Basic Information

This section contains basic information about the dataset, suitable for a minimal metadata entry.

Title: Deep substrate model (100m) of the Pacific Canadian shelf

Dataset ID: substrate100m-data

Status: Completed

Quality Control: Completed

Summary: This deep water substrate bottom type model was created to aid in habitat modeling, and to complement the nearshore bottom patches. It was created from a combination of bathymetrically-derived layers in addition to bottom type observations. Using random forest classification, the relationship between observed substrates and bathymetric derivatives was estimated across the entire area of interest.

The raster is categorized into: 1 - Rock 2 - Mixed 3 - Sand 4 - Mud

Maintainer Email: Joanne.Lessard@dfo-mpo.gc.ca

Cite this data as: Dr. Dana Haggarty, Edward Gregr, Joanne Lessard, Cole Fields and Sarah Davies. 2018. Deep substrate (100 m) for the Pacific Canadian shelf. Published Sept 16, 2018. Data Distributor: J. Lessard, Fisheries and Oceans Canada, Nanaimo, BC.

Start Date: 1984-01-01

End Date: 2018-03-30

Contact Information

This section contains contact information for the data creator and program manager.

Data Creator:

Name: Dr. Dana Haggarty
Email: Dana.Haggarty@dfo-mpo.gc.ca
Position: Inshore Rockfish and Lingcod Program Head
Organization: Government of Canada; Fisheries and Oceans Canada; STAR
Address: 3190 Hammond Bay Rd, Nanaimo, British Columbia, V9T 6N7, Canada
Phone: 250-756-7386

Co-Creators: Cole Fields, Edward Gregr, Joanne Lessard, Sarah Davies, Jessica Finney, Jessica Nephin

Program Manager:

Name: Joanne Lessard
Email: Joanne.Lessard@dfo-mpo.gc.ca
Position: Program Head
Organization: Government of Canada; Fisheries and Oceans Canada; MSEA Section
Address: 3190 Hammond Bay Road, Nanaimo, British Columbia, V9T 6N7, Canada
Phone: 250-729-8364

General

General metadata compatible with the Canada Open Data metadata standard.
Topic Category: Oceans

Date Completed: 2018-03-30

Date Published: 2018-09-17

Update Frequency: Irregular

Dataset Level: Dataset

Keywords (GoC Thesaurus): ocean floor, seabed, sea bed

Science

This section contains metadata specific to the Science branch at DFO.

Science Keywords: rugosity, sediment, substrate

Theme: Base Mapping

Methods: Substrate observations were assembled from the following sources: Canadian Hydrographic Service (CHS), Dive survey data, CHS marsh data, Natural Resources Canada (NRCan), and Remotely Operated Vehicle (ROV) surveys. These observations were mapped to common bottom type classes: rock, mixed, sand, or mud. Using ArcGIS, the points were classified into training (2/3) or testing (1/3) data. The training subset is fully withheld from the model-building process and is used to evaluate the model's performance. The random partitioning of the data into training and test subsets does not address the issue of spatial autocorrelation between observations.

The code to create the model follows the following steps:
A raster stack of environmental predictors is created. Point observations are overlayed onto the raster stack, and the values from the predictors are extracted to the points. Records with invalid BType4 values or with NA values from the predictors are removed. The training data are weighted according to their prevalence and are used to fit a random forest model using the ranger package (which supports case weighting). The fitted model is used to predict to the extent of the input environmental raster stack, classifying the entire area into rock, mixed, sand, or mud. The predicted surface is exported as a GeoTIFF raster file in the same resolution as the raster stack predictors (in this case, 100 metre resolution) and with the same projection. Evaluation statistics including Kappa, and accuracy by predicted class are generated using the withheld test data.

Manual steps after the model has been generated:
1) Reproject layer to EPSG: 3005 (R does not support writing the top-level EPSG code to the coordinate reference system information).
2) Create a raster attribute table with SUBSTRATE field for text classes.
3) Add a field for prevalence in the raster attribute table.

