



usSEABED: Atlantic Coast Offshore Surficial Sediment Data Release

By Jamey M. Reid, Jane A. Reid, Chris J. Jenkins, Mary E. Hastings, S. Jeffress Williams, and Larry J. Poppe

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government

U.S. Geological Survey Data Series 118
2005, Version 1.0

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior

Gale A. Norton, Secretary

U.S. Geological Survey

P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia

For Additional Information:

<http://woodshole.er.usgs.gov/project-pages/aggregates/>

<http://woodshole.er.usgs.gov/project-pages/stellwagen/>

<http://walrus.wr.usgs.gov/nearshorehab/>

Telephone: 508-548-8700

Feedback on usSEABED is appreciated, both in usefulness and in error detection. Please use the following contact information for issues, questions and (or) data to contribute to the growing usSEABED information system for the U.S. EEZ.

Jamey M. Reid: *for adding Atlantic Coast and (or) Gulf Coast data*

USGS, Woods Hole Science Center

384 Woods Hole Road

Woods Hole, MA 02543-1598

Tel: 508-548-8700

jreid@usgs.gov

Jane A. Reid: *for information about the Benthic Habitats project and (or) adding Pacific Coast, Alaska or Hawaii data*

USGS, Pacific Science Center

400 Natural Bridges Drive

Santa Cruz, CA 95060

Chris J. Jenkins: *for information about the dbSEABED program as well as adding global data*

University of Colorado

Institute of Arctic and Alpine Research

1560 30th Street

Campus Box 450

Boulder CO, 80309-0450

S. Jeffress Williams: *for information about the Marine Aggregates project*

USGS, Woods Hole Science Center

384 Woods Hole Road

Woods Hole, MA 02543-1598

Larry J. Poppe: *for information about sediment data collection; laboratory and analysis techniques*

USGS, Woods Hole Science Center

384 Woods Hole Road

Woods Hole, MA 02543-1598

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:

World Wide Web: <http://www.usgs.gov>

Telephone: 1-888-ASK-USGS

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted material contained within this report.

Contents

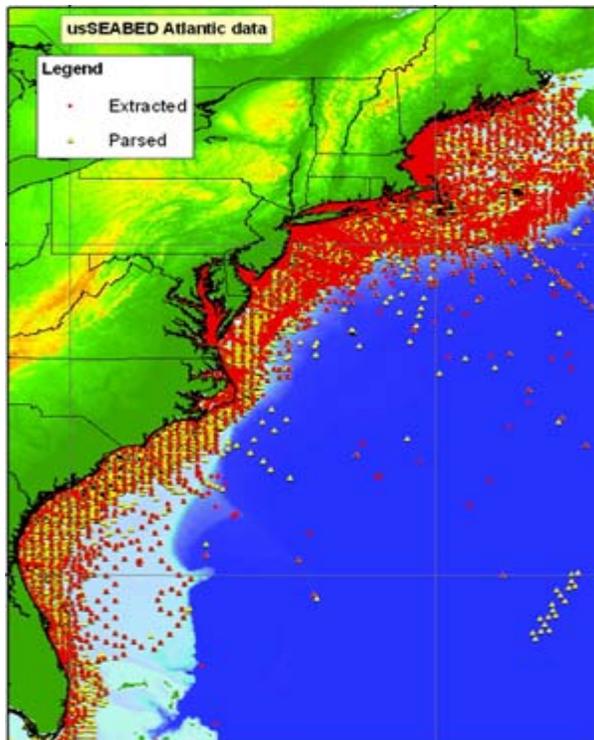
Abstract	6
Contents	7
Introduction	7
Future Plans, Updates and Online Usage in a GIS	8
Applications	8
Marine Aggregate Resources	8
National Benthic Habitats Studies	9
usSEABED Data	10
How usSEABED is Built	10
Sources of Data	11
Data Themes and Output Files	11
Output files	11
<i>Source data</i> (<i>_SRC</i>)	12
<i>Par</i>	12
<i>Calculated data</i> (<i>_CLC</i>)	13
<i>C</i>	13
Quality Control	14
Spatial and Temporal Uncertainties	14
dbSEABED	14
Introduction	14
Data Import Methods	15
The Numeric Data Type	15
The Linguistic Data Type	16
Calibration	16
Expansion of Data Coverage	19
Summary of the onCALCULATION methods used in dbSEABED	19
Data Catalog	23
Data Files	25
Sources:	25
usSEABED Atlantic Data:	26
Base Maps:	26
Legends	26
Color	26
Roughness	27
References	27
Acknowledgments	28
Frequently Asked Questions (FAQs) about dbSEABED	28
dbSEABED	28
Output files	29
Linguistic data versus numeric data	29

Grain size.....	29
Color.....	30
Data import methods.....	30
Data processing methods.....	30
How to.....	30
dbSEABED.....	30
What is dbSEABED?.....	30
What kind of outputs does dbSEABED produce? [top].....	30
What quality control measures are in place for dbSEABED?.....	31
Output files.....	31
What is the SRC file output?.....	31
What is the function of the output CMP and FAC files?.....	31
What is the CMP file output for?.....	31
What is the meaning of the data in the FAC output file?.....	32
What do all the “-99”’s mean?.....	32
Why are there negative values for seabed strength and critical shear strength.....	32
Compared to values given in some scientific papers, the output values from dbSEABED are quite rounded. Why?.....	32
What is the Rock Membership?.....	32
What is the meaning of the SeaBedClass and ClassMembership?.....	33
What is the Weed Membership?.....	33
Hydrographic Bottom Type (HBT) codes can be output. What do they mean?.....	33
What does the Roughness code mean?.....	33
Why do some points plot on the land?.....	33
Can the same sample or analysis be represented more than once in the data collection?.....	33
Why are the Kurtosis and Skewness not reported from the database?.....	34
Linguistic data versus numeric data.....	34
How does dbSEABED make word data conformable with numerical data?.....	34
What if a user doesn’t want to use the word-based data for their mappings?.....	34
Which form of data should I trust more? Numeric-value analyses or word-based descriptions?.....	34
How does word-based descriptive data relate to numeric-value analytical data?.....	34
Grain size.....	35
Why are values of grain size only given to 1 decimal place, and percent sand etc given as integers?.....	35
How are the grain sizes extracted from data?.....	35
Are detailed grain size analyses used to output textural statistics like mean and sorting?.....	35
Why would a sample that is known to have had a detailed grain size analysis not appear in the output data?.....	35
How are the outputs for the grain size fractions gravel, sand, mud extracted?.....	35
Are grain size scales other than Wentworth used?.....	35
How is one central grain size recovered from diverse data that may present mean, median and other graphical grain sizes?.....	36
How are the central grain size and sorting estimated in the CLC process of the database?.....	36
Data processing methods.....	36

What is Fuzzy Logic and how does it work?	36
What is the method of attributing sub-bottom depths to samples from cores?.....	36
How are the shear strengths derived?	36
What is the method for arriving at the SHEPARD and FOLK classes?.....	37
What relationships are used to estimate the Critical Shear Stress values?	37
How are the porosity values obtained?	37
What is the basis for the data on compressional (p-wave) sound speeds?.....	37
How are the sorting values arrived at?.....	38
Carbonate values: how are they arrived at? And Organic Carbon?.....	38
Data import methods.....	38
How does dbSEABED hold the basic data?	38
Why does dbSEABED hold its underlying data in documents rather than a relational database like Oracle?	38
What is involved in importing data sets into dbSEABED? [top]	39
How is metadata treated? [top]	39
Why are sediment descriptions and biological names held as abbreviations?.....	39
Color	39
Where can I find more information on the Munsell color code?.....	39
How are the colour Munsell codes derived?.....	39
How to	40
How can I map the coded information on Color and Roughness in a GIS?	40
How can I use the Critical Shear Stress in a practical application?.....	40
References:	40
Tables A-E.....	41

U.S. Geological Survey Data Series 118, usSEABED: Atlantic Coast Offshore Surficial Sediment Data Release

By Jamey M. Reid, Jane A. Reid, Chris J. Jenkins, Mary E. Hastings, S. Jeffress Williams, and Larry J. Poppe



usSEABED data for the Atlantic coast showing locations of the extracted (EXT) and parsed (PRS) outputs. The PRS data consist of word-based descriptions. The EXT data consist of numeric, lab-based analyses.

Abstract

Over the past 50 years there has been an explosion in scientific interest, research effort and information gathered on the geologic sedimentary character of the U.S. Atlantic Coast continental margin. Data and information from thousands of publications have greatly increased our scientific understanding of the geologic origins of the shelf surface but rarely have those data been combined and integrated.

This publication is the first release of the Atlantic coast data from the usSEABED database. The report contains a compilation of published and unpublished sediment texture and other geologic data about the seafloor from diverse sources. usSEABED is an innovative database system developed to bring assorted data together in a unified database. The dbSEABED system is used to process the data. Examples of maps displaying attributes such as grain size and sediment color are included. This database contains information that is the scientific foundation for the USGS Marine Aggregate Resources and Processes Assessment and Benthic Habitats projects, and will be useful to the marine science community for other studies of the Atlantic coast continental margin.

The publication is divided into ten sections: Home, Introduction, Content, usSEABED (data), dbSEABED (processing), Data Catalog, References, Contacts, Acknowledgments and Frequently Asked Questions. Use the navigation bar on the left to navigate to specific sections of this report. Underlined topics throughout the publication are links to more information. Links to specific and detailed information on processing and those to pages outside this report will open in a new browser window.

CITATION: This document should be cited as:
Reid, J.M., Reid, J.A., Jenkins, C.J., Hastings, M.E., Williams, S.J., and Poppe, L.J., 2005, usSEABED: Atlantic Coast offshore surficial sediment data release: U.S. Geological Survey Data Series 118, version 1.0. Online at <http://pubs.usgs.gov/ds/2005/118/>

Contents

The Introduction provides a brief overview of usSEABED and its purposes and describes two USGS projects (Marine Aggregates and Benthic Habitats) for which usSEABED data are currently used.

The section on usSEABED (data) describes the database, including its construction, sources of data, data themes, output files, quality control, information on spatial and temporal uncertainties within the data, as well as future plans and updates. The types and contents of the output files are explained in detail.

The dbSEABED (processing) section explains the dbSEABED information processing system as it relates to data import and processing methods. It describes the differences between the numeric and linguistic data types as well as information on how accuracy can be assessed through calibrations and statistical analyses.

The Data Catalog is the heart of the publication and contains the six output files (_EXT, _PRS, _CLC, _CMP, _FAC and _SRC). The data are presented within compressed (zip format) with the data held as shapefiles and comma delimited text files. It also includes several basemap layers for use with the data. The Data Catalog Section includes examples of how to visualize the data and two ArcView legends: Munsell color and Seabed Roughness.

The References section contains the references cited.

The Contacts section lists the names, and contact information for various issues such as how to contribute data, data processing questions, project related information and questions about sediment data collection or laboratory and analysis.

The Acknowledgments section recognizes the many people and organizations who have contributed to this publication.

The Frequently Asked Questions section provides a list of questions commonly asked by users of the data.

Introduction

This data release publication provides an improved and robust integrated database (usSEABED) of seabed characteristics for the Atlantic continental margin of the United States that fulfills a need for information about seabed characteristics for use by geologists, ecologists, biologists, resource managers, and national defense investigators. usSEABED provides a digital, integrated database of existing physical data and information from the seafloor, including textural, statistical, geochemical, geophysical, and compositional information. It uses the dbSEABED data mining and processing software to extend the coverage of information in areas where data coverage is more descriptive than quantitative. The data coverage includes the U.S. Atlantic coast from northern Maine to Key West, Florida, including some lakes, Lake Erie and Lake Ontario, beaches, rivers and estuaries, and extends seaward across the continental shelf and slope. More than 150 different data sources containing over 200,000 data points are currently contained in usSEABED for the Atlantic.

This data-series publication is the first in a set of similar publications that covers the entire Exclusive Economic Zone of the United States (U.S. EEZ). Companion publications will include data for the Pacific (California, Oregon, Washington) and Gulf of Mexico margins as well as Alaska and Hawaii. These publications

will be updated as significant amounts of new data are included in usSEABED. This publication contains information on the usSEABED data collection, the dbSEABED program and processing, as well as a Data Catalog where the data are included (within zipfiles) as GIS layers and comma-delimited text files.

The overall usSEABED database holds data for the entire U.S. EEZ and is an ongoing task of the Marine Aggregates Resources and Processes and National Benthic Habitats Studies (Pacific) projects, USGS Coastal and Marine Geology teams in Santa Cruz, CA, Woods Hole, MA, and St. Petersburg, FL, and the University of Colorado.

Future Plans, Updates and Online Usage in a GIS

It is expected that usSEABED will continue to expand throughout the U.S. EEZ by the incorporation of new data sets and by the utilization and mining of the data in new and different ways. As significant changes are made, we expect to reissue usSEABED as updated publications.

Data contributions and (or) additional partners are welcomed. For information, please see the contacts page.

Published usSEABED data from this region and others are accessible online (using ESRI® ArcIMS) at <http://coastalmap.marine.usgs.gov/regional/contusa/index.html>. Data layers will also be submitted for viewing and download through Geodata.gov.

Applications

The usSEABED database is a very large compilation, containing complex assortments of data and geologic information on the geology of the seafloor. Although this database was developed for use in conducting studies of offshore sedimentary character for assessing marine aggregates and characterizing benthic habitats, it has much greater potential for application by the marine science community and other users. Users are encouraged to generate their own queries and extract information to meet specific needs. Other potential applications for data and maps from usSEABED follow:

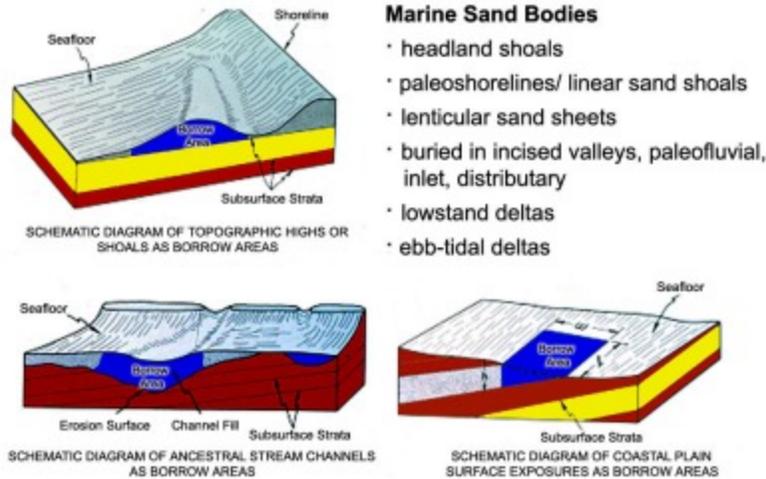
- Research ocean observation and monitoring
- Coastal zone/ocean management and planning
- Homeland security, military applications
- Seafloor engineering planning and design
- Ocean disposal site placement, monitoring
- Cultural resources
- Fisheries management, marine protected areas
- Seabed roughness, bedform distribution, critical shear stress, sediment transport flux
- Public education
- Seafloor bottom friction values for calibration of modeling processes, such as the effects of storm waves on sediment mobility and transport

Marine Aggregate Resources

Continental margins are products of complex geologic processes. They comprise submerged landforms that offer a variety of natural functions such as: benthic habitats for fisheries navigation, home-land security, and engineering activities. Continental shelves also contain hard-mineral deposits, such as sand and gravel that are potential aggregate resources.

Coastal erosion, resulting from a combination of natural processes (storms, sea-level rise, sediment starvation, land subsidence) and anthropogenic factors (dams, dredging, coastal engineering structures), is pervasive in most U.S. coastal regions. Development in the coastal zone continues to increase and demographic projections show that the trend of people moving to the coast will likely continue, placing more people and development at increasing risk from coastal hazards. With the prospects of future global climate change likely to cause changes in storminess and accelerated global sea-level rise, coastal regions are likely to experience even greater erosion, inundation, and storm-surge flooding in the near future.

