Data Quality Index for usSEABED Atlantic Coast Offshore Surficial Sediment
Atlantic Coast
January 13, 2016

Prepared for*:*
Northeast Regional Ocean Council (NROC)
www.northeastoceandata.org

Prepared by:
Kate Longley-Wood
SeaPlan
89 South Street, Suite 202
Boston, MA 02111

1. INTRODUCTION

This dataset is intended to be a companion layer to the U.S. Geological Survey (USGS) Data Series usSEABED Atlantic Coast Offshore Surficial Sediment extracted data for the entire U.S. Atlantic Coast ("usSEABED Extracted Data"). The usSEABED Extracted Data for the Entire U.S. Atlantic Coast depicts the point locations of known sediment samplings, inspections and probings from the usSEABED data collection which were integrated using the software system dbSEABED. These point locations were extracted from the database through data mining. The resulting dataset contains information on samples from 58 individual datasets collected by the USGS and other research projects, and includes data processed by the USGS sediment laboratory, in addition to datasets compiled from gray literature or unpublished sources. More information on this dataset can be found in the metadata document.

Because the usSEABED Extracted Data covers data collected and published between 1934 and 2004 using a variety of sampling and analysis techniques, data quality within the dataset varies widely. This Data Quality Index product was developed to serve as a guide on the quality and reliability of the usSEABED Extracted Data by assigning data quality values to variables including the age of the sample, the sampling device, and the analytical technique used.

The data quality index is informed by a methodology developed and employed by the Massachusetts Division of Marine Fisheries (Ford & Voss 2010) and the Massachusetts Office of Coastal Zone Management (Sampson & Huntley 2015). In this methodology, sample age is considered the most important factor because of changes in positioning methods over time, and because of the variability of sediment stability, particularly in coastal

environments. Sampling device was considered the second determinant of quality, using the assumption that grabs designed for surface sampling are the most accurate in determining surficial sediment composition when compared to other devices which may not accurately capture surface sediment. Analytical technique was the third determinant of quality, with laboratory techniques assumed to provide more precise descriptions of sediment composition than visual methods. Data quality values were assigned within these ranked categories following Sampson & Huntley (2015) and are as follows:

Age Quality Values	**Year**	**Data Quality Score**
	2000 - present	12
	1985 - 1999	11
	1960 - 1984	7
	pre-1960	1
Sampling Device Quality Value	**Sampler**	**Data Quality Score**
	Grab	4
	Photo	4
	Core	3
	Dredge	2
	Lead Line	1
Analytical Technique Quality	**Technique**	**Data Quality Score**
	Laboratory	2
	Visual	1

The data quality index value was determined using the following equation from Sampson & Huntley (2015):

Data quality index = ((Age Value/12) + (Sampling Device Value/4) + (Analytical Technique Value/2))

Data quality values were informed by the metadata for individual datasets that make up the usSEABED Extracted Dataset. Metadata for individual datasets within the usSEABED Extracted Dataset are located here. In the event that the age of the dataset spanned multiple age quality bins, the most recent year (if available) or the dataset publication date was used to assign the age value unless there was additional information present in the metadata that suggested that a value for older data should be assigned (e.g. if there were doubts about the positional accuracy of the instrumentation used in the sampling process). If the years in which the samples were collected were otherwise unknown, the publication date was used to assign a data value score, as described in Ford and Voss (2010). Similarly, if no description of sediment analysis technique was present in the metadata, it was assumed the samples were analyzed visually if the % gravel, sand, mud, or clay values were less than 100 and divisible

by 5 or 10 (Ford & Voss 2010). If there were no % gravel, sand, mud or clay values, the analytical technique quality value was zero.

No data quality values or data quality index scores were assigned if the data met one or more of the following criteria:

- the metadata did not contain sufficient information to assign data quality values
- the sample information did not contain data on the Folk code or data type
- phi values and percent composition values of the sample were both omitted

Data quality index scores ranged from 1.33 to 3 and were divided into quantiles yielding three data quality confidence levels of high, medium, and low.

2. PURPOSE

The purpose of this dataset is to assist in ocean planning activities in the Northeast by providing data quality reference information about a commonly-used sediment dataset.

3. SOURCES AND AUTHORITIES

- United States Geological Survey (USGS)
- usSEABED Extracted Data for the Entire U.S. Atlantic Coast (metadata document)
- usSEABED Metadata Sources
- Ford, K.H. & Voss, S. (2010). *Seafloor Sediment Composition in Massachusetts Determined Using Point Data* (Massachusetts Division of Marine Fisheries Technical Report TR-45). New Bedford: Massachusetts Division of Marine Fisheries.
- Sampson, D. & Huntley, E. (2015, April). *Creating a comprehensive seafloor sediment map in Massachusetts.* Presentation at Coastal Geotools, Charleston, SC.

