REGIONAL HABITAT MAPPING

Standard Approaches to Synthesizing, Visualizing, and Disseminating High-Resolution Geophysical Data to Advance Benthic Habitat Mapping in the Wind Energy Areas of the Northeast

Prepared for:

MASSACHUSETTS CLEAN ENERGY CENTER®

Submitted to:
Nils Bolgen
Massachusetts Clean Energy Center
63 Franklin St. 3rd Floor
Boston, MA 02110
NBolgen@masscec.com

Submitted by:

INSPIRE ENVIRONMENTAL

INSPIRE Environmental
513 Broadway, Suite 314
Newport, Rhode Island 02840

Final Report
April 28, 2023

Table of Contents

Page

Appendices

List of Tables

List of Figures

List of Acronyms

AUC	area under curve
BOEM	Bureau of Ocean Energy Management
CMECS	Coastal and Marine Ecological Classification Standard
COP	Construction and Operations Plan
INSPIRE	INSPIRE Environmental
IT	information technology
MaxEnt	Maximum Entropy
MDE	Marine Data Exchange
NAMERA	Northwest Atlantic Marine Ecoregional Assessment
NOAA	National Oceanic and Atmospheric Administration
NOS	National Ocean Service
NROC	Northeast Regional Ocean Council
Portal	Northeast Ocean Data Portal
ROC	receiver-operating characteristic curves
RWSC	Regional Wildlife Science Collaborative
TNC	The Nature Conservancy
WEA	Wind Energy Area

1.0 INTRODUCTION

Since the first offshore wind farm, the Block Island Wind Farm, became operational in 2016, the Northeast region has been a pioneering force for offshore wind energy development in the United States (US). With more than 15 offshore wind farms projects anticipated to be built by 2026 (reviewed by Degraer et al. 2020), there is a substantial need to characterize the distribution of benthic habitats in designated lease areas and along proposed cable routes. Understanding benthic habitat distribution is necessary not only for engineering suitability and planning but also to properly assess the value that these habitats provide to ecosystem services, specifically supporting economically important commercial and recreational fisheries (Costanza et al. 1997; Liquete et al. 2013). In particular, hard bottom habitats (e.g., gravel, cobble, boulder) are limited in their spatial distribution and known to support a diverse and abundant assemblage of regionally important biological resources, including American lobsters (*Homarus americanus*) (Wahle and Steneck 1991), longfin squid (*Loligo pealei*) (Griswold and Prezioso 1981; Roper et al. 1984), and Atlantic cod (*Gadus morhua*) (Fahay et al. 1999). Understanding the distribution of these habitats across the region, not just within areas proposed for wind development, will contextualize the proposed activities and support cumulative impacts assessment and development of hypotheses for future research and mapping efforts.

Additionally, while government agencies provide general guidance and recommendations to wind developers regarding the collection of geophysical and benthic assessment data (BOEM 2019, 2020; NOAA Fisheries 2021), no standard protocols exist for sharing or reporting these data, such as standard file and data types or detailed data sharing agreements. Standards are vital to facilitating knowledge exchange and to establishing more stable, coherent, and predictable analytical environments that allow stakeholders to evaluate how complex systems change over time (Jackson and Barbrow 2015). Although all wind developers collect seafloor data for proposed projects, a lack of standard protocols can impact timelines and decision-making processes as government agencies and stakeholders cannot readily access these data for review. These accessibility challenges are often related to inconsistent data formatting and file types or, for geophysical data, proprietary information that results in restricted access until wind developers can review more thoroughly. Standardized protocols will provide resource management agencies, industry entities, and other stakeholders with a predictable set of tools to ensure the informed and efficient development and review of future offshore wind projects.

In the context of offshore wind development, standard protocols should focus on survey data required by the Bureau of Ocean Energy Management (BOEM) prior to submission of a Construction and Operations Plan (COP), which includes geophysical data and in-situ sediment data that are collected, respectively, during geophysical and benthic assessment surveys. A pilot project was therefore conducted by INSPIRE Environmental (INSPIRE) and the Northeast Regional Ocean Council (NROC) to address this need for standard protocols when mapping and disseminating habitat distributions. The project identified recommended best practices for: (1) integrating geophysical and in-situ sediment data to map benthic habitat distributions; and (2) streamlining the public dissemination of high-resolution geospatial data to numerous stakeholders on a vetted and established forum, i.e., the Northeast Ocean Data Portal (Portal). Accomplishing the proposed goals will aid in characterizing benthic environments that may be affected by wind development, ensuring regional sediment products are compatible with existing mapping standards, and enhancing data exploration techniques to better serve stakeholders and future research pursued by the Regional Wildlife Science Collaborative (RWSC).

To better serve the needs of entities and stakeholders in the Northeastern US, INSPIRE and NROC have conducted several activities to date. For instance, the project team originally sought to develop a standard means of classifying benthic habitat distributions using high-resolution geophysical and benthic assessment data from wind developers and publicly accessible sources in the region. Although initial discussions with wind developers aided in crafting a data sharing memorandum of understanding (MOU) template (Appendix A) for future consideration, data formatting and sharing restrictions prevented accessibility to wind developer data, resulting in the strict use of publicly available data sources for creating regional mapping products under this pilot project. The Seafloor Habitat Data Work Group, a collection of federal and state regulators as well as offshore wind personnel with expertise in offshore wind regulatory processes, seafloor imagery, and interactive data exploration, was established and has met throughout the project's duration to inform and recommend data sources and methods for the pilot project. Such activities include defining the viability and expected outputs of the pilot project (June 2021 project update) as well as comments on project updates and requested features or research tasks during RWSC Habitat & Ecosystem Subcommittee meetings (August – October 2022).

1.1 Objectives

Given the critical importance of benthic habitat distribution to the engineering design of offshore wind developments and the ecological, economic, and social resources of the region, a shared understanding of these habitats by all stakeholders, such as developers, regulators, and fishers, is paramount. The overall goal of this project was to develop a standardized approach for integrating geophysical and benthic assessment data into a habitat data product and disseminating to regional stakeholders on the Portal, NROC's central data repository that facilitates marine decision-making activities, to support responsible wind development and natural resource protection. The results of this project will establish a framework for the best use of these data to produce regional-scale habitat mapping products that are most relevant to all stakeholders. These data will represent pre-construction conditions and can be used as baseline data against which post-construction monitoring and impact data collections can be compared.

The overall objectives of the project were to:

- Elicit input and guidance from the multiple stakeholders involved with the use of the natural resources associated with the Massachusetts and Rhode Island Wind Energy Areas (WEA) to vet habitat mapping products and use of the Portal for data exploration and analysis (Task 1).

- Use current data to develop a standard approach to integrate geophysical data with benthic assessment data (i.e., sediment profile images and plan view images, grab sample results) to classify a gradient of benthic habitat types found across the Northeastern US (Task 2).
- Develop a standard approach for delivering interactive habitat data products to the Portal for use by offshore wind development stakeholders (Task 3).
- Conduct a series of research and development tasks to guide future research efforts pursued by the RWSC:
    - Create a streamlined method to supply data to the Portal that can also be used by other entities and groups; and work with the multiple stakeholders to disseminate and display benthic habitat data products for a variety of potential uses (Tasks 4 & 5).
    - Identify and recommend (i) standards for data delivery formats and metadata for ground-truth data for benthic habitat mapping and (ii) procedures for future data sharing and periodic updates of the regional sediment model (Tasks 4 & 5).

2.0 METHODS

2.1 Data Collection and Formatting

A desktop study was conducted to collect existing and publicly available datasets within the designated study area off the Northeastern US (Figure 2-1). Search queries were limited to georeferenced (i.e., spatially defined) datasets, specifically bathymetric raster products for geophysical data and raw sediment observations for benthic assessment data, across all study years to maximize data coverage (Appendix B). Point-based sediment observations were preferred over other popular geospatial file types (e.g., vector polygons, raster imagery) to exclude interpolated sediment results. Raw sediment assessment results were primarily collected through in-situ sampling efforts and imagery, including sediment profile/ plan view imaging, sediment grabs, and video transects.

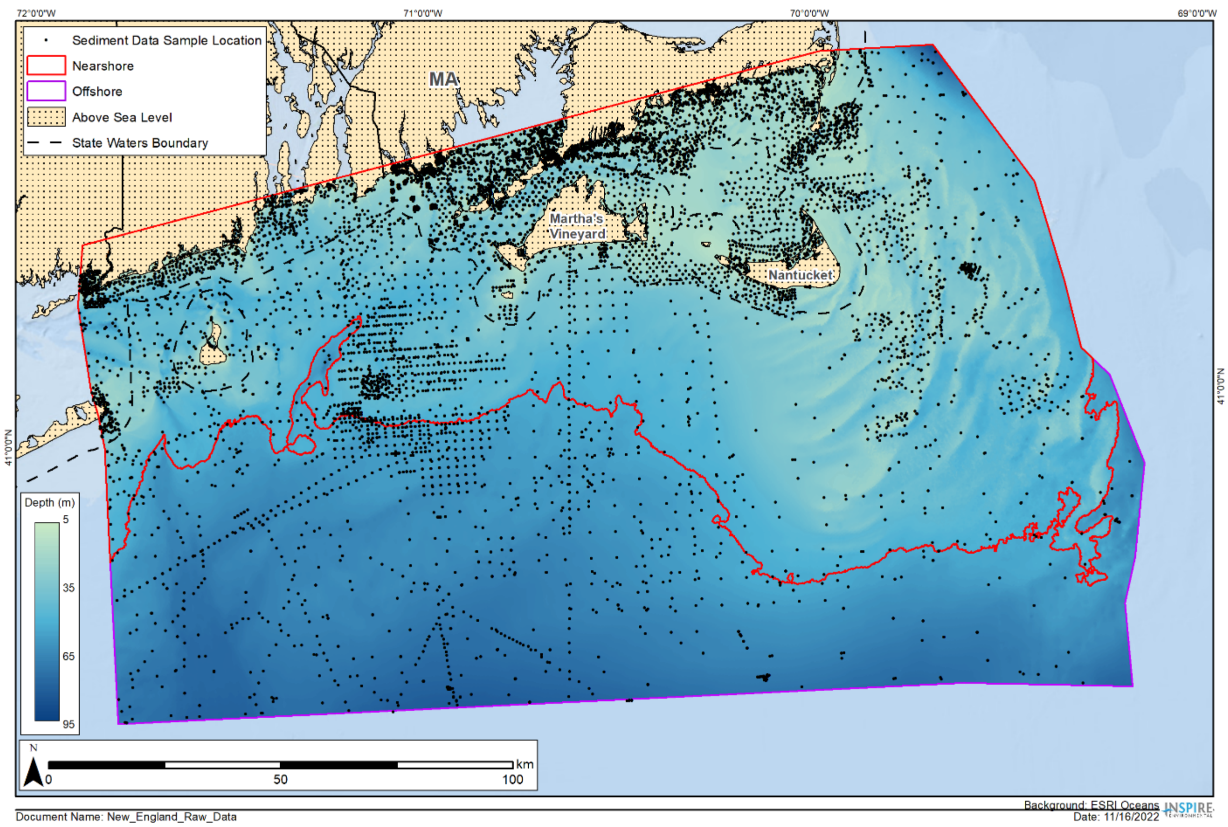


Figure 2-1. Data collected within the study area off the Northeastern United States in the waters south of Rhode Island (RI) and Massachusetts (MA). Bathymetric data are visually depicted using a color gradient in the background whereas in-situ sediment sampling data are shown with black dots in the foreground. Sample locations and bathymetric features occurred more frequently in shallow water, which biased the analysis. As such, "nearshore" and "offshore" areas (using the 45 m bathymetric contour as a divider) were established to enhance the analysis (see Section 2.2.3).

Collected data were then processed based on their geospatial data structures to aid in modeling workflows and identify recommendations for data preparation. For benthic assessment datasets, raw sediment results were reclassified using Substrate Group and Substrate Subgroup classifications defined by the Coastal and Marine Ecological Classification Standard (CMECS; FGDC 2012). For instance, sediment results presented as grain sizes were converted to CMECS sediment classes using grain size descriptor crosswalks (Table 2-1; Wentworth 1922; FGDC 2012). CMECS sediment class definitions were also used to reclassify sediment observations presented as proportions of each major Substrate Group.

High-resolution geophysical raster data files (i.e., 2-m resolution, NOAA National Ocean Service [NOS]) were down-sampled to a resolution of 8-m to balance the competing needs of data quality and computational performance (Appendix C). Preliminary investigations confirmed that a decrease from the initially targeted 4-m (see June 2021 project update) to 8-m resolution reduced computing time and preserved patterns in predicted results (Figure 2-2). Processed geophysical data were finally merged into a single raster that covered the study area through a process known as “mosaicking” (Appendix D).

Table 2-1. Sediment Grain Size Classification Descriptors for Wentworth (1922) and the Coastal and Marine Ecological Classification Standard (CMECS; FGDC 2012)

Wentworth (1922)			CMECS (FGDC 2012)[1]		
Phi Size (Φ)	**Size Range (mm)**	**Size Class**	**Substrate Group (Subgroup)[2]**	**Grain Size (mm)**	**Class Sizes (phi)**
			Gravel[3]	2 to <4,096	-1 to <-12
<-8	>256	Boulder	(Boulder)	256 to <4,096	-8 to <-12
-7 to -8	128 to 256	Cobble	(Cobble)	64 to <256	-6 to <-8
-6 to -7	64 to 128	Cobble			
-5 to -6	32 to 64	Very coarse pebble	(Pebble)	4 to <64	-1 to <-6
-4 to -5	16 to 32	Coarse pebble			
-3 to -4	8 to 16	Medium pebble			
-2 to -3	4 to 8	Fine pebble			
-1 to -2	2 to 4	Very fine pebble	(Granule)	2 to <4	-1 to <-2
			Sand	0.0625 to <2	4 to <-1
0 to -1	1 to 2	Very coarse sand	(Very Coarse Sand)	1 to <2	0 to <-1
1 to 0	0.5 to 1	Coarse sand	(Coarse Sand)	0.5 to <1	1 to <0
2 to 1	0.25 to 0.5	Medium sand	(Medium Sand)	0.25 to <0.5	2 to <1
3 to 2	0.125 to 0.25	Fine sand	(Fine Sand)	0.125 to <0.25	3 to <2
4 to 3	0.0625 to 0.125	Very find sand	(Very Fine Sand)	0.0625 to <0.125	4 to <3
>4	<0.0625	Silt/clay	**Mud**	<0.0625	>4
			Silt	0.004 to <0.0625	>4 to 8
			Clay	<0.004	>8

[1] CMECS uses the term Mud to describe all particles smaller than sand (less than 0.0625 mm).
[2] Values in parentheses represent Subgroups of the overarching Substrate Group (e.g., Boulder is a Subgroup of the Substrate Group Gravel).
[3] The term Gravel is used to describe all rock fragment particles that are 2 mm or larger.

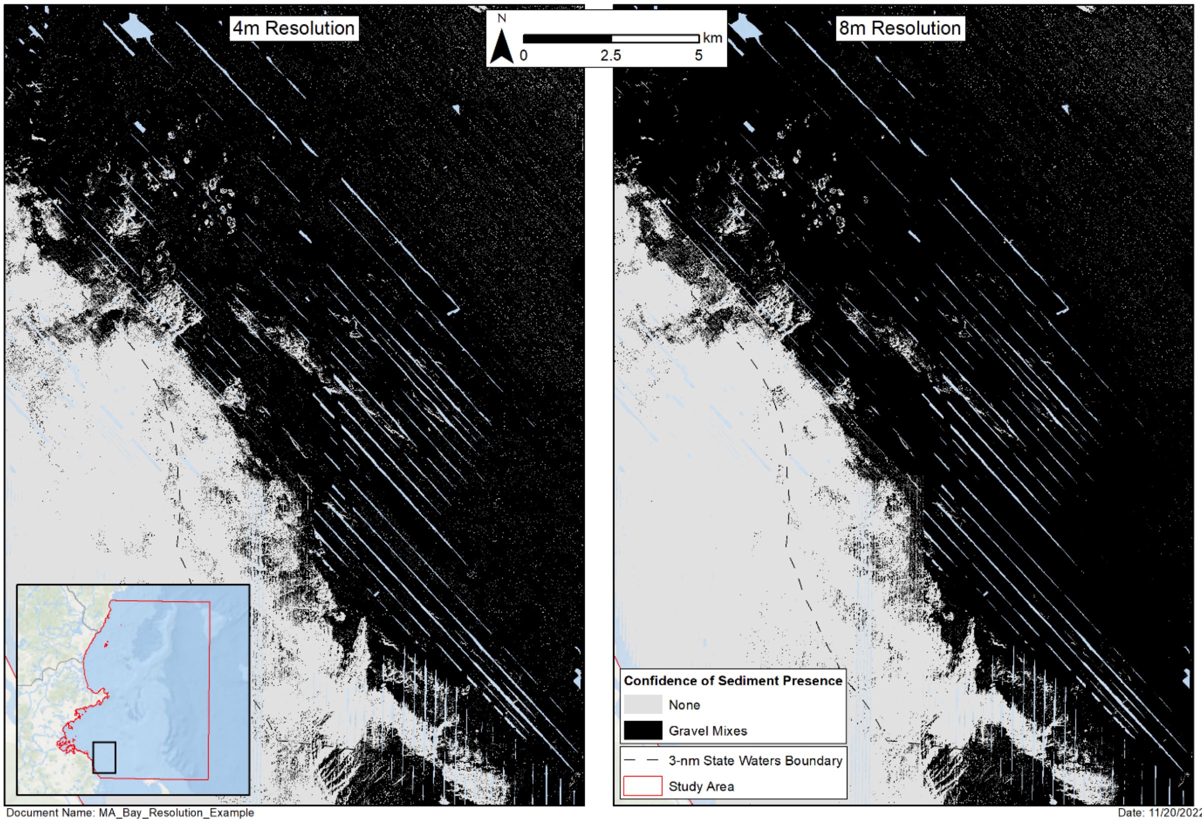


Figure 2-2. Comparison of sediment presence confidence between bathymetric resolutions of 4 m (left) and 8 m (right). The current example shows the predicted presence of Gravel Mixes in the western Gulf of Maine using publicly available datasets.

2.2 Data Analysis

A multifaceted protocol was developed to produce a habitat delineation data product from geophysical and benthic assessment data (Task 2; Figure 2-3). This protocol consisted of the following steps:

1) the selection of variables from the collected datasets,
2) the evaluation of bathymetric characteristics for predicting the presence of individual habitat delineations (i.e., sediment types), and
3) the generation of a sediment presence composite map from individual model outputs.

Geospatial processing and manipulations were conducted using Esri's ArcGIS software suite and the Benthic Terrain Modeler extension (Walbridge et al. 2018; requires Spatial Analyst extension). Statistical analyses were performed with the computing software R (R Core Team 2022) and the packages "dismo" (Hijmans et al. 2021, v. 1.3.5) and "rJava" (Urbanek 2021, v. 1.0.6).

2.2.1 Variable Preparation

Geophysical and benthic assessment datasets were further processed to identify appropriate variables for properly modeling the influence of bathymetric characteristics on sediment type presence. A set of independent (i.e., predictor) variables were initially considered and extracted from the 8-m bathymetric data described in Section 2.1 using the Benthic Terrain Modeler extension. Pairwise Pearson correlation coefficients were calculated and variables with $r > 0.6$ were assessed spatially to determine which variables needed to be removed to limit model multicollinearity (Figure 2-4). In the end, seven of the nine predictor variables were retained from the mosaiced bathymetric dataset to classify the benthic environment. These seven variables included depth, slope, aspect in both north-south and east-west components, plan (positive values, convergence) and profile (positive values, divergence) curvatures, and fine-scale bathymetric position index (inner radius of 8 m and outer radius of 25 m) (Table 2-2).

To characterize the habitat, these seven independent variables were used to predict a single response variable, i.e., sediment class. Independent models were to be generated for each sediment class. To reduce the number of modeling sessions, CMECS sediment groups from raw benthic assessment datasets (see Section 2.1) were simplified into five “determined sediment classes”: Gravel, Gravel Mixes, Gravelly, Sand, and Sand-Mud Mix.