Beach nourishment, a practice of placing sand dredged from offshore area onto eroding beaches, is increasingly viewed as a cost-effective and environmentally acceptable method for mitigating coastal erosion, reducing storm and flooding risk, and restoring degraded coastal barrier island ecosystems for developed coasts. For beach nourishment to be viable, however, it is necessary to locate high-quality sand; the sand bodies must ideally be reasonably close to beaches being considered for nourishment, and the sand volumes must be sufficient to meet recurring nourishment requirements for 50 years or longer. Sand bodies on inner continental shelf regions are often the most suitable sand sources for beach nourishment. Examples of marine sand bodies are shown in the following figures:



Marine sand bodies, having diverse origins and evolutionary histories, can be buried or exposed on continental shelves and often have been greatly modified by marine processes associated with the Holocene transgression. Nearshore marine sand bodies of the types shown above may offer the best potential sources for high quality sand for beach nourishment. (Williams and others, 2003)

Because offshore areas are increasingly important, comprehensive, up-to-date and integrated databases that can be used as sources for modern Geographic Information Systems (GIS) are needed to produce base maps displaying thematic information such as seafloor geology, sediment character and texture, seafloor roughness, and engineering properties. Digital geologic maps, based on unified national data sets and showing the sedimentary character of U.S. continental margins, are critical for scientists to better understand and interpret the geologic history and evolutionary processes of continental margins. These products are also useful to managers for protecting and managing coastal and marine environments

The USGS, in collaboration with other Federal agencies (Navy/Office of Naval Research (ONR), Minerals Management Service (MMS), U.S. Army Corps of Engineers (USACE), National Oceanic and Atmospheric Administration (NOAA)), coastal states, and universities, is leading a nationwide effort to gather legacy marine geologic data for use in conducting assessments of offshore sand and gravel resources and for producing GIS map products of seafloor geology that can serve many additional needs. Assessments are being conducted in offshore Louisiana, the New York Bight (New York-New Jersey), and the Gulf of Maine regions. The GIS maps and usSEABED database from this study are providing fresh scientific insights into the geologic character and development of U.S. continental margins and useful information about the quality and potential availability of offshore sand and gravel aggregates. Additional details are available at: <http://woodshole.er.usgs.gov/project-pages/aggregates/index.htm>.

National Benthic Habitats Studies

The quality and protection of natural environments is a high priority for the Nation. This includes Federal lands in the coastal and marine realm where a variety of human activities stress natural systems and resources. The

USGS is a lead scientific agency conducting research on the Atlantic, Gulf of Mexico, and Pacific margins to better understand marine seafloor environments. This research on benthic habitats is conducted under the general topic of “habitat geoscience,” which is defined as the study and classification of seabed habitats in the context of their geologic framework, their response to seabed processes, and their function as substrate for invertebrates and fish, is underway in all three regions. This research requires close collaboration between geologists, biologists, and oceanographers. High-resolution and detailed knowledge of bottom characteristics and sediment distribution, a goal of the usSEABED database, is fundamental to these studies.

Habitat research on the Atlantic margin is presently conducted off the east coast of Florida, in the New York Bight region, in Long Island Sound, and in coastal and offshore areas of New England in collaboration with many Federal and State agencies and academic institutions. Habitat studies are based on high-resolution seafloor imagery (bathymetry, backscatter) and groundtruth data (video and photo imagery, and geological and biological samples). These data sets are used to produce interpretive maps that show surficial and subsurface geology, habitats as classified into different ecologic types, the distribution of fauna and flora, and the effects of seabed processes.

Applications of research results include fishery management, habitat disturbance and recovery studies, impacts of invasive species, studies of deep-water coral habitats, the fates of contaminants from coastal sources, location and monitoring of disposal sites, the routes of cables and pipelines on the seabed, and the location of aggregate deposits having potential for use for beach nourishment. For more information on benthic habitats studies, see: <http://marine.usgs.gov/>, <http://woodshole.er.usgs.gov/>, and <http://walrus.wr.usgs.gov/nearshorehab/>.

usSEABED Data

The usSEABED database differs from a traditional relational database (RDB) because the data are processed and extended to maximize density and usability, making them more comprehensive for mapping and analysis. A traditional RDB, often creates simplistic and sparse data summary coverages with thinly populated and unwieldy tables. The usSEABED database not only treats the usual forms of numerical data, but it also contains a vast store of data about the seafloor in word-based descriptions that can be rich in information, but difficult to quantify, map, plot, or use in comparative analyses or models. The usSEABED database provides numeric values for typical seabed characteristics that are based on these descriptive data as well as numeric analytical data.

The usSEABED database also differs from other marine databases in that it incorporates a wide variety of information about seafloor sediment texture, composition, color, biota, and rocks, seafloor characteristics such as hardness or sediment ripples, acoustic properties, and geochemical and geotechnical analyses. The usSEABED output files are produced in comma-delimited text for ease of use in many software applications, and are ready for inclusion into many different GIS, RDB and other software applications.

How usSEABED is Built

The usSEABED database is built using the dbSEABED processing software created at the University of Sydney (Australia) and the University of Colorado. It has companion databases built along similar lines: auSEABED for Australia, balticSEABED, and a global database, goSEABED. Each of these databases relies on pre-existing data from a variety of sources to mine and extrapolate useful information about the seabed.

The dbSEABED program allows these source data sets to be compiled in a standardized format and integrates information across a series of data themes (Table A) and equipment: physical sampling equipment (grabs, cores, or probes) or remotely sensed sampling (photographs, videos, geophysics, soundings). These data may be numeric lab- or instrument-based textural, acoustic, geochemical, and geophysical data and (or) verbal (linguistic) descriptions of grabs, cores, or photographs, or a combination of any of these.

In the usSEABED database, most data held in these reports are queried for additional information that increases the data density over the seabed, allowing for more complete information. Few source reports contain all data reportable in usSEABED; null values are given in those fields that do not contain reliable information.

Sources of Data

usSEABED relies on previously existing and newly collected data, both published and unpublished, from federal, state, regional, and local agencies and consortiums, as well as research institutions. For the Atlantic coast, many of the data are from the USGS, including the previously published continental margins data from the early 1960s, and more recent (1990s, 2000s) data published by the USGS.

Data gathered by the National Ocean Service (NOS) during their many sounding surveys during the 1960s to 1990s are included, as archived by the Smithsonian Institution and provided by the National Geophysical Data Center (NGDC). USACE reports, including several from the USACE Coastal Hydraulics Laboratory (formerly CERC), Environmental Protection Agency (EPA) data sets, educational institutions' reports and theses, State geological survey reports, and NOAA reports are also included. Two large data compilations archived at NGDC are included: the '073' data compilation, which is primarily from U.S. Navy reports, and the Deck41 data set, which is from a wide variety of data sources. A complete reference list (*atl_sources.html*) of information for the Atlantic coast is included.

Data duplication may exist within the input Atlantic data sets. Data from the same cruise or site may be published in more than one report or data compilation. Efforts have been made to reduce data duplications in usSEABED. In other instances, data from different sources for a given site may be included if significant additional data are included. (For example, one source may report only grain size for a particular site, but another source may include geophysical properties for the same sites/sample)

Data Themes and Output Files

Seabed data come in a variety of forms, which have information in different parameters. For example, textural analyses may have information for percentages of gravel, sand, mud (silt and clay); statistical measurements such as mean, median, sorting, skewness, and kurtosis often accompany the textural information. Acoustical measurements include various velocities and derived densities. Benthic habitats studies may include a short description of the seafloor sediment type and a numerical survey of animals and plants on the seafloor, or evidence of them.

The original seabed information in usSEABED is entered into different data themes. A list of data themes is given in Table A. The thematic basis of the values found in the outputs can be found in field 11 ('DataType') of the extracted (_EXT), parsed (_PRS), and calculated (_CLC) output files. Information on contribution of each source report is in the accompanying metadata files.

Output Files

This publication provides six usSEABED output data files for the Atlantic coast.

usSEABED Output files

_EXT	Extracted (numeric, lab-based)
_PRS	Parsed (word-based)
_CLC	Calculated (calculated variables)
_CMP	Components (content and features)
_FAC	Facies (components only)
_SRC	Source

These files are downloadable from the Data Catalog. An additional output file type, although unpublished, provides quality control for the data and was used extensively prior to publication to debug and test the data. Field parameters for the data files are listed in Table B.

Relational keys

The usSEABED data file types are linked relationally by the foreign keys: DataSetKey (for individual data sets), SiteKey (for individual sites), and the SampleKey (for individual analyses). The DataSetKey field gives the relationship of the data to the original source. The tables can be loaded into a RDB, relationships may be constructed, and the tables may be joined using the keys.

Source data (*_SRC*)

Information about the original data are in the source (*_SRC*) file, including links to metadata about the original data. Each of the output data files discussed below is linked to the *_SRC* file by the DataSetKey field.

Textural and other basic information (*_EXT*, *_PRS*, *_CLC*)

Textural, statistical, geochemical, geophysical, dominant component, and color information are held in three separate, but similar, data files, based on the type of data: *_EXT*, *_PRS*, *_CLC*. The three data file types have the same fields (Table B) and can be combined for more extensive coverage of the seafloor. It is important for users to understand the inherent limitations of each type of file in order to choose the best data file, or combination of data files appropriate for a particular use. Other dbSEABED programs can combine the three files in a variety of ways, by concatenation or by telescoping, before they are mapped or used for other types of analysis. For access to these files, please see contact page.

Extracted data (*_EXT*)

The data file with the *_EXT* tag is the "extracted" data: those data from strictly performed, lab-based, numeric analyses. Most data in this file are listed as reported by the source data report; only minor unit changes are performed or assumptions may be made about the thickness of the sediment analyzed based on the sampler type. Typical data themes include textural classes and statistics (TXR: i.e., gravel, sand, silt, clay, mud, and various statistics), phi grain-size classes (GRZ), chemical composition (CMP), acoustic measurements (ACU), color (COL), and geotechnical parameters (GTC). The *_EXT* file is based on rigorous lab-determined values and form the most reliable data sets. Limitations however exist due to the uncertainty of the sample tested. For example, were the analyses performed on whole samples or only on the matrix, possibly ignoring larger particles?

Parsed data (*_PRS*)

Numeric data obtained from verbal logs from core descriptions, shipboard notes, and (or) photographic descriptions are held in the parsed data set (*_PRS*). The input data are maintained using the terms employed by the original researchers and are coded using phonetically sensible terms for easier processing by dbSEABED. Longer descriptions may have the data divided by theme (Table A). The descriptions often include information on associated biota, seafloor features, and structure. Typical data themes for the parsed data set are lithologic descriptions (LTH), biology (BIO), color (COL), and (or) seafloor type (SFT, descriptions from photos or videos). The values in the parsed data file are calculated using the dbSEABED parser that assigns field values based on the form and content of a description. See the section on dbSEABED processing and fuzzy set theory for a more complete explanation.

The parsing process has been tested and calibrated by comparing the outputs against analytical results for the same samples. Due to the nature of visual descriptions by observers and the use of fuzzy set theory in the parser, the output data variously show the degree of representation in the sample, or percent abundance values. An assumption in the process is that the output degrees of representation reflect absolute abundances to some degree of accuracy. The calibrations provide information on that accuracy. Although at first sight the descriptive results in the parsed file may seem less accurate than measured values in the extracted file, they are frequently more representative of the sample and seabed as a whole, as they include description of objects such as shells, stones, algae, and other objects (Table C) that are a textural component of the seabed and which are often left out of laboratory analyses, particularly when a machine analysis is employed.

Calculated data (_CLC)

For the extracted and parsed data, some values are not reported by the original source, but can be calculated directly or estimated by standard derivative equations using assumptions (See Frequently Asked Questions) about the conditions or variables. These values are reported in the calculated (_CLC) data files. Although the calculated (_CLC) data can be combined with the extracted and the parsed (Table B), they are the least reliable of the three data file types and should be used with caution.

Component/feature and facies data (_CMP, _FAC)

Two usSEABED data files contain information about the presence of certain seafloor features, compositional content, biota, and sediment structure. These use major synonyms defined by the thesaurus in the dbSEABED parsing software, which clusters comparable descriptive terms together, i.e., granite represents granite, aplite, granodiorite, pegmatite, while laminated represents laminated, laminations, or lamina. Individual components and features (i.e., terms like feldspar, phosphorite, bivalves, seagrass, and wood) are held in the _CMP data file (Table D). Appropriately combined components are held in the 'facies' (_FAC) data files (Table E). As with the parsed data files, the values held within the _CMP and _FAC files are the results of filters based on fuzzy set membership to chosen sets, and represent a measure of truth about the attribute, not percentages or defined values. These files only indicate presence, not absence, of material; e.g., it is rare that a report might state, "no bivalves" or "no phosphorite".

The _CMP file contains information about compositional content (individual minerals, rocks), genesis (terrigenous, carbonate), and certain biota. These components are internally evaluated and the value for each attribute is based solely on the relationships of attributes within the original description. The flora and fauna included in the compositional components are those that may have an effect on textural determinations in the _PRS data file, such as halimeda, bivalves, or foraminifera (Table C). The values within these attribute fields range between 0 (no membership, probably due to no information), to 100 (complete membership, i.e., shell hash = 100 to the shell debris set).

The _CMP file also includes information on seafloor features such as bedforms, fissures, internal structure (bedding, bioturbation), and other flora and fauna. Unlike the compositional content information, which is construed as an abundance within the sample, these attributes are an intensity of development or density of occurrence relative to scales of development or density of occurrence observed elsewhere. The flora and fauna included in the feature category are soft-bodied, for example those that do not have an input on the textural determination within the _PRS data files, such as kelp, ophiuroids, or annelids. Values within the attribute fields range from 0 (no membership, possibly due to no information) up to 100% (maximum development). In contrast to the situation with component abundances, the sum of feature intensities in a sample is allowed to exceed 100%.

The 100 most common components (number limited by dbSEABED processing software) in the U.S. EEZ are given in the _CMP file, and those attributes with "_F" denote features. Table D lists the components and gives basic forms of descriptive terms that may trigger membership for each. Included in this file are 27 components that are included in the facies (FAC) file only. The dbSEABED thesaurus used for usSEABED is also used for the sister data compilations (auSEABED, BalticSEABED, goSEABED), and the list of trigger terms may include some that are not known in U.S. waters.

The second file, the facies file (_FAC), is created from components only, similar to the _CMP file. This file configures multiple components into appropriate groups or facies, such as igneous, metamorphic, ooze, foraminifera, and others. The dbSEABED processing software is restricted to a maximum of six components per facies. Table E lists the facies type and the components that comprise each facies group.

Again, these files only indicate presence, not absence, of material; e.g., it is rare that a report might state, "no bivalves" or "no phosphorite". The values within this attribute field range between 0 (no membership, probably due to no information), to 100 (complete membership, for example, schist = 100 to the metamorphic set).

Relationship between the _PRS and _CMP outputs

The dbSEABED processing software recognizes that many skeletonized biota, such as halimeda, rhodoliths, shells (broken and unbroken), and others often comprise a sediment sample. Such biological terms are included in the parsing of the textural values. To see the selected biota with textural implications, see Table C . When using the parsed data, it may be important to cross-check with the component file using the relational foreign keys (SiteKey, SampleKey) to determine if biota are to be included in the textural outputs.

Within the _PRS file, the "seabed class" and "class membership" fields indicate the dominant compositional class and the fuzzy set membership of a sample to that class. Other components and mined information may also be listed for that sample in the _CMP file, linked by the relational keys.

Quality Control

Quality control over the data is an iterative process implemented using criteria in the following steps. First, graphical plots of site locations and parameter values are used to detect outliers and edit them appropriately. Each data set is viewed in a GIS to ensure that data locations are reasonable relative to survey extents; those sites with unresolvable location issues or known incomplete analyses are deactivated and are not included in the usSEABED output files. (Note: usSEABED does contain a small number of onshore samples.) This step may be optional depending on the data set. Older sets may require more scrutiny at this step whereas newer or well exercised data sets require less.

Second, built-in filters in the dbSEABED processing software detect implausible values for numeric fields, unknown verbal terms, incomplete analyses (i.e., Gravel-Sand-Silt Clay (mud) (GSSC(m)) greater than 100% or less than 95%), and incorrect field types (string or number). The software also detects samples that seem to belong to a core though they are described as independent samples. For the parsing of verbal descriptions, all terms must be known to the dbSEABED data processing program, with values assigned; those analyses that fail this test have null values given to all appropriate fields. Edits are made to the data (i.e., at the level of the usSEABED input data files) and metadata are entered explaining the changes. The edits (or deactivations) are then taken into account in the next dbSEABED program run.