4. DATABASE DESIGN AND CONTENT

Native storage format: ArcGIS File Geodatabase – simple feature class

Feature Types:

Sampling points

High: A classification used to describe points where the data quality scores are between 2.68 and 3, indicating high confidence in the quality of the data collected at that point

Medium: A classification used to describe points where the data quality scores are between 2.43 and 2.67, indicating a moderate level of confidence in the quality of the data collected at that point

Low: A classification used to describe points where the data quality scores are between 1.33 and 2.42, indicating a low level of confidence in the quality of the data collected at that point

No quality value: A classification used to describe points where there was insufficient information to assess the quality of the data collected at that point

Data Dictionary:

Line	Name	Definition	Type	Size
1	OBJECTID	Uniquely identifies a feature	OBJECTID	*
2	Shape	Geometric representation of the feature	geometry	*
3	datasetKey	Unique numerical key to original dataset component of usSEABED Extracted dataset	short	3
4	siteKey	Unique numerical key to site from usSEABED Extracted dataset	long	6
5	sampleKey	Unique, sequential, numerical key to sample from usSEABED Extracted dataset	long	6
6	sampleYear	Year(s) of sample collection or publication date	text	25
7	yearValue	Age quality value assigned to the sample	short	4
8	sampleDevice	Device used in sampling	text	50
9	deviceValue	Sampling device quality value assigned to the sample	short	4
10	sampleAnalys	Technique used to analyze the sample	text	50
11	analysisValue	Analysis technique quality value assigned to the sample	short	4
12	qualityScore	Data quality index score for the sample	float	7
13	dataConfidence	Qualitative ranking of data quality for the sample	text	20

Feature Class Name: usSEABED Data Quality

Total Number of Unique Features: 49257

Dataset Status: Complete

5. SPATIAL REPRESENTATION

Geometry Type: vector point
Reference System: GCS North American 1983

Horizontal Datum: North American Datum 1983
Ellipsoid: Geodetic Reference System 1980

XY Resolution: XY Scale is .000000001
Tolerance: 0.000000008983153

Geographic extent: -81.93 to -65.00, 23.02 to 45.91

IS0 19115 Topic Category: environment, oceans, geoscientificInformation

Place Names: Place Names:

Androscoggin River, Atlantic Ocean, Atlantic margin, Baltimore Canyon, Boston Harbor, Block Island Sound, Buzzards Bay, Cape Cod Bay, Cape May, Chesapeake Bay, Cobscook Bay, Georges Bank, Gilbert Canyon, Great Bay Estuary, Great South Bay, Gulf of Maine, Hudson Shelf Valley, Kennebec River, Long Island, Lydonia Canyon, Maine Inner Continental Shelf, Massachusetts Bay, Merrymeeting Bay, Narragansett Bay, New Jersey Continental Shelf, New Jersey Harbor, New Jersey margin, New York Bight, New York Harbor, North Carolina Outer Banks, Northern Maryland Shoal Fields, Oceanographer Canyon, Sandy Hook, Scotian Shelf, Southern New England Shelf, Upper Delaware Estuary, Wrightsville Beach

Recommended Cartographic Properties:

(Using ArcGIS ArcMap nomenclature)

Simple Fill Symbol, Circle 1 no outline, size = 8.0, color model: HSV
High: color: 100-100-66
Medium: color: 60-100-100
Low: color: 0-100-90
No data quality: color: 0-0-51

Scale range for optimal visualization: 100,000 to 15,000,000

6. DATA PROCESSING

Processing environment: ArcGIS 10.2, Windows 7 Professional, Intel Core i5 CPU

	Process Steps Description
1	The atl_ext dataset was downloaded from the USGS data catalog and uploaded into ArcMap
2	Data quality and value fields were added to the dataset
3	Some source datasets were identified within the usSEABED dataset using Select by Location queries

4	Data quality and value fields were filled in using the field calculator
5	Non-null data quality values were classified by quantiles in 3 classes, and assigned descriptors of high, medium, and low, to indicate data quality confidence, using the field calculator

7. QUALITY PROCESS

Attribute Accuracy: Attribute values are derived from authoritative metadata sources. In the case of attribute value uncertainty, attribute values were assigned using the methodology described in the introduction.

Logical Consistency: None

Completeness: The completeness of the data reflects the feature content of the data source, usSEABED Extracted Data, and its associated metadata. Not all source datasets had complete information, and data value scores reflect that uncertainty.

Positional Accuracy: Positional accuracy may vary according to positioning methodology. Positional uncertainty is accounted for in the data quality index by accounting for sample age, as described in the introduction. Additional information may be found in the usSEABED Extracted Data metadata.

Timeliness: Based on samples collected since 1934 and dataset publication dates through 2004.

Use restrictions: Data is presented as is. Users are responsible for understanding the metadata prior to use.

Distribution Liability: All parties receiving these data must be informed of all caveats and limitations.