud	79.40	0.18
Sand	45.47	0.27
Gravel	19.54	0.26

These classes were differentiated by bathymetry and bathymetric mean. The deep channel floor class was characterised by a bathymetric low with the greatest median depth of any morphology class (407 m). The plane class was a bathymetric high (median cluster depths 57 m and 76 m), and the shallow channel floor was between the deep channel floor and the plane classes (median depth 276 m). The shallow channel floor class was always bordered by morphological classes that, by definition, involve a changing of depth, such as footslopes and slopes (Figure 8).

The ridge cluster is considered a bathymetric high based on the high median value for bathymetry (114 m) and bathymetric mean (111 m; Figure 7). Ridges also have a high median slope value (2.64°) similar to the escarpment clusters (1.96° and 2.84°; Figure 7). Ridges and escarpments are distinguished by bathymetric variance. The two escarpment clusters indicated higher bathymetric variance than any other clusters, while the median bathymetric variance of the ridge classification was lower.

Footslope and shoulder clusters had similar median values for bathymetric variance, bathymetric standard deviation, and slope. Median bathymetry characterised shoulders as bathymetric highs (68 m), with footslopes being deeper (276 m). This is reflected in Figure 8, where shoulders mainly appear along the edge of planes, a bathymetric high, and footslopes appear along the edges of channel floors, which are bathymetric lows.

3.3 Sediment distribution by morphology class

Mud was the dominant grain size class present in the channel floors, with a median value of approximately 87.5% in both shallow and deep channels (Figures 9, 5A). The channel floor classes contained the lowest proportions of gravel and sand out of any morphological class, with median values of 0.45% for gravel on the deep channel floor and 0.33% for the shallow channel floor (Figures 9, 5C). For sand, median percentage was approximately 11.8% for both floor classes (Figures 9, 5B). Gravel had the highest median proportion on planes, with median values of 27.6%. By contrast, planes had the lowest proportion of mud out of all morphological classes, with median values of 10.5%.

The "gravelly sand" Folk class was present in every morphological class of the GSL, except for the two channel floor classes (Figure 10). Gravelly sand was most common on shoulders, ridges, and escarpments. All three of these morphological classes start at a bathymetric high and slope downward on one side. The two channel floor classes were approximately 97% covered by the "slightly gravelly mud" Folk class. In both channel floor classes, the other Folk classes were "gravelly mud" and "slightly gravelly muddy sand".

4 Discussion

4.1 Sediment distribution modelling

Gravel most commonly occurred on the southern plateau, with the random forest model predicting up to approximately 74% gravel

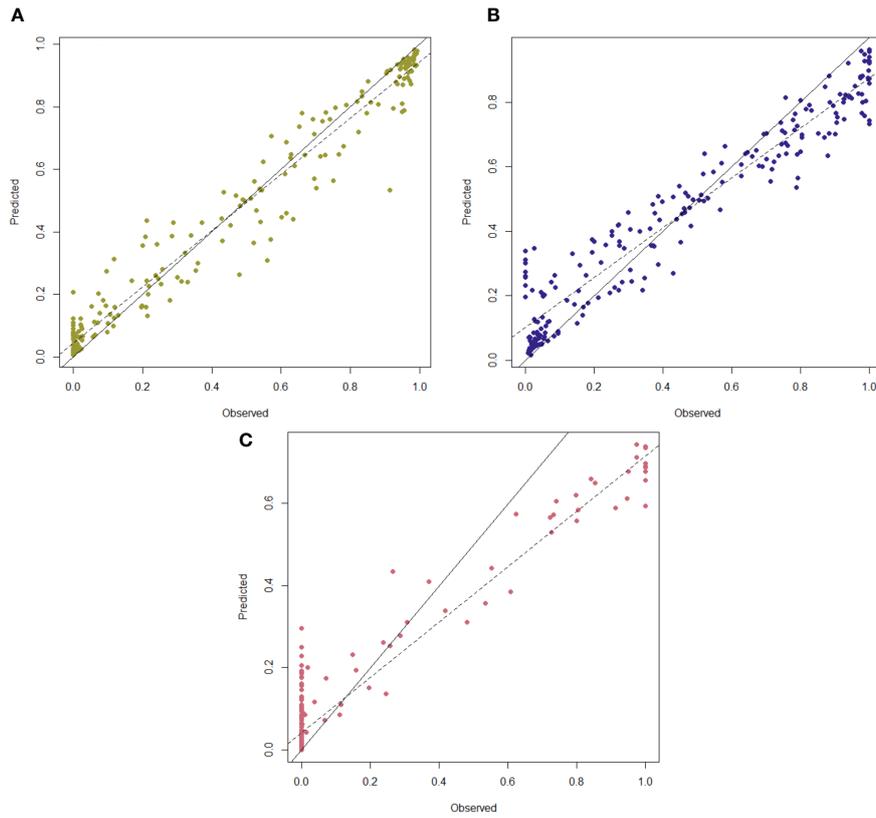


FIGURE 3 Observed and predicted values for the data points in the three random forest sediment models: mud (A), sand (B), and gravel (C). The black line is given by $y=x$, where the predicted and observed values are the same. The dashed line is the line of best fit between observed and predicted values.

in parts of this region. Sand was also predicted here in high proportions, reaching 98% at some locations. Folk classes in the area were mixtures of sand and gravel (Figure 5D). While the data density in the southern plateau is relatively low compared to the rest

of the study area (Figure 2), Loring and Nota (1973) also indicated mixtures of gravel and sand, which they had sampled comprehensively. This provides greater confidence in that region despite the lower number of samples available for modelling.