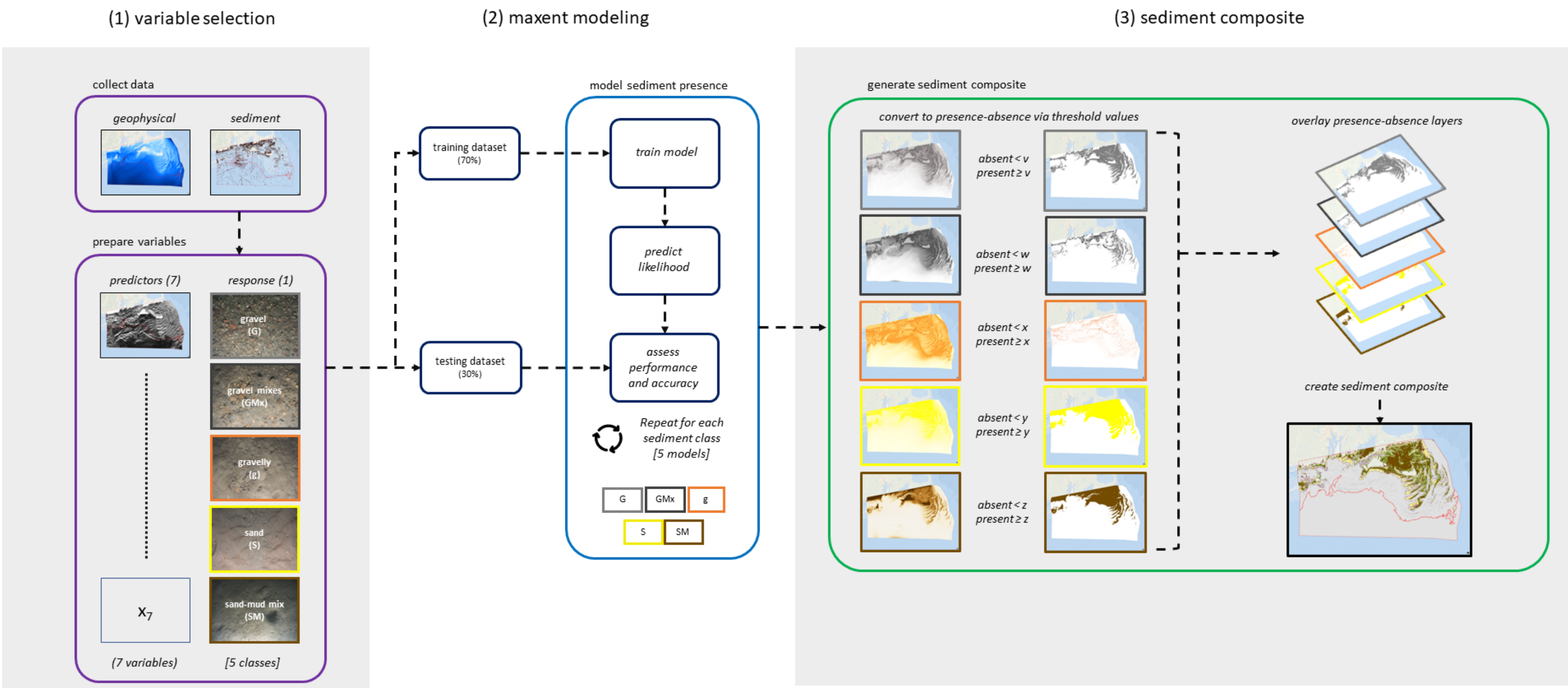


Figure 2-3. Established workflow for generating a sediment composite using a sediment-specific maximum entropy (MaxEnt) modeling approach. The workflow includes (1) the extraction and preparation of predictor and response variables from collected data, (2) the building and testing of MaxEnt models for each sediment class, and (3) the generation of a sediment composite by converting predicted likelihoods to presence-absence outputs and overlaying them. The number of predictor variables are indicated within parentheses whereas number of categories for the single response variable are shown inside brackets.

Table 2-2. Independent variables extracted from bathymetric data using Benthic Terrain Modeler (Walbridge et al. 2018). Fields include variable name, definition, and whether it was retained in the final maximum entropy model. All variables possess a resolution of 8 m.

Variables	Definition	Final Model?
Depth (m)	Water depth in meters	Yes
Geodesic Slope	Measure of gradient	Yes
Aspect – N/S	Gradient in N/S direction	Yes
Aspect – E/W	Gradient in E/W direction	Yes
Curvature, Profile	Measure of 'exposure' - Parallel direction, benthic flow	Yes
Curvature, Planar	Measure of 'exposure' - Perpendicular direction, benthic convergence	Yes
Bathymetric Position Index, Fine (8, 10)*	Measure of relative surrounding elevation, fine (peaks +, depressions -, plateau 0)	No
Bathymetric Position Index, Broad (8, 25)*	Measure of relative surrounding elevation, broad (peaks +, depressions -, plateau 0)	Yes
Bathymetric Position Index, Broad (8, 75)*	Measure of relative surrounding elevation, broad (peaks +, depressions -, plateau 0)	No

* Inner and outer radii used to calculate bathymetric position indices

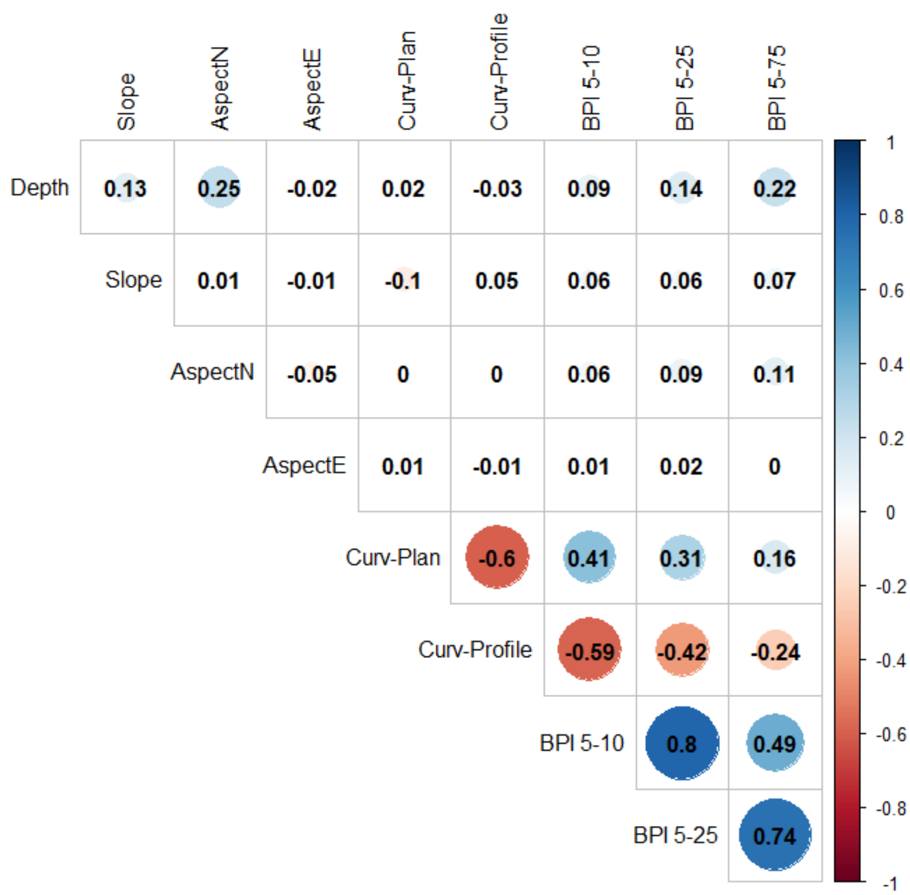


Figure 2-4. Correlation plot matrix displaying Pearson correlation coefficient values between variables extracted from bathymetric data using the Benthic Terrain Modeler extension (see Table 2-2 for details).

2.2.2 MaxEnt Modeling

While many approaches exist for modeling the geographic occurrence of a given species or feature, like sediment class, the type of model used depends heavily on what kind of data are available. The publicly available datasets collected for this study, for instance, only specify the locations where a sediment type was observed (i.e., presence-only) and rarely specify absence. Because the intent and methods for collecting these data are rarely known, presence-only data lack explicit information needed to infer the absence of sediment types, and such inferences would contain errors and biases (Elith et al. 2006). For this reason, a Maximum Entropy (MaxEnt) model was used to predict the likelihood of a sediment class occurring as a function of bathymetric characteristics within the study area (Phillips et al. 2006; Phillips and Dudík 2008). In brief, the MaxEnt modeling method estimates the likelihood of a feature's occurrence in space (between 0 and 1) by finding the maximum entropy probability distribution (i.e., most uniform distribution) given a set of constraints (e.g., bathymetric attributes) (Phillips et al. 2006). As a machine learning approach, MaxEnt models are typically trained (i.e., built) using a random subset of background data and then validated for predictive performance by testing predictions on the remaining data (Elith et al. 2011). In comparison to other novel methods, the MaxEnt algorithm was chosen due to its accessibility, speed, and predictive accuracy.

For each sediment class, MaxEnt models were built using a random subset (70%) of presence-only data and the seven bathymetric variables prepared (see Section 2.2.1). Model building protocols resulted in five sediment occurrence models as well as post-hoc assessments to evaluate variable importance (i.e., the predictive importance of each variable) and response (i.e., how variables affect model predictions). A test dataset, which included the remaining presence-only data (30%) for each sediment class, was then used to assess the performance for each model. One such model performance metric included assessing variation among the independent runs using the area under curve (AUC) value from the receiver-operating characteristic curves (ROC), which measures the ability of the model to discriminate between a sediment class being present or absent on a scale from 0 to 1 (Melo 2013). An AUC value of 0.5 is considered poor predictive performance (i.e., the model predicts outcomes no better than random), 0.7 to 0.8 is acceptable, 0.8 to 0.9 is excellent, and 0.9 and greater is outstanding (reviewed by Mandrekar 2010). Additional model performance metrics were calculated from the model specific confusion matrix to evaluate a model's performance (i.e., accuracy, sensitivity, specificity) in predicting sediment occurrence in each grid cell. Model uncertainty in occurrence data was estimated by comparing model prediction variation (standard deviation) between the randomly partitioned test datasets, essentially quantifying where sediment occurrence data disagreed with prediction outputs (i.e., high levels of uncertainty equal more disagreement). The test set was also used to calculate model threshold values for binary classification (see Section 2.2.3).

2.2.3 Sediment Composite

To visualize the presence of multiple sediment classes in a single composite image, thresholds were first calculated for each of the five modeled sediment classes using Cohen's kappa maximum to classify sediment likelihoods as either present or absent. For example, sediment presence likelihoods were converted to a value of 0 (i.e., absent) if less than the calculated threshold or a value or 1 (i.e., present) if above said threshold. In the offshore region of the study area, however, the presence of sediment classes was sparsely predicted (i.e., low sediment class likelihood) due to the disparity in bathymetric features and sediment sampling between nearshore and offshore areas. Therefore, to improve the prediction of sediment class presence offshore, thresholds were estimated for each sediment class in areas less (i.e., nearshore) and greater (i.e., offshore) than 45-m in depth (Table 2-3; Figure 2-1). In other words, two thresholds were calculated for each threshold based on area, resulting in a total of ten thresholds. When a threshold was calculated below 0.5, a default threshold of 0.5 was used to maximize the accuracy of sediment occurrence.

Table 2-3. Estimated threshold values for converting the predicted likelihood of each sediment class occurring within each nearshore and offshore location into a binary presence-absence output. In instances where estimated thresholds were low, a default value of 0.50 was used to increase accuracy (identified in cells using parentheses).

	Region	
Sediment Class	**Nearshore**	**Offshore**
Gravel Mixes	0.74	0.67
Gravel	0.46 (0.50)	0.49 (0.50)
Gravelly	0.81	0.71
Sand	0.45 (0.50)	0.40 (0.50)
Sand-Mud Mix	0.39 (0.50)	0.76

Presence-absence distributions for each sediment class were then merged and summarized to properly characterize benthic habitat complexity for each grid cell. To do so, presence values for each sediment class were reclassified as unique non-zero values (e.g., Gravel = 1, Gravel Mixes = 2, etc.). By stacking reclassified sediment class distributions and summing unique non-zero presence values in ArcGIS, a sediment composite was generated where each grid cell contained mutually exclusive scores and thus sediment class combinations. The sediment composite was qualitatively compared between regions with high and low bathymetric sources as well as against publicly available regional sediment data products to gauge composite accuracy and performance.

3.0 RESULTS

3.1 Sediment Class Predictions

Overall, the generated sediment composite indicated high confidence in sediment occurrence for coastal areas, especially on bathymetric relief associated with Nantucket Shoals, and low predictive confidence in offshore areas (Figure 3-1). Given the overlay protocol applied to the MaxEnt modeling outputs for the original five sediment classes, nine unique sediment class combinations were used to characterize benthic habitat composition (Table 3-1; Figure 3-1). For instance, Sand-Mud Mix with Gravel (10.1%) and Gravel & Mixed Gravel Classes (8.7%) represented most of the study area. When examining the nearshore and offshore regions established in Section 2.2.3, Sand-Mud Mix with Gravel (17.4%) and Gravel & Mixed Gravel Classes (14.9%) were still the most common classes in the nearshore region as compared to Gravelly (0.2%) and Sand-Mud Mix (0.1%) in the offshore region.

Due to the paucity of data, the model successfully characterized sediment in only 27.9% (884.3 km^2) of the entire study area, failing to characterize the remaining 72.1% (2,284.73 km^2) of the study area given low confidence in sediment predictions. When separated by region, the model confidently characterized a larger percentage of the nearshore area (48.8%) as compared to the offshore region (~0.2%) (Table 3-1; Figure 3-1).

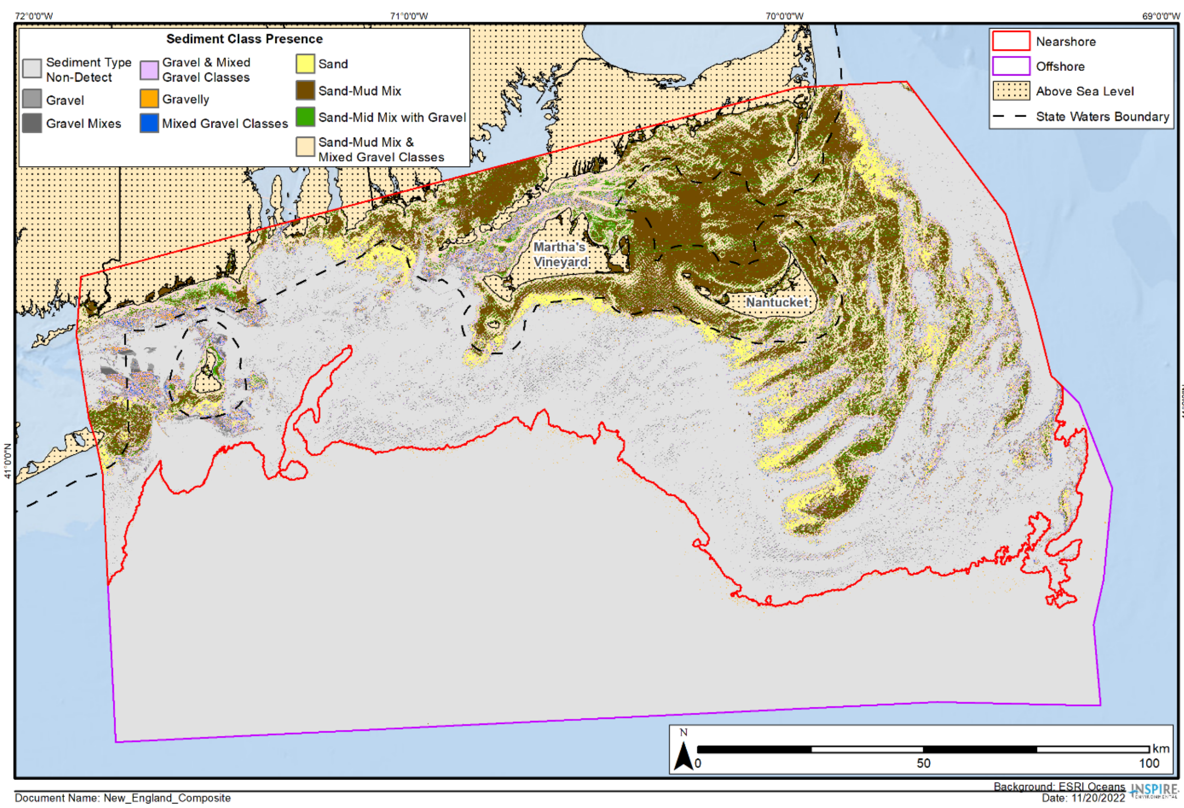


Figure 3-1. Predicted occurrence of nine sediment class combinations within the study area using a multifaceted modeling protocol. The nearshore-offshore region boundary identifies where different sets of thresholds were used to classify sediment occurrence likelihoods as either present or absent prior to raster calculations.

Table 3-1. Estimated Area Occurrence of Nine Sediment Class Combinations within the Study Area, Specifically Nearshore, Offshore, and Overall

	Nearshore		Offshore		Overall	
Sediment Class	**Total Area (km^2)**	**Percent Area**	**Total Area (km^2)**	**Percent Area**	**Total Area (km^2)**	**Percent Area**
Not Classified	962.38	52.20	1322.34	99.76	2284.73	72.10
Gravel	6.67	0.36	0	0	6.67	0.21
Gravel Mixes	3.14	0.17	0.02	0	3.16	0.10
Gravel & Mixed Gravel Classes	274.15	14.87	0.01	0	274.17	8.65
Gravelly	52.25	2.83	2.24	0.17	54.49	1.72
Mixed Gravel Classes	4.55	0.25	0.03	0	4.58	0.14
Sand	94.22	5.11	0	0	94.22	2.97
Sand-Mud Mix	56.32	3.05	0.69	0.05	57.01	1.80
Sand-Mud Mix with Gravel	320.85	17.40	0	0	320.85	10.12
Sand-Mud Mix & Mixed Gravel Classes	68.97	3.74	0.17	0.01	69.14	2.18
TOTAL	**1843.50**	**100.00**	**1325.50**	**100.00**	**3169.02**	**100.00**

3.2 Model Performance

The bathymetric predictor variables of depth (80.5%) and geodesic slope (16.5%) were deemed most important, on average, when predicting sediment class occurrence (Table 3-2). Variable responses curves indicated that, in general, sediment occurrence predictions decreased with increasing depths except for Gravelly, which had an increase in prediction at approximately 40 m. All sediment types, except for Sand-Mud Mix, reached peak prediction at or slightly greater than a slope of 0° (i.e., no seafloor gradient). In contrast, Sand-Mud Mix was more likely to occur at slightly elevated seafloor slopes. AUC values for all sediment classes ranged from 0.74 (Gravelly) to 0.85 (Gravel), which are considered acceptable in terms of model predictive capability. Model prediction uncertainty in occurrence data ranged from 0 to 0.6 (standard deviation) across the study area for all sediment classes with greater uncertainty occurring in the coastal areas of the nearshore region.

Table 3-2. Variable importance of bathymetric variables for predicting sediment occurrence in terms of percent contribution. A mean percent contribution identifies the average contribution of each variable across sediment classes.

	Percent Contribution					
Variable	**Mean**	**Gravel Mixes**	**Gravel**	**Gravelly**	**Sand**	**Sand-Mud Mix**
Depth (m)	80.48	68.2	67.9	77.9	94.4	94
Geodesic Slope	16.52	28.8	31.8	11.8	5.2	5
Aspect – N/S	2.06	1.2	0.2	8	0.3	0.6
Aspect – E/W	0.70	1.3	0.1	1.9	0.1	0.1
Curvature, Profile	0.00	0	0	0	0	0
Curvature, Planar	0.08	0.4	0	0	0	0
Bathymetric Position Index (8, 25)	0.12	0	0	0.3	0	0.3

4.0 DISCUSSION and RECOMMENDATIONS

4.1 Limitations

The paucity and resolution of publicly available data in the region present a major challenge to confidently predicting and classifying sediments in the study area. For instance, the distribution of in-situ sampling sediment data was unbalanced across the study area due to individual project objectives (Figure 2-1), preventing the proper characterization and eventual prediction of specific sediments using bathymetric features like depth. Publicly available geophysical data primarily consisted of bathymetry as compared to other types of acoustic information, such as backscatter, a reflectivity measurement that aids in understanding seafloor hardness and roughness (Lurton and Lamarche 2015). In contrast, bathymetry only registers depth and slope, so the inclusion of other geophysical data types would provide additional explanatory power when predicting sediment occurrence.

Varying bathymetric resolutions affected sediment occurrence predictions and overall patterning. As observed in Figure 4-1, more natural sediment occurrence patterns were predicted in Nantucket Sound (10-m resolution) as compared to coarser categorizations in areas south of Martha’s Vineyard (250-m resolution). With independent variables such as geodesic slope (i.e., a measure of seafloor gradient), lower-resolution data cannot capture fine-scale differences that define specific sediment types. Therefore, when only low-resolution bathymetry data are available in areas with small changes in slope, the model fails to predict the actual diversity of benthic habitats.

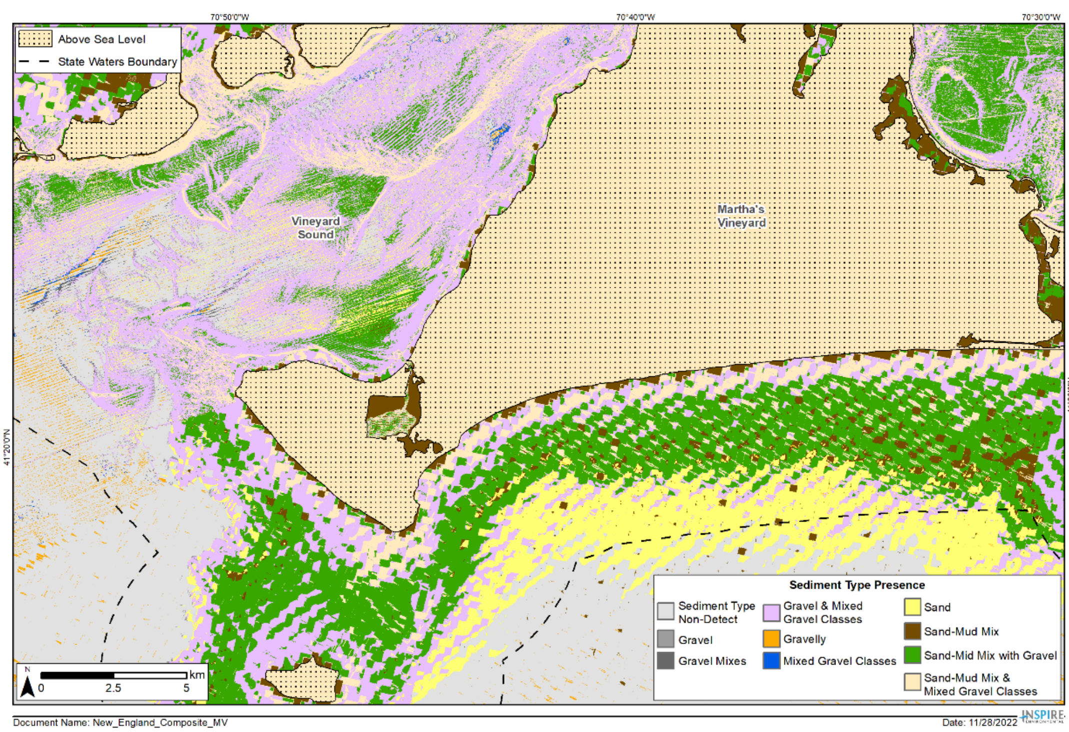


Figure 4-1. Comparison of predicted sediment occurrence patterns between Vineyard Sound (10-m resolution) and the coastal waters south of Martha’s Vineyard (250-m resolution).