Finally, output data are analyzed in a GIS to test whether the data outputs "make sense" for a given geographic area. Users of the output data should, however, note the limitations imposed by the source data sets as to navigational precision, sampler type, and analytical technique.

See the dbSEABED section and the Frequently Asked Questions for details about the usSEABED data mining program and the application of fuzzy set theory.

Spatial and Temporal Uncertainties

Metadata are available for each source report and are linked through the ATL_SRC.txt which gives information about navigational, sampling, and processing techniques where known. In cases where no original metadata were available, metadata were created based on existing available information accompanying the data. Of particular importance, site locations are as given in the original sources, with uncertainties due to navigational techniques and (or) datums ignored in the usSEABED compilation. As many reports are decades old, users of usSEABED should use their own criteria to determine the appropriateness of data from each source report for their particular purpose and scale of interest.

dbSEABED

Introduction

The dbSEABED processing system was developed by Jenkins (1997, 2002, 2003) in collaboration with the USGS and others over the past decade. Currently, there is no open-source code for the program. This explanation of the system is intended to give information and guidance about how the data are compiled, integrated, and processed in the usSEABED database.

The dbSEABED system aims to produce a unified mappable database from the multitude of data sets dealing with the seabed. The primary objective of the dbSEABED system is to produce integrated data that can be mapped, analyzed, and visualized. Data sets include both legacy and modern collections, involving data from samplings, and visual inspections. Filtering routines within dbSEABED unify marine geologic data that originally may be disparate in purpose, function, style, and collection or analytical techniques. It works on data files that hold the source data in their original values (except for minor unit changes and phonetically sensible word codes) and provides standardized output data. It is important that users of the `_PRS` and `_CLC` output files understand the parsing process, the meanings of field values, and the limitations of the usSEABED output.

More information about the dbSEABED software can be found on the dbSEABED Web site or in the Frequently Asked Questions section of this publication.

Data Import Methods

Incoming data files number in the hundreds and are diverse in content and format. The process of import begins with manual re-formatting which usually involves rearranging the data in columns specific to each parameter, such as color, percent sand, seabed description, or multisensor core logger acoustic velocities. These columns are arranged according to a template specific to dbSEABED. Most data are pre-arranged by columns, but in some cases sections of prose may need to be cut into their constituent parts.

New parameters are sometimes encountered as a data set is imported. These new parameters are added to the template at the ends of the appropriate data theme, and the dbSEABED processing software is modified to take the new parameter(s) into account if possible. For future reference and to help editing, the original data are often held as commentary metadata alongside the active data. Some data that are not useful to dbSEABED are held only as commentary metadata.

After import, the data are held in a type of written log arranged according to the nested sequence: data set / site / subbottom depth / subsample. Sites are specified by each new sampling operation. The written log structure is unusual for a database, having more in common with XML-format structure than relational databases. It has distinct advantages for dealing with seafloor sampling data sets, such as:

- (i) an algorithm can perform highly useful calculations on the data for each sample, which has made it possible to meet user demands in a timely way despite the complexity and size of the data holdings
- (ii) data that are human readable, especially if metadata are interspersed
- (iii) it conforms with data structures that are generated by core-loggers at sea and in the laboratory
- (iv) it is efficient to import
- (v) it is able to cope with variable data quality and incompleteness
- (vi) it is low maintenance, nonproprietary, and programs that address it have low cost of entry and are highly adaptable.

The disadvantage of the written log structure is that specialized programs such as those in dbSEABED are needed for conversion to the flat-file formats that most users require. These flat-file formats are provided in this publication.

The Numeric Data Type

A primary function of the processing programs is to read, quality-check, and then report numerical data that have been obtained from laboratory analyses of grain size, composition, color, shear strength and other parameters. Although we describe these data as "numeric," it also includes coded data such as Munsell color codes. In many cases, the numeric data can be echoed unchanged to outputs (in `_EXT` files), for instance, in percent sand, average grain sizes, carbonate, and porosities. Checks are performed, however, on whether a value is properly numeric or string, and if it is within plausible ranges. Problems are reported to a diagnostics file that is a basis for quality and completeness checks, with possible corrective edits to the data file (along with explanatory metadata). Data items are often deactivated if they are suspected as incorrect.

The numeric data output to `_EXT` files have had minimal manipulation. The data in grain size analyses (held at their original phi intervals) are summed into gravel, sand, silt, and clay percentages; the median, average

and standard deviations are calculated. If grain density is available, bulk densities and water contents are converted to porosities, with the porosity parameter adopted by dbSEABED. Many parameters that are available in the data are not reported to the `_EXT` files, for instance skewness and kurtosis. They may however be obtained from RDB renderings of the data (not available on this publication). The dbSEABED output of Central Grain size is a composite of median (preferred), moment average, and graphical averages. Currently, only the second moment (standard deviation) is directly transferred to the sorting field.

The reporting of consistent mappable values for geotechnical and acoustic parameters is not an easy task. The results of physical property tests are very dependent on experimental setup, such as strain rates, sample preparation, equipment dimensions, and detection of behavioral thresholds for the materials. The shear strength reported from dbSEABED is a composite of penetrometer, vane shear values (undrained, unconfined) in the unremolded states (i.e., for initial failure). Also included for the sake of maximizing mappability are the cohesions from shear box and low-pressure triaxial experiments. P-wave acoustic velocities are reported without regard to the frequencies of measurement. In both cases investigators wanting more specific information on the analyses can resort to the original data and metadata.

The extracted outputs based on numeric and coded data are put out separately from the parsed and calculated results of dbSEABED. It is recognized that some investigators will choose one over the other - or may wish to combine them in different ways. It must also be recognized that rarely can a sensible coverage of the seabed be obtained from the extracted data alone, as it is too sparse.

The Linguistic Data Type

A feature of dbSEABED is its ability to parse word-based descriptive data such as "brown fine sand with abundant shells; seagrass and some pebbles; whiff of h2s". These types of data are held using their original terms although some abbreviation and coding is necessary. Thus dbSEABED is not a natural language parser even for the noun phrase constructions, such as the above description. The ability to handle word-based data greatly extends the power of the system to map the seabed, because on a global average, approximately 85% or more of data characterizing the seabed are word based. Calibrations are performed to validate this process relative to analytical data on the same sediments. A simplified description of the parsing functions is included in this publication.

The dbSEABED program applies these concepts to geological descriptions, using:

- A parser that divides the descriptions into arithmetic equations
- A thesaurus that attaches meanings and memberships to the quantifiers, modifiers, and Objects
- A linear weighted assembly of the numerical totals

In the dbSEABED program, word memberships can be defined across many parameters, not just grain size. Fuzzy memberships are best thought of as a measure of truth or possibility (note: not probability).

The outputs are fuzzy memberships of parameters such as mud, grain sizes, carbonate, organic carbon, grain types, sedimentary features, rock and weed coverages, and engineering strengths.

Statistical comparisons can be made between the `_EXT` and `_PRS` data outputs, resulting in calibrations which are an overall guide to the accuracy of the regional mappings, and a highlighting of areas and issues in the data where improvements can be made.

Calibration

dbSEABED is an information processing system, that can perform statistical and individual tests of accuracy across the range of output parameters.

Statistical Tests

In the case of the thousands of samples where both analytical and descriptive data exists, a statistical comparison can be made between the `EXT` and `PRS` data outputs. The results of this calibration are an overall guide to the accuracy of the regional mappings, and a highlighting of areas and issues in the data where improvements can

be made. Those improvements involve both the analytical and descriptive raw input data. For example, grain size analyses that appear to be the whole sediment, but are really only of the sand fraction or analyses where gravel / shell have been omitted from an analysis.

The EXT and PRS outputs are imported into MS Access and links are created between the two files (based usually on the SampleKey). Entries with null values (-99) in either EXT or PRS are eliminated through a query. This query is brought into MS Excel and used to calculate the frequency distribution of deviations (+ and absolute) and plotted for inspection. Percentile statistics are calculated using the absolute deviation at the 50% (Median Absolute Deviation (MAD)), 68% and 95% percentiles (1s, 2s). Examples of the outputs are shown in the description of usSEABED. For most data sets the percentile statistics are 0.4, 0.8 and 4 phi for the 50, 68, and 95% levels, which may be acceptable over such a diverse set of input data sets but can be improved. An example of this analysis is shown in Figure 1 below, for a data set which is under improvement.

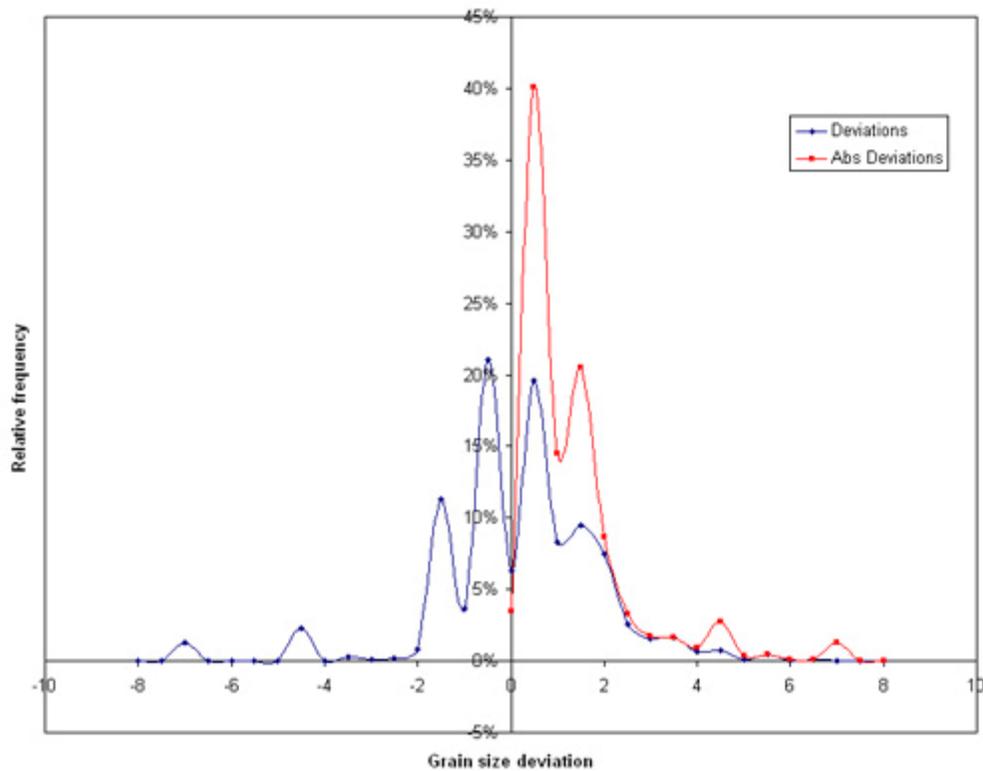


Figure 1. Statistical calibration of outputs for grain sizes, shown using a frequency plot of the deviations between PRS and EXT data, using the improving Atlantic data set. Deviations are the result of inaccuracies in EXT and PRS input data, as well as to identifiable issues in the data as highlighted below and in Fig. 2. The 50, 68, and 95 percentile confidences are 0.8, 1.3, 4.3 phi, respectively.

A second way of statistically evaluating the results uses a cross-plot between the EXT and PRS output data in the figure below. This type of plot serves to highlight some of the issues that may reduce the accuracy of dbSEABED with incoming data sets. At the locations A-D these common issues are identified in populations of points:

- _PRS coarser than _GRZ, apparently due to outsized shells / clasts being omitted from lab grain size analyses;
- sediment described as very fine in PRS, but only the sand fraction is represented by the EXT analysis data;
- detailed analyses of grain sizes does not go beyond coarsest class of about -2.5 phi;
- descriptive PRS data does not distinguish grades of sand, and is apparently dominated by reports of very large clasts, such as cobbles or shell, that were not analyzed.

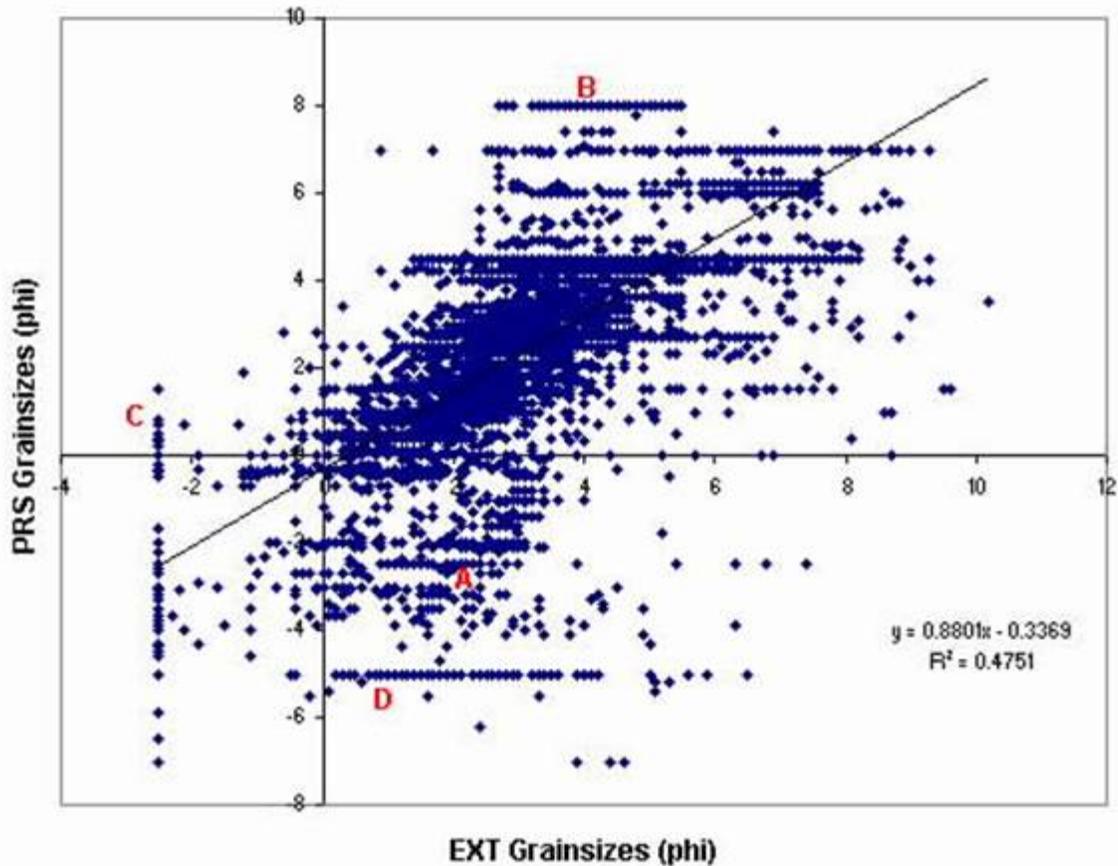


Figure 2. Cross-plot of the EXT and PRS output data for a region, not in US waters. A-D represent populations of data that have commonly encountered problems. With the detection and fixing of these problems the accuracy of outputs is substantially improved over that shown here. Notice that overall PRS extend further in coarse grades, and EXT further in fine grades, reflecting their common observational biases.

Individual Tests

The programs of dbSEABED have been equipped to detect problematic data, whether by values falling outside plausible limits, or by mismatches between EXT and PRS results. These tools normally do not prevent the problem values being output, but they do report detections to a diagnostics file that is particularly useful in the preparation and cleaning of incoming data sets. The statistical data shown in Figure 1 is employed to set the filters, usually at the 68% (1s) level. The original data can then be revisited, checked for issues such as those shown in Figure 2, and can be corrected, deactivated or left alone as appropriate.

Conclusion

Issues of accuracy and reliability become apparent as soon as data are integrated. Tools for monitoring the integration process are required, with feedback to the input data, so that improvements can be made in the system.

Basic uncertainties exist in all the incoming data that cannot be reduced and integrative systems cannot proceed past that uncertainty. Parallel studies in dbSEABED have determined on the basis of replicate analyses that analytical data such as grain size analyses (Syvitski and others, 1991) has 1-sigma uncertainties on the order of 4% of the total parameter range, or 0.8 phi. With good maintenance of the data, the outputs from dbSEABED approach those levels of reliability.

Expansion of Data Coverage

A summary of the theoretical or empirical relationships that are used by dbSEABED to expand the coverages of seabed parameters (often not directly measured or calculated in individual reports) is given in the onCALCULATION document.