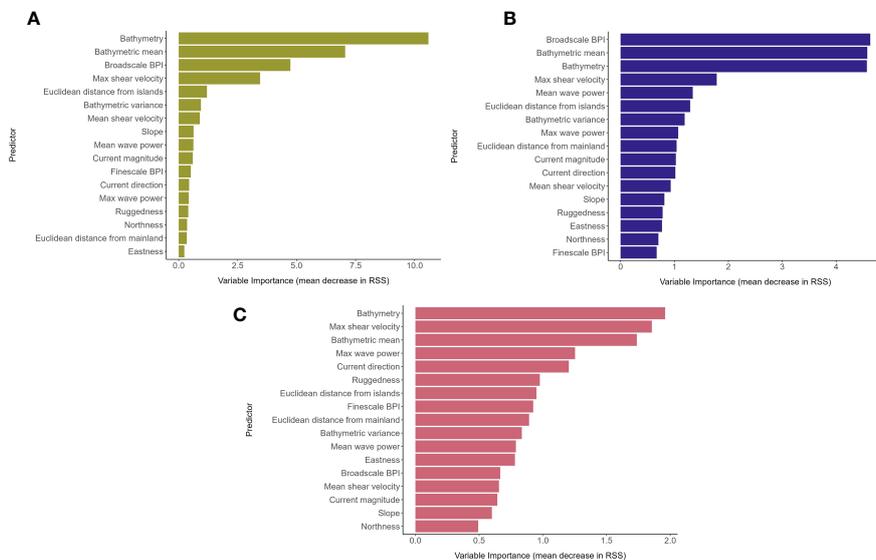


FIGURE 4 Variable importance, presented as mean decrease in residual sum of squares (RSS), for the mud (A), sand (B), and gravel (C) random forest models.

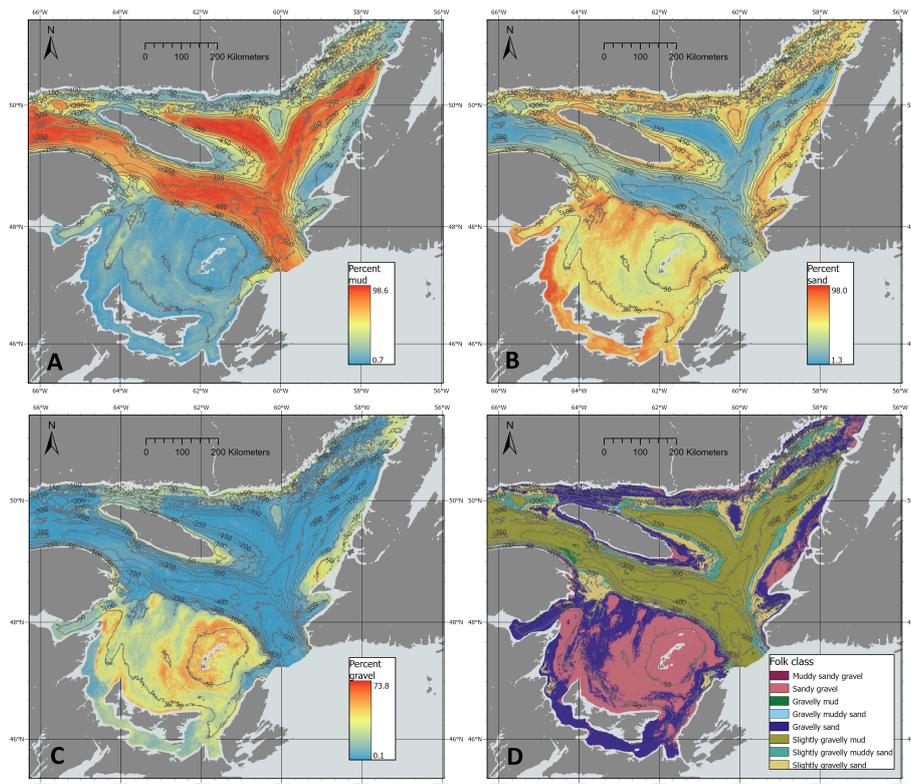


FIGURE 5 Grain size predictions for mud (A), sand (B), and gravel (C) fractions, with Folk classification (D). Contour lines are at 50 m intervals.

The gravel predictions demonstrated the weakest performance of the three grain size models, with a VE of 19.5%. However, gravel also contained only 50 non-zero samples, while the other two sediment types had over 200, and the RMSE for sand was higher than that of gravel. The VE of the gravel predictions is affected by the high proportion of zero values, which lowers the variance of the dataset. The high number of zeros also skewed the spread of residuals, as these data points could only have positive residuals and no negatives (Figure 3).

Maximum seafloor shear velocity was consistently one of the most important predictor variables in all three models. Shear velocity is

known to be influenced by morphology and influences morphology in turn through erosion (Stow et al., 2009; Breitzke et al., 2017). The erosion of morphological features on the seafloor can also contribute to sedimentation rate and thus the sediment class (Stow et al., 2009). In the GSL, the highest values for max shear velocity occurred where the terrain was classified as “plane”, such as the southern plateau, peaking at 0.225 cm/s (Supplementary Figure 13). Areas classified as planes were almost always comprised of sand and gravel Folk classes (e.g., gravelly sand, sandy gravel, etc.). High current velocities directly influence shear velocities, and only sediments of larger grain sizes are deposited under these conditions (Stow et al., 2009). Max shear velocity was reduced in the channels identified here, with values as low as 0.017 cm/s in some places. Slower velocities allow for smaller grain sizes to settle (Stow et al., 2009), corroborating random forest models here that predicted up to 98% mud composition in the channels.

The presence of hard substrate is an important consideration from a benthic ecological perspective, which may support different benthic assemblages (e.g. primarily epifauna) compared to unconsolidated, finer-grained substrata which are dominated by infaunal species (Harris and Baker, 2011). Data used here for sediment grain size models were obtained by physical sampling (e.g., grabs), which limited model predictions to size fractions smaller than cobble. Loring and Nota (1973) noted the presence of outcropping bedrock in the GSL but there were insufficient data on the presence of hard substrata (e.g., exposed bedrock, boulders) for geospatial modelling in our analyses. Future work could aim to model hard substrates in the GSL by obtaining ground truth seafloor imagery data, potentially coupled

TABLE 4 Total area and percent cover for each Folk class.