4.2 Data Utility

The generated sediment composite accurately predicts the occurrence of sediment classes in the nearshore areas of southern New England where data are both diverse and abundant. The sediment composite provides fine-scale patterning in coastal areas, identifying complex patterns of habitat composition and distribution in greater detail, which are not identified by The Nature Conservancy's (TNC) Northwest Atlantic Marine Ecoregional Assessment (NAMERA) interpolated soft sediment data product (Anderson et al. 2010). For example, Nantucket Shoals, an area of large distinct bathymetric relief composed mainly of sand, is well represented in the composite compared to the TNC data (Figures 4-2 and 4-3). However, the TNC product describes areas that are poorly described by the sediment composite because the TNC product is based on interpolated grain size data and does not set thresholds for data inclusions; therefore, interpolated data are available for offshore areas where sediment occurrences could not be confidently predicted in the current sediment composite. In summary, the sediment composite offers better resolution in coastal areas whereas the TNC product offers a broader spatial scope.

Given the observed strengths and limitations of the current product, the generated sediment composite is intended to complement other regional data (e.g., TNC's NAMERA, MA Coastal Zone Management interpolated surficial sediment data, etc.), and should thus be used in concert with them for drawing conclusions. For example, the higher uncertainty around sediment distributions offshore made explicit by the results of this pilot project provide useful context for stakeholders using the TNC data to further investigate features and questions of interest.

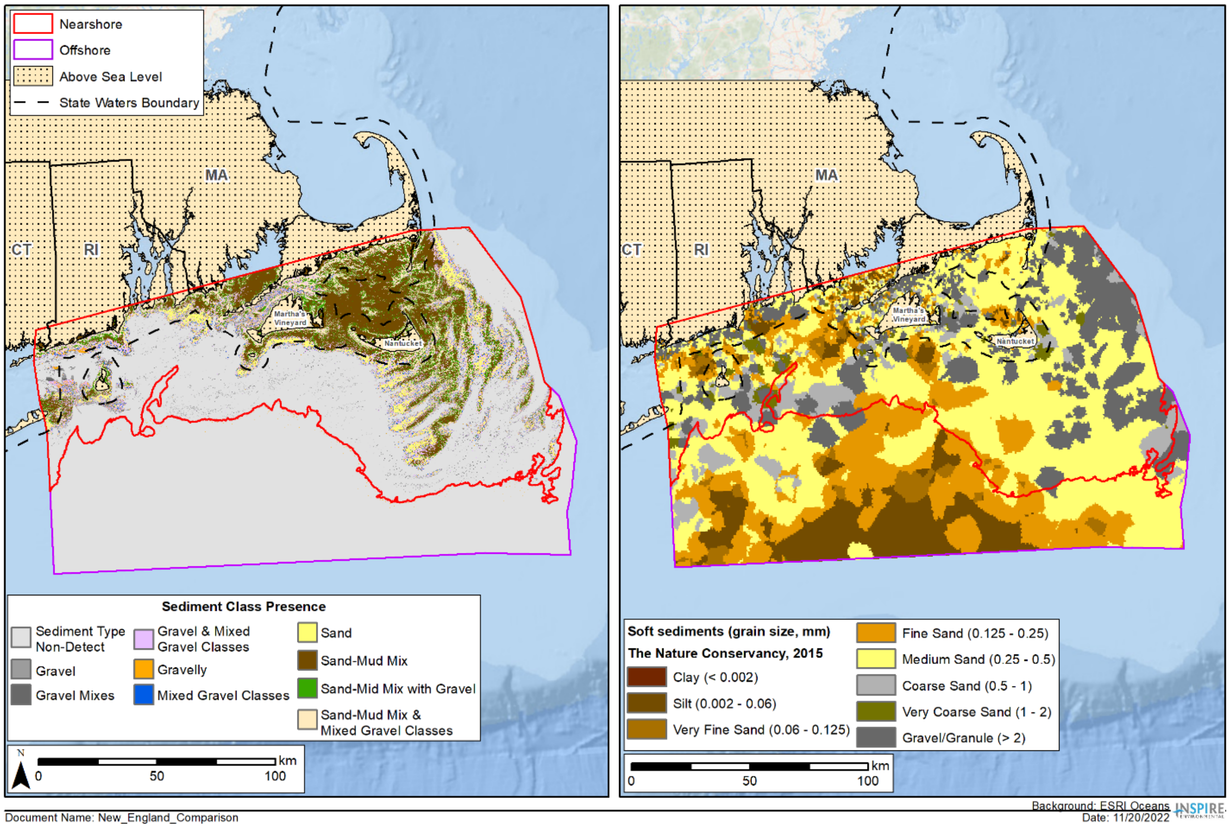


Figure 4-2. Regional comparison of this project's sediment composite (left) with The Nature Conservancy's soft sediment data product based on interpolated grain size data (right).

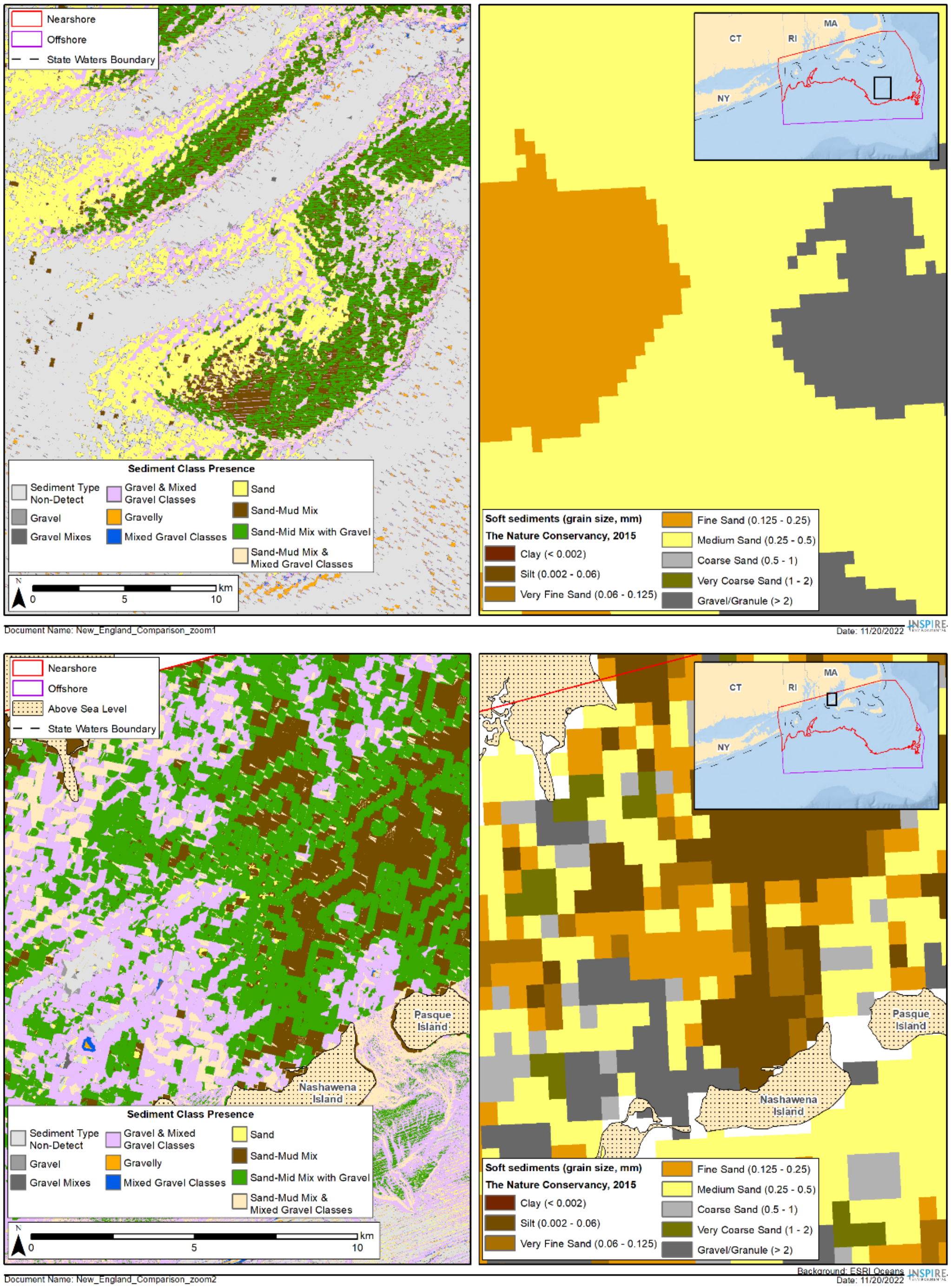


Figure 4-3. Example comparisons of this project's sediment composite (left) with The Nature Conservancy's soft sediment data product based on interpolated grain size data (right). The top and bottom row of images compares these data products in different geographic areas, specifically Nantucket Shoals (top) and within Buzzards Bay, MA (bottom).

4.3 Recommendations

4.3.1 Data Availability and Quality

Access to a greater diversity of data sources, including geophysical and habitat assessment data, would aid in addressing the challenges mentioned previously. Although demand and available funding limit data collection by government agencies and academic institutions, offshore wind energy developers collect geophysical and in-situ sediment data to fulfill COP requirements and, at later intervals, to inform engineering and construction design decisions. This suite of developer geophysical data (e.g., bathymetry, backscatter, side-scan sonar) are collected at resolutions finer than many of the publicly available datasets used in the present study (e.g., <1 m). In addition to higher resolution geophysical datasets, the use of other geophysical data can improve the predictive accuracy of the MaxEnt modeling protocol by increasing the number of situations that separate unique sediment classes. For example, to examine the added value of backscatter to the MaxEnt modeling protocol, a small case study was conducted that used boulder pick data derived from geophysical contact data, backscatter (absolute dB, 8-m resolution), and the prediction outputs for the five sediment classes described in Section 3.0 to predict boulder presence. Results showed that backscatter was the most significant contributor in predicting boulders (variable importance over 50%) and generally confirmed boulder presences aligned with higher model predictions (likelihood between 0.5 and 0.78).

In addition to federal and state regulators, developers would greatly benefit from a sediment composite data product with high predictive accuracy. It is therefore recommended that offshore wind energy developers continue to be included in future conversations about data accessibility and sharing. These discussions will highlight data needs and address developer concerns with sharing proprietary data among other regional stakeholders. These conversations are expected to be moderated through the RWSC.

4.3.2 Data Processing and Formatting

Based on protocols and products developed from the current project, a set of technical recommendations was generated to expedite data processing, modeling, and online dissemination. To start, data providers should provide geophysical and benthic assessment data based on recommendations highlighted in Table 4-1.

Table 4-1. Technical Recommendations for Preparing and Sharing Bathymetry and Sediment Collection Data

	Bathymetry		Sediment
Data Type	Raster		Vector (point)
File Type	GeoTIFF		Spreadsheet
File Extensions	.tiff		.xlsx, .xls, .csv
Coordinate System	WGS 84		WGS 84
Resolution	≤8 m		
Data Format	XYZ		Wide[1]
Compression[2]	.zip		*(optional)*
Metadata	Yes		Yes

[1] Wide data format refers to a spreadsheet with one unique record per row and columns to identify characteristics.

[2] Compressing raster files into a .zip file is recommended to keep all relevant files in one directory.

For the sake of completeness, the bulleted list below offers data providers additional insight for achieving these recommendations and streamlining the data intake process.

- All data
  - Providers are encouraged to **format geospatial data in the WGS84 reference coordinate system** (i.e., longitude-latitude), however, other coordinate systems like NAD83 can be utilized if they are properly defined and noted.
  - **Metadata should be appended to all datasets**, including title, data description, source, author, last modified date, thumbnail, and tags.
- Bathymetry
  - Providers can down-sample raster files to a **maximum of 8-m resolution** if they cannot send high-resolution bathymetry data (<1 m) due to data confidentiality or file size (Appendix C).
    - Case studies conducted during the present study confirmed that down-sampling from 4-m to 8-m resolution maintained predictive power while decreasing computational time during model building (45 minutes to 15 minutes per sediment class) (Figure 2-2).
    - Down-sampled data should be reviewed by data providers prior to dissemination to confirm confidential information is properly masked. Data providers should contact recipients if sensitive features are not properly masked at the 8-m resolution to identify potential solutions (i.e., additional down-sampling simulations and reviews).
- Sediment

- Sediment observations should be **classified using CMECS Substrate Group definitions** (FGDC 2012) or measurements (phi size, grain size) that can be easily crosswalked to CMECS Substrate Groups.
- Collected data should be recorded in a wide, or row-based, data format; **one unique sediment record per row** and columns identifying measurements and point locations (Appendix E).
    - Column, or header, titles should not have any spaces (e.g., “Sediment_Class” and “SedimentClass” are preferred over “Sediment Class”)
    - Latitude-longitude coordinates should be provided in decimal degrees (DD.dd).

4.3.3 Data Hosting and Delivery

As the quality and volume of benthic characterization data grows, so too will the demand for online storage and delivery options to share said information with a wider audience. Data managers and organizations have a suite of options available to them when considering online data storage and hosting technologies, including on-premises vs. cloud computing services, cloud providers, and the overall functionality and tools provided by specific services. Therefore, Task 4 sought to strategically consult with organizations with demonstrated data hosting and sharing expertise to develop technical recommendations for future consideration. Conversations with the Marine Data Exchange (MDE), an established online data repository managed by The Crown Estate of the United Kingdom for publicly sharing industry- and research-derived seafloor data, and a review of available services yielded the following set of baseline recommendations. Additional details regarding these recommendations can be found in Appendix F.

- **Cloud computing services should be considered** over on-premises (i.e., local) computing services.
    - Cloud computing services provide predictable pricing plans, reduce reliance on physical hardware, software, and associated information technology (IT) personnel, and allow for quick scaling to meet data storage and processing demands.
- **Microsoft Azure is the recommended cloud computing provider** given its simplicity, integration with Microsoft services, and use by many enterprise clients; this provider comes highly recommended by managers and product owners from the MDE.
- **Capitalize on the collective experience of IT professionals and Microsoft Azure features** to create a product that supports both data providers and public end users.

4.3.4 Data Visualization and Analysis

In addition to data hosting and delivery, data visualization is vital to translating complex datasets and results into digestible messages for a variety of stakeholders and use cases. The current sediment composite (and other data products) should not only be visualized appropriately but also contain supplementary datasets and metadata to articulate to end users how the data should be used. Using the existing data product as a case study, Tasks 3 and 5 established a set of technical recommendations for effectively visualizing and communicating spatial data on the Portal. Additional details regarding these recommendations can be found in Appendix G.

- **Develop clear visuals and provide detailed metadata** to maximize comprehension.
    - Data layers should include necessary legend entries as well as metadata, including title, description, source, author, last modified date, thumbnail, and tags.
- **Include background and supplementary data** to aid end users in evaluating the data product.
    - Suitable data layers include raw input sources (Figure 4-4) as well as interactive features that effectively summarize multiple findings (Figure 4-5).
- **Provide different data product versions** to satisfy an assortment of stakeholders and specific use cases.
    - In instances where data layers contain multiple dimensions (e.g., temporal), data providers and visualizations specialists should consider crafting and displaying other versions so multiple stakeholder groups can benefit.

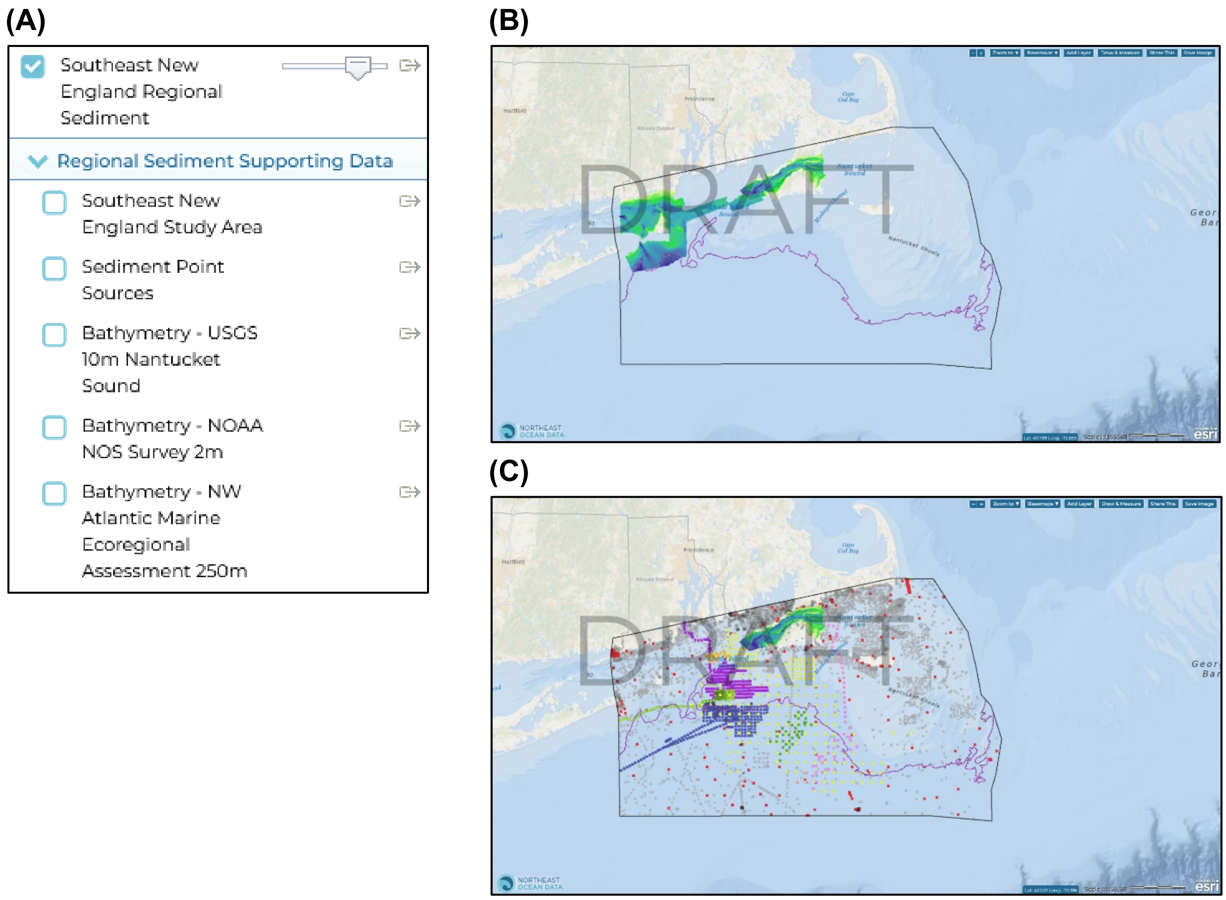


Figure 4-4. (A) Example of supporting datasets that users can access on the Portal development site for the sediment composite, including (B) bathymetric sources and (C) sediment sampling locations.

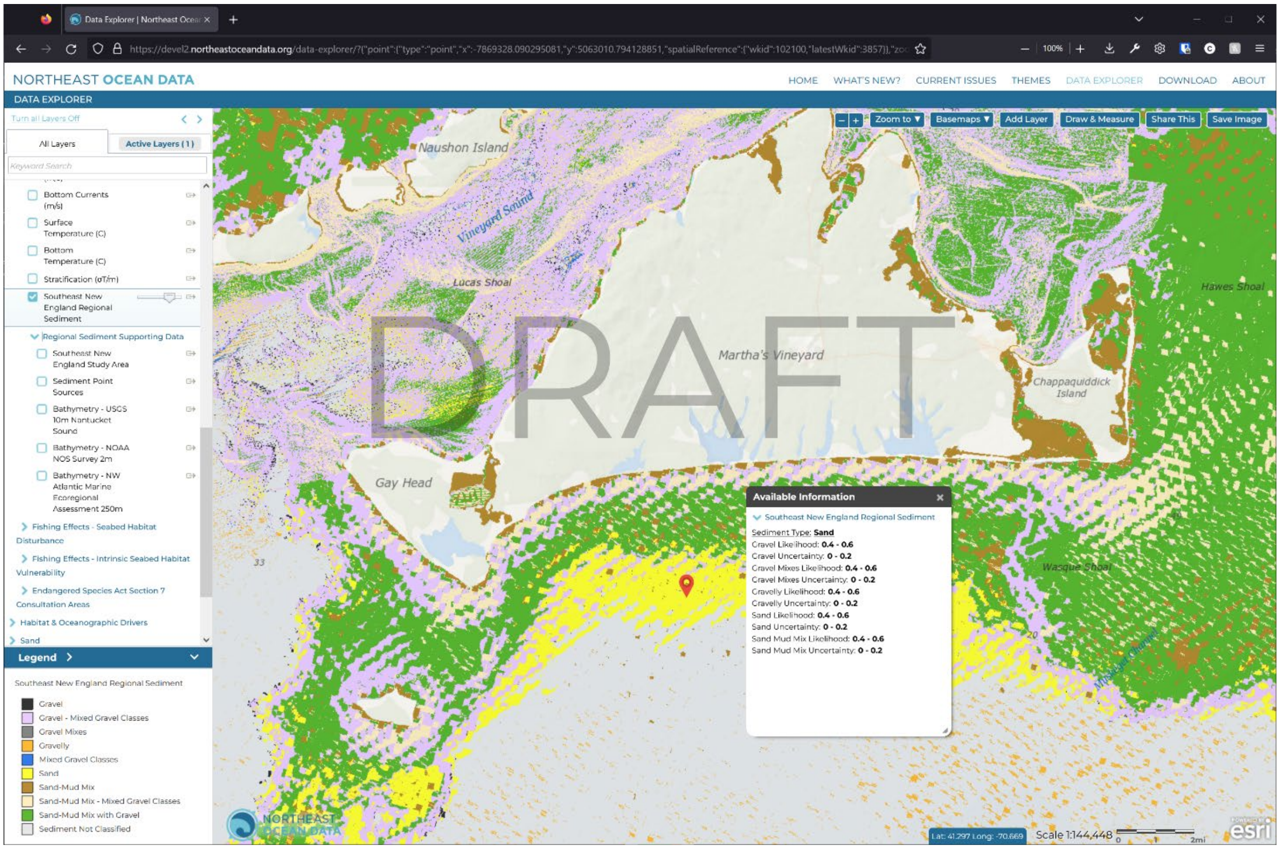


Figure 4-5. Example of an interactive window that users can toggle to view supporting information within the sediment composite, specifically model likelihood and uncertainty for each of the five sediment classes.