Summary of the onCALCULATION Methods Used in dbSEABED

The module onCALCULATE performs a small amount of modeling to extend the outputs of dbSEABED to reasonable estimates of seafloor properties. This is particularly useful for geoacoustic and other physical properties which have no practical mappable data distribution based on actual analyses. The modeling is based on theoretical or empirical relations that are documented below. The chosen relationships may be replaced by others in the future as further physical properties research is published and onCALCULATION results are validated. Users of dbSEABED may choose whether to use these lower accuracy estimates in mappings or to work with map data coverages that are less reliable because they contain fewer data.

A couple of principles have guided the selection of relationships and their implementation: (a) the methods should be transparent and well documented, not complex, (b) they may be built on extracted or parsed parameter inputs but not on values that are themselves onCALCULATED, (c) they might apply only to certain sections of the input parameter range, (d) they preferably are published relationships, but if not then (e) should be supported by the analysis of a substantial amount of data from a wide range of sediments.

The methods that are currently implemented follow.

1. Grain size / Sorting from Gravel:Sand:Silt:Clay and Gravel:Sand:Mud Ratios

Gvl:Snd:Slt:Cly and Gvl:Snd:Mud (GSSC, GSM) ratios are in effect short grain size-fraction histograms for the sediments. An estimate of AV and SD grain size can be made from them as follows. Each class is assigned a central grain size value based on examinations of large USGS and other data sets in Excel. These GSSC(M) class central values result: -3, 2, 5, 7, (8) in phi.

A weighted mean and weighted standard deviation are formed across the GSSC or GSM classes, leading to an estimate of the average grain size and sorting of the sediment. A method of validation and uncertainty calculation is available by comparison of the results for samples where a mean/sorting is already measured.

2. Hydrographic Chart Bottom Type code

This code, described in UKHO (2005) and NOS (1997), has the form "Cy.S.Co". It is essentially the same as the US NOS codes. The calculation is mostly a matter of assigning textural classes in front-significant order based on the GSSC(M) ratios, but also with special classes "R", "Wd" where rock and weed memberships are significant. Thus, the output codes are minimal codes.

3. Folk and Shepard Classifications

These have been implemented following the schemes in Poppe and others (2000).

4. P-wave Velocities for Consolidated Materials Based on Time-Average Model

This calculation is performed only where there is an indication of cementation or consolidation in the material, usually expressed by measured or parsed Shear Strength >50 kPa or porosity <35%, and where the porosity has been measured. Then:

$$mVp = (1 - Poros) / VelSol + Poros / VelFlu$$

where Poros is the fractional porosity of the material and VelSol and VelFlu are the solid and fluid phase P-wave velocities. The constants are: VelSol = 5000, VelFlu = 1520 (these are measured values, different from those optimized for models such as Biot Theory; see Thorsos and others (2001)). The relation is associated with data over the whole range of porosities (Fig. 3). The time average model is attributed to Wyllie and others (1963).

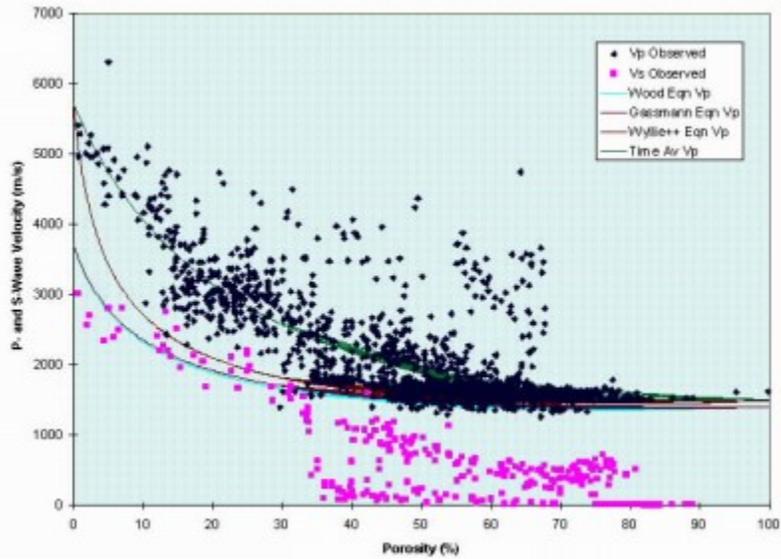


Figure 3. The distribution of sediments by their Porosity and P- and S-wave Velocity values. Several separate populations are apparent. For P-wave velocities, the loose sediments with low velocities from 34% porosity and following the Gassmann function; consolidated sediments following the Time-average relation; apparently cemented sediments with anomalously high velocities for the porosity. The S-wave populations have yet to be explained.

5. Porosity Based on Mud Content of Loose Sediments

A compilation of many published analysis (Fig. 4) results supports the empirical relationship:

$$mPor = 0.4 * mud + 43$$

for mud% > 7%. It appears to hold equally for terrigenous and carbonate sediments. Figure 5 illustrates the relationship for sediments of the Mississippi-Alabama-Florida (MAFLA) shelf.

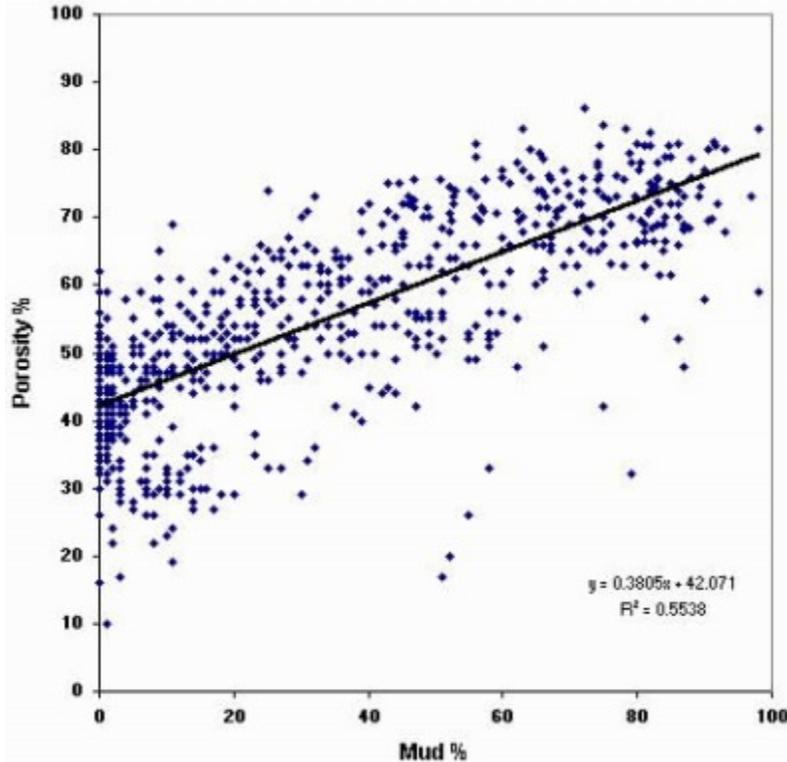


Figure 4. Empirical data supporting the Mud%-Porosity relation where mud fraction is >7%. The plotted data are a mix of terrigenous, carbonate, loose and consolidated samples, Mississippi-Alabama-Florida (MAFLA) shelf.

6. Porosity Based on Average Grain size

Richardson and Briggs (1993) proposed a relationship between porosity and average grain size (AvGrsz, phi units) based on their measurements of muddy and sandy sediments. The relation is inverted for the onCALCULATION, and is applied only in the range AvGrsz > 0 phi:

$$\text{Por} = 26.92 + 5.92 * \text{AvGrsz}$$

The form is less accurate than methods where the percent of mud is known, and is not used in those cases.

7. Coarse Fraction and the P-wave Velocity

Related to the porosity-mud fraction function is another between coarse fraction and Vp:

$$mVp = 0.0009 * SpG^3 + -0.14 * SpG^2 + 8.56 * SpG + 1512.76$$

where SpG is the percent coarse fraction. This polynomial is a poor fit and further work is required.

8. Wood-Gassmann Equation for P-wave Velocity in Loose Sediments

This method of estimating Vp is applied to sediments with no evidence of consolidation. It depends on assumed values for some acoustic constants of the sediments: (a) the Bulk Moduli K for solid, fluid, frame (Ko1Sol, Ko1Flu and Ko1Fra 3.6E+10, 2E+09 and 4E+08, MKS units); and (b) the Rigidities for solids and frame (RigidSol, RigidFra 2.2E+07, 1E+07).

The sediment Bulk Modulus and Density are:

$$Ko1 = (1-fPor)/Ko1Sol + fPor/Ko1Flu$$

$$\text{SedDens} = (1-fPor)*\text{DenSol} + fPor*\text{DensFlu} .$$

And the p-wave velocity is calculated as:

$$Qgass = Ko1Flu * (Ko1Sol - Ko1Fra)$$

$$(fPor * (Ko1Sol - Ko1Flu))$$

$$Kgass = Ko1Sol * (Ko1Fra + Qgass) / (Ko1Sol - Qgass)$$

$$VPgass = \sqrt{(Kgass + 4/3 * RigidFra) / \text{SedDens}}$$

Gassmann (1951). VPgass is output as the estimate of Vp.

The Wood-Gassmann relation is one of several that have been proposed between porosity and the acoustic velocities. As can be seen on Fig. 1, the relation has an associated population of data ranging only between 35-80% porosity. In the dbSEABED on-calculation it is applied only where porosity is known and >35% and where there is no indication of consolidation.

9. Roughness from Grain Protrusion for Gravels and Coarser

Kirchner and others (1990) offer a method for the calculation of grain protrusion (PnKir) above a sediment surface:

$$EnKir=0.5*(DnKir-D50Kir+(DnKir+D50Kir)*\cos(F100nKir))$$

$$PnKir=EnKir+\pi*D50Kir/12$$

where D50Kir and DnKir are the median and nth percentile grain sizes, and F100nKir is the Friction angle with a test grain of the 100-nth grain size percentile. In onCALCULATION, the estimation is done only using the central and the coarsest grain sizes (CSESTsz; either _PRS or _EXT) for D50Kir and DnKir. PnKir is output as an estimate of roughness.

10. Roughness Metric from Outsized Clasts

This metric was employed as an early measure of seabed roughness; it is based on an idealized arrangement of the sediment clasts and particles. The coarsest grain size is logged from previous processing of grain size analysis and descriptive data inputs or is estimated from the average plus 2 times the SD (sorting) where both are known. The vertical roughness is estimated as half the clast size (D) with allowance for natural oblateness:

$$dZ=0.5 * D * CSF$$

where CSF is the Corey Shape Factor to account for non-sphereicity. (In naturally worn materials CSF is about 0.7; e.g., Jimenez and Madsen 2003.) The spacing of the clast grain size is assessed as half the repeat distance implied by the fractional linear abundance P_L of the outsized clast with size D. Linear abundance is related to areal and volume (most common) abundances P_A and P_V as: $P_L=P_A^{1/2}$, $P_L=P_V^{1/3}$. The clast protrusion and half-spacing are output as the vertical and horizontal roughness scales.

11. Critical Shear Stress

If the material shows evidence of consolidation, then the Critical Shear Stress (CSS, N/m²) is set equal to the reported Shear Strength (kPa). Whitehouse and others (2000, p. 27) discuss the relationship, which is interim in the onCALCULATION.

For loose sediments, the functions related to grain size (AvGRZ) were investigated on the basis of data from many studies (Fig. 3). This work was done in conjunction with IOW, in Germany. The conclusions were: (i) with fine grained loose sediments where density or porosity are known, use the relationship of Mitchener, and others (1996, in Whitehouse, and others (2000)); (ii) else for those sediments use a generalized value of 0.5 N/m²; (iii) for loose coarse grained sediments use a log-linear relationship as shown in Fig. 3, $\log_{10}(CSS)=\log_{10}[1.04-AvGRZ*0.6]$.

Bioturbation and bioconsolidation were not recognized in the estimation process for fine sediments, though they can be important (see Black, and others (2002)). A correction of minor importance compared to the overall uncertainties is applied in the onCALCULATION.

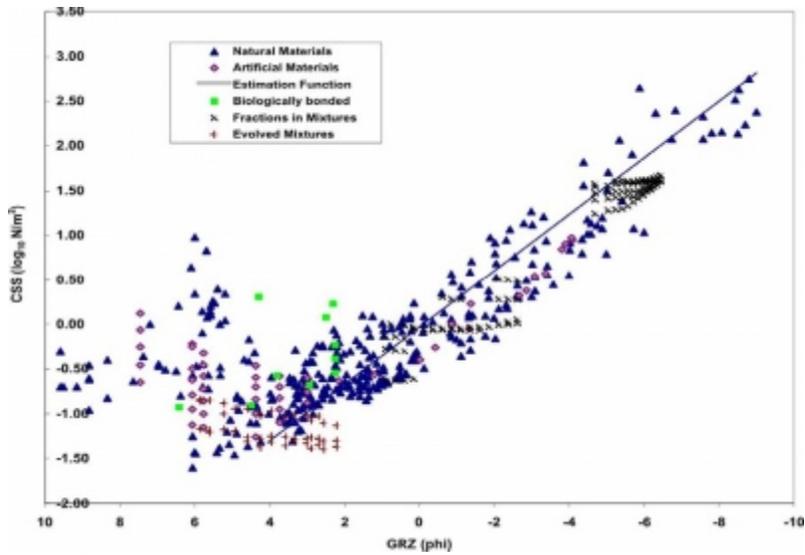
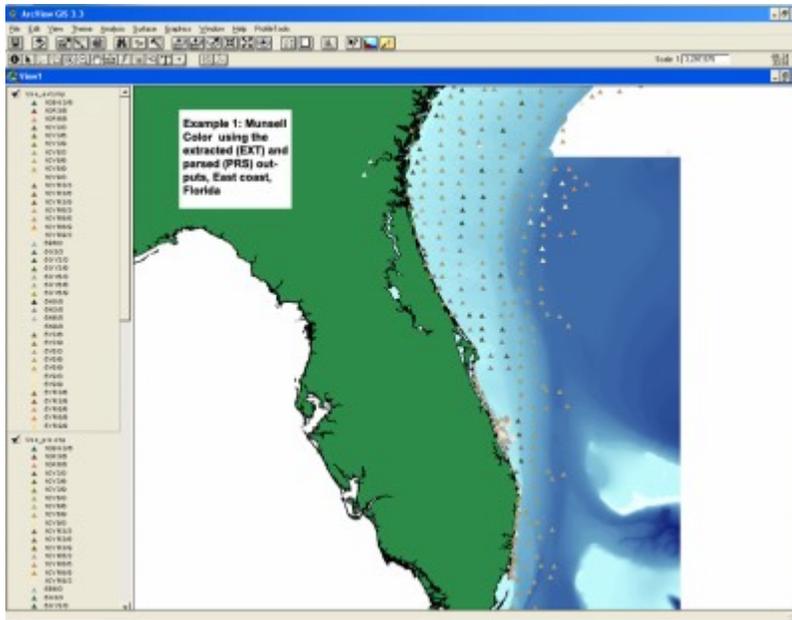


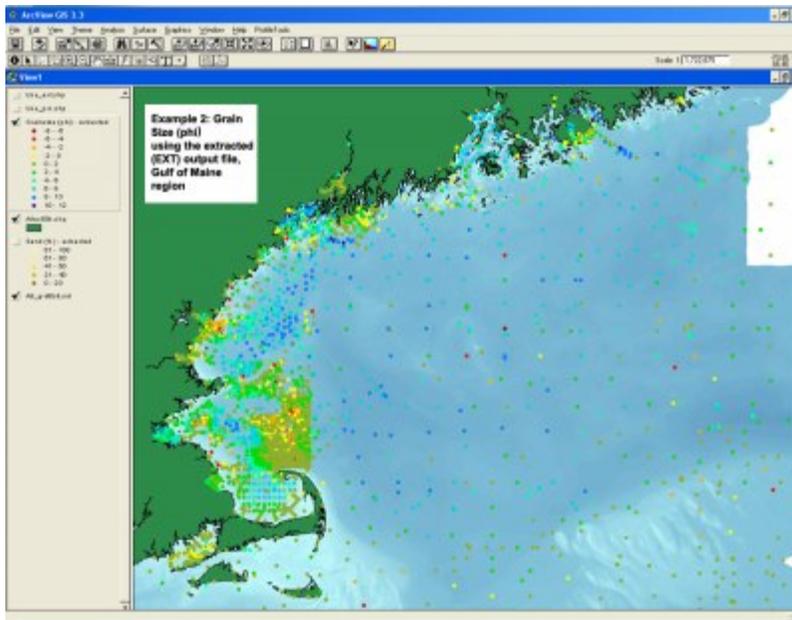
Figure 5. A compilation of Critical Shear Stress results by sediment grain size over the gravel to clay range. Over 24 references were used referring to marine and river sediments, field and laboratory experiments, on unaltered and manipulated natural sediments