Folk class	Area (km ²)	Percent of total area
Muddy sandy gravel	3.22	1.64x10 ⁻³
Sandy gravel	41992.03	21.38
Gravelly mud	817.72	0.42
Gravelly muddy sand	835.72	0.43
Gravelly sand	60371.49	30.74
Slightly gravelly mud	58270.88	29.67
Slightly gravelly muddy sand	11907.03	6.06
Slightly gravelly sand	22202.99	11.31

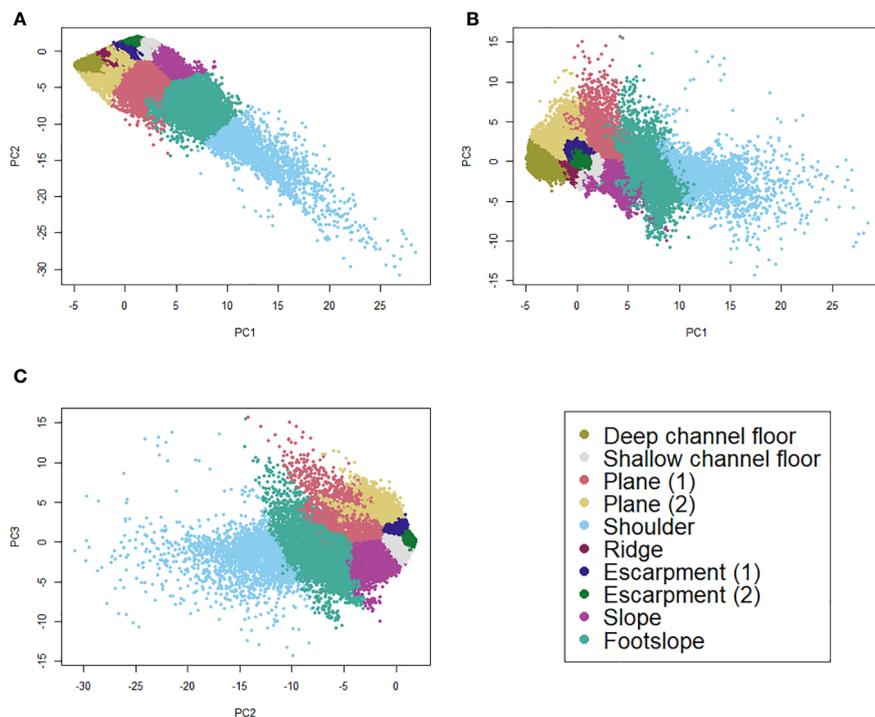


FIGURE 6

The first 3 principal components (PC1, PC2, PC3) of seafloor morphology variables, coloured according to k-means cluster, in a three dimensional plot. (A-C) represent the same plot from three different angles.

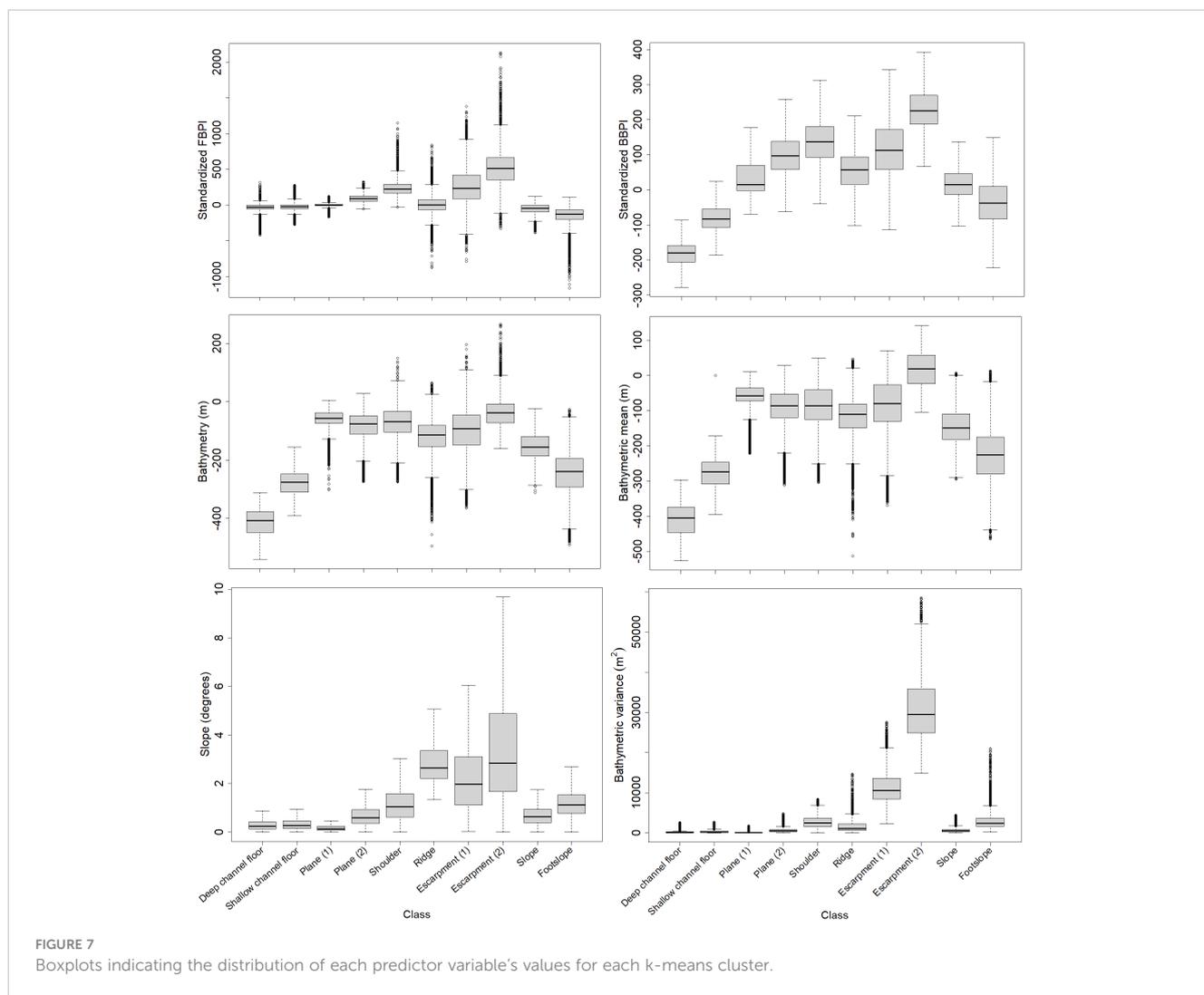
with additional remote sensing data such as acoustic backscatter. Presence/absence models could then be used to predict hard substrata, and outputs could be integrated with sediment predictions presented here to provide a more comprehensive understanding of substrate distribution in the GSL (e.g., Misiuk et al., 2019).