5.0 SUMMARY

The current project demonstrates the value of integrating publicly available geophysical and habitat assessment data to derive a regional sediment occurrence composite. As a result, the project identified key takeaways to aid in understanding the sediment composite's overall utility and technical recommendations for future consideration.

- The generated sediment composite does well at predicting the occurrence of various sediment classes in the nearshore areas of the southern New England study area.
    - Data paucity and quality and conservative thresholds impacted the model's ability to predict the occurrence of sediment classes with high confidence, especially in areas that exceed 45 m depth (i.e., offshore region).
    - The sediment composite is intended to complement, not replace, existing regional sediment data products given its observed strength and limitations, e.g., the sediment composite offers better resolution in coastal areas while other products offer a broader spatial scope.
    - Due to its expected benefit to regional stakeholders, continued conversations with offshore wind energy developers are crucial to sharing high-resolution geophysical data and generating more accurate sediment composites.
- Technical recommendations were established for enhancing future implementations, including requirements for generating sediment composites, hosting, and delivering online data, and visualizing and analyzing data.
    - Sediment composites depend on high-resolution geophysical and habitat assessment data, so it is important to continue to engage offshore wind energy developers and encourage data sharing.
    - Cloud computing services simplify hosting, maintaining, and delivering large datasets to numerous stakeholders.
    - Data visualization should leverage unique geoprocessing tools to integrate supporting data layers and increase data transparency and reproducibility.

6.0 REFERENCES

Anderson, M.G., Greene, J., Morse, D., Shumway, D. and M. Clark. 2010. Benthic Habitats of the Northwest Atlantic in Greene, J.K., M.G. Anderson, J. Odell, and N. Steinberg, eds. The Northwest Atlantic Marine Ecoregional Assessment: Species, Habitats and Ecosystems. Phase One. The Nature Conservancy, Eastern U.S. Division, Boston, MA.

Bureau of Ocean Energy Management (BOEM) Office of Renewable Energy Programs. 2019. Guidelines for Providing Benthic Habitat Survey Information for Renewable Energy Development on the Atlantic Outer Continental Shelf Pursuant to 30 CFR Part 585. June 2019.

Bureau of Ocean Energy Management (BOEM) Office of Renewable Energy Programs. 2020. Guidelines for Providing Geophysical, Geotechnical, and Geohazard Information Pursuant to 30 CFR Part 585. May 27, 2020.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P. and M. van den Belt. 1997. The value of the world's ecosystem services and natural capital. Nature 387: 253-260.

Degraer, S., Carey, D.A., Coolen, J.W.P., Hutchison, Z.L., Kerckhof, F., Rumes, B. and J. Vanaverbeke. 2020. Offshore wind farm artificial reefs affect ecosystem structure and functioning: A synthesis. Oceanography 33(4): 48–57.

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

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E. and C.J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17(1): 43–57.

Fahay, M.P., Berrien, P.L., Johnson, D.L. and W.W. Morse. 1999. Essential Fish Habitat Source Document: Atlantic Cod, *Gadus morhua*, Life History and Habitat Characteristics. NOAA Technical Memorandum NMFS-NE-124.

Federal Geographic Data Committee (FGDC). 2012. FGDC-STD-018-2012. Coastal and Marine Ecological Classification Standard. Reston, VA.

Griswold, C.A. and J. Prezioso. 1981. In-situ observations on reproductive behavior of the long-finned squid, *Loligo pealei*. Fishery Bulletin 78: 945–947.

Hijmans R.J., Phillips, S., Leathwick, J. and J. Elith. 2021. dismo: Species Distribution Modeling. R package version 1.3.5, https://CRAN.R-project.org/package=dismo.

Jackson, S.J. and S. Barbrow. 2015. Standards and/as innovation: Protocols, creativity, and interactive systems development in ecology. In Proceedings of the 33rd annual ACM conference on humans factors in computing systems. doi: 10.1145/2702123.2702564.

Liquete, C., Piroddi, C., Drakou, E.G., Gurney, L., Katsanevakis, S., Charef, A. and B. Egoh. 2013. Current Status and Future Prospects for the Assessment of Marine and Coastal Ecosystem Services: A Systematic Review. PLoS ONE 8: e67737.

Lurton, X. and G. Lamarche (Eds). 2015. Backscatter measurements by seafloor-mapping sonars. Guidelines and Recommendations. 200p. http://geohab.org/wp-content/uploads/2014/05/BSWGREPORT-MAY2015.pdf.

Mandrekar, J.N. 2010. Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology 5(9), 1315-1316. doi: 10.1097/JTO.0b013e3181ec173d.

Melo, F. 2013. Area under the ROC Curve. In: Dubitzky, W., Wolkenhauer, O., Cho, KH., Yokota, H. (eds) Encyclopedia of Systems Biology. Springer, New York, NY. doi: 10.1007/978-1-4419-9863-7_209.

National Marine Fisheries Service (NOAA Fisheries) Greater Atlantic Regional Fisheries Office Habitat Conservation and Ecosystem Services Division. 2021. Recommendations for Mapping Fish Habitat. March 2021. https://media.fisheries.noaa.gov/2021-03/March292021_NMFS_Habitat_Mapping_Recommendations.pdf?null.

Phillips, S.J., Anderson, R.P. and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modeling 190: 231 – 259.

Phillips S.J. and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161–175.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roper, C.F.E., Sweeney, M.J. and C.E. Nauen. 1984. FAO Species Catalogue, Vol. 3 Cephalopods of the world. An annotated and illustrated catalogue of species of interest to fisheries. FAO Fisheries Synopsis 125 (3):1–277.

Urbanek, S. 2021. rJava: Low-Level R to Java Interface. R package version 1.0.6, https://CRAN.R-project.org/package=rJava.

Wahle, R.A. and R.S. Steneck. 1991. Recruitment habitats and nursery grounds of the American lobster *Homarus americanus*: a demographic bottleneck? Marine Ecology Progress Series 69: 231–243.

Walbridge, S., Slocum, N., Pobuda, M. and D.J. Wright. 2018. Unified Geomorphological Analysis Workflows with Benthic Terrain Modeler. Geosciences 8(3):94. doi: 10.3390/geosciences8030094.

Wentworth, C.K. 1922. A Scale of Grade and Class Terms for Clastic Sediments. The Journal of Geology 30: 377–392.

REGIONAL HABITAT MAPPING

Standard Approaches to Synthesizing, Visualizing, and Disseminating High-Resolution Geophysical Data to Advance Benthic Habitat Mapping in the Wind Energy Areas of the Northeast

APPENDICES

Prepared for:

MASSACHUSETTS CLEAN ENERGY CENTER®

Submitted to:

Nils Bolgen
Massachusetts Clean Energy Center
63 Franklin St. 3rd Floor
Boston, MA 02110
NBolgen@masscec.com

Submitted by:

INSPIRE ENVIRONMENTAL

INSPIRE Environmental
513 Broadway, Suite 314
Newport, Rhode Island 02840

Final Report
April 28, 2023

Appendices

Appendix A - Data Sharing MOU

Pilot Studies - Regional Fisheries Monitoring in Relation to Southern New England Offshore Wind Projects - REGIONAL HABITAT MAPPING

November 10, 2021

Memorandum of Understanding (MOU) for Data Sharing – Regional Habitat Mapping MassCEC Project

Under funding from the Massachusetts Clean Energy Center (MassCEC), the Bureau of Ocean Energy Management (BOEM), and the Rhode Island Department of Environmental Management (RIDEM), **INSPIRE Environmental (INSPIRE)** and the **Northeast Regional Ocean Council (NROC)** are working together to ***develop standard approaches to synthesizing, visualizing, and disseminating high-resolution geophysical data to advance benthic habitat mapping in the wind energy areas of the Northeast***.

Parties

INSPIRE Environmental (INSPIRE) is a private marine science consulting company specializing in the collection of imagery data to assess seafloor condition. INSPIRE has extensive experience integrating benthic ground-truth data with geophysical data to map benthic sediment types and habitats in the Northeast.

The Northeast Regional Ocean Council [NROC] is a state and federal partnership that facilitates the New England states, federal agencies, regional organizations, and other interested regional groups in addressing ocean and coastal issues that benefit from a regional response. It is NROC's mission to provide a voluntary forum for New England states and federal partners to coordinate and collaborate on regional approaches to support balanced uses and conservation of the Northeast region's ocean and coastal resources.

[DATA PROVIDER] is an entity that wishes to share seafloor data it has collected in the New England Area.

Overview

MassCEC, with additional financial support from BOEM and RIDEM, has awarded several contracts to fund pilot studies in support of advancing science to support offshore wind development off southern New England. INSPIRE and NROC are under contract with MassCEC to conduct the project "Developing Standard Approaches to Synthesizing, Visualizing, and Dissemination High-Resolution Geophysical Data to Advance Benthic Habitat Mapping in the Wind Energy Areas of the Northeast." The project will be implemented in collaboration with the Northeast Regional Ocean Council (NROC) and NROC has entered a sub-contract with INSPIRE. INSPIRE and NROC have assembled the Seafloor Habitat Data Work Group[1], a regional work group of experts and stakeholders across disciplines and organizations, to inform the development of best practices for integrating high resolution geophysical and ground-truth

[1] https://neoceanplanning.org/issues/seafloor-habitat-data/

data to improve maps of benthic habitats and make those data available through the Northeast Ocean Data Portal[2] to inform future planning and analysis.

Following the initiation of this project, NROC, together with the Mid-Atlantic Regional Council on the Ocean (MARCO) and the Coastal States Stewardship Foundation (CSSF), were selected to host the Regional Wildlife Science Entity (RWSE) - a forum for sharing information, standardizing data collection and monitoring protocols, defining scientific/research needs, and amplifying the results of ongoing research. Many members of the Seafloor Habitat Data Work Group are involved in guiding and operating the RWSE through one of the four sectors (federal agencies, state agencies, offshore wind developers, and non-governmental organizations), including the funders of this pilot project (BOEM, MassCEC, RIDEM). The workflows, methods, and outputs of this pilot project will be leveraged as the RWSE begins coordinating research activities and data sharing for wildlife, ecosystem, and offshore wind issues along the US Atlantic coast.

Data providers, including representatives from offshore wind developers and lease holders in southern New England, are members of the work group. Two work group meetings have been held and members agreed on the types of data that would be useful to request from data providers in support of this project. Formal requests for data have been made to all work group members. Only data that are shared through this work group forum or are otherwise publicly available will be included in the project's final data products provided through the Northeast Ocean Data Portal (Portal).

Scope

The following project outputs represent a workflow that is being tested through the pilot project with INSPIRE and NROC. In conversations since the January 22, 2021 Work Group kickoff meeting, we asked Work Group members if this pilot project workflow is viable, and if it is realistic to repeat and build on this workflow into the future as data are collected. There was broad agreement that this pilot project presents an opportunity to understand the available data, establish relationships for longer-term sharing, and recommend a workflow to be iterated and implemented in the future.

Project Output 1: Standard data request for a set of specific habitat data products relative to site characterization

> This output refers to a list of habitat variables, filetypes, and resolutions that would be included in a request to entities collecting seafloor habitat data. This output assumes that data collectors are using methods and approaches in a consistent fashion to obtain, interpret, classify, and map data. Work Group members have indicated that this assumption has not held true in the past for the existing data throughout the region that have been used to build current regional habitat products. A subset of the Work Group with familiarity with applying the Coastal and Marine Ecological Classification Standard (CMECS) could help address this issue with further discussion and consideration of ongoing CMECS implementation guidance. This project will document best practices applied in the development of pilot project products (e.g., discussions with the Work

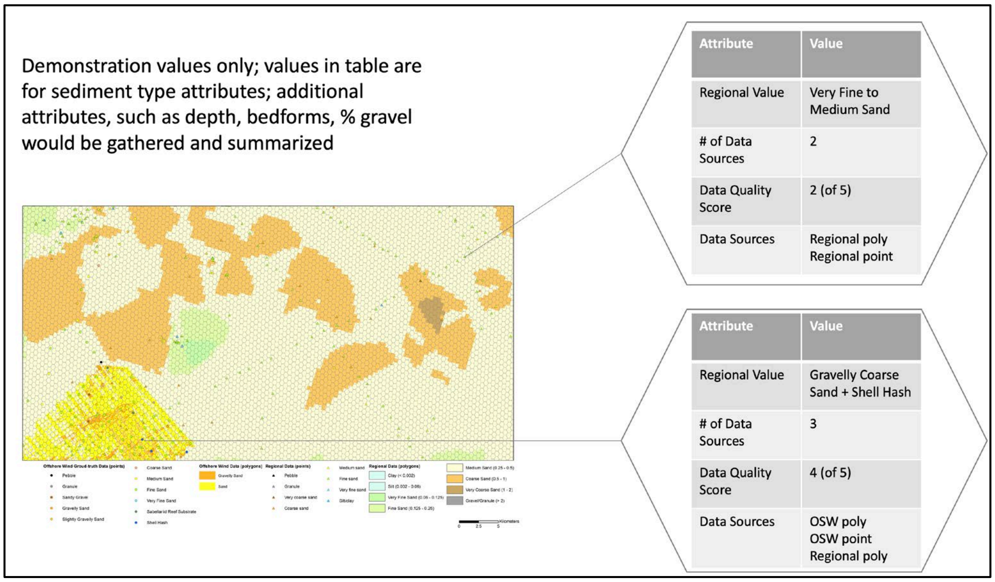
[2] https://www.northeastoceandata.org

Group on thresholds, data types that do not yield useful information for habitat mapping such as vibracores).

<u>Project Output 2: Improved understanding of habitat distributions across the region</u>

<u>Project Output 3: Use Northeast Ocean Data Portal to disseminate standard set of intermediate scale regional habitat composites that improve upon current region-scale data; possibly pilot a system to provide password-protected access to higher resolution underlying data</u>

These outputs refer to improved regional scale data products derived from the data requested and obtained via Project Output 1. A conceptual mockup of a regional scale data product and how it could be queried via Portal tools is shown below.



<u>How could these Outputs be used?</u>

- *Future planning and siting, that may include offshore wind but also aquaculture, submarine cables and pipelines, sand extraction, dredged material disposal, artificial reefs etc.*
- *Assessing cumulative impacts and monitoring*
- *Identifying data gaps; responding to frequent federal requests to prioritize areas within the region for future mapping and characterization*
- *Inform communication/public awareness around offshore wind projects*
- *Could pilot password-protected access to higher-resolution underlying data from the developers*
- *If timely, data products could be used for regulatory review of future projects, but note that regional-scale products cannot replace or negate the need for project-specific data*

<u>Project Status:</u>

Following the January 22, 2021 Work Group kickoff meeting, the project team met individually with representatives of each offshore wind company to hear their direct feedback on the project's potential outputs. Subsequently, NROC convened Work Group members from just state agencies, federal agencies, and the fishery management council to provide a status update for the project and hear feedback. The following points capture the major discussion themes.

Highlights common to conversations with both developers and agencies:

- All are supportive of the three project outputs
- All highlighted an interest in, and the importance of, data sharing and contributing to a regional understanding of habitat
- Greater understanding of regional habitat gives greater certainty (lower risk) in development and review processes
- Consider how these data could be used for other planning/management issues besides offshore wind, e.g., sand resource assessment, aquaculture, paleoarchaeological resources
- Consider developing data quality/data density/confidence metrics to help evaluate whether additional data are needed to support a regional interpretation

Highlights from conversation with the developers

- Discussions focused on two tracks for data sharing – 1) existing data, and 2) future data. Existing data have already been collected and are more difficult to adapt/convert to “new” standards. All agreed that future data will be much easier to share with respect to new standards/formats
  - Sharing existing data will be easier if the INSPIRE/NROC project team can do any down-sampling/conversions
  - INSPIRE/NROC project team needs to describe desired format(s) and resolution(s) so that new data can be delivered accordingly
- Data is shareable once a Construction and Operations (COP) is complete and sufficient/Notice of Intent (NOI) stage
- Requested data would not compromise propriety information, interfere with the development process, or be made available to stakeholders prematurely
- Consider a public license agreement, which would satisfy some concerns from companies’ legal/management teams

Highlights from conversation with the agencies

- The pilot project workflow is adjacent to, but not a part of, regulatory processes
- Like other ocean use and resource datasets on the Portal, regional habitat products provide context and a starting point for additional analyses or data collection that would be required as part of an individual project/action
- The three project outputs will help compare across projects and regionally, provide context for impact assessment and monitoring, and be used to inform communication with the public

Seafloor Data Request

The following datasets are requested to serve as "building blocks" of regional habitat products that would address regional management/regulatory questions of interest and be continually updated as new data are collected and received. This request also includes metadata/reports associated with each variable; accompanying tabular data may be provided (CSV is preferred but Excel will also be accepted). Each dataset will be integrated and down-sampled into regional-scale data products. Vertical datum and horizontal coordinate system of received datasets will be converted to a common framework by the INSPIRE/NROC project team. The desired resolution of the multibeam data is set to be compatible with the finest resolution of the NOAA National Bathymetric Source. The requested data do not include side scan sonar backscatter, archaeological targets, or UXO targets. The INSPIRE/NROC project team is aware that the point/polygon variables listed below may be available under different names or in different format; any and all related data (e.g., grain size analysis results) are welcome. A mutually agreeable file transfer process will be used to provide data from [DATA PROVIDER] to INSPIRE and NROC; INSPIRE can set up a file share option and credentials as needed.

Seafloor variable(s)	**Type/format**	**Desired resolution**
Multibeam bathymetry	Raster/tiff	8m*
Multibeam backscatter	Raster/tiff	8m*
Boulder fields and/or picks (identification of boulder locations)	Vector-points/polygons ESRI shapefile	N/A
Seabed interpretation – CMECS Substrate Group, Subgroup	Vector-points/polygons/ESRI shapefile	N/A
Seabed interpretation – Shell substrate	Vector-points/polygons/ESRI shapefile	N/A
Seabed interpretation – Bedforms	Vector-polygons/ESRI shapefile	N/A
Seabed interpretation – CMECS Biotic Subclass	Vector-points/ESRI shapefile	N/A

*Bathymetry and backscatter data sets may be provided at high resolutions and will be down-sampled to the desired 8-m resolution by the project team; alternatively, down-sampling instructions will be provided to [DATA PROVIDER] to ensure consistent methodology.

Assumptions

Data shared under this MOU will be kept confidential and secured and will not be shared beyond the parties to this agreement and their subcontractors without explicit, written permission from [DATA PROVIDER]. Data shared may be preliminary data that will be updated with a new data delivery. The models and data products built under this pilot study will be updated with new data when provided. Data products generated during this pilot study will be vetted and approved by the Seafloor Habitat Data Work Group, including the [DATA PROVIDER] before they are shared publicly. This agreement shall not create legally enforceable rights and cannot be the basis of any legal claim between INSPIRE, NROC, and [DATA PROVIDER].

Duration and Termination of Agreement

The duration of this agreement will be for five years from the signature date of [DATA PROVIDER] and may be extended or superseded by future agreements (through the RWSE or otherwise). Termination of the agreement before this time may be initiated by either party in writing with a notice period of 30 days.