Data Catalog

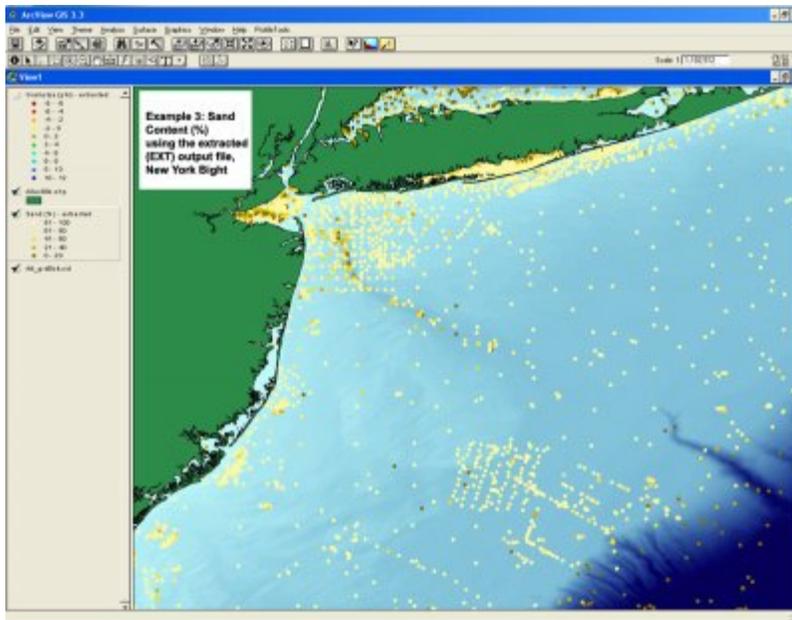
The data supplied on this CD-ROM are made available with geographic coordinates to allow the data to be incorporated into a Geographic Information System (GIS). The data layers along with additional base map layers have been compiled into an ArcView™ project file (usSEABED_Atlantic.apr), which is located at the top-level directory of this publication. The project file serves to provide an example of how the data can be displayed in a GIS. It contains several views including one which shows the Munsell Color legend on the _PRS data. A variety of basemap layers that can be used to accompany these data can be found on the Coastal and Marine Geology Program's U.S. Atlantic East Coast Map Server. Several have been included below and are used in the project file. Other examples of ways to visualize these data are also included below:



Example 1. Munsell Color using the extracted (EXT) and parsed (PRS) outputs, East coast, Florida.



Example 2. Grain size (phi) using the extracted (EXT) output file, Gulf of Maine region.



Example 3. Sand Content using the extracted output file, New York Bight.

For those who do not have the ESRI software or a compatible GIS data browser available on their computer, a free viewer, ArcExplorer, is available from ESRI . Please note that the ArcExplorer software is limited to the Microsoft Windows operating systems.

Clicking on the layer name under the column header "Data Layer Name & Description" in the table below will open a new window with a graphical representation of that layer. Federal Geographic Data Committee (FGDC) metadata are included with data layers in three formats in the table below: (HTML, FAQ, and text).

A downloadable zip archive file containing the elements that comprise the ArcView shapefile for each data layer is also provided. Each zip file includes:

- ArcView shapefile for each layer (with associated files)
- comma-delimited text version of the data file
- all three versions of the metadata
- browse graphic of the data layer

The zip files were created using WinZip v. 8.1. Users may obtain a free copy of the software from www.winzip.com. In addition to the ArcView shapefile, the usSEABED data layers are available in an ASCII text format as an alternate way to view and examine the data sets. The first record of the ASCII file contains the name of the data fields for that file.

Data Files

Sources

ATL_SRC – A comma-delimited text file containing a list of sources included in the output files. Also included is a list of all sources of data with links to associated metadata where available. Clicking on the DataSetKey will open a new window showing a browse graphic of that particular data set. Clicking on the DataSet name will open a new window showing the HTML metadata for that particular data set.

usSEABED Atlantic Data

Data Layer Name & Description	Metadata	Files	File Size
atl_ext- usSEABED extracted data for the entire U.S. East Coast	HTML FAQ txt	atl_ext.zip	3.13 MB
atl_prs - usSEABED parsed data for the entire U.S. East Coast	HTML FAQ txt	atl_prs.zip	4.50 MB
atl_clc - usSEABED calculated data for the entire U.S. East Coast	HTML FAQ txt	atl_clc.zip	6.58 MB
atl_cmp - usSEABED component data for the entire U.S. East Coast	HTML FAQ txt	atl_cmp.zip	0.98 MB
atl_fac - usSEABED facies data for the entire U.S. East Coast	HTML FAQ txt	atl_fac.zip	1.00 MB

Base Map Layers:

Data Layer Name & Description	Metadata	Files	File Size
NOS80K - Medium Resolution Digital Vector Shoreline (Source: NOAA)	HTML FAQ txt	nos80k.zip	25.9 MB
Coastal Relief Model Bathymetry - Image representation of CRM bathymetry (MrSid format) (Source: NOAA)	HTML FAQ txt	atl_grd.zip	1.91 MB
U.S. EEZ Boundary - Boundaries of the Exclusive Economic Zone (EEZ) (Source: NOAA)	HTML FAQ txt	useez.zip	156 KB
State Boundaries - internal U.S. State boundaries (Source: USGS/CMGP)	HTML FAQ txt	state_bounds.zip	1.76 MB

Legends

To map the coded information on Color and Roughness in a GIS, load the ArcView legends "munsell.avl" or "rgh_pt.avl" which are available with the database. ArcView legends may be imported into ArcGIS. To make your own legends for other applications, employ a classification that uses a "unique value" process

Color

Color of sediment is described either in terms (brown, light greenish gray) or in Munsell color codes given in values of hue (spectral content), value (lightness), and chroma (saturation). Munsell codes are explained in a publication of the Geological Society of America (Goddard, and others, 1951). The dbSEABED program converts the former into average values of Munsell codes, rounded to increments of 5 in hue, 3 in value, and 3 in chroma (Jenkins, 2003). An ESRI® ArcView legend is included for ease of mapping.

Roughness

This is a coded output representing the V:H of the seabed roughness element which is observed with greatest aspect ratio. That feature may be fixed roughness like a cobble, or moveable roughness like ripples. The outputs can only report observed roughness elements, so are influenced by the size scales of samplers and observations.

The V and H values are the centimeter values of the height and horizontal dimensions written in integer log 2 (base 2). For example "4:6" represents 16 cm height over length scale of 64 cm. Powers <0 are set to zero (i.e., scales <1 cm are not considered). The horizontal length H are the length of expression of a feature, rather than wavelength of repetition. Where a feature is elongated, H is taken normal to elongation (i.e., equals ripple wavelength).

ArcView Legends:

Munsell Color
Roughness: AVL | JPG

References Cited

- Black, K.S., Tolhurst, T.J., Paterson, D.M., and Hagerthey, S.E., 2002, Working with natural cohesive sediments: *Journal of Hydraulic Engineering*, v. 128, no. 1, p. 2-8.
- Folk, R.L., 1954, The distinction between grain size and mineral composition in sedimentary rock nomenclature: *Journal of Geology*, v. 62, no.4, p. 344-359.
- Folk, R.L., 1974, *The petrology of sedimentary rocks*, Austin, Tex., Hemphill Publishing Co., 182 p.
- Gassmann, F., 1951, Elastic waves through a packing of spheres: *Geophysics*, v. 16, no. 673-685.
- Goddard E.N., Trask, P.D., de Ford, R.K., Rove, O.N., Singewald, J.T. and Overbeck, R.M., 1951, *Rock Color Chart*: Geological Society of America. 6 p.
- Jenkins, C.J., 1997, Building offshore soils databases: *Sea Technology*, v. 38, no. 12, p. 25-28.*
- Jenkins, C., 2002, Automated digital mapping of geological colour descriptions: *Geo-Marine Letters*, v. 22, no. 4, p.181-187.*
- Jenkins, C., 2003, Data management of MARGINS geologic data, with emphasis on efficiency, quality control and data integration: *MARGINS Newsletter 10*, Spring 2003, p. 8-10.*
- Jimenez, J.A., and Madsen, O.S., 2003, A simple formula to estimate settling velocity of natural sediments: *Journal of Waterway, Port, Coastal and Ocean Engineering*, March/April 2003, p. 70-78.
- Kirchner, J.W., Dietrich, W.E., Iseya, F., and Ikeda, H., 1990, The variability of critical shear stress, friction angle, and grain protrusion in water-worked sediments: *Sedimentology*, v. 37, p. 647-672.
- Mott, J.L., Kandel, A., and Baker, T.P., 1986, *Discrete mathematics for computer scientists and mathematicians* (2nd ed.): Reston Publishing Co., 751 p.
- Nafe, J.E. & Drake, C.L., 1960, Physical properties of marine sediments, in Hill, M.N., ed., *The Sea*, v. 3, p. 794-815.
- National Ocean Service (NOS), 1997, Chart no. 1, United States of America, *Nautical chart symbols, abbreviations and terms*: US National Ocean Service [Chart].
- Poppe, L.J., Eliason, A.H., Fredericks, J.J., Rendigs, R.R., Blackwood, D., and Polloni, C.F., 2000, Grain size analysis of marine sediments: methodology and data processing, in Poppe, L.J., and Polloni, C.F., eds., *USGS east-coast sediment analysis: procedures, database, and georeferenced displays*: U.S. Geological Survey Open-File Report 00-358, CD-ROM.
- Richardson, M.D. and Briggs, K.B., 1993, On the use of acoustic impedance values to determine sediment properties: *Proceedings of Instruments of Acoustics*, v. 15, p. 15-24.
- Shepard, F.P., 1954, Nomenclature based on sand-silt-clay ratios: *Journal of Sedimentary Petrology*, v. 24, p. 151-158.
- Syvitski, J.P.M., LeBlanc, K.W.G. and Asprey, K.W., 1991, Interlaboratory, interinstrument calibration experiments, in Syvitski, J.P.M., ed., 1991, *Principles, methods, and application of particle size analysis*: Cambridge, Cambridge University Press, Cambridge, p. 174-193.

- Thorsos, E.I., Williams, K.L., Jackson, D.R., Richardson, M.D., Briggs, K.B., and Tang, D., 2001, An experiment in high-frequency sediment acoustics; SAX99: IEEE Journal of Oceanic Engineering, v. 26, p. 4-25.
- UK Hydrographic Office (UKHO), 2005, Chart symbols & abbreviations: Chart BA 5011 [Booklet], UK Hydrographic Office, 38 p.
- Wentworth, C.K., 1922, A scale of grade and class terms for clastic sediments: Journal of Geology, v. 30, p. 377-392.
- Whitehouse, R., Soulsby, R., Roberts, W., and Mitchener, H., 2000, Dynamics of estuarine muds; A manual for practical applications: London, Thomas Telford, 209 p.
- Williams, S.J., Jenkins, C., Currence, J.M., Penland, S., Reid, J., Flocks, J., Kindinger, J., Poppe, L., Kulp, M., Manheim, F., Hampton, M., Polloni, C., and Rowland, J., 2003, New digital maps of U.S. continental margins; Insights to seafloor sedimentary character, aggregate resources and processes: Proceedings of the International Conference on Coastal Sediments 2003: Corpus Christi, Tex., World Scientific Publishing Corporation and East Meets West Productions, CD-ROM.
- Wyllie, M., Gardner, G., and Gregory, A., 1963, Studies of elastic wave attenuation in porous media: Geophysics, v. 27, p. 569-589.

*Please contact Chris Jenkins at chris.jenkins@colorado.edu for reprints of these articles.

Acknowledgments

This publication is dedicated to Ms. Polly Hastings, (1964 - 2005) who made significant contributions to the usSEABED data collection effort, and to the Marine Aggregates project as well as many other USGS research activities. She was a valued colleague and friend who will be greatly missed.

We thank the following agencies for their contribution of data to usSEABED on the Atlantic coast:

Christina Rasmussen, U.S. Army Corps of Engineers, NY District
 Jane Uptegrove and Andrea Freidman, New Jersey Geological Survey
 Roger Amato and John Rowland, Minerals Management Service, Herndon, VA

We thank the following interns for their assistance in data entry, coding, and data testing: Viness Ubert, Emma Mitchell, Emily Denham and Shea Quinn.

We thank VeeAnn Cross, Jim Flocks, Jodi Harney, Guy Cochrane, Debbie Hutchinson, Don Woodrow, Rob Thieler, and Kathy Scanlon for their thoughtful and thorough reviews; and Valerie Paskevich and Carolyn Degnan for their reviews of the metadata; and Page Valentine for his contribution and information on benthic habitats on the Atlantic coast; and Brian Buczkowski for the 11th hour changes.

dbSEABED has benefited from the contributions of many people and institutions. It is a community structure, currently managed from the University of Colorado. Funding from the Australian Department of Defence, Commonwealth Scientific and Industrial Research Organisation (CSIRO) Australia, Geosciences Australia, Institute of Arctic and Alpine Research (INSTAAR) / University of Colorado, Institute für Ostseewissenschaften-Warnemünde (IOW, Germany), Lamont Doherty Earth Observatory, NOAA National Geophysical Data Center (Boulder), ONR: Office of Naval Research, and Victorian DNR (Australia). Mike Field and Jim Gardner who first arranged to apply dbSEABED to the US EEZ, in 1999. Ideas for development of dbSEABED have been contributed in discussions by L. Hamilton and P. Mulhearn (Defence Science and Technology Organization (DSTO)), G. Rawson, A. Short (Univ. Sydney), P. Sliogeris (Royal Australian Navy METOC Services (Australia)), T. Wever (Forschungsanstalt der Bundeswehr für Wasserschall und Geophysik, Germany (FWG)), J. Harff, B. Bobertz and B. Bohling (IOW), P. Morin (Univ. Minnesota), M. Kulp and S. Briuglio (Univ. New Orleans), J. Goff (Univ. Texas), G. Sharman and C. Moore (NOAA/NGDC), J. Flocks, C. Polloni (USGS) and the coauthors of this report.

Frequently Asked Questions (FAQs) about dbSEABED

dbSEABED

What is dbSEABED?

What kind of outputs does dbSEABED produce?

What quality control measures are in place for dbSEABED?

Output files

What is the SRC file output?

What is the function of the output CMP and FAC files?

What is the CMP file output for?

What is the meaning of the data in the FAC output file?

What do all the "-99"'s mean?

Why are there negative values for seabed strength and critical shear strength?

Compared to values given in some scientific papers, the output values from dbSEABED are quite rounded. Why?

What is the "Rock" membership?

What is the meaning of the SeaBedClass and ClassMembership?

What is the "Weed" membership?

Hydrographic Bottom Type (HBT) codes can be output. What do they mean?

What does the Roughness code mean?

Why do some points plot on the land?

Can the same sample or analysis be represented more than once in the data collection?

Why are the Kurtosis and Skewness not reported from the database?

Linguistic data versus numeric data

How does dbSEABED make word data conformable with numerical data?

What if a user doesn't want to use the word-based data for their mappings?

Which form of data should I trust more? Numeric-value analyses or word-based descriptions?

How does Word-Based Descriptive data relate to numeric-value analytical data?

Grain size

Why are values of grain size only given to 1 decimal place, and percent sand etc given as integers?

How are the grain sizes extracted from data?

Are detailed grain size analyses used to output textural statistics like mean and sorting?

Why would a sample that is known to have had a detailed grain size analysis not appear in the output data?

How are the outputs for the grain size fractions gravel, sand, mud extracted?

Are grain size scales other than Wentworth used?

How is one central grain size recovered from diverse data that may present mean, median and other graphical grain sizes?

How are the central grain size and sorting estimated in the CLC process of the database?

Color

Where can I find more information on the Munsell color code?