Legacy data used here were collected over 50 years ago; it is therefore important to consider the possibility of temporal variability in the sediment distribution of the GSL. Geological processes are slow, with sediment accumulation rates in the ocean typically measured at rates of metres per thousand years (Sadler, 1981; Gingerich, 2021). However, anthropogenic disturbance may modify the benthos over shorter time periods (Houziaux et al., 2011; Oberle et al., 2016). Trawlers may dispose of collected sediment in different locations to facilitate future trawling activities (Houziaux et al., 2011). Larger clasts, such as gravels, may thus be replaced over time by finer-grained sediment such as sand. Trawling may also resuspend fine-grained sediment and induce off-shelf sediment transport from continental shelves on par with the volumes transported by river-supplied sediment (Oberle et al., 2016). Dredging may also be conducted, either to maintain proper depth to ensure safe passage of vessels or to collect materials such as gravel and sand for construction (de Groot, 1986). This leads to mass displacement and removal of sediments; in the Canadian Atlantic region, which includes but is not limited to the GSL, 5.7 million m^3/yr of sand and gravel were extracted between 1979 and 1983 (de Groot, 1986). Anthropogenic impacts such as these were not considered in our sediment models, and are often neglected when modelling sediment distribution and transport (Oberle et al., 2016).

Correspondence between sediment type and morphology predictions were observed spatially over the GSL. Folk classes that were mixtures of sand and gravel were predominantly associated with planes. Channel floors (both shallow and deep) were dominated by high percentages of mud and the “slightly gravelly mud” Folk class. Sloped bathymetric highs, such as shoulders and ridges, contained high percentages of sand. Many of the predictor variables in the grain size models provided morphological information pertaining to the shape of the seafloor (e.g., broadscale BPI, bathymetric variance). One of the most important predictors, maximum shear velocity, is not a measure of seafloor morphology but is heavily influenced by it. Previous studies have identified the importance of morphological information in sediment distribution models (Stephens and Diesing, 2015; Misiuk et al., 2018; Wilson et al., 2018; Misiuk et al., 2019), but this trend has not previously been formally identified in the GSL with respect to morphological classification.

4.2 Morphology classification

DEMs are frequently used to apply morphological classification schemes to the seafloor and to land, often with the aid of machine learning (e.g., Ismail et al., 2015; Iwahashi et al., 2018; Maschmeyer et al., 2019; Barbarella et al., 2021; Lin et al., 2021), but the application of PCA followed by k-means is relatively new in the context of morphology classification. Expert interpretation may be performed to classify DEMs according to morphology or



geomorphology, but can often be both subjective and time-consuming (Barbarella et al., 2019, 2021). Results from this paper demonstrate the efficiency of the PCA k-means clustering method as an objective alternative to expert morphology interpretation and classification.

In the case of the GSL, 10 k-means clusters reduced to 8 classes provided the best results based on localised knowledge of the study area. Reclassifying outputs from an unsupervised k-means clustering in this way, where clusters are grouped together based on predictor statistics, has been done successfully prior to this study (Iwahashi et al., 2018). Because of varying global seafloor morphological complexity, 10 clusters may not be universally applicable for morphological classification schemes. Classes should be selected with care based on peer-reviewed definitions of morphological features and knowledge of the local geological setting, which includes the formational processes that the morphology has undergone in the geologic past. In the GSL, much of the geology is based on the glacial history of the region. Knowledge of the past ice cover in the area and how it evolved explains many features, such as the north shore submarine canyons and the size and orientation of the channels (Loring and Nota, 1973).

Submarine valleys and canyons are both characterised by elongated bathymetric lows that slope upward on both sides to bathymetric highs (Table 3). The valleys and canyons within the study area were often associated with a ridge feature on either side, bordering the bathymetric high of the valley/canyon (Figure 8 inset). The final version of the PCA k-means model was unable to detect valleys and canyons, instead classifying them as alternating slopes and ridges. However, submarine canyons on the north shore of the GSL are well-described (Loring and Nota, 1973; Normandeau et al., 2015) and are visible in bathymetry raster images of the area. It is important to correctly identify these features, as they can act as channels for sediment and nutrient transport into deep water and are therefore crucial to benthic communities (Kenchington et al., 2014). To correct for the inability of the model to detect valleys and canyons, the valley classification from r.geomorphons supplanted the classification from the k-means model. One predictor variable that can provide the model with the necessary information to detect the valley/canyon class is curvature. Curvature can be used to describe concave or convex features and has successfully been used in seafloor classification before (Mitchell and Clarke, 1994; Ismail et al., 2015; Koop et al., 2021). It also has a strong correlation to submarine canyon morphology (Goff, 2001). Introducing