___________________________________ ___________________

[DATA PROVIDER] Date

___________________________________ ___________________

Drew A. Carey, CEO, INSPIRE Environmental Date

___________________________________ ___________________

NROC Date

Appendix B - Data Sources

Title	Author	Date	Type	Format	Hyperlink
Sediment					
usSEABED Database	U.S Geological Survey (USGS)	1960 - 2020	Sediment grabs, images, box cores	Excel Spreadsheet	https://www.usgs.gov/publications/sediments-and-sea-floor-continental-shelves-and-coastal-waters-united-states-about
East Coast Sediment Texture Database	USGS	2014	Sediment grabs, images, box cores	GIS feature class (.shp)	https://woodshole.er.usgs.gov/project-pages/sediment/
New England (Park City/Commonwealth) Wind Farm	Epsilon Associates INC & RPS	2019	Sediment grabs	Table 8 in report	https://www.boem.gov/sites/default/files/documents/renewable-energy/VW-South-COP-Volume-II-H.pdf
Sea-Floor Geology in Central Rhode Island Sound South of Sakonnet Point, Rhode Island	Kate McMullen & USGS	2012	Sediment grabs and images	GIS feature class (.shp)	https://pubs.usgs.gov/of/2012/1004/title_page.html
Revolution Wind Farm	Orsted & INSPIRE Environmental	2019	Sediment Profile Imagery & Plan View (SPI/PV) images	Attachment A in report "*Benthic Assessment Technical Report*"	https://www.boem.gov/sites/default/files/documents/renewable-energy/state-activities/App_X1%20BenthicAssessment.pdf
South Fork Wind Farm	Orsted & INSPIRE Environmental	2017	Plan View (PV) images	Attachment A in report "*Sediment Profile and Plan View Physical Ground-Truth Survey in Support of the South Fork Wind Farm Site Assessment*"	https://www.boem.gov/sites/default/files/renewable-energy-program/State-Activities/NY/App-H4_SFWF_SPI_Geophysical_DataReport_2019-05-15.pdf
Sunrise Wind Farm	Orsted & INSPIRE Environmental	2020	Sediment Profile Imagery & Plan View (SPI/PV) images	Attachment A in report "*Appendix M1 Benthic Resources Characterization Report – Federal Waters*"	https://www.boem.gov/sites/default/files/documents/renewable-energy/state-activities/SRW01_COP_AppM1_Fed_BenthicResCharacReportandFigures_2022-08-19_508.pdf
SMAST Drop Camera Survey	Umass Dartmouth	2019	Towed Video	Excel Spreadsheet	https://static1.squarespace.com/static/5a2eae32be42d64ed467f9d1/t/5efe5321529d2c2a9da96665/1593725732956/SMAST+Drop+Camera+501S+2019+Final+Report.pdf
Mayflower Wind Farm	AECOM	2020	Sediment Profile Imagery & Plan View (SPI/PV) images	Appendix A1 within report "*Appendix M. Benthic and Shellfish Resources Characterization Report*"	https://www.boem.gov/sites/default/files/documents/renewable-energy/state-activities/Appendix%20M_Benthic%20and%20Shellfish%20Resources%20Characterization%20Report.pdf

Title	Author	Date	Type	Format	Hyperlink
Geophysical					
NOAA NOS surveys: H11920, H11922, H11995, H11996, H12009, H12010, H12011, H12015, H12023, H12137, H12139, H12296, H12386, H12429, H12430, H12431, H12700, H12702, H12707, H12801, H12802, H12811	NOAA National Centers for Environmental Information	2008, 2009, 2011, 2012, 2013, 2014, 2015, 2021	Bathymetry (2-m resolution); Bathymetric Attributed Grids (BAGs)	Data available in multiple formats for download	Bathymetric Data Viewer; https://www.ncei.noaa.gov/maps/bathymetry/
Bathymetry grid of Vineyard and western Nantucket Sounds	Wayne Baldwin, USGS	2016	Bathymetry (10-m resolution); Single and multibeam sonar	ZIP file	https://pubs.usgs.gov/of/2016/1119/GIS_catalog/SourceData/bathy/vns10m_navd88.html
NAMERA Benthic Data	The Nature Conservancy	2010	Bathymetry (250-m resolution)	ZIP file	https://www.sciencebase.gov/catalog/item/5408c9b7e4b0621a5983c2a7

Appendix C - Down-Sampling Protocol

Regional Habitat Mapping Down-Sampling Protocol

August 22, 2022

1.0 Introduction

Funded by the Massachusetts Clean Energy Center (MassCEC), INSPIRE Environmental and the Northeast Regional Ocean Council (NROC) are working together to develop standard approaches for synthesizing, visualizing, and disseminating high-resolution geophysical and imagery data to advance benthic habitat mapping in the wind energy areas off the Northeastern United States (US).

Federal agencies and other management bodies require the synthesis and visualization of diverse datasets to accurately assess and mitigate the potential impact of human activities in ecologically and economically important ecosystems. As offshore wind development gains momentum in the Northeastern US, there is a substantial need to characterize and display the distribution of benthic habitats in wind farm lease areas and cable route corridors. Understanding benthic habitat distribution is necessary not only for engineering logistics but also for the important ecological value these environments provide, for example, to commercial fisheries.

This project will develop best practices for 1) integrating geophysical and benthic assessment data to map benthic habitats; and 2) making those habitat data available to federal and state regulators and stakeholders in a vetted and established forum, specifically the Northeast Ocean Data Portal.

The protocol detailed below outlines how geophysical data are prepared and standardized prior to their incorporation into regional composite data products. The requested geophysical data include multibeam bathymetry and backscatter. If data providers prefer to down-sample their own high-resolution data before sharing them to the project team, the following protocol should be followed to ensure consistency to the extent feasible.

2.0 Methods

Down-sampling is the process of averaging raster cell values over larger areas, effectively creating a coarser raster dataset with fewer cells and a smaller file size. Such procedures are vital to combining multiple datasets that span large spatial extents while preserving enough data resolution for regional assessment purposes. As such, the goal herein is to detail the process of reducing raster data file sizes, which promotes more efficient data processing routines and online portal usage, without greatly compromising the resolution of said data. Input raster data will be down-sampled to a resolution of 4 meters.

The Resample tool available in Esri's ArcMap and ArcPro converts the resolution of a raster dataset by interpolating values across new pixel cells. There are four techniques for resampling (down-sampling) spatial data in ArcGIS: nearest neighbor, majority, bilinear, and cubic. Bilinear interpolation was determined to be the best option for this project because it is the most commonly used interpolation technique for continuous data, i.e., raster datasets. Additional tests on sample data with varying cell sizes over a range of depth values indicated the bilinear approach produced down-sampled raster files with the best data quality overall (i.e., minimized overestimation, data range preservation) and required minimal computational effort. The Resample tool can be utilized in ArcMap Versions 10.6, 10.7, and 10.8, and in ArcPro Versions 2.7, 2.8, and 2.9 with basic, standard, and advanced licenses.

3.0 Protocol – ArcMap (Version 10.7.1)

3.1 Input Data and Setup

Input raster datasets should be delivered with depths (z-axis) as negative values and specific mention of the coordinate system and vertical units applied to each dataset. These datasets can be delivered in any commonly used, industry standard format such: *.TIF/TIFF; *.GRD; *.FLT. Please do not submit a rendered RGB or hill-shaded format.

1. File name and directory (i.e., folder) structure:

   (a) Files should be named using the following naming convention: “EntityCode_DataYear_DataTypeCode_Resolution”

For example: OR_21_BT_50cm

Where EntityCode = OR (Orsted), DataYear = 21 (2021), DataTypeCode = BT (Bathymetry), Resolution = 50cm

These project codes will be provided to developers or INSPIRE Environmental personnel will rename data upon receipt.

   (b) Folders should be organized as below (Figure C-1):

- One main folder: Project_fromEntity

- One subfolder for the source bathymetry data: Project_DataYear_Bathy_Resolution
- One subfolder for the source backscatter data: Project _DataYear_BS_Resolution
- One subfolder for the down-sampled bathymetry raster: Project _DataYear_Bathy_Resolution
- One subfolder for the down-sampled backscatter raster: Project _DataYear_BS_Resolution

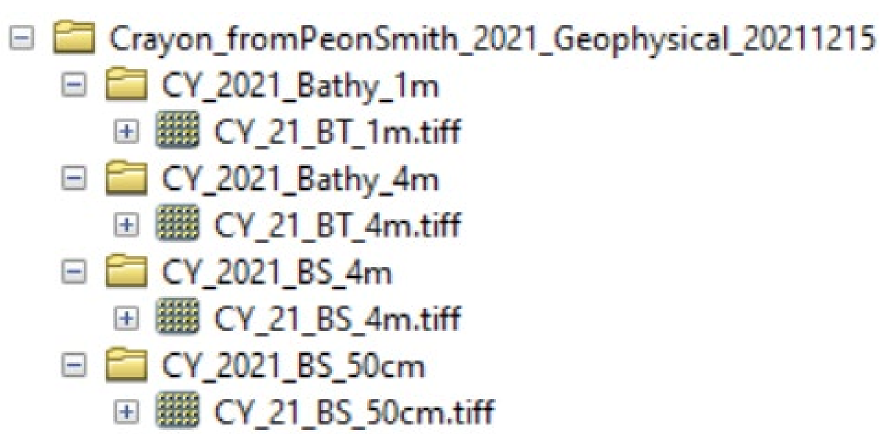


Figure C-1. Preferred folder structure (completed in ArcCatalog or File Explorer).

2. Map document and geoprocessing properties:

   (a) Open a new, blank ArcMap .MXD file

   (b) Leave all geoprocessing environment settings as default

3.2 Procedure

1. Adding raster files to the map:

   (a) Add the original raster files to the map using either the Add Data tool or the Catalog Tree.

   (b) Symbolize the raster files using the below properties:

Bathymetry: Minimum-Maximum stretch type, with the Invert box checked as below. These settings are within the Symbology tab of the Layer Properties (Figure C-2A).

Backscatter: Standard Deviation stretch type (n = 2). These settings are within the Symbology tab of the Layer Properties (Figure C-2B).

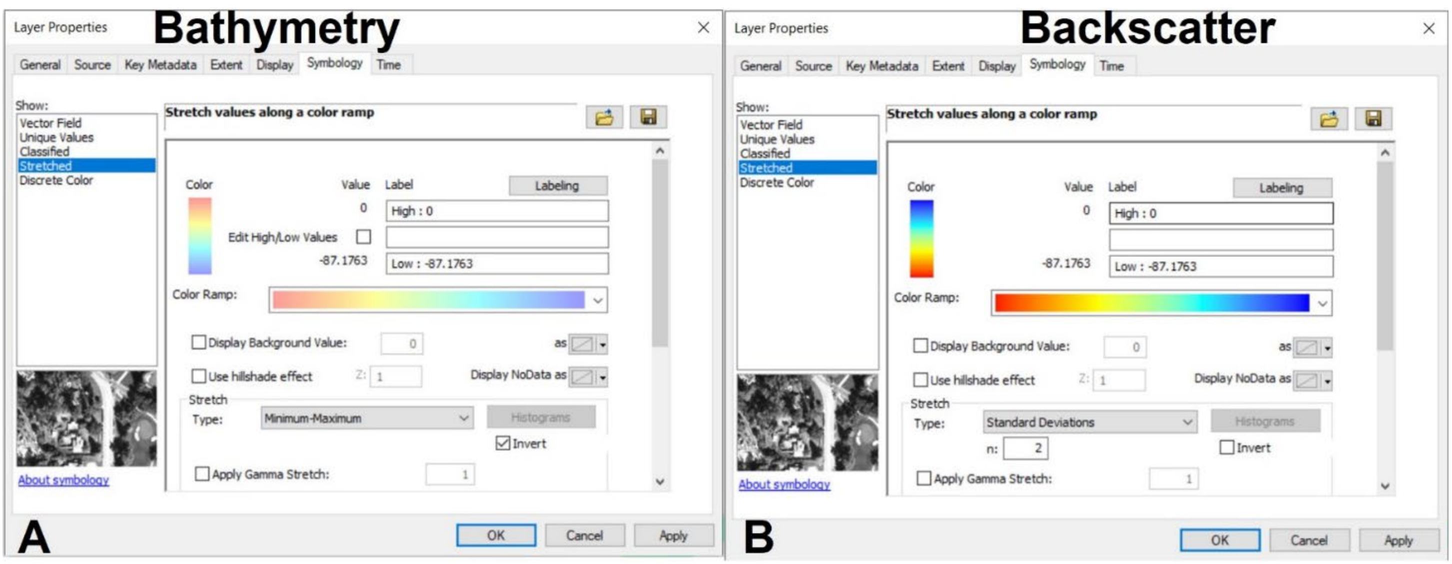


Figure C-2. Preferred symbology for (A) bathymetry and (B) backscatter raster file datasets.

2. Resample

(a) Search for the Resample (Data Management) tool using the search function in ArcMap (Figure C-3).

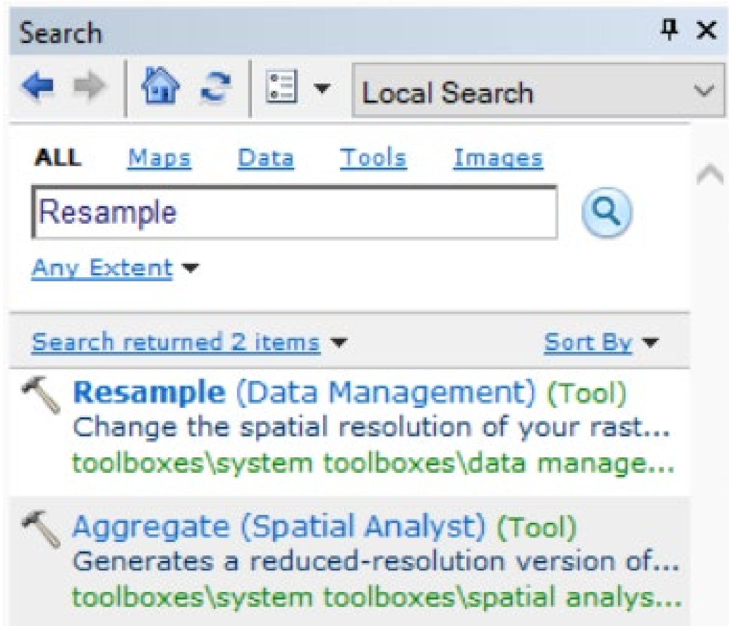


Figure C-3. The Resample tool can be found using the Search function in ArcMap.

(b) Select the Resample tool to open the tool's dialog box.

(c) Fill out the tool's parameter fields using the guidance below and Figure C-4.

i. Input Raster: the original data, stored in the folder created within Step 1(b) of Section 3.1. If the raster was added to the .MXD file, it can also be selected by using the dropdown menu (blue circle, Figure 4).

ii. Output Raster Dataset: location and name of the desired down-sampled raster. Use the folder icon (green circle, Figure 4) to navigate to the folder created in Step 1(b) of Section 3.1.
Note: *The output raster's name can only be 13 characters long. Therefore, it is essential that the file is named appropriately. The tool will signal if the name is too long with a red X icon by the Output Raster Dataset line.*

iii. Cell Size: enter the desired cell size (e.g., 8) for X and Y.

iv. Resampling Technique: the default is set to Nearest Neighbor. Change to Bilinear.

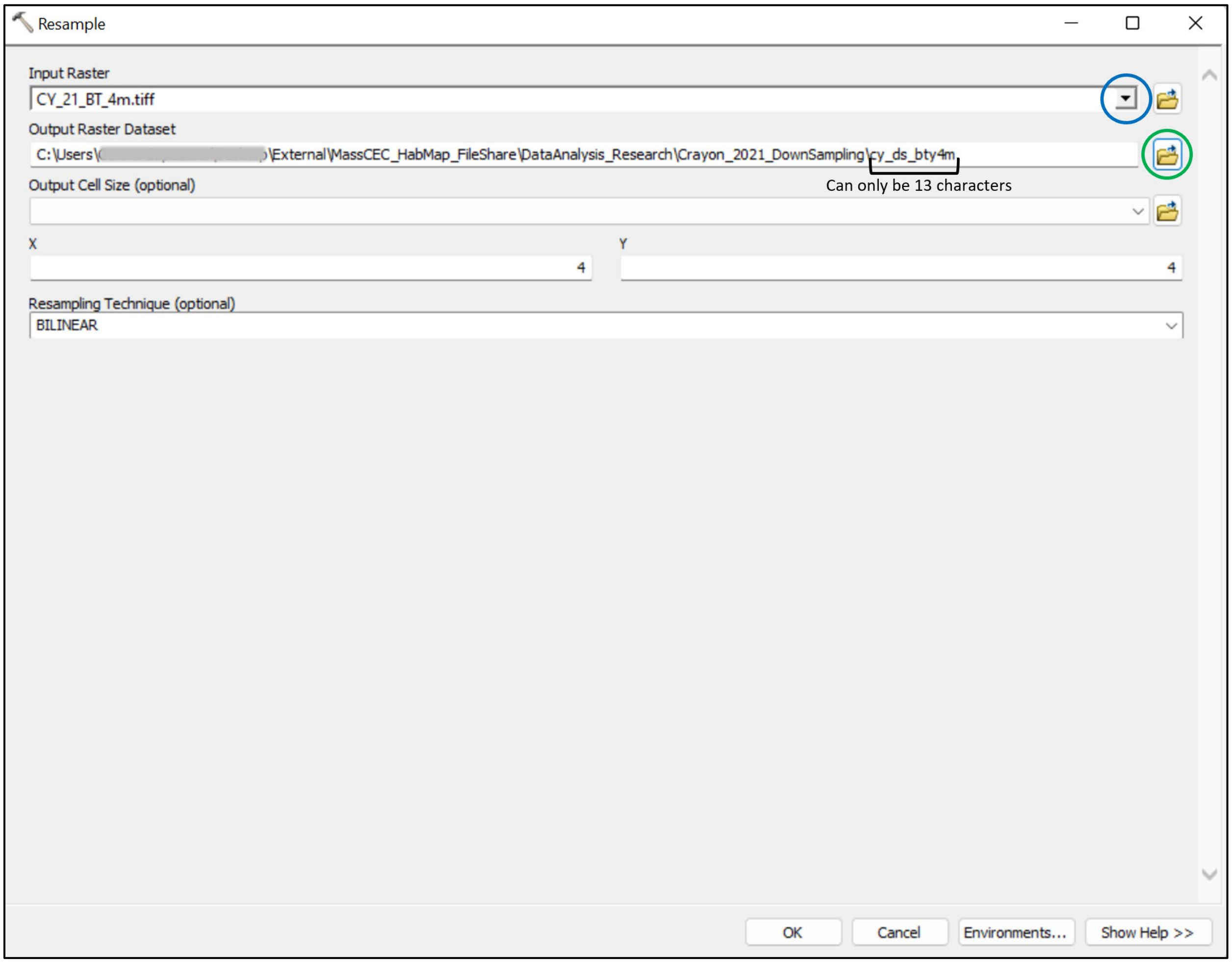


Figure C-4. The appropriate parameters and settings for the Resample tool.

(d) Symbolize the down-sampled raster using the same symbology identified in Step 1 of Section 3.2.

(e) Repeat the process for each bathymetry and backscatter raster file dataset as needed.

3. Creating a down-sampled hill-shaded relief model raster.

   (a) To create a hill-shaded raster (i.e., hill-shaded relief model) from bathymetry data, follow these steps:

      i. Under the Windows tab, select Image Analysis (Figure C-5).
      ii. Select the down-sampled bathymetry raster in the Image Analysis side panel (Figure C-6, blue rectangle).
      iii. Under the Display section, change the interpolation method to Bilinear Interpolation (Figure C-6, red rectangle).
      iv. Under the Processing section within the panel, select a black to white gradient color ramp (Figure C-6, green rectangle).
      v. Select the Shaded Relief icon next to the color ramp to create a hill-shaded relief model (Figure C-6, purple circle).
      vi. Symbolize the hill-shaded relief model using the symbology show in Figure C-7. When the Stretch Type is changed to Minimum-Maximum, allow ArcMap to calculate statistics if prompted.

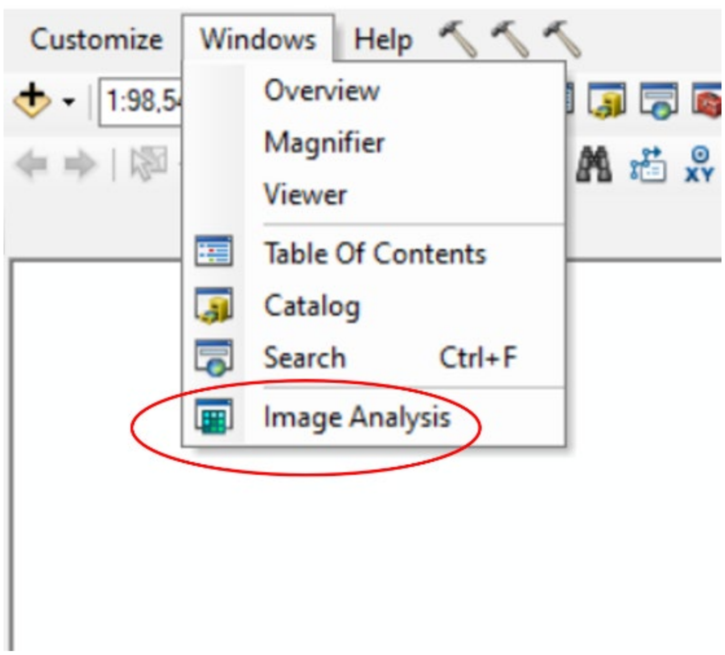


Figure C-5. The Image Analysis tool can be found under the Windows tab.