Predictor Layers: Using GDAL, the source bathymetry layers (NOAA and BCMCA) were resampled and mosaicked to produce a 100 m bathymetry raster layer. Where bathymetry rasters overlapped in space (the majority of the region except for Dixon Entrance, Chatham Sound and Portland Canal areas) the NOAA bathymetry layer overwrote the BCMCA layer. The derivatives of bathymetry were created using the Arcpy module in python (see Scripts section in the metadata). The bathymetry layer was smoothed using a focal mean prior to generating the following derivatives: slope -> standard deviation of slope, and curvature. This step was required as a method to reduce artefacts found in the derivatives. During development, these artefacts from the non-smoothed bathymetric derivatives carried through to the predicted substrate models. The non-smoothed bathymetry was used to generate the standardized BPI layers because this processing already involves processing with a neighbourhood of cells. A categorical layer for rugosity was used from the B.C. Marine Conservation Atlas (BCMCA). It was converted to a 100 m raster grid to align with the other predictor layers. It was decided to use this layer because it had been derived from bathymetry and then manually edited. See the methods described in the data sources link. As input to the random forest model, the original non-smoothed bathymetry was used. Ocean energy layers were also included – mean bottom ocean currents (Masson and Fine 2012), and mean tidal speed on the bottom (Masson and Fine 2012). The original data for the Regional Ocean Modeling System (ROMS) has a 3 by 3 km grid resolution. These data were interpolated using Spline with Barriers (ESRI) and resampled to 100 m resolution rasters. See the scripts section for a link to the ROMS data processing. The bathymetry was primarily sourced from NOAA, with a small portion mosaicked to the northwest from SciTech bathymetry (see Data Sources for links). The bathymetric position index (BPI) layers were created using the Benthic Terrain Modeler toolbox and were standardized after being calculated. It is important to note that when calculating BPI, the tool expects bathymetric data to have negative values associated with depths rather than positive.

Predictor Layers:
1: Bathymetry
2: Slope (bathymetric derivative) – degrees
3: Standard Deviation of Slope (bathymetric derivative)
4: Broad Bathymetric Position Index (bathymetric derivative) - Inner Radius: 25 - Outer Radius: 250
5: Medium Bathymetric Position Index (bathymetric derivative) - Inner Radius: 10 - Outer Radius: 100
6: Fine Bathymetric Position Index (bathymetric derivative) - Inner Radius: 3 - Outer Radius: 25
7: Curvature (bathymetric derivative; slope of slope)
8: Rugosity (BCMCA)
9: Circulation
10: Tidal

Data Sources:

Source: Bathymetry:
https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ngdc.mgg.dem:4956 (3 arc-second resolution)
https://bcmca.ca/datafeatures/eco_physical_bathymetry/ (100 m resolution)

Source: Coastline:
https://www.gis-hub.ca/dataset/chs_hwl

Source: Rugosity:
https://bcmca.ca/datafeatures/eco_physical_highrugosity/

Source: Oceanographic:
https://www.gis-hub.ca/dataset/bc-roms-bottomwaters (3 km resolution)

Scripts or Software Routines:

Data preparation/creation: https://gitlab.com/dfo-msea/environmental-layers/bathy-derivatives
Random Forest model using ranger package: https://gitlab.com/dfo-msea/environmental-layers/random-forest-substrate/-/releases/v1.1
ROMS processing: https://gitlab.com/dfo-msea/environmental-layers/bc-roms-climatologies
Spline with Barriers Interpolation https://gitlab.com/dfo-msea/environmental-layers/spline-barriers-resample

Aligning layers: https://gitlab.com/dfo-msea/environmental-layers/SnapExtractMask

Spatial Data Quality: Bottom substrate sample points are considerably denser near-shore as opposed to deeper areas, due to data collection priorities and ease of collection. As such, the model may be biased towards shallower observations. Additionally, the CHS samples used are likely to be biased towards 'hard' substrate, since the safe navigation mandate of the CHS requires a focus on rocks and reefs. The prevalence