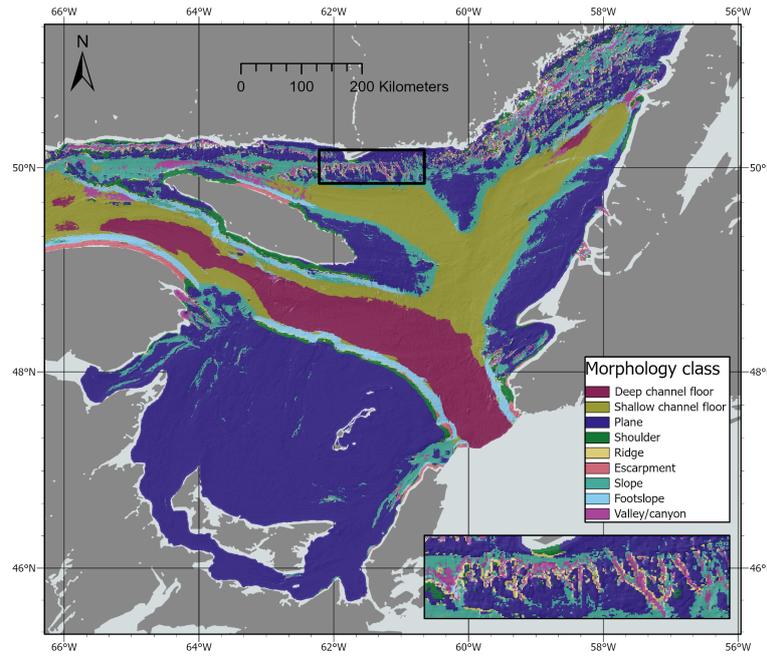


FIGURE 8 Morphology of the GSL. Inset shows a section of the system of canyons and valleys that make up much of the north shore.

curvature to the principal component analysis allowed for the k-means clustering to detect valleys and canyons. However, when trialled here, curvature also amplified data artefacts and classified many “valleys/canyons” that were the size of a single cell scattered throughout the entire GSL. For this reason, curvature was removed from the model. The bathymetric data available for the GSL exists as a mosaic of different data collection methods at different resolutions (GEBCO Compilation Group, 2021) and a shortcoming of deriving curvature from such a DEM is that compilation artefacts may propagate to bathymetric derivatives (Iwahashi et al., 2018). The use of curvature may work in an area where bathymetric data collection is more

uniform and from a single source. In other parts of the global ocean, seafloor curvature might also prove useful for classifying morphological features defined as a bathymetric high surrounded by bathymetric lows, such as cones, knolls, or mounds (Dove et al., 2020).

5 Conclusions

Sediment grain size models based on legacy substrate data were developed here for the entire GSL by utilising a machine learning framework. This enabled quantitative geospatial predictions of grain

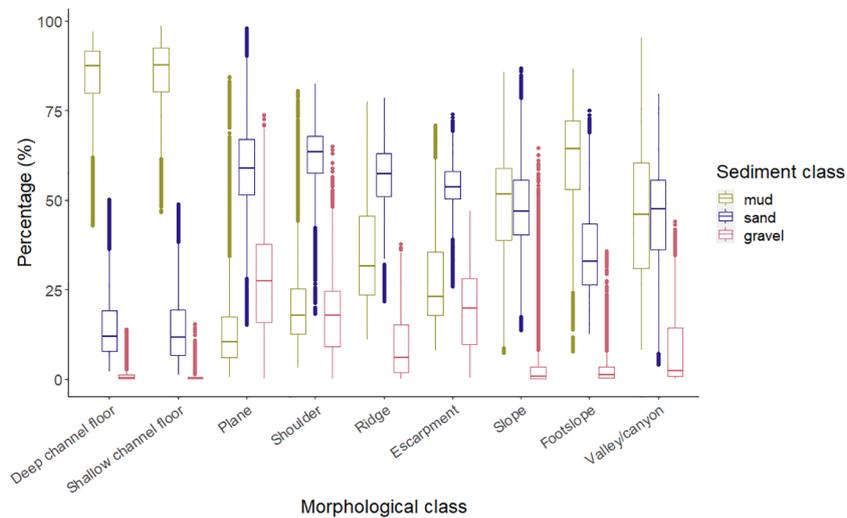
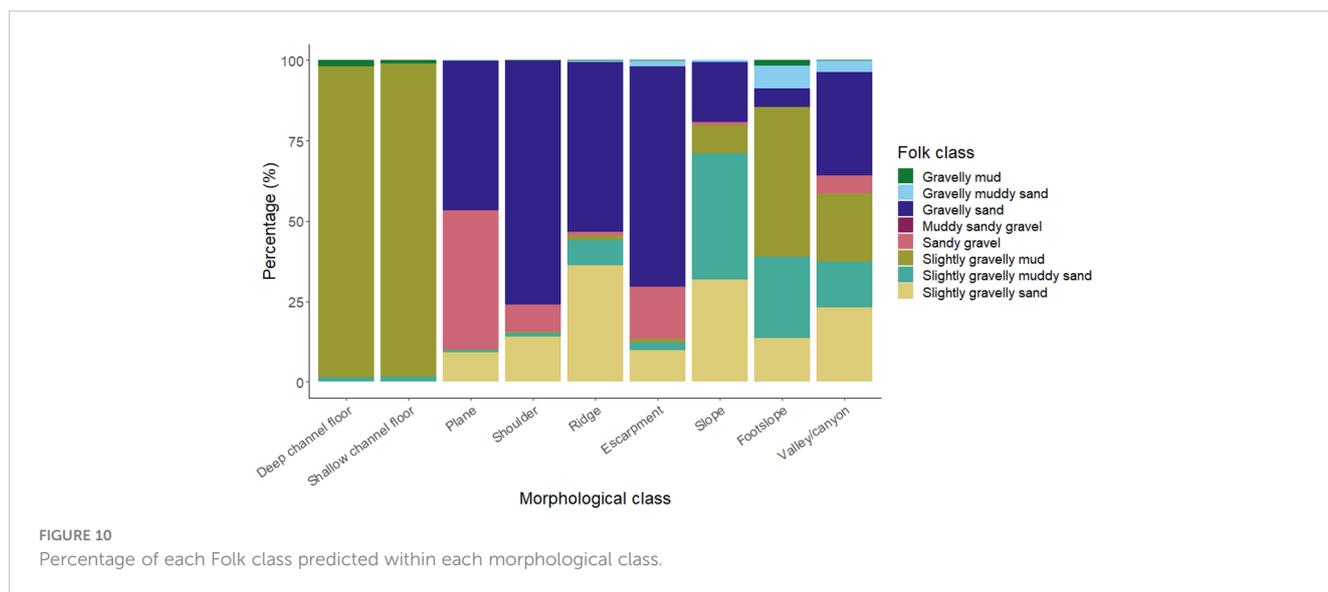