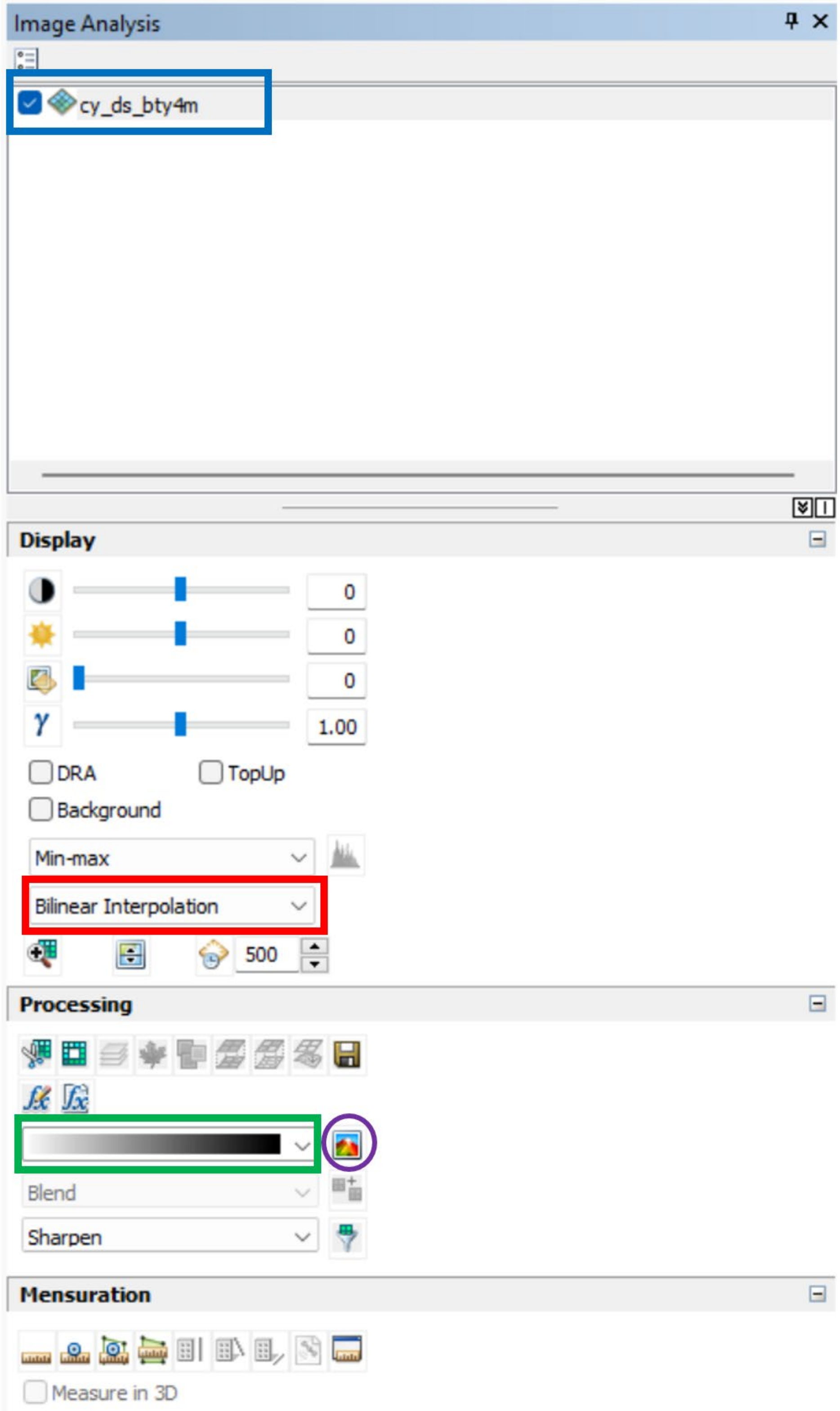


Figure C-6. Parameters within the Image Analysis tool to create a hill-shade relief model.

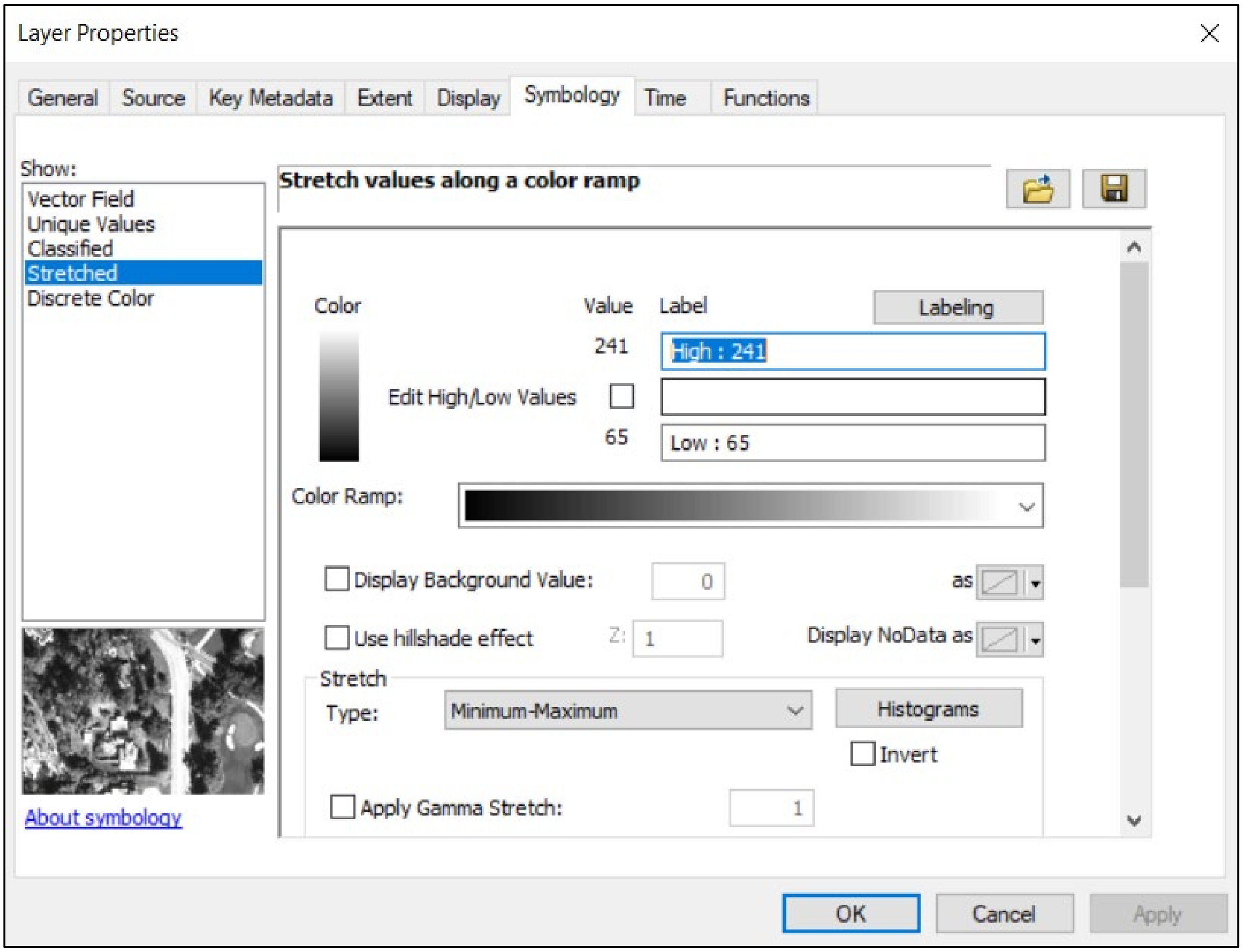


Figure C-7. The preferred symbology for bathymetry hill-shaded relief models.

4. Exporting Data

   (a) Down-sampled raster files are already saved within the folders created in Step 1(b) of Section 3.1.

   (b) Created hill-shaded relief models will need to be exported because they originally exist as temporary files (Figure C-8). This can be confirmed in the Source tab of the hill-shaded relief model's layer properties.

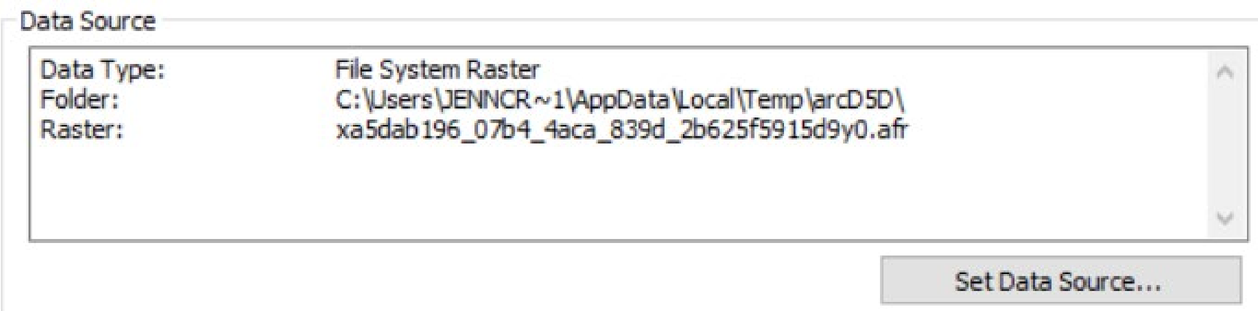


Figure C-8. Layers created in the Image Analysis tool are stored in a temporary area by default.

   (c) To export a hill-shaded relief model, right click on the layer name in the Table of Contents and select Data, then select Export Data.

   (d) Navigate to the folder with the down-sampled bathymetry raster file and set that folder as the Location (Figure C-9).

   (e) Name the hill-shaded relief model (remember, the name has a limit of 13 characters).

   (f) Select *.TIFF as a Format.

   (g) Save the file.

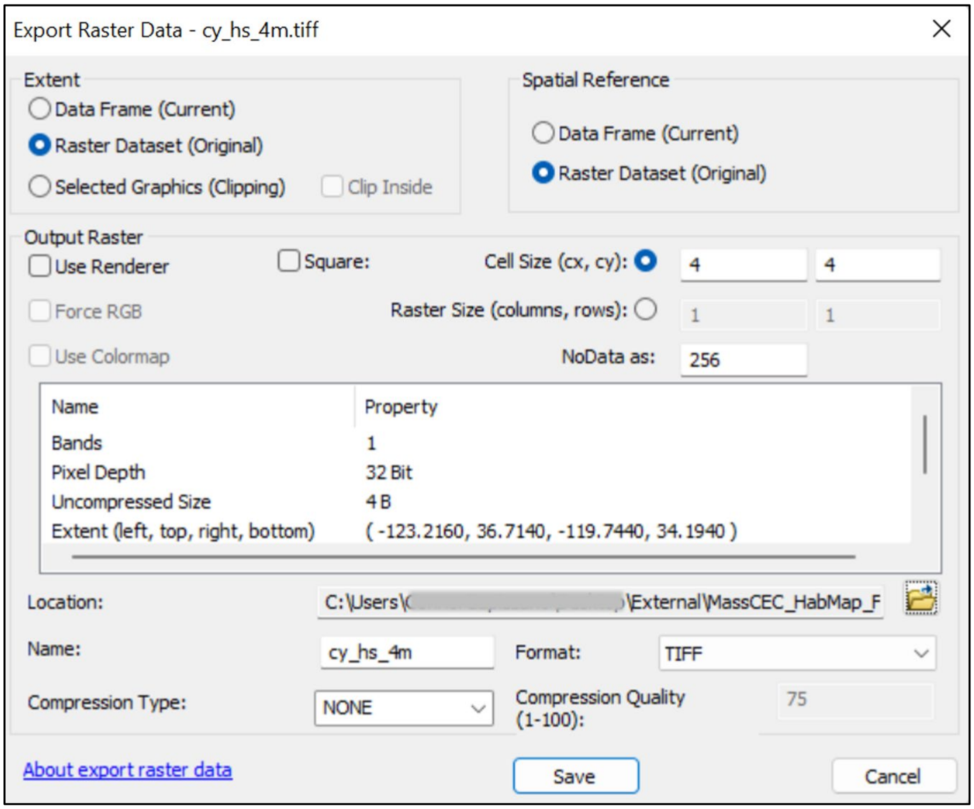


Figure C-9. The settings to properly export a temporary raster as a *.TIFF.

Appendix D - Mosaic Protocol

Mosaic Protocol

The current protocol was conducted in Esri's ArcMap Desktop 10.8.2, but it can be replicated in most versions of ArcMap. Although mosaicking is also possible in ArcPro, it is not covered in this document.

Creating Mosaics

1. Import all raw bathymetry raster files (i.e., GeoTIFFs; .tif) into the ArcMap document. Please review Esri's online documentation if issues arise during the raster file import.

2. Navigate to the "Image Analysis" panel by selecting the "Windows" tab from the main navigation ribbon at the top of the ArcMap document (Figure D-1).

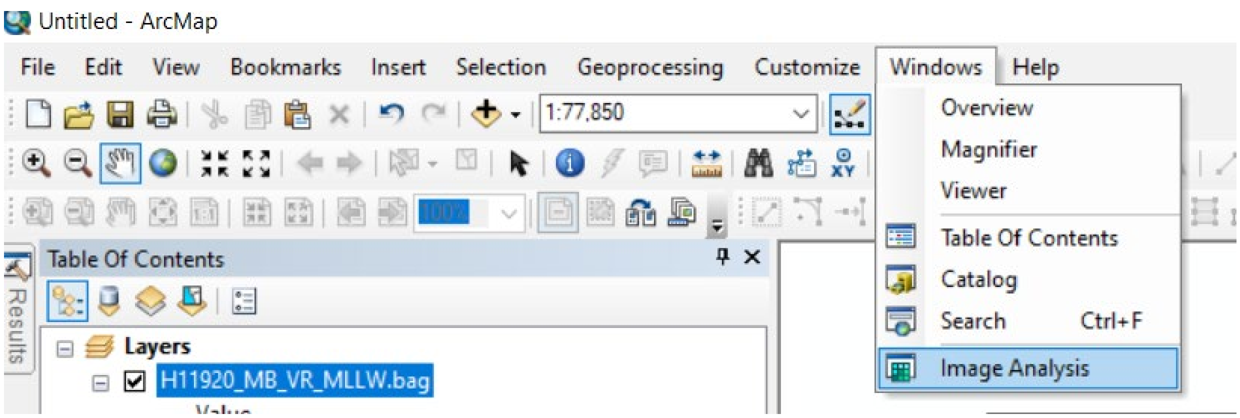


Figure D-1. The Image Analysis tool can be accessed under the Windows tab.

3. Within the "Image Analysis" panel, choose all input raster files (Figure D-2) and ensure that the *Blend* option is selected under the "Processing" section. To create the mosaic, select the mosaic icon from the panel (red circle, Figure D-2).

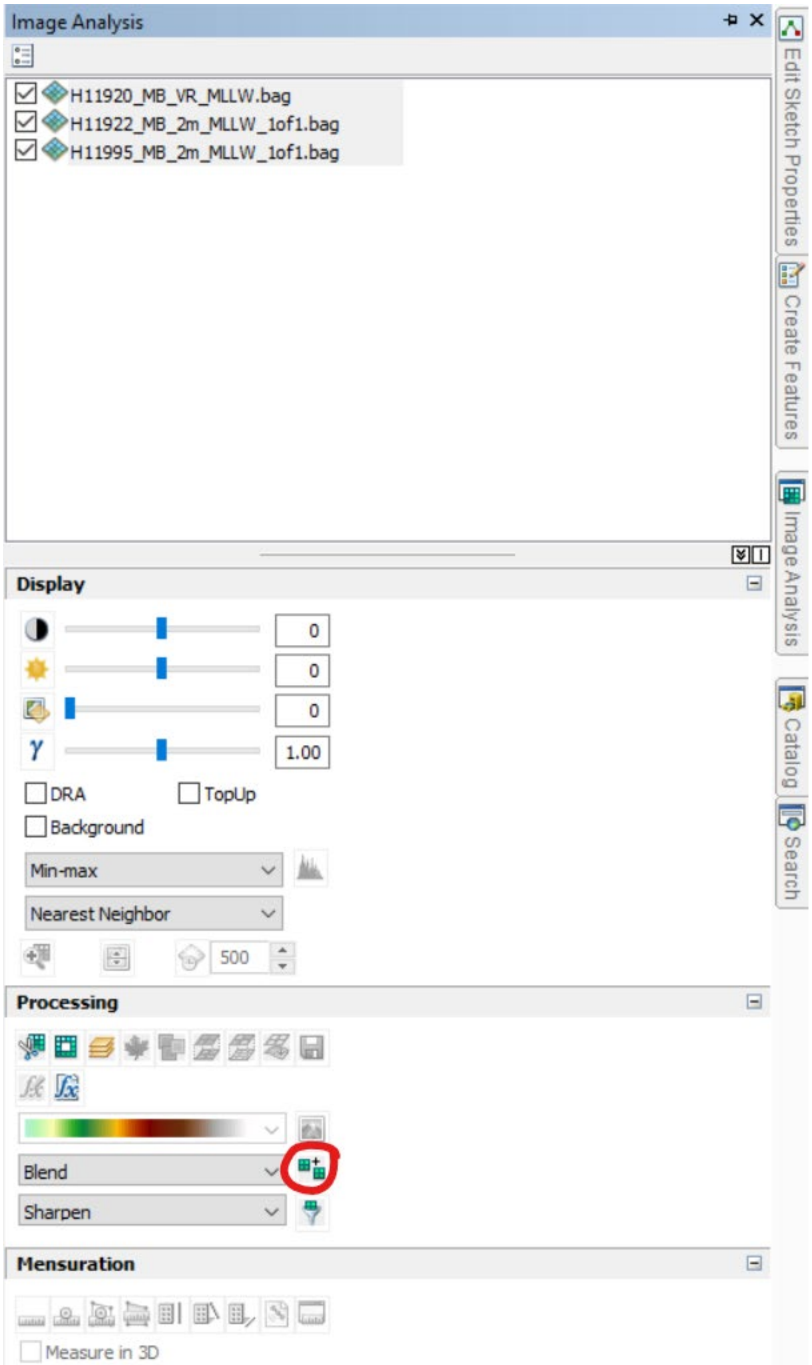


Figure D-2. Settings for creating a mosiac within the "Image Analysis" panel.

5. The mosaic will automatically generate and appear in the "Table of Contents", where the name usually duplicates the name of the first input raster file.

6. Define the mosaic's symbology by right-clicking the mosaic and selecting the "Layer Properties" option. Ensure the symbology of the mosaic is "Stretched" and the "Stretch

Type” is *Minimum-Maximum*. The type of color that is stretched between the values is not relevant at this time, but it typically defaults to a black and white color ramp (Figure D-3).

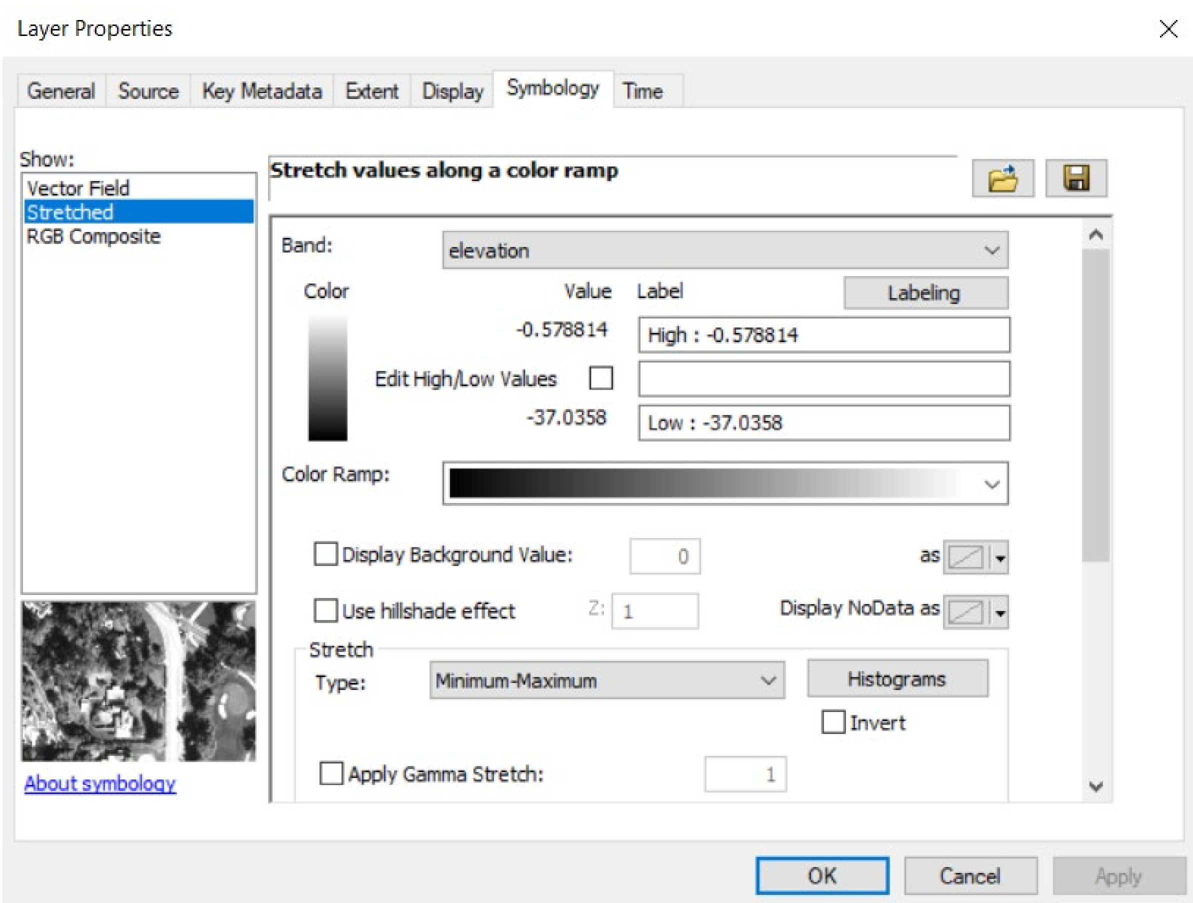


Figure D-3. The suggested symbology settings within the “Layer Properties” dialog window for the input raster files.

7. Although visible in the document, the mosaic is a temporary file and needs to be saved for future use.

   To save, right click the temporary mosaic in the “Table of Contents” and select “Data” > “Export Data…” (Figure D-4).