How are the Munsell color codes derived?

Data import methods

How does dbSEABED hold the basic data?

Why does dbSEABED hold its underlying data in documents rather than a relational database like Oracle?

What is involved in importing data sets into dbSEABED?

How is metadata treated?

Why are sediment descriptions and biological names held as abbreviations?

Data processing methods

What is Fuzzy Logic and how does it work?

What is the method of attributing sub-bottom depths to samples from cores?

How are the shear strengths derived?

What is the method for arriving at the SHEPARD and FOLK classes?

What relationships are used to estimate Critical Shear Stress values?

How are the porosity values obtained?

What is the basis for the data on compressional (p-wave) sound speeds?

How are the sorting values arrived at?

Carbonate values: how are they arrived at? And Organic Carbon?

How to

How can I map the coded information on Color and Roughness in a GIS?

How can I use the Critical Shear Stress in a practical application?

dbSEABED

What is dbSEABED?

- dbSEABED makes a single integrated data set from seabed sample information that has been collected over many years by ocean expeditions and research programs.
- dbSEABED is an “Information Processing” system that produces outputs that are compatible with Geographic Information System (GIS), Relational Database (RDB), as well as several other highly useful kinds of outputs. The outputs can be used in many end-user software applications.

What kind of outputs does dbSEABED produce? [top]

- For mapping in Geographic Information Systems, dbSEABED produces a number of flat-files that are comma delimited. Those files give the 20 most useful parameters, and information on grain types and

sedimentary features. These outputs can be enhanced in various ways, for instance by combining analytical and descriptive outputs, assigning gridded water depth values where not present, and re-arranging the data by its position within the seafloor.

- For use in Relational Database Systems like Oracle, MySQL or Access, dbSEABED produces a set of tables that are linked at the levels of seafloor site or sample using a system of numeric keys.

What quality control measures are in place for dbSEABED?

Quality control is practiced at many stages in dbSEABED:

- initially, by selecting which data sets to enter;
- by personnel scrutinizing the data at time of import, virtually item-by-item, for example for implausible values and positions not conforming to a report's survey pattern on a map; in some cases special filtering programs are written, for example to detect implausible ship speeds between stations;
- by testing in programs whether the data entries are text (string), numeric or a special formatted code as specified in the data model;
- in programs, testing if values lie within their plausible ranges, for example between -8 and 12 phi for grain sizes;
- by comparing site locations to land areas;
- in a more sophisticated way, comparing reported site water depths to a regional or global topography (such as etopo2);
- screening the program output values, also using plausibility filters;
- by having end-users of the data report any problems and having them fixed at source; this is exercising ('working') the data sets.

As each new data set is entered, it is tested in programs of dbSEABED to seek errors, etc. The program detections of problems are highlighted on screens during run time and are also reported to logs. As errors are detected, edits are made in the structured data files, complete with metadata explaining the edits. In some cases data will be deactivated (flagged out), again with metadata explaining.

Output files

What is the SRC file output?

- The "Source" (_SRC) file gives basic information on the survey and data set that a sample belongs to, including the collector and institution, confidentiality, number of samples, dates, region, etc.

What is the function of the output CMP and FAC files?

- The _CMP files provide information on the abundances (or prominence) of grain types, sedimentary features, and other components of the seafloor, for example of Halimeda algal grains, of ripples, and of hydrogen sulfide odor.
- The _FAC file synthesizes these data by applying a Fuzzy Set operator and creating a superset across a selection of components/features – a geologic facies. For example, a calcareous pelagic facies is defined as the FST membership across samples that show nannoplankton, planktic foraminifera, or pteropod remains, or are described as calcareous ooze.

What is the CMP file output for?

- Geologists and others frequently describe the presence and abundance of components in a sediment, like shell debris, quartz, heavy minerals, etc. They do this by grain counts or in their descriptions. dbSEABED assesses the abundances of these components and outputs some of them in the CMP file. Users of the software are able to say which components will be output. Counting grains (and gauging

their volumes or weights) is not an accurate process, and neither is conveying the abundance of a component in a description. The results in the CMP file are therefore only an approximate guide to abundances.

- The CMP file also carries fuzzy membership values for some features of the seabed, like odors (e.g., H₂S), and bioturbation and ripples.

What is the meaning of the data in the FAC output file?

Once the Memberships (Abundances) of components are calculated, they can be combined into an assessment of the membership a sample shows to a range of sedimentary, rock or ecologic facies. Each Facies is typified by components Af,Bf,... with memberships af,bf,..., for example a CalcareousPelagic facies as "pfrm nan ptr cal_ooz ; 1 1 1 1 0 0". The components are senior synonyms from the parsing dictionary. Given a set of components in a sample As,Bs,... with memberships as,bs,..., the sample membership of each facies is calculated as $\text{MIN}(ac,af)+\text{MIN}(bc,bf)+\dots$ for each case of a coinciding component. In Fuzzy Set Theory this is a set intersection - an AND.

Of course, the Facies output is possible only from the parsed word-based data. Because not all studies report grain components, it is advisable to plot these results as point symbols only, not areal griddings.

What do all the "-99"'s mean?

- "-99" is the null value for numeric fields output by dbSEABED. Null values will be found only in the output tables.
- We have to specify null values because otherwise some end-user mapping and analysis applications make decisions about what a blank field in the table represents.
- The value "-99" is used because it almost never occurs in real data. It does for longitudes however, and for latitude/longitude the null is "-999". The null for output string (text) data are "-".

Why are there negative values for seabed strength and critical shear strength

- These parameters are given in terms of their logarithm (base 10). A negative value implies a strength or shear stress less than 1.

Compared to values given in some scientific papers, the output values from dbSEABED are quite rounded. Why?

- Studies of uncertainty show that accuracies on most sediment parameters are of order 1-3% of the total parameter range even under favorable laboratory conditions. The significant figures in dbSEABED outputs reflect the observed uncertainties.

What is the "Rock" membership?

- Many observations of the seafloor report the presence of rock, loose and as bedrock. This output statistic is meant to convey the degree of exposure of rock which is coarser than cobble (-8 phi). Loose rocks and bedrock which are partly covered by sediment give memberships of less than 100%.
- Terms which have rock memberships include: basalt, granite, limestone, rock, boulders, lithified, hard, hard bottom, reef, rock platform, greywacke, where they occur in LTH or SFT themes, that is are not part of a grain-type description as in a PET theme line. For output data on specific rock types like basalt, refer to the CMP output file.
- Note that samples which have only loose sediments recorded are given a value of zero rock membership in CLC outputs.

What is the meaning of the SeaBedClass and ClassMembership?

These columns put out the facies showing the largest membership value for each sample, and that largest membership. The memberships are calculated as described in the FAQ about the FAC file. Output occurs only provided that the membership is greater than or equal to 0.33. Of course, this output is possible only from the parsed word-based data.

What is the "Weed" membership?

- Many observations of the seafloor report the presence of soft algae, seagrasses in general terms or specific to a taxon. Depending on the associated abundance terms like sparse, abundant, meadows, scattered, etc, the weed is given a membership which is reported to outputs (PRS). There are no EXT outputs for weed membership. Calcareous algae like Halimeda and Lithothamnion are not included.
- Terms which have weed memberships include: Zostera, seagrass, soft algae, kelp, Cladophora, weed. These data are usually in a SFT theme line. For specific data on seagrasses, refer to the CMP output file.

Hydrographic Bottom Type (HBT) codes can be output. What do they mean?

- Many navies and engineering groups use these codes. The British Admiralty set used here (British Admiralty, 1973) are not very different from US usage within USCG and NOAA. Some users of dbSEABED will find the codes familiar and helpful. Note that the terms at the front of the code are the most significant (abundant). For the clearest plottings, use the no-overlap options in your GIS and plot only the surficial seabed samples of an area.
- The codes output in EXT are either passed direct from naval/engineering HBT codings or re-codings of the lithological types that scientists have described in sediment samples. The codes output in CLC refer only to the gravel, sand, mud, rock and weed components present in a sediment or at a site.

Note: Data for Hydrographic Bottom Type have not been included in this publication.

What does the Roughness code mean?

- This is a coded output representing the V:H of the seabed roughness element which is observed with greatest aspect ratio. That feature may be fixed roughness like a cobble, or moveable roughness like ripples. The outputs can only report observed roughness elements, so are influenced by the size scales of samplers and observations.
- The V and H values are the centimeter values of the height and horizontal dimensions written in integer log 2 (base 2). For example "4:6" represents 16 cm height over length scale of 64 cm. Powers <0 are set to zero (i.e., scales <1 cm are not considered). The horizontal length H is the length of expression of a feature, rather than wavelength of repetition. Where a feature is elongated, H is taken normal to elongation (i.e., equals ripple wavelength).

Why do some points plot on the land?

- Some of the data collections we used contain data in estuaries, rivers and lakes; others contain data for coastal dunes, beaches, even coal mine pits. We could have culled these data, especially the ones that are obviously not marine. But the accuracy of maps of the national shoreline is not good enough to do a cull based on position of samples alone. And some people will find the dune, river, and lake data useful. Therefore we left all the data in.

Can the same sample or analysis be represented more than once in the data collection?

- Yes, when there is a good reason. For instance, the same sample may be described by two different labs, or be re-analyzed years later. Or an analysis may be expressed in a different way by later work. These

double-up results are all valid, and should be included in the database for mapping. (Note that the duplicate analyses may not carry exactly the same sample names.)

Why are the Kurtosis and Skewness not reported from the database?

- These higher order moments are not only rarer in data sets, but to be accurately reported require a higher standard of data and calculation than lower moments. Statistical moment and graphical measures are also very difficult to reconcile, exacerbating the shortage of available data on which to base outputs.

Linguistic data versus numeric data

How does dbSEABED make word data conformable with numerical data?

- dbSEABED software attaches Fuzzy Set Theory memberships to word terms through a look-up dictionary table. The memberships are summed across a description and the totals are output. Some words like “slightly” only adjust the memberships. Other words adjust the character of another, for instance “fine” in “fine sand”. For more detail, see Jenkins, 1997.
- dbSEABED also recognizes if a word is not in its dictionary, if the meaning is unknown. For example: “zyzgy” is not in the dictionary and will be labeled an ‘unknown’. It also recognizes if a word is neutral for a question. For example: “sediment” is neutral on the question of color.

What if a user doesn't want to use the word-based data for mappings?

- Users have a choice: if only numerical analytical data are required, then the ‘extracted’ (_EXT) form of output should be used; if word data only, then the ‘parsed’ (_PRS) form. Some users will not want to use the ‘calculated’ (_CLC) outputs which have a higher level of uncertainty than these, and that is possible too. If a user wants to integrate these to achieve the best possible coverage in a region, then they can be added or telescoped together using two other formats (_ALL and _ONE formats respectively; not included in this publication).

Which form of data should I trust more? Numeric-value analyses or word-based descriptions?

There are several issues that users must consider in deciding which form of data to use:

- Issue 1. Descriptive data are more representative of the seabed, the shells, coral, rocks and other items that will not fit in laboratory bottles or analytical instruments but which have a power effect on erodibility, acoustics and habitat. It is also essential for dealing with the creatures of the bottom, burrows and other structures, sediment strength, odors, color, etc.
- Issue 2. To get enough data to make a map with a useful number of points, you will have to use word-based data, which account for 85% of all available data in usSEABED.
- Issue 3. Analytical data are more precise, often precise as to within 2% of the full parameter range. And it is repeatable, given the same instrument unlike descriptions which may vary between observers, even trained ones. However it may not be accurate for the seafloor, as it usually only deals with the seabed matrix – those materials fine enough to fit inside the instrument tubes / apertures.

How do word-based descriptive data relate to numeric-value analytical data?

- The values of grain size, percent gravel, carbonate, etc., that are calculated from word-based data are based on Fuzzy Set Theory calculations. They are memberships, reduced to a single value. An analysis is also a single realized value of the Fuzzy Set.
- Sometimes word-based and numeric-value results are available for the same sample, and then an inter-calibration of the two forms of output can be made.

Grain size

Why are values of grain size only given to 1 decimal place, and percent sand, etc., given as integers?

- Work in the dbSEABED project has established the typical accuracies on measurements of these parameters. Those accuracies, even for careful work are of the order of 2-5% of half the total range of the parameter. The table outputs are designed to be brief to restrict data volumes and make mapping faster and easier, and to carry the data at a precision that is appropriate to measurement accuracies.

How are the grain sizes extracted from data?

- By a simple reporting of the average grain size, median grain size, Inman Mean, and Folk Graphic Mean grain sizes. A comparison of data sets shows that these reflections of 'central' grain size are not significantly different within usual error bounds on sampling then analysis. (On the other hand, sorting values are significantly different and cannot be combined in the same way.)

Are detailed grain size analyses used to output textural statistics like mean and sorting?

- Yes. The percentages in each fraction are summed to output the percent gravel, sand, mud and clay.
- A weighted average and standard deviation is also calculated, leading to output of central and sorting grain sizes. Standard moment statistics (see Blatt and others, 1980) are used.

Why would a sample that is known to have had a detailed grain size analysis not appear in the output data?

- Grain size analyses are scanned to check that they are in good order, and some are rejected. An analysis with any phi interval is acceptable, provided it is fine enough to resolve those fractions. Analyses that have a significant weight percent in the finest and coarsest classes are treated as suspect because this implies that the part of the sediment that was analysed probably does not represent the whole sediment.

How are the outputs for the grain size fractions gravel, sand, mud extracted?

- Many data sets contain these values (based on Wentworth scale), and in those cases the values are passed through to outputs (EXT). In some cases the data are presented in the form of a detailed grain size analysis – such as at $\frac{1}{2}$ phi intervals. dbSEABED assembles grain size analysis streams into G:S:M fractions by assigning each analysis class to its fraction (or proportioning if the class straddles a fraction boundary).

Note: Many grain size analyses techniques range only through sand and mud, and gravel is not analyzed. In these cases the database reports the gravel percent as null in outputs.

Are grain size scales other than Wentworth used?

- Generally no. However, an option does exist for users to report percent mud from values of 'engineering' grain size mud (i.e., finer than #200 Sieve, 75 μ m).

How is one central grain size recovered from diverse data that may present mean, median and other graphical grain sizes?

- The database adopts mean moment grain size as the standard for its measure of a sediment's central grain size. Studies show that the median, Inman 'mean', and Folk 'mean' grain sizes fit this quite well for a wide variety of sediments. However mode grain sizes do not, and are not included in 'central grain size'.

How are the central grain size and sorting estimated in the CLC process of the database?

- Many sediments have gravel:sand:silt:clay (g:s:s:c) or gravel:sand:mud (g:s:m) ratios reported, but are without basic grain size statistics. Some users of dbSEABED requested that 'best estimates' of the statistics be made by modeling grain size distributions. This is done by creating a weighted mean and standard deviation from the g:s:s:c and g:s:m ratios of the samples.
- The class statistics underpinning the modeling were arrived at by examining the most common values for single phase sediments in USGS data sets (e.g., just silt). The values were for central grain sizes g:s:(m):s:c -3.0: 2.0: (7.0): 5.0: 8.0 . Typical values of sorting were also applied: for g:s:(m):s:c 1.4: .0.9: (2.1): 3.0: 5.8. These values may be adjusted in the future.

Data processing methods

What is Fuzzy Logic and how does it work?

- "Fuzzy Logic" (more properly "Fuzzy Set Theory") allows an object to belong partially to a set. In classical 'Crisp' Set Theory objects are either in a set or not. FST suits words because they are often partial carriers of meaning. For example, 'warm' is partially hot and partially cold. A formal arithmetic for Fuzzy Sets was discovered by Zadeh, (1965). A good reference is Mott, and others, (1986).

What is the method of attributing subbottom depths to samples from cores?

- Ideally a sample from a core will have a subbottom depth assigned in terms of meters below the sediment surface. However this is not always the case. While this information may have to be left unknown if not given by the original researchers, some limits can be placed if the sampler type is known. For example a Shipek Grab has only 5 cm of penetration usually. Then a subbottom range of 0-0.05m can be assigned. In cores that may be several meters long, the limitation is less strict.

How are the shear strengths derived?

Shear strength values are obtained either from actual measurements held in the database, or assessed from descriptions that convey lithification or consolidation.