FIGURE 9 Box plots depicting the distribution of grain size fractions for each morphological class.



size fractions for the first time in this region, including at areas of scarce ground-truth data. Results from an objective and data-driven morphological classification demonstrated apparent correspondence with predicted sediment classes. Channels were predicted to primarily comprise muds, while planes are likely composed of sand or a sand/gravel mix. The use of *r.geomorphons* was effective at supplanting the PCA k-means morphology classification where the model failed to correctly identify submarine canyons and valleys. The PCA k-means approach provided a fast and objective method to classifying submarine morphology of the GSL, however, some expert interpretation was still required to assign class labels and assess the feasibility of the model output.

Data availability statement

Publicly available datasets were analysed in this study. This data can be found here: https://www.gebco.net/data_and_products/gridded_bathymetry_data/#global, https://ed.marine-geo.canada.ca/index_e.php, <https://search.open.canada.ca/opendata/>; contains information licensed under the Open Government Licence – Canada. <https://open.canada.ca/en/open-government-licence-canada>.

Author contributions

ES: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. EB: Conceptualization, Methodology, Writing – review & editing. BM: Methodology, Writing – review & editing. GM: Conceptualization, Funding acquisition, Writing – review & editing. CB: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by MITACS grant number FR58582, under the project “Geospatial prediction, prioritisation and impact assessment of ghost fishing gear in the Gulf of St Lawrence”, in partnership with CIDCO.

Acknowledgments

The authors would like to thank the Bedford Institute of Oceanography for their help in locating the sediment data, and Vicki Gazzola for assistance with visualisation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2024.1306396/full#supplementary-material>