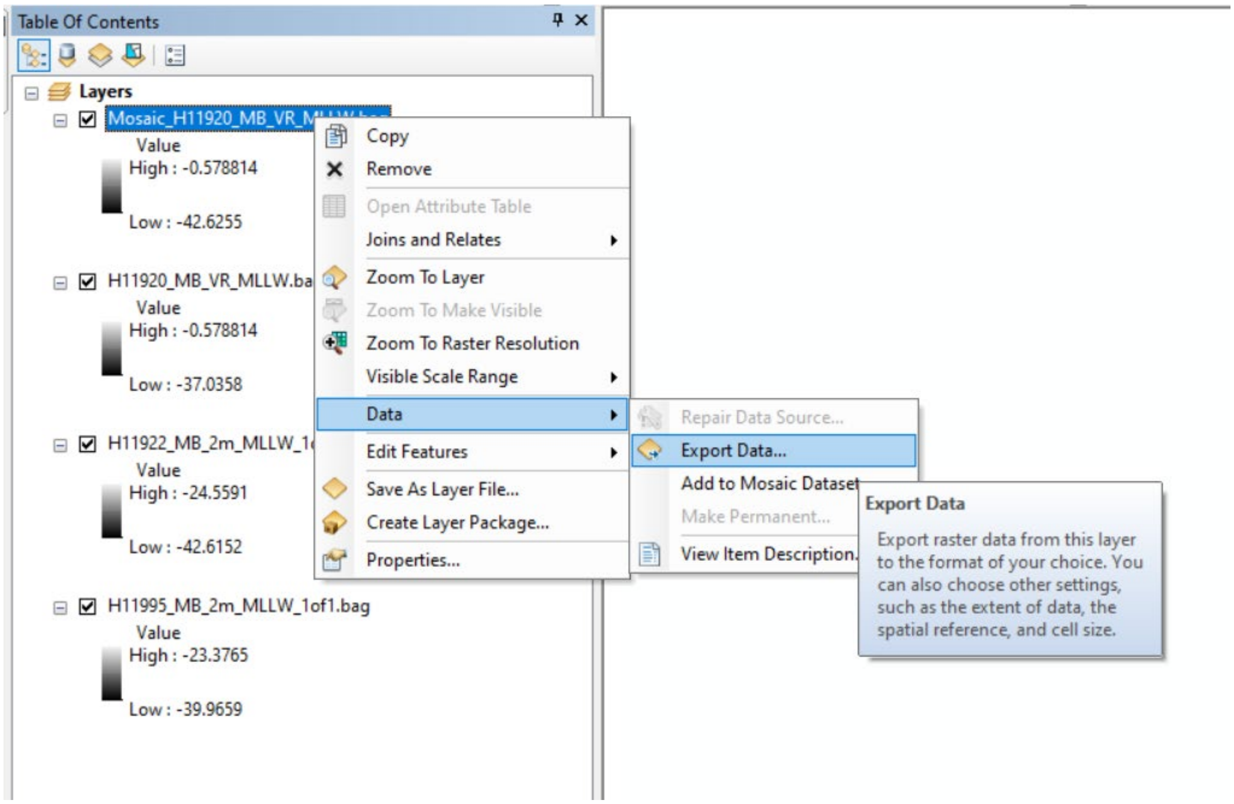


Figure D-4. Navigating to the Export Data feature to save the temporary mosaic file.

8. Within the “Export Raster Data” dialog window, specify the file’s “Location”, “Name”, and “Format”. The remaining fields should use the default settings (Figure D-5).

    - **Note:** By default, the character limit of names is 13 characters. This limit can be bypassed by adding “.tif” to the end of the “Name” field.

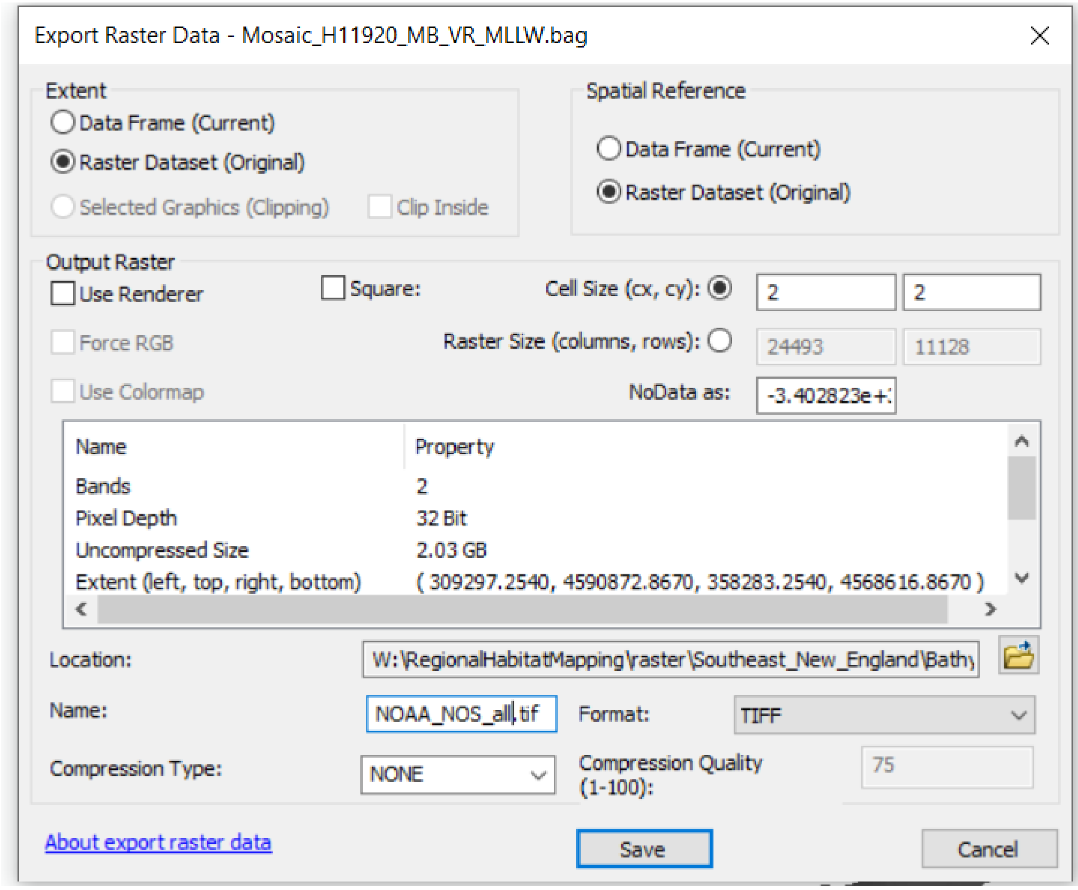


Figure D-5. Recommended settings for using the "Export Raster Data" dialog window.

9. The exported mosaic file will now serve as the final raster product for sharing and analyses.

Appendix E - Sediment Sampling Spreadsheet Template

Project_ID	Survey_ID	Station_ID	Replicate	Date	Time	Latitude_ WGS84	Longitude_ WGS84	Depth	Sample_Type	CMECS_Group*	Grain_Size*	Phi_Size*
EXAMPLE	EXAMPLE_1	101	A	11/30/2022	8:02:20	40.98333	-70.67367	49	SPI/PV	Gravel	2 to <4,096	-1 to <-12
EXAMPLE	EXAMPLE_1	101	B	11/30/2022	8:04:46	40.99333	-70.68322	52	SPI/PV	Sand	0.0625 to <2	4 to <-1
EXAMPLE	EXAMPLE_1	101	C	11/30/2022	8:06:33	40.97333	-70.66872	46	SPI/PV	Mud	<0.0625	>4

Notes:

Column	**Definition**	**Accepted Format**
Project_ID	Project abbreviation/ID	Alphanumeric
Survey_ID	Survey ID for project	Alphanumeric
Station_ID	Station ID for survey	Alphanumeric
Replicate	Replicate ID for station	Alphanumeric
Date	Date sample was collected (MM/DD/YYYY)	Month-Day-Year with leading zero
Time	Time sample was collected (HH:MM:SS)	24-hour time notation
Latitude_WGS84	Latitude where sediment was collected (DD.dd)	Decimal degrees (DD.dd)
Longitude_WGS84	Longitude where sample was collected (DD.dd)	Decimal degrees (DD.dd)
Depth	Depth at which sample was collected (m)	Numeric
Sample_Type	The type of sampling method used (e.g., Sediment Profile Imaging [SPI], Plan View [PV], Sediment Grab, Drop Camera, etc.)	Category
CMECS_Group*	Substrate Group as defined by the Coastal and Marine Ecological Classification Standard (CMECS)	Category
Grain_Size*	Measured sediment grain size major mode from sample (mm) or Folk classification derived from grab sample [if available]	Numeric
Phi_Size*	Measured sediment grain size major mode from sample (phi scale, Wentworth 1922) [if available]	Category

*For sediment classifications, the CMECS_Group column (green highlight) is the recommended field for reporting sediment type; the Grain_Size and Phi_Size columns (yellow highlight) are secondary if data are not available/formatted for the CMECS_Group column.

Appendix F - Task 4: Data Hosting and Delivery Repository; Data Delivery Protocols

Task 4: Data Hosting and Delivery Repository; Data Delivery Protocols

Prior to meetings with the RWSC Habitat and Ecosystem Subcommittee, participants were asked to complete an online survey to better understand their beliefs and attitudes towards hosting and sharing site characterization data. In the context of this appendix, participants were asked what they believed were the greatest barriers to hosting and interacting with site characterization data. In brief, participants collectively believed that “archiving and transferring large volumes of data” was the most significant obstacle to sharing these data (Figure F-1).

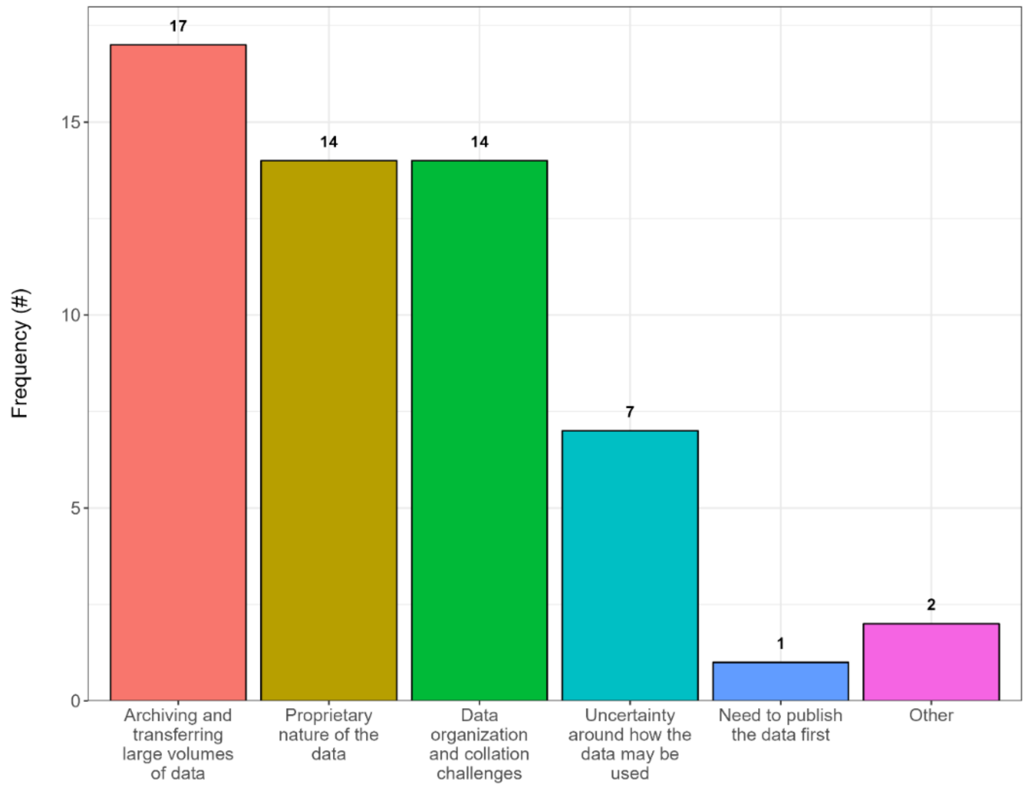


Figure F-1. Responses of RWSC Habitat & Ecosystem Subcommittee members towards the biggest barriers to sharing site characterization data (e.g., seafloor data, benthic community data).

The following appendix was therefore crafted to provide stakeholders with technical material and recommendations for guiding future decisions in data hosting, maintenance, and sharing. Such materials include (1) the strengths and limitations of on-premises and cloud computing services, (2) cloud computing service providers, (3) lessons learned from an established geospatial data repository, and (4) technical recommendations for future consideration.

1. On-Premises vs. Cloud Computing Services

The fundamental difference between on-premises and cloud computing services is where they reside and who is responsible for maintenance. On-premises computing services, for instance, involve a group of servers that a company privately owns and controls locally. In contrast, cloud computing services lease data center resources (e.g., storage, processing) from a third-party servicer provider online. Although on-premises computing services boast absolute control over a system, they pose a significant upfront cost, require maintenance and dedicated staff, lack scalability, and have increased security risks. In today's digital age where speed and flexibility are vital, cloud computing allows companies and organizations to dynamically scale data storage and applications to changing demands and establish new infrastructure without the cost of software, hardware, and information technology (IT) personnel. Furthermore, cloud computing services are not only for storage but also serve as a valuable tool that can aid in unit and integration testing, delivery systems, and live tools and web applications.

Additional comparisons between on-premises and cloud computing services can be found below in Figure F-2 and Table F-1.

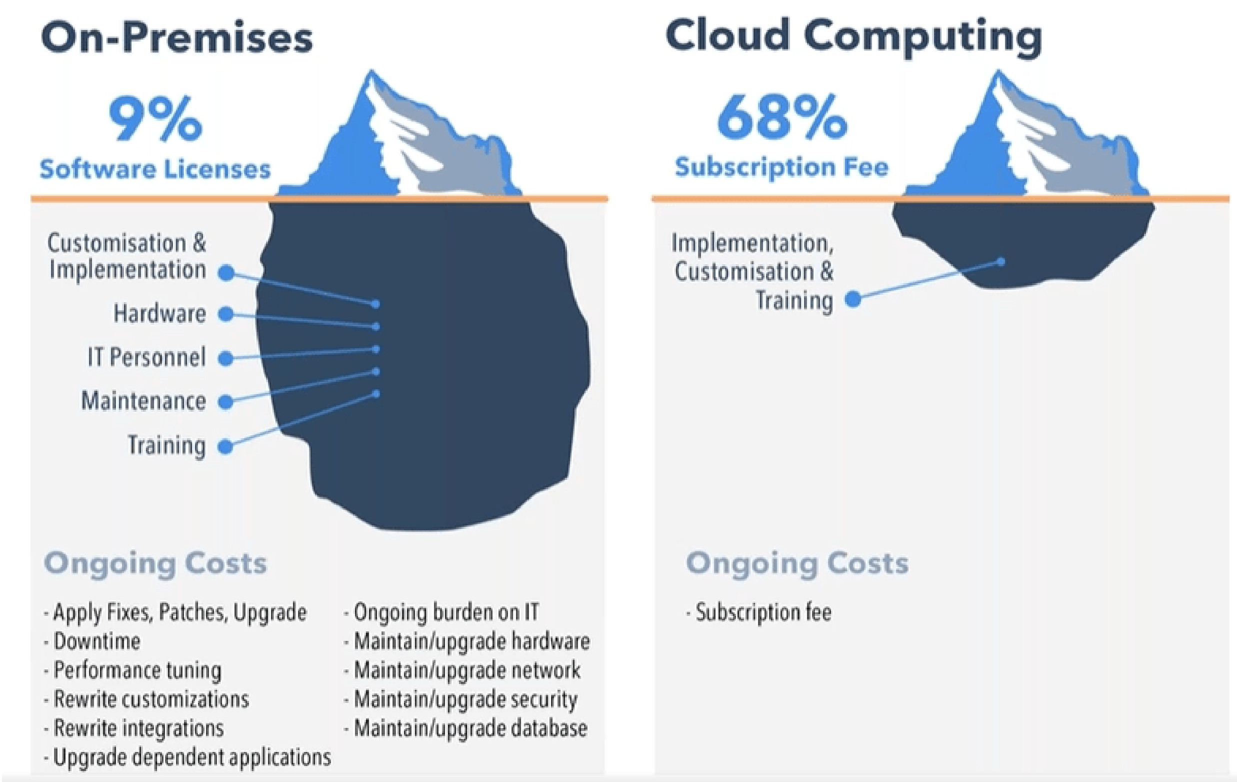


Figure F-2. Infographic comparing on-premises (left) and cloud computing (right) services. Source: https://www.peoplehr.com/blog/2015/06/12/saas-vs-on-premise-hr-systems-pros-cons-hidden-costs/.

Table F-1. Detailed Comparison Between On-premises and Cloud Computing Services

	On-premises	Cloud
Costs and maintenance	• Large up-front cost in hardware, installation, software licenses, data backup, extra IT services, support, etc. • Ongoing IT support after setup for maintenance, energy costs, physical storage space, software/hardware updates • Replacement and new equipment costs • Will not have to pay subscription fees for cloud service	• No enormous upfront costs or those associated with maintenance, up-to-date software, security and support; only subscription fee • No costs associated with troubleshooting issues • Base subscription plans include many benefits (e.g., large storage, security, file sharing) • Additional fees can occur due to activating new features, scaling up performance
Security/ threat protection	• Required to create a security system and responsibility is yours alone • Require additional staff to monitor and maintain network, firewall, encryption, and secure access control • Can still suffer cyber-attacks as well as physical attacks	• Security is better but not infallible • Subscriptions include cyber security and access to experts and multi-layer security options (e.g., threat monitoring, mass file deletion protection, etc.)
Compliance	• Will need to maintain systems so they are compliant with industry-specific regulations • Need to have necessary employees and resources on hand • Failure to meet compliance rules falls on responsibility of organization • Audits and fines fall on organization if not compliant	• Can select cloud providers that can assist with achieving compliance requirements
Scalability	• Cannot handle changing workloads and will need to add resources (e.g., hardware and software) • Requires money, labor, expertise, installation, hardware, software, monitoring systems, and time • If processing demand is short-lived, your spending will be highly inefficient	• Scalability can be done instantly using built-in features (scale up, scale down, auto scale) • Processing flexibility cuts out overhead costs associated with monitoring and scaling resources manually
Reliability	• Available to organization on-premises • Limits access to those in the office, as compared to those working remotely • Requires power, backup power, and storage backup system, which add to costs	• Accessible to on-premises and remote individuals • Requires fast, reliable Internet connection • Breaks in connectivity can delay operations
Data backup	• Requires space, power, hardware for efficient backups • Can have backups on-premises or off-site • Many organizations use the cloud for at least a portion of their data protection strategy	• Reliable cloud storage provider offers many features to avoid data loss, such as built-in redundancy, failover, backup, automatic logging, monitoring • Allows for shorter recovery time compared with on-premises

	On-premises	Cloud
Access	• Good choice if most workers are in the office and rarely mobile • Effective option if working with large files and video • VPN is an option but can be overwhelmed	• Virtual desktop as compared to Virtual Private Network (VPN) • Can access anything in the office as long as have Internet connection • Perfect for situations where applications need to be running 24/7 and be available • Allows for real-time collaboration, better version control, and easy file sharing
Integration of apps and systems	• Legacy line of business apps that directly access local file server	• More suitable for using modern applications for accessing data • Maintained by provider and scalable for mobile use

2. Cloud Computing Service Providers

Given the numerous benefits associated with cloud computing services, the question no longer fixates on whether to opt into cloud computing but what service provider to use. Like on-premises solutions, a suite of global service providers exists to support cloud computing goals. Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) are the top three cloud computing service providers based on popularity and market share. Each provider offers organizations an excellent array of online storage, computing, and networking services, but some are better than others and the decision will depend on an organization's specific requirements.

AWS, for instance, is the most comprehensive and established cloud platform that offers smooth integration with a variety of third-party services. Despite its broad adoption by organizations, AWS requires specialists to run properly and has a complicated pricing plan. Azure is an extensive and simple-to-use cloud solution used by many enterprise clients as it easily integrates with Microsoft services already used by many and offers hybrid cloud infrastructure. Due to favoring Microsoft products, Azure has a difficult time integrating with open-source products and has limited customization. Finally, GCP offers a simple pricing model and integrates well with other Google services. Although considered a leader in AI and ML-powered cloud data processing, GCP does not provide an extensive selection of cloud services or compatible programming languages.

Overall, AWS takes the lead for global availability and long-term status in the cloud market, Azure as the champion for number of services offered and integration within a Microsoft environment, and GCP for its customer-friendly pricing models.

3. Lessons Learned: The Crown Estate's Marine Data Exchange

Given the number of available cloud computing service providers, discussions with organizations with successfully deployed data repositories were essential. To this end, our project team met with personnel from The Crown Estate on several occasions to discuss the Marine Data Exchange (MDE), an online data repository for sharing industry- and research-derived seafloor data to a wide audience of end users. In addition to navigating relationships with data providers and developing standardized protocols, discussions with MDE's Marine Evidence Manager Chelsea Bradbury and Senior Product Owner Tazz Shaw have yielded valuable recommendations for establishing a similar data repository in the United States.

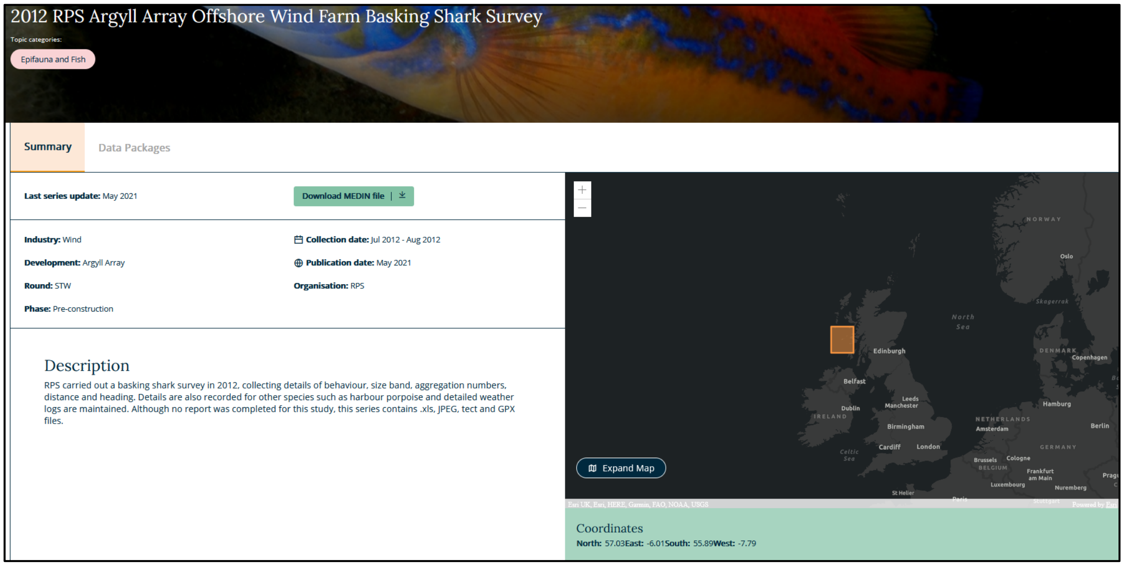
Briefly, the MDE utilizes a Microsoft Azure cloud computing environment to maintain two sites, a client-side data portal and a public-side data repository, that facilitate the uploading, hosting, and delivery of high-resolution spatial data to multiple end users. The client-side data portal allows data providers/ users to view, update, and download project reports, datasets, and metadata in a secure online environment. While users can preview the presentation of their datasets before being published online, internal quality assurance (QA) personnel can validate metadata and file structure and stability. After confirming with the user, the internal QA team will publish (i.e., deploy) the user's data to the public-side data repository where it can be found and downloaded by whomever (Figure F-3).