- These measures of strength are accepted: hand penetrometer strength, (maximum, unremolded) vane shear strength, cohesion values from triaxial and shear box tests, compressive shear strength at low confining pressures, cone resistances. Obviously, uncertainties arise not only from measurements themselves but by combining different measures like this.
- Many descriptions refer to a sediment or rock being "soupy", "loose", "soft", "cohesive" ("firm"), "stiff", "friable", "cemented", "lithified" or "hard". These terms are indications of the mechanical state of the materials and are transferred to a Shear Strength value in PRS outputs. They are (respectively): 2, 25, 50, 100, 500, 1000, 5000, 8000, 10000.

What is the method for arriving at the Shepard and Folk classes?

- The original Shepard (1954) and Folk (1954) ternary classifications have had to be modified in a couple of ways: (a) Where the silt and clay breakdown of mud is not available, the silt-clay domains in the classifications are merged under one name; and (b) Because these schemes refer only to sediments, an extra class “SOLID” is added for dbSEABED to classify all lithified materials (rock). The aim of (a) was to have a class attached to a greater number of samples. Very few samples in the database have silt: clay specified.

What relationships are used to estimate the Critical Shear Stress values?

Depending on the quality of inputs and nature of the sediment, several relationships are used to predict Critical Shear Stresses. The details and supporting references are given in the onCALCULATION document. The relationships include:

- Where the consolidation is known and substantial, the CSS is set equal to the value of the Shear Strength.
- For unconsolidated granular sediments a compilation using hundreds of experimental results from the field and laboratories suggests a linear relationship with phi grain size.
- For unconsolidated fine-grained sediments (>5phi grain size) where there is information on density or porosity, the CSS is based on the density using published relationships.
- For unconsolidated fine grained sediments without information on density or porosity, a general value is used.

<>No regard to the effects of bio-consolidation or bioturbation is made.

How are the porosity values obtained?

- First, values that are in the database are passed to the EXT outputs. In addition, where bulk or dry densities, void ratios or moisture contents are present, they are transferred to a porosity value. Where this requires a grain density, this may be available, or is assumed to be 2.5 g/cm³ which is suitable for a wide range of materials.
- To predict porosities for unconsolidated sediments we use an empirical relationship between porosity and grain sizes described by Richardson and Briggs' (1993): $AvGrsz = -4.55 + 0.169 * Por$, with R² of 0.81 in the range 0-12 phi. We use the inverted form: $Por = 26.92 + 5.92 * AvGrsz$. These results appear in the CLC outputs. Other unpublished relationships of higher accuracy are available.

What is the basis for the data on compressional (p-wave) sound speeds?

Although there are some direct measurements of compressional sound speeds for areas of the seabed, they are few and far between. dbSEABED reports any direct measurements in the _EXT outputs, including the average of those data which recognize velocity anisotropy. One difficulty with sound speeds is that they do depend on the frequency, pressure and temperature of the measurement. The outputs make no allowance for these factors, though conditions of the original measurements are recorded in the underlying raw data of the database.

Several users of dbSEABED requested that estimated sound speeds be provided wherever possible. This is done as follows.

- Where the material is consolidated a measured porosity is available, then the time average equation is employed (Nafe and Drake (1960)). Consolidation is judged to be when the ShrStrs based on measurement or a parsing of compaction/cementation is > 50kPa, or the porosity is < 33%. These values for constants were adopted: solids and fluid densities 2500 and 1025 kg/m³, respectively; sound speed of solids and fluids 5000 and 1520 m/s, respectively.
- Alternatively, where the material is a loose sediment the velocity is calculated based on the relationships of Richardson and Briggs (1993): $Vb/Vw = 1.18 - 0.034 * Mz + 0.0013 * Mz^2$ with an R² of 0.82, where Vb/Vw is the ratio of seafloor sediment and bottom water p-wave velocities. For an absolute output velocity, this result is combined with the bottom water velocity (assumed 1520 m/s).

Other unpublished relationships of higher accuracy are available. The method makes no allowance for the consolidation of a sediment within the seafloor, for example in a core.

How are the sorting values arrived at?

- Values of moment sorting are simply passed through to output. Graphical measures of sorting such as those of Inman and Folk are not generally compatible with moment measures and at present do not contribute to outputs.

Carbonate values: how are they arrived at? And Organic Carbon?

- Carbonate measurements that have been made on the complete sediment are simply reported in the EXT outputs. Many sediments have had carbonate determined on just the sand or mud fractions: those results are not output because they do not represent the whole sediment.
- Carbonate values are also output from the parsing process. The carbonate parts of sediment are summed in quantity, as are the non-carbonate and unknown carbonate parts (like 'mud'). If the sum of unknowns is less than 5% then a carbonate membership is reported in PRS outputs. The accuracy of this result is not as good as for analysis data.
- Organic carbon data are treated the same way. However, because OC values are typically small in sediments (~1%), the accuracies in the PRS results are not good and should be treated with caution at this stage.

Data import methods

How does dbSEABED hold the basic data?

- Data sets for processing are held in structured documents that are called "Data Resource Files" (DRF). They set out data in the form of a written geological core-log, in a tree structure that nests the data according to expedition, the sample site, the sample, and finally phase inside the material. This format is processed by programs to produce other formats that are compatible with GIS and RDB.

Why does dbSEABED hold its underlying data in documents rather than a relational database like Oracle?

This is one of the most successful features of dbSEABED.

- First, it is more efficient to bring in data sets from geologists in the form of written core-logs rather than tables, especially as tables and sub-tables. This is how dbSEABED has been able to quickly absorb a million data sites.
- Second, it is not possible (easily, inexpensively) to do some sedimentological operations on the data in software like Oracle. For example, parsing the word-based descriptions.
- Third, in the documents, it is possible to have data which was obtained at the same time (such as replicates, suites of analyses) to be kept right next to each other. The same is true for the metadata; a person can read the data and the metadata about it easily and at the same time.
- Fourth, it keeps the data very faithfully in its original form. Relational databases require that data be shuffled and diced, and to obey sets of rules before can be entered. Data are kept faithful to that of the original observers.
- Relational Databases are just one type of data structure. There is no reason why dbSEABED shouldn't be able to generate others too, like XML, and Spreadsheets and GIS tables.

What is involved in importing data sets into dbSEABED? [top]

- This depends on the structure and idiosyncrasies of the original data. Many data sets are simple to import into the DRF format. Others have proven to be most difficult. It is usually done by arranging data items by column in EXCEL according to a template that sets out the field locations for each data item (such as latitude, sampler type, phi grain size, Munsell color code). Different types of data are divided between themes such as geotechnical, texture, or color. Some themes allow for a sample to be from the subbottom (in a core). Others such as seafloor type which is dedicated to surficial sediment surveys such as those by divers and remote vehicles, treat only the seafloor surface.
- Although dbSEABED tries to hold data as close as possible in its original terms, this is not always possible. In the case of numeric data items, if a simple conversion of units is desirable – for instance from cm into m – then that can be done at data import stages. In the case of word descriptions, it would be unwieldy to hold or process the descriptions in original long words and words are therefore converted to abbreviations. Note that this does not change the terms, it is NOT a re-classification. Terms such as “lithothamnion”, “scattered”, “low water content” are abbreviated to “lthmn”, “scatrd” and “lo_watr_cntnt”. This speeds processing, allows importing staff to check the data for spelling errors, allows us to use foreign languages (e.g. “sabb(IT) for “sabbiosi”), and permits homonyms to be distinguished (eg: dens_ab” and “dens” for abundance and geotechnical densities). It also allows the data to be read more easily on computer screens.

How is metadata treated? [top]

- Metadata are treated at several levels. First, over-arching metadata have been compiled for the processed outputs of dbSEABED (see “http://instaar.colorado.edu/~jenkinsc/dbseabed/resources/db9_MetadataFGDC.txt”). Second, the USGS and others have compiled metadata for individual data sets as they are entered into dbSEABED. Third, metadata describing measurement techniques, features and problems in the data, edits that we have made in dbSEABED, who collected and analysed the data, etc is held in direct association with the data in the DRF files. This most detailed metadata are best viewed in the RDB structure, which also puts out reports when data values are beyond plausible limits, for example a grain size of 14 phi.

Why are sediment descriptions and biological names held as abbreviations?

- To make the parsing (and dictionary lookup) computationally faster, to make long descriptions shorter and more readable, to give flexibility in handling homonyms, and to better distinguish active data from metadata in the structured documents.

Color

Where can I find more information on the Munsell color code?

- Refer to the Geological Society of America “Color Rock Chart,” or to the company GretagMacbeth which is the modern-day custodian of Munsell's color technologies.

How are the colour Munsell codes derived?

- This process is fully described in Jenkins (2002). Munsell codes are explained in a publication of the Geological Society of America (Goddard, and others, 1951). The essential steps in the treatment of colour data are as follows:
- If data are available in Munsell Code form then that is transferred to outputs.
- Where a color description like “light greenish gray” is available, then that is parsed by a weighted summing of the hue (color), chroma (colour intensity) and value (greyiness) of the terms. From these components a Munsell Code is formed. The process is calibrated using the GSA scales and others.
- Outputs are stepped at intervals of 5 in hue, and 3 in chroma and value. Otherwise the full range of >3240 possible natural codes would not be mappable. This stepping arrangement reduces the number

to about 40, as seen in the ready-to-use legend that is available with this publication. That legend was constructed by adjusting the RGB values of the symbols to match their Munsell colors as seen on a computer screen.

How to

How can I map the coded information on Color and Roughness in a GIS?

- Load the ArcView legends “munsell.avl” or “rgh_pt.avl” which are available with the database. ArcView legends may be imported into ArcGIS.
- To make your own legends for other applications, employ a classification that uses a “unique value” process.

How can I use the Critical Shear Stress in a practical application?

- The CSS and grain size are essential components for calculating the Shields criterion of sediment erosion, between the Shields Parameter and the Grain Reynolds Number. Those parameters can be calculated once the flow characteristics, fluid velocities and densities, etc are known. See standard textbooks on sediment erosion for more information (e.g., Soulsby (1997)).

References:

- Blatt, H., Middleton, G.V. and Murray, R.C., 1980, Origin of sedimentary rocks (2nd ed), Prentice-Hall, Inc., 782 p.
- British Admiralty, 1973, Chart 5011. Symbols and abbreviations used on Admiralty Charts (new series). The British Hydrographic Office, Taunton, UK.
- Folk, R.L., 1954, The distinction between grain size and mineral composition in sedimentary rock nomenclature: *Journal of Geology*, v. 62 n.4, p. 344-359.
- Folk, R.L., 1974, The petrology of sedimentary rocks, Austin, TX, Hemphill Publishing Co., 182 p.
- Goddard E.N., Trask, P.D., de Ford, R.K., Rove, O.N., Singewald, J.T. and Overbeck, R.M., 1951, Rock Color Chart. Geol Soc Am, Colorado.
- Mott, J.L., Kandel, A., and Baker, T.P., 1986, Discrete mathematics for computer scientists and mathematicians (2nd ed.), Reston Pub Co, 751 p.
- Nafe, J.E. & Drake, C.L., 1960, Physical properties of marine sediments. In: Hill, M.N. (ed), *The Sea*, v. 3, p. 794-815.
- Poppe, L.J., Eliason, A.H., Fredericks, J.J., Rendigs, R.R., Blackwood, D., and Polloni, C.F., 2000, Grain size analysis of marine sediments: methodology and data processing. In: Poppe, L.J., and Polloni, C.F. (eds.) USGS East-Coast sediment analysis: Procedures, database, and georeferenced displays, U.S. Geological Survey Open-File Report 00-358, CD-ROM.
- Richardson, M.D. and Briggs, K.B., 1993, On the use of acoustic impedance values to determine sediment properties: *Proc. Inst. Acoust.*, v. 15, p. 15-24.
- Shepard, F.P., 1954, Nomenclature based on sand-silt-clay ratios: *Journal of Sedimentary Petrology*, v. 24, p. 151-158.
- Soulsby, R., 1997, Dynamics of marine sands, Thomas Telford, London, UK, 249 p.
- Zadeh, L.A., 1965, Fuzzy sets: *Information and Control*, v. 8, p. 338-353

Tables A-E

Table A. Key to data themes in usSEABED output files

Acronym Meaning

ACU	Acoustic properties
BIO	Biota
CMP	Sediment composition analyses
COL	Color
GRZ	grain size analysis results
GTC	Geotechnic properties
LTH	Lithology
MSL	Multisensor core logger
PET	Grain petrology
SFT	Seafloor type descriptions
TXG	Graphical texture statistics
TXR	Texture statistics

Table B. Field parameters, format, units, range, meaning, and comments for _EXT, _PRS, _CLC data files

Field Name	Parameter	Data Format	Units, Range, Meaning	Comment
Latitude	Latitude	Decimal 00.00000	Decimal degrees, 90° to -90° range	WGS 84 Spheroid.
Longitude	Longitude	Decimal 000.00000	Decimal degrees, - 180° to 180° range	WGS 84 Spheroid.
WaterDepth	Water depth	Integer 00000	Meters	Not always tidally correct.
SampleTop	Sample top	Decimal 000.00	Meters below seabed surface	Sample top as noted in source report.
SampleBase	Sample base	Decimal 000.00	Meters below seabed surface	Sample bottom as noted in source report.
SiteName	Site name	Character XXX: XXX	Survey or laboratory code for the sampling site	Not unique; site name as given in report; sometimes linked to cruise name or other information to decrease site name overlap.
DataSetKey	Data set number key	Integer 000	For audit only	Relational key to _SRC file; _SRC file contains links to source metadata.
SiteKey	Site number	Integer	For audit only	Relational key to other data

	key	0000000		files. Each site counted sequentially as total output; core data may have more than one sample per site.
SampleKey	Sample number key	Integer 0000000	For audit only	Relational key to other data files. Each site counted sequentially as total output; Multiple samples may be at each site (i.e., in core).
Sampler	Sampler type	Character XXXXXXXX....	Type of sampling device	As given in source report; recovery (rcvy) or penetration ('pen') length appended if given in source report. For more complete information on sampler, see source metadata.
DataType	Data types	Character XXX: XXX	For audit principally	Source data types (Table A).
Gravel	Gravel	Integer 000	Gravel grain size fraction, %	Textural class.
Sand	Sand	Integer 000	Sand grain size fraction, %	Textural class.
Mud	Mud	Integer 000	Mud grain size fraction, %	Textural class.
Clay	Clay	Integer 000	Clay grain size fraction, %	Textural class; output for '_EXT' only, as clay value can be determined only by analysis.
grain size	grain size	Decimal 00.00	Phi characteristic grain size	Consensus of mean and median grain sizes.
Sorting	Sorting	Decimal 0.00	Phi grain size dispersion	Standard deviation, sorting only.
SeafloorClass	Seafloor class	Character XXXXX...	That class (or 'facies') with the maximum fuzzy membership, if above 30%	Output for '_PRS' table only.
ClassMbrshp	Class membership	Decimal 000	Fuzzy membership (%) of the class (or 'facies'), noted above	Output for '_PRS' table only.
Folk classification; Shepard classification	Folk classification; Shepard classification	Character xx.XX...		
RockMbrshp	Rock index	Integer 000	Fuzzy membership (%)	Membership of sample to 'rock' fuzzy set; reported only in _PRS data
WeedMbrshp	Weed index	Integer 000	Fuzzy membership (%)	Membership of sample to weed fuzzy set; reported only in _PRS data
Carbonate	Carbonate	Integer 000	%; may be Fuzzy	

			membership (_PRS)	
MunsellCode	Munsell Color code	Character XXXXX	Standard alphanumeric coding of color partitioned into Hue, Value, and Chroma	Ex: '5YR 6/4', See "Rock-Color Chart", Geological Society of America, 1991.
OrganicCarbon	Organic carbon	Integer 000	%	Minimum value from descriptions (PRS tables) is 0.1%.
ShearStrength	Log shear strength	Decimal 00.0	kiloPascals, undrained, unconfined	From a variety of instrumentation.
Porosity	Porosity	Decimal 00.00	%	
P-waveVelocity	P-wave velocity	Decimal 00.0	m/sec	Usually not corrected for P/T effects.
Bottom roughness	Roughness	Decimal 0000.00	Coded to express the height and length of the bottom feature with greatest aspect ratio.	In a coding that expresses the height and length of the bottom feature with greatest aspect ratio; a coded output representing the V:H of the roughness element with greatest aspect ratio, values expressed as (rounded) integer log2.
Critical shear stress	Log critical shear stress	Decimal 0000.00	Log10 of Tau in kPa,	Log 10 of Tau in kPa, being the shear stress required to initiate easily observable erosion and transport, whether by traction or suspension; taken from a compilation of published relationships ranging from large boulder to muds, through a range of grain shapes (eg. shell).
Sample phase	Where in sample the data are from	Character Xxxxx....	Where sample is from	Records whether the results are for the whole, bulk sediment or just to some special part like: inside a nodule, burrow-infill, the sand fraction, porewater (chemistry), a layer in the core, a gradient observed in the core, badly preserved, a layer that is not properly located, or sample with questionable location; also may report a type of analysis or observation if that is special or potentially unrepresentative (e.g., XRD, smear slide). Output is as a word-based description that may involve numerics. This field will need to be selected against when GIS mappings of the bulk sediment characters are

				being mapped. Only blank entries should be included in such a mapping.
--	--	--	--	--

Table C. Most frequently occurring biological components that may have textural implications (U.S. waters only)

barnacles	coralline algae	fish debris_F	pteropods	serpulids
bivalves	corals	forams	radiolaria	shells
brachiopods	crabs	halimeda	razor clams	sponges_F
bryozoa	crustaceans	molluscs	reefs	worm tubes_F
calcareous algae	diatoms	nannofossils	scaphopods	
clypeasts	echinoids			

Table D. Components (features*) processed within usSEABED

Only the descriptive terms found in source reports are defined in the dbSEABED thesaurus. Conversely, as usSEABED uses the same thesaurus as its sister data compilations (auSEABED, goSEABED), some terms listed below may not occur within U.S. waters. Only one of possible variations is listed below, e.g., laminated (laminae, lamination); mollusc (mollusk, mollusca), etc.