References

- Barbarella, M., Cuomo, A., Di Benedetto, A., Fiani, M., and Guida, D. (2019). Topographic base maps from remote sensing data for engineering geomorphological modelling: An application on coastal mediterranean landscape. *Geosciences* 9, 500. doi: 10.3390/geosciences9120500
- Barbarella, M., Di Benedetto, A., and Fiani, M. (2021). Application of supervised machine learning technique on LiDAR data for monitoring coastal land evolution. *Remote Sens* 13, 4782. doi: 10.3390/rs13234782
- Breiman, L. (2001). Random forests. *Mach. Learn* 45, 5–32. doi: 10.1023/A:1010933404324
- Breiman, L. (2002) *Manual on setting up, using, and understanding random forests V3.1* (Berkeley (CA: Department of Statistics at UC Berkeley). Available online at: https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf (Accessed 2022 June 6).
- Breitzke, M., Wiles, E., Krockner, R., Watkeys, M. K., and Jokat, W. (2017). Seafloor morphology in the Mozambique Channel: evidence for long-term persistent bottom-current flow and deep-reaching eddy activity. *Mar. Geophys. Res.* 38, 241–269. doi: 10.1007/s11001-017-9322-7
- Bushuev, E., Misiuk, B., Gazzola, V., and Brown, C. J. (2023). “Machine learning approaches for broad-scale characterization of seafloor geology on the Northwest Atlantic Shelf,” in *The Atlantic Geoscience Society (AGS) 49th Colloquium and Annual Meeting*, Truro, NS, February 2023. 3–4.
- Casse, M., Montero-Serrano, J. C., and St-Onge, G. (2017). Influence of the Laurentide Ice Sheet and relative sea-level changes on sediment dynamics in the Estuary and Gulf of St. Lawrence since last deglaciation. *Boreas* 46, 541–561. doi: 10.1111/bor.12230
- CHS. (2022). *Canadian Hydrographic Service Non-Navigational (NONNA) Bathymetric Data* (Ottawa (ON: Department of Fisheries and Oceans Canada). Available online at: <https://open.canada.ca/data/en/dataset/d3881c4c-650d-4070-bf9b-1e00aabf0a1d> (Accessed 2022 Oct 11).
- de Groot, S. J. (1986). Marine sand and gravel extraction in the North Atlantic and its potential environmental impact, with emphasis on the North Sea. *Ocean Manage.* 10, 21–36. doi: 10.1016/0302-184X(86)90004-1
- DFO. (2021). *Seafisheries landed value by regio* (Ottawa (ON: Department of Fisheries and Oceans Canada). Available online at: <https://www.dfo-mpo.gc.ca/stats/commercial/land-debarq/sea-maritimes/s2020av-eng.htm> (Accessed 2022 June 21).
- 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. *Cont. Shelf Res.* 84, 107–119. doi: 10.1016/j.csr.2014.05.004
- Dove, D., Nanson, R., Bjarnadóttir, L. R., Guinan, J., Gafeira, J., Post, A., et al. (2020). *A two-part seabed geomorphology classification scheme (v.2). Part 1: morphology features glossary* (Nottingham, UK: British Geological Survey), 22 p. doi: 10.5281/ZENODO.4075248
- Dufour, R., and Ouellet, P. (2007). *Estuary and Gulf of St. Lawrence marine ecosystem overview and assessment report*. vii+112 p. Mont-Joli, QC: Fisheries and Oceans Canada
- Environment Canada (2013). *Canada's National Programme of Action for the protection of the marine environment from land-based activities (NPA)* (Ottawa (ON: Environment Canada), 125 p.
- Folk, R. L. (1954). The distinction between grain size and mineral composition in sedimentary rock nomenclature. *J. Geol* 52, 344–359. doi: 10.1086/626171.
- GEBCO Compilation Group. (2021). *GEBCO 2021 Grid* (Paris (FR: Intergovernmental Oceanographic Commission). doi: 10.5285/c6612cbe-50b3-0cfe-e053-6c86abc09f8f
- Gingerich, P. D. (2021). Rates of geological processes. *Earth-Sci Rev.* 220, 103723. doi: 10.1016/j.earscirev.2021.103723
- Goes, E. R., Brown, C. J., and Araújo, T. C. (2019). Geomorphological classification of the benthic structures on a tropical continental shelf. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00047
- Goff, J. A. (2001). Quantitative classification of canyon systems on continental slopes and a possible relationship to slope curvature. *Geophys. Res. Lett.* 28, 4359–4362. doi: 10.1029/2001GL013300.
- Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., et al. (2008). A global map of human impact on marine ecosystems. *Science* 319, 948–952. doi: 10.1126/science.1149345
- P. T. Harris and E. Baker (Eds.) (2011). *Seafloor geomorphology as benthic habitat: GeoHAB atlas of seafloor geomorphic features and benthic habitats* (Waltham (MA: Elsevier Inc).
- Houziaux, J. S., Fettweis, M., Francken, F., and Van Lancker, V. (2011). Historic, (1900) seafloor composition in the Belgian–Dutch part of the North Sea: A reconstruction based on calibrated visual sediment descriptions. *Cont. Shelf Res.* 31, 1043–1056. doi: 10.1016/j.csr.2011.03.010
- IHO. (2019). *Standardization of undersea feature names: Guidelines, proposal form, terminology. Edition 4.2.0. IHO Publication B-6*. 43 p. Monaco: Intergovernmental Oceanographic Commission
- Ismail, K., Huvenne, V. A. I., and Masson, D. G. (2015). Objective automated classification technique for marine landscape mapping in submarine canyons. *Mar. Geol* 362, 17–32. doi: 10.1016/j.margeo.2015.01.006
- Iwahashi, J., Kamiya, I., Matsuoka, M., and Yamazaki, D. (2018). Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification. *Prog. Earth Planet Sci.* 5, 1. doi: 10.1186/s40645-017-0157-2
- Jasiewicz, J., and Stepinski, T. F. (2013). Geomorphons - a pattern recognition approach to classification and mapping of landforms. *Geomorphology* 182, 147–156. doi: 10.1016/j.geomorph.2012.11.005
- Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: A review and recent developments. *Phil. Trans. R. Soc. A* 374, 20150202. doi: 10.1098/rsta.2015.0202
- Kenchington, E. L., Cogswell, A. T., MacIsaac, K. G., Beazley, L., Law, B. A., and Kenchington, T. J. (2014). Limited depth zonation among bathyal epibenthic megafauna of the Gully submarine canyon, northwest Atlantic. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 104, 67–82. doi: 10.1016/j.dsr2.2013.08.016
- Koop, L., Snellen, M., and Simons, D. G. (2021). An object-based image analysis approach using bathymetry and bathymetric derivatives to classify the seafloor. *Geosciences* 11, 26. doi: 10.3390/geosciences11020045
- 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. *Sediment Geol* 281, 35–49. doi: 10.1016/j.sedgeo.2012.07.009
- Leutner, B., Horning, N., Schwalb-Willmann, J., and Hijmans, R. (2022) *RStoolbox: Tools for remote sensing data analysis v. 0.3.0*. Available online at: <https://cran.r-project.org/package=RStoolbox>.
- Lever, J., Krzywinski, M., and Altman, N. (2017). Principal component analysis. *Nat. Methods* 14, 641–642. doi: 10.1038/nmeth.4346.
- Li, M. Z., Wu, Y., Hannah, C. G., and Perrie, W. A. (2021). Seabed disturbance and sediment mobility due to tidal current and waves on the continental shelves of Canada. *Can. J. Earth Sci.* 58, 1209–1232. doi: 10.1139/cjes-2020-0139
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.
- Lin, S., Chen, N., and He, Z. (2021). Automatic landform recognition from the perspective of watershed spatial structure based on digital elevation models. *Remote Sens* 13, 3926. doi: 10.3390/rs13193926
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inf Theory* 28, 129–137. doi: 10.1109/ITIT.1982.1056489.
- Loring, D. H., and Nota, D. J. G. (1973). *Morphology and sediments of the Gulf of St. Lawrence* (Ottawa (ON: Fisheries and Marine Service), 147 p.
- Malik, A., and Tuckfield, B. (2019). *Applied unsupervised learning with R: Uncover hidden relationships and patterns with k-means clustering, hierarchical clustering, and PCA* (Birmingham (UK: Packt Publishing), 296 p.
- Maschmeyer, C. H., White, S. M., Dreyer, B. M., and Clague, D. A. (2019). High-Silica Lava Morphology at Ocean Spreading Ridges: Machine-Learning Seafloor Classification at Alarcon Rise. *Geosci J.* 9, 21. doi: 10.3390/geosciences9060245
- McArthur, M. A., Brooke, B. P., Przeslawski, R., Ryan, D. A., Lucieer, V. L., Nichol, S., et al. (2010). On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuar. Coast. Shelf Sci.* 88, 21–32. doi: 10.1016/j.ecss.2010.03.003
- Miramontes, E., Garreau, P., Caillaud, M., Jouet, G., Pellen, R., Hernández-Molina, F. J., et al. (2019). Contourite distribution and bottom currents in the NW Mediterranean Sea: Coupling seafloor geomorphology and hydrodynamic modelling. *Geomorphology* 333, 43–60. doi: 10.1016/j.geomorph.2019.02.030
- Misiuk, B., Diesing, M., Aitken, A., Brown, C. J., Etinger, E. N., and Bell, T. (2019). A spatially explicit comparison of quantitative and categorical modelling approaches for mapping seabed sediments using random forest. *Geosci J.* 9, 34. doi: 10.3390/geosciences9060254
- Misiuk, B., Lecours, V., and Bell, T. (2018). A multiscale approach to mapping seabed sediments. *PLoS One* 13, e0193647. doi: 10.1371/journal.pone.0193647
- Misiuk, B., Lecours, V., Dolan, M. F. J., and Robert, K. (2021). Evaluating the suitability of multi-scale terrain attribute calculation approaches for seabed mapping applications. *Mar. Geod* 44, 327–385. doi: 10.1080/01490419.2021.1925789
- Misiuk, B., and Brown, C. J. (2024). Benthic habitat mapping: A review of three decades of mapping biological patterns on the seafloor. *Estuarine, Coastal and Shelf Science*, 296, 108599. doi: <https://doi.org/10.1016/j.ecss.2023.108599>
- Mitchell, N. C., and Clarke, J. E. H. (1994). Classification of seafloor geology using multibeam sonar data from the Scotian Shelf. *Mar. Geol* 121, 143–160. doi: 10.1016/0025-3227(94)90027-2.
- Mohn, C., Rengstorf, A., White, M., Duineveld, G., Mienis, F., Soetaert, K., et al. (2014). Linking benthic hydrodynamics and cold-water coral occurrences: A high-resolution model study at three cold-water coral provinces in the NE Atlantic. *Prog. Oceanogr* 122, 92–104. doi: 10.1016/j.pcean.2013.12.003
- Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37, 17–123. doi: 10.1093/biomet/37.1-2.17