Although a variety of cloud computing service providers exist, Microsoft Azure was selected by the MDE due to its vast integrations with Microsoft products (favorable for those with existing enterprise accounts) and existing partnership with Microsoft to support software development. To recreate a similar user experience for both data providers and public users, MDE personnel summarized their recommended best practices below:

- Establish a product owner mindset
  - Prepare to evaluate the needs of both data providers and the public.
  - Employ developers, test engineers, user interface designer, and an overall product owner.
- Use Angular User Interface (UI)
  - A front-end web development framework that offers rich user experiences, fast responsiveness, and code maintainability.
  - Said framework allows data providers to upload their data on the client-side data portal.
- Employ Azure data storage and databases
  - SQL databases store project metadata and allow public users to search for them when made available.

    - Azure’s Data Lake storage allows developers to store large amounts of data while retaining their file structure for easy access and processing.
    - Azure’s Data Blob storage is a cost-effective solution to store massive amounts of unstructured data in a cloud environment.
- Deploy Google Analytics
    - The freemium online service can inform data providers how their data are being accessed by public users to encourage continued data sharing or refine their current approach.
- Take advantage of Microsoft Azure’s vast features
    - Azure offers a suite of processing and analytical tools to improve workflows that can be quickly tested or implemented without inflating processing costs.
    - The use of built-in features can prevent an organization’s reliance on developing custom functionality that can hinder debugging and efficiency.

(A)



(B)

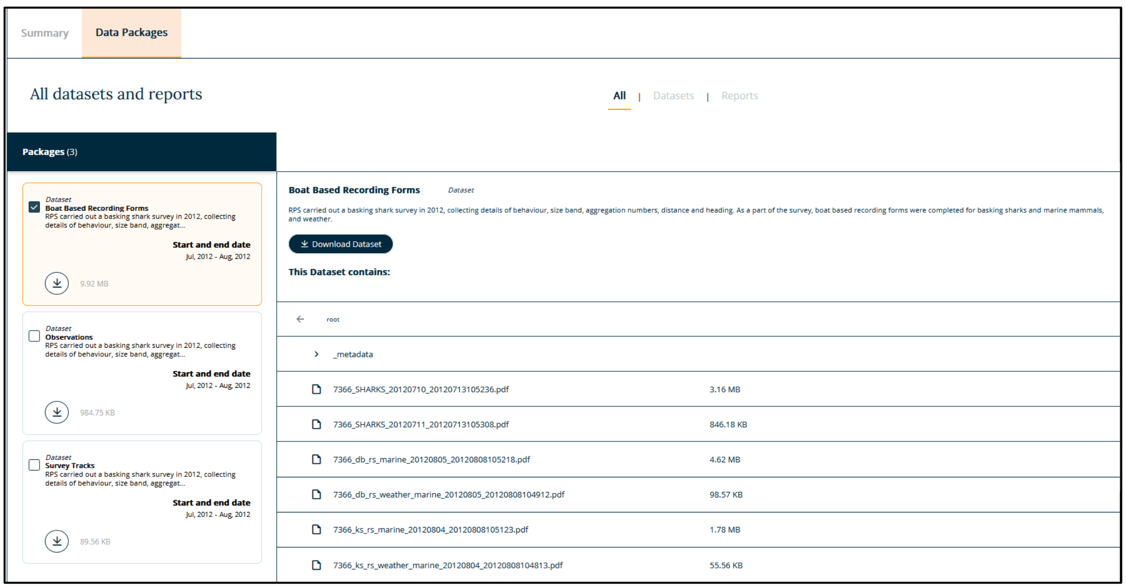


Figure F-3. Screenshots from The Crown Estate's Marine Data Exchange (MDE) demonstrating the functionality of the public-side data repository where users can explore and download publicly available datasets. Either through the MDE's search feature or overall file structure, public users can (A) navigate and view project metadata and (B) download associated data in a streamlined and user-friendly format.

4. Technical Recommendations*

- **Cloud computing services should be considered** over on-premises (i.e., local) computing services.
    - o Cloud computing services provide predictable pricing plans, reduce reliance on physical hardware, software, and associated IT personnel, and allow for quick scaling to meet data storage and processing demands.
    - o On-premises computing services are completely owned by the organization and allow for both local and remote access, however, they require a significant investment in building, updating, and maintaining.
- **Microsoft Azure is the recommended cloud computing provider** given its simplicity, integration with Microsoft services, and use by many enterprise clients; this provider comes highly recommended by the MDE.
    - o Amazon Web Service (AWS), Microsoft Azure, and Google Cloud Platform (GCP) are the top three cloud computing service providers, in that order, based on popularity and market share.
    - o Despite AWS's status as the most comprehensive and established service, its pricing plans are complex and requires specialists to run efficiently. Moreover, MDE personnel commented that AWS frequently forces system updates that require processing time and results in increased spending.
    - o GCP offers a simple pricing model, integrates well with other Google services, and excels at artificial intelligence and machine learning-powered cloud processing, but it lacks an extensive collection of cloud services and programming languages.
- **Capitalize on the collective experience of IT professionals and Microsoft Azure features** to create a product that supports both data providers and public end users.
    - o Begin scouting for a team of developers, test engineers, and user interface designers to establish proper infrastructure and functionality.
    - o Microsoft Azure hosts of suite of features for establishing responsive and streamlined tasks, which prevents the use of custom coding that can bottleneck future requests.

*Please note, the listed recommendations are based on current technologies and services. Users should evaluate the existing material with recent online sources to ensure consistency.

Appendix G - Task 5: Analysis Mode Portal Interface

Task 5: Analysis Mode Portal Interface

The "Analysis Mode Portal Interface" is a collection of data sources and user interface features that allow end users to properly explore and evaluate the presented data, from input data to final modeling outputs. Due to the publicly available nature of data presented on the Northeast Ocean Data Portal (Portal), providing users with appropriate information is vital to overall transparency and preventing misinterpretation of results. The following document, therefore, identifies lessons learned from processing this set of data and tools as well as overall recommendations for future consideration.

- Lessons Learned
  - In addition to vector and raster files, the use of designated symbology layers (.lyr files in ArcGIS) were important for standardizing visual aesthetics and legend entries across similar data layers.
  - A raster "brick" (i.e., advanced raster overlay) was proposed for allowing users to access geospatial supporting information, such as model likelihood and uncertainty for each of the five sediment classes, within the sediment composite using mouse gestures (Figure G-1).
    - To properly retrieve data necessary for the user accessible window, supporting raster information (i.e., model likelihood, uncertainties) needed to be concatenated into a single file (Figure G-2).
    - The ability for the Portal to simultaneously extract values from multiple files, however, is influenced by file size and number of grid cells.
    - Given the number of grid cells across the study area, the Portal would not be able to efficiently deliver a streamlined and responsive user experience.
    - Model likelihood and uncertainty values for each sediment class were therefore reclassified, i.e., categorized into sensible groupings, to reduce their complexity and file size, thereby improving the Portal's ability to collect and display supporting information for end users.

- Technical Recommendations
  - **Develop clear visuals and provide detailed metadata** to maximize comprehension (Figure G-3).
    - Data layers should include necessary legend entries as well as metadata, including title, description, source, author, last modified date, thumbnail, and tags.

        - Given the ability to quickly toggle data layer descriptions on the Portal, this is a suitable area to include limitations of the data and recommended use cases.
    - **Include background and supplementary data** to aid end users in evaluating the data product (Figure G-4).
        - For example, the sediment composite includes a "Supporting Data" subsection that provides essential information for understanding input sources and the modeling protocol.
            - Raw bathymetric and sediment input files allow users to view the distribution of data across the study area. Such capabilities may allow users determine where the predictions in the composite should be used with confidence.
            - Study area boundaries aid users in understanding where different thresholds were applied to predicting sediment occurrence in nearshore and offshore areas.
    - **Provide different data product versions** to satisfy an assortment of stakeholders and specific use cases.
        - Either due to multiple dimensions (e.g., time) or estimated statistics (e.g., threshold), complex data layers are difficult to visualize in a single layer.
        - Data providers and specialists should therefore develop data product variations to highlight these differences and ultimately benefit multiple stakeholder groups. For example,
            - The current sediment composite could be displayed with sediment occurrence predictions based on low, medium, and high thresholds to satisfy the needs of government agencies, academic institutions, and non- and for-profit organizations.
                - High thresholds (used in the current project) predict the occurrence of sediment classes with high confidence and are valuable to offshore wind energy developers and regulatory agencies when surveying and evaluating marine environments. However, they create large swaths of area where sediment classes could not be confidently classified
            - Low to medium thresholds, in contrast, would produce a more complete, yet less strict, sediment composite that may support academic institutions and non-profit organizations when conducting regional studies (e.g., species habitat selection).

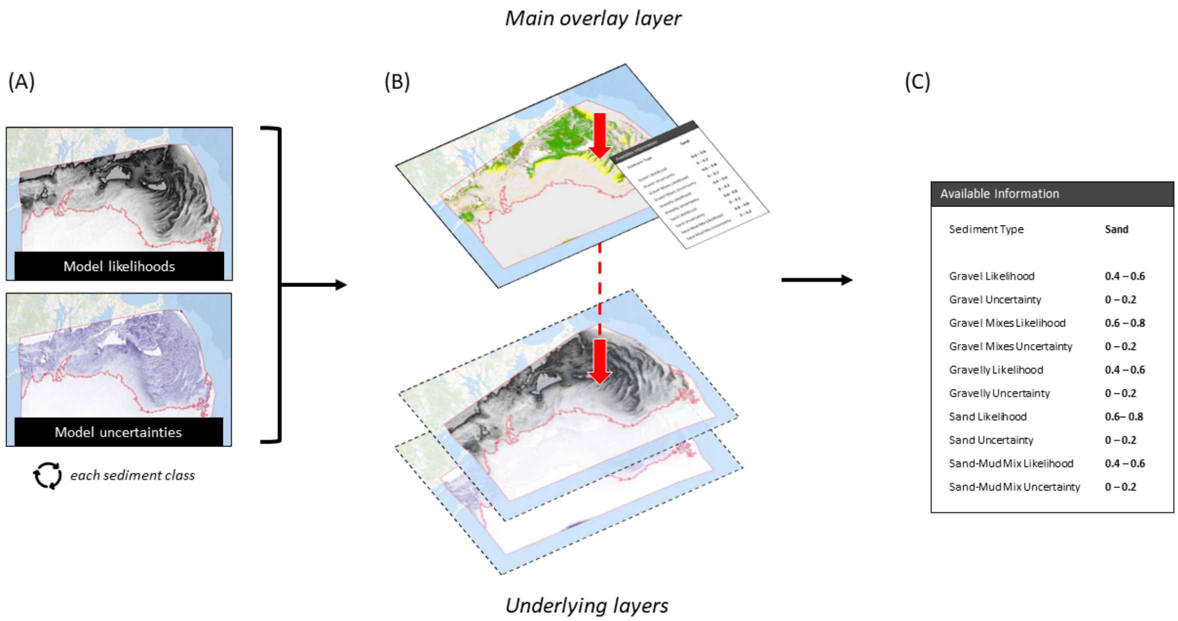


Figure G-1. Geoprocessing protocol for assimilating the sediment composite and supporting data layers into a raster "brick". In doing so, users can select any location within the sediment composite to toggle an interactive window that details various information, including model likelihood and probability for each of the five sediment classes.

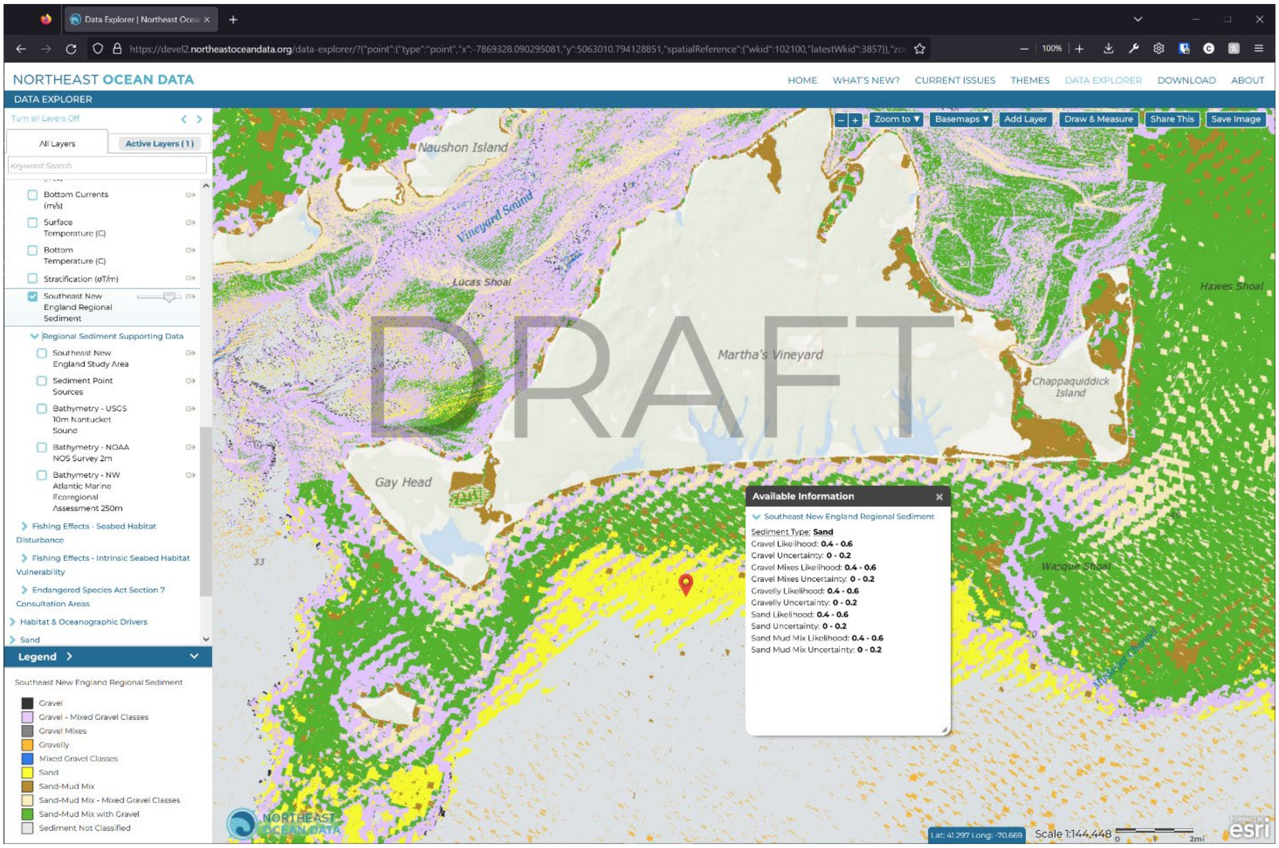


Figure G-2. Example of an interactive window that users can toggle to view supporting information within the sediment composite, specifically model likelihood and uncertainty for each of the five sediment classes.

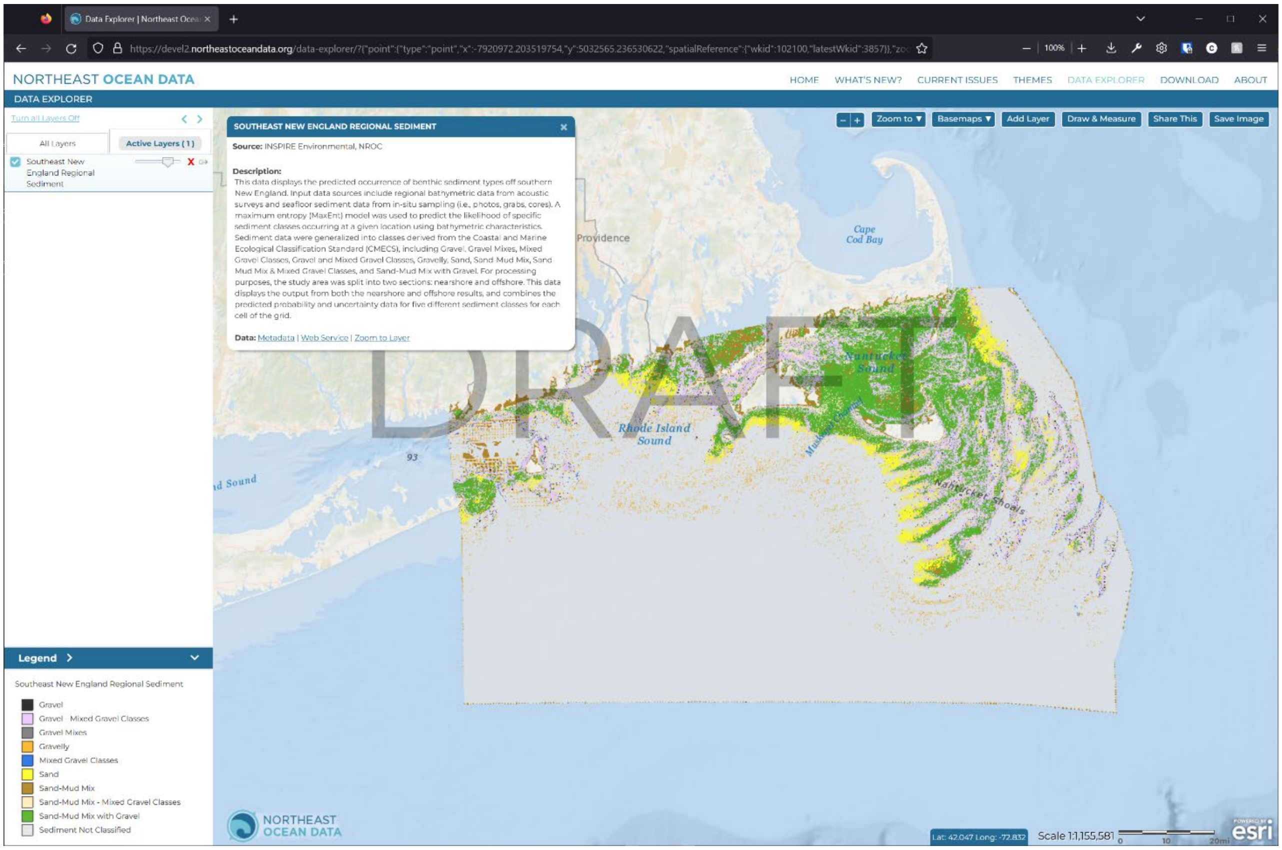
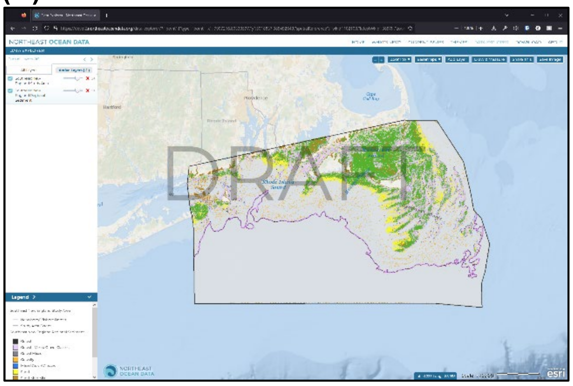
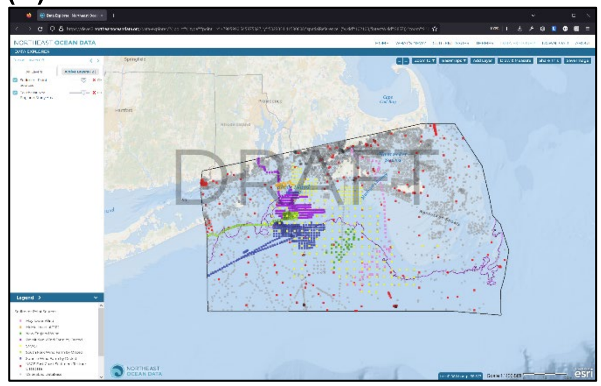


Figure G-3. Example of the sediment composite on the NROC Data Portal development site displaying the data product as well as associated legend and metadata.

(A)



(B)



(C)

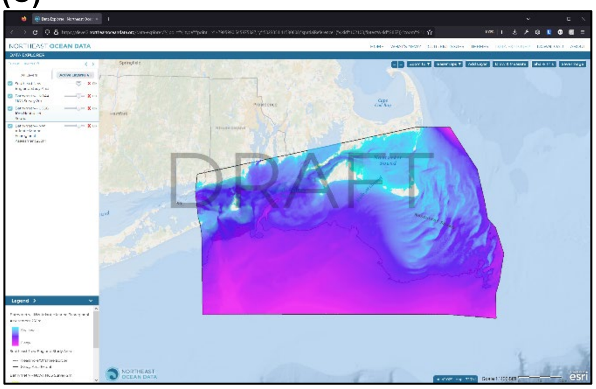


Figure G-4. Supporting datasets for the sediment composite that users can access on the NROC Data Portal development site, including (A) study area boundaries, (B) sediment sampling locations, and (C) bathymetric sources.