Major synonym	Triggering words (<i>word variations not included</i>)
andest	andesite, augite andesite, benmoreite, trachyandesite
anmne_F	anemone, tube anemone, cerinth, cerianthid, coryanactid
annld_F	annelid, arenicola, beachworm, bloodworm, bristleworm, funnelworm, nereid worm, polychaete, polynoid
aren_frm	arenaceous foraminifera, agglutinated foraminifera, ammobaculite, ammodiscus, textularid foraminifera
artif_F	artificial, soot, anchor, brass, cinder, coal, contaminated, lumber, obstruction, petroleum, oil-gas, rubber band, snag, tar, wood chip, wreck
asterd_F	asteroid, basket star, briseaster, sea star, starfish
barit	barite (-concretion -vein)
baslt	basalt, diorite, metabasalt, scoria, trap rock, trachybasalt
bioturb	bioturbation
bitumn	bitumin
biv	bivalve, arctica, astarte, cardium, chama, chione, chlmys, clam (-shell -flat material -hash -valves), cockle (-anadara -shell), donax, glycymeris, katalysia, lamellibranch, macoma, mercenaria, mulinia, mussel (-bed -bank -shell), mya, mytilus, nucula, pelecypod, quahog, rangia, seep mytilid, slipper shells, surf clam, tellina, tellinid, venerid, venus clams, vesicomysid, yoldia
bluschst	blue schist, crossite-albite schist, crossite-quartz schist, glaucophane,

	quartz crossite schist, quartz glaucophane schist
bnth_frm	benthic foraminifera, archaias, bolivina, bulimina, coralline forams, discorbis, eponides, homotrema, hyaline, lenticulina, loxostema, miliolid, nodosirid, nonien, notosirid, peneroplis, porcellanous, rotaiid, uvigerina
borng_F	boring, bioeroded
brach	brachiopod, lingula
brncl	barnacle
bryz	bryozoa, polyzoa
burw_F	burrow, chondrite, clam siphon, crab hole, lebensspuren, Thalassinoides
c_alg	calcareous algae, purple algae, red algae
calc_ooz	calcareous ooze, nannofossil -mud -ooze, pteropod -mud -ooze, foraminiferal -marl -ooze -mud, globigerina -mud -ooze
calcrst	calcareous crust, tufa
calct	calcite (-cement -core -filling -veinlets)
carb	allogenic grain, authigenic carbonate, biogenic, calcareous, calcilutite, calcarenite, calcirudite, calcareous biogenic, carbonate, limey, marl, skeletal micrite
chrcoal_F	charcoal, fire debris
chrt	chert, flint, porcellanite
claymin	clay mineral, bentonite, chlorite, collophane, illite, kaolinite
clypeast	clypeasteriod, sand dollar
coal	coal, lignite, bituminous
coralgl	algal coral, coralgal
crab	crab, hermit crab, sand crab, spider crab, swimming crab
crinod_F	crinoid, basket star
crl	coral, Acropora palmata, brain coral, Dendrophyllia, Madrepore, Manicina, Porite, sea twig
crl_dbr	coral debris
crlrf	coral reef, coral heads, shingle bank, reefal shoal
crnalg	coralline algae, calcareous algae, lithothamnion
crustac	crustacea, decapods, lobster, shrimp shell
defrm_F	deformation, convolute, flame structure, flow structure, load -cast -structure, pull apart
diat	diatom, diatomite/diatomaceous
dolmt	dolomite, ankerite, molar magnesium carbonate
echnd	echinoid, heart urchin, keyhole urchin, sea urchin, spiny urchin
echndrm_F	echinoderm

fault_F	fault
fces	feces, coprolite
ferug	ferruginous, iron fragment, iron (-cement -streak -flake -stain), iron stone, laterite, limonite
flasr_bed_F	flaser bed
fld	feldspar, albite, andesine, anorthite, K-feldspar, labradorite, orthoclase, plagioclase
frm	calcareous foraminifera, foraminifera, globigerina bit, planktonic
gas	foamy, gas
gbbro	gabbro, diabase, diorite, dolerite, meta-dolerite, monzodiorite, monzonite, quartz diorite
glacl	glacial, diamicton, erratic, moraine, till
glauc	glauconite, greensand
gniss	gneiss, diorite gneiss, granite gneiss
gradd_F	coarsening upward, fining upward, increasing grain size, normally graded, reverse graded
granit	granite, aplite, granodiorite, pegmatite
grnschst	greenschist
gstrpd	gastropod, cerithium, conch, turritella, snail, nassarius, olivella, tenebrae, turritella
h2s	hydrogen sulfide, hydrogen sulfide -odor -smell, sulfur odor
halmda	halimeda, Peyssonnelia
holoth	holothurian, sea cucumber
hvy_min	heavy mineral, anatase, andalusite, apatite, black sand, brookite, cassiterite, clinozoisite, corundum, dumortierite, epidote, garnet, ilmenite, jadeite, kyanite, leucoxene, magnetite, monazite, ore mineral, piedmontite, rutile, sillimanite, sphene, spinel, staurolite, titanomagnetite, titanite, tourmaline, topaz, zircon, zoisite
hydrt	hydrate, gas hydrate
ign_rck	igneous rock, acidic rock, alkali basinite, augite plagioclase porphyry, augite porphyry, basic rock, dacite, felsite, olivine plagioclase clinopyroxene, olivine plagioclase phyric, plagioclase andesite porphyry, plagioclase augite porphyry, plagioclase porphyry, plutonic rock, porphyry, rhyolite, syenite, trachyte
klp_F	kelp, brown algae, ecklonia, M. vertebralis, red brown algae
lamintd_F	laminated
lenticlr_bed	lenticular bed, lenticular mass
limstn	limestone, beach rock, bioclastic -floatstone -grainstone -limestone -rudstone, boundstone, bryozoan -floatstone -grainstone -rudstone, calcareous -chip -rubble -rock, coral limestone, floatstone, grainstone, packstone, rudstone, wackestone
lmp_F	lump, aggregate, ball, cast, clump, compact clot, intraclast, pellet, pisolitic, peloid

lrg_frm	large foram, foraminiferal gravel, amphistegina, heterostegina, macro foraminifera, marginopora
maf	mafic, actinolite, aegirite, amphibole, augite, (brown- green- basaltic-) hornblende, bronzite, clinopyroxene, ferromagnesian, hypersthene, olivine, orthopyroxene, oxyhornblende, pyroxene, titanaugite, titaniferous, tremolite
met	metamorphic, calcsilicate, granitized, mylonite, porphyroblast, saussurite
methne	methane
metlif	metalliferous
mica	mica, biotite, chlorite, muscovite, sericite, talc
mlsc	mollusc
mn_crst	manganese crust, manganese iron oxide crust, manganese nodule, manganese pavement, manganese phosphate material
mn_nod	manganese nodule, iron manganese nodule, manganese concretion, micronodule
mnoxid	manganese oxide, iron-manganese (-coat -stain -vener), iron-manganese oxide
motl_F	mottle, chickwire mottle
mudImp_F	mud lump, armored mud ball, silty lump, clay (-aggregate -ball -chip -clast -gall -lump -mass -pod), mud (-aggregate -ball -chunk -clump -clot -clast -lump -pebble -pellet, claystone -chip -gall), sandstone fragment, soft pebble, shale (-fragment -concretion)
mudstn	mudstone, calcareous (-mudstone- siltstone), clay (-rock -shale -stone), marlstone, mud (-rock -stone), pelite, shale, siliceous shale, siltstone
nan	nannofossil coccolith, nannofossil (-coccolith -ooze), silicoflagellate
nod_F	nodule, concretion
odr	odor, anoxic, fetid, foul, fishy, organic, sewage, smell
oil	oil, oil glob, oil sheen
ool	oolite, sporbo
ooz	ooze
ophiurd_F	ophiuroid, brittle star, ophiomusium
orgcbn	organic carbon, carbonaceous, organic (-streak -detritus -matter -mud -content -enriched -material -part -paricle -rich -rimmed), sapropel
oyst	oyster, jingle shell
peat	peat, lignite
pelag	pelagic, hemipelagic, planktic, planktonic
phspht	phosphate, phosphorite
pit_F	pit, crater, feeding depression, hole, pockmark, mound, resting trace
planr_bed_F	planar bed
plnk_frm	planktonic foraminifera, globulina, globorotalid, planktic foraminifera

plnt_F	plant, mangrove, root, vegetation, weed, root
ptr	pteropod
pumc	pumice, ash
pyrt	pyrite, marcasite
qtz	quartz, arkosic sand, calcareous quartz sand, milky vein quartz, quartz (-content -fragment -grain -granule -groundmass -mass -rich -vein -veinlet -crystal), quartzose, quartzite (-cobble -gravel -pebble), sandstone (-chunk -fragment), silica
rad	radiolaria
ripl	ripple, bedform
rlct_F	relict
root_struct	root structure, root clast, root trace
rzr_clm	razor clam, ensis, pinna, pteria, pelecypod
schst	schist, (-albite -chlorite -epidote), (-albite -chlorite), (-albite -mica), (-albite -quartz -chlorite), (-biotite -quartz), (-chlorite -albite), (-chlorite -quartz -albite), (-chlorite), (-phyllitic), (-quartz -albite), (-quartz -albite -chlorite), (-quartz -biotite), (-quartz -chlorite)
scllp	scallop, astropecten, pecten, placopecten
scour	scour, crag and tail, erosional, lag deposit
scphpd	scaphopod, dentalium
seagrs_F	seagrass, acetabularia, amphibolis, batophora, eel grass, grass, halophila, heterozostera, phyllospadix, posidonia, sargassum, strap grass, syringodium, manatee grass, thalassia, zostera
sft_alg_F	soft algae, algal (-scum -filament -mush -strand), Bossea, green algae, macrophytic, red algae
sfterl_F	soft coral, alcynacian, ascidian, coelentrates, gorgonian, hydrozoa
shl	shell, shell (-bed -bank -carpet -fraction -content -material), shellfish, valves
shl_dbr	shell debris, shell hash, coquina, shell (-bit -conglomerate -fragments -festoon -grit -lag -mash -material -piece -particle)
shrmf_F	shrimp, amphipod, ampelisca, copepod, isopod, tanid
sidrt	siderite, siderite nodule
sil_ooz	siliceous ooze, siliceous mud
slte	slate, phyllite, quartzite, metamorphic rock
sndstn	sandstone, gritstone, graywacke, labile sandstone, sandstone reef, wacke
sol_crl	solitary coral, cup coral, disc coral, horn coral, lophelia coral
spng_F	sponge, calcareous sponge, glass sponge, hard sponge, hexactinellid sponge, porifera, Thalassodendron sponge
spoil_F	spoil, brick, coke, dumped sediment
srpul	serpulid, serpulid tube, serpulid worm tube
sulf	sulfide, chalcopyrite

trail_F	trail, trace (animal), track
trrg	terrigenous, lithic, inorganic
umafic	ultramafic, amphibolite, anorthosite, dunite, greenstone, harzburgite, lherzolite, norite, orthopyroxenite, periodotite, picrite, pyroxenite, serpentinite, troctolite, wehrlite
vol_rck	volcanic rock, volcanic (-cobble -pebble)
volgls	volcanic glass, obsidian, hyaloclastite, pyroclastic, quenched, vitric, subvitreous
volrck	volcanic rock, welded tuff
volsed	volcanic sediment, ash, tuff, lapilli
wood	wood, bark, twig
wrm_F	worm, chordate, echiurid, fan (-flat -glob -juicy -long -sand -tube worm), maldanid, pogonophora, priapulida, sliverfish, siphunculid, tunicate
wrm_tbe_F	worm tube, agglutinated worm tube, amphipod tube, annelid worm tube, arenicola, chitinous worm tube, diopatra worm tube, polychaete worm tube, pogonophoran tube worm, vestimentiform tube worm
wavy_bed_F	wavy bed

Table E. Facies and their component makeup

Facies values are determined by a combination of components and their mined values from word-based descriptions. Numeric textural, geochemical, and geophysical information held in _PRS data files. Values represent memberships to fuzzy sets, given as percents. A minimum of 30% component presence is required to trip a given facies, and a component may trip more than one facies. See _FAC files for actual data, and table C for component information. Facies notes presence only, not absence.

Field name	Parameter	Data format, units	Information/triggering components
Latitude*	Latitude	Decimal 00.00000	Decimal degrees, 90° to -90° range
Longitude*	Longitude	Decimal 000.00000	Decimal degrees, -180° to 180° range
WaterDepth*	Water depth	Integer 00000	Meters
SampleTop*	Sample top	Decimal 000.00	Meters below seabed surface
SampleBase*	Sample base	Decimal 000.00	Meters below seabed surface
SiteName*	Site name	Character XXX: XXX	Survey or laboratory code for the sampling site
DataSetKey*	Data set number key	Integer 000	Relational key to _SRC file; _SRC file contains links to source metadata
SiteKey*	Site number key	Integer 0000000	Relational key to other data files. Each site counted sequentially as total output; core data may have more than one sample per site.
SampleKey*	Sample number key	Integer 0000000	Relational key to other data files. Each site counted sequentially as total output; Multiple samples may be at

			each site (i.e., in core).
Terrigenous	Terrigenous	(%)	Fld, hvy_min, maf, mica, qtz, trrg
Carbonate	Carbonate	(%)	Calcrst, calct, carb, dolmt, limstn, sidrt
Igneous	Igneous	(%)	Andest, baslt, gbbro, granit, ign_rck, umafic
Volcanic	Volcanic	(%)	Baslt, pumc, vol_rck, volgls, volrck, vlosed
Metamorphic	Metamorphic	(%)	Bluschst, gniss, grnschst, met, schst, slte
Mineralized	Mineralized	(%)	Barit, metlif, phspht, pyrt, sulf
AuthFeMn	Authigenic Fe Mn	(%)	Mn_crust, mn_nod, mnoxd, ferug
Ooze	Ooze	(%)	Ooz, calc_ooz, sil_ooz
Carbon	Carbon	(%)	Coal, bitumn, orgcbn, peat
GeochemSignal	Geochemical signal	(%)	Gas, h2s, hydrt, methne, odr, oil
Forams	Forams	(%)	Aren_frm, bnth_frm, frm, lrg_frm, plnk_frm
OtherCalcPelag	Other calcareous pelagics	(%)	Nan, ptr, calc_ooz
SilcPelag	Siliceous pelagics	(%)	Diat, rad, sil_ooz
Shell	Shell	(%)	Shl, shl_dbr
Coral	Coral	(%)	Cr1, cr1_dbr, c1rf, sol_cr1
HardPlant	Hard plants	(%)	C_alg, coralgl, crnalg, halmda