- Mudryk, L. R., Dawson, J., Howell, S. E. L., Derksen, C., Zagon, T. A., and Brady, M. (2021). Impact of 1, 2 and 4 °C of global warming on ship navigation in the Canadian Arctic. *Nat. Clim Change* 11, 673–679. doi: 10.1038/s41558-021-01087-6.
- Normandeau, A., Lajeunesse, P., and St-Onge, G. (2015). Submarine canyons and channels in the Lower St. Lawrence Estuary (Eastern Canada): Morphology, classification, and recent sediment dynamics. *Geomorphology* 241, 1–18. doi: 10.1016/j.geomorph.2015.03.023
- Oberle, F. K. J., Storlazzi, C. D., and Hanebuth, T. J. J. (2016). What a drag: Quantifying the global impact of chronic bottom trawling on continental shelf sediment. *J. Mar. Syst.* 159, 109–119. doi: 10.1016/j.jmarsys.2015.12.007
- Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., et al (2020). geoBoundaries: A global database of political administrative boundaries. *PLoS ONE*, 15(4), e0231866. doi: <https://doi.org/10.1371/journal.pone.0231866>
- Sadler, P. M. (1981). Sediment accumulation rates and the completeness of stratigraphic sections. *J. Geol.* 89, 569–584. doi: 10.1086/628623.
- Stephens, D., and Diesing, M. (2014). A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. *PLoS One* 9, e93950. doi: 10.1371/journal.pone.0093950
- Stephens, D., and Diesing, M. (2015). Towards quantitative spatial models of seabed sediment composition. *PLoS One* 10, e0142502. doi: 10.1371/journal.pone.0142502
- Stewart, K. R., Lewison, R. L., Dunn, D. C., Bjorkland, R. H., Kelez, S., Halpin, P. N., et al. (2010). Characterizing fishing effort and spatial extent of coastal fisheries. *PLoS One* 5, e14451. doi: 10.1371/journal.pone.0014451
- Stow, D. A. V., Hernández-Molina, F. J., Llave, E., Sayago-Gil, M., Díaz del Río, V., and Branson, A. (2009). Bedform-velocity matrix: The estimation of bottom current velocity from bedform observations. *Geology* 37, 327–330. doi: 10.1130/G25259A.1
- Tecchiato, S., Collins, L., Parnum, I., and Stevens, A. (2015). The influence of geomorphology and sedimentary processes on benthic habitat distribution and littoral sediment dynamics: Geraldton, Western Australia. *Mar. Geol.* 359, 148–162. doi: 10.1016/j.margeo.2014.10.005
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika* 18, 267–276. doi: 10.1007/BF02289263
- Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. J. (2018). Unified geomorphological analysis workflows with Benthic Terrain Modeler. *Geosciences* 8, 24. doi: 10.3390/geosciences8030094
- Wang, Z., Lu, Y., Greenan, B., Brickman, D., and DeTracey, B. (2018). BNAM: An eddy resolving North Atlantic Ocean model to support ocean monitoring. *Can. Tech Rep. Hydrogr Ocean Sci.* 327, vii + 18.
- Wilson, R. J., Speirs, D. C., Sabatino, A., and Heath, M. R. (2018). A synthetic map of the north-west European Shelf sedimentary environment for applications in marine science. *Earth Syst. Sci. Data* 10, 109–130. doi: 10.5194/essd-10-109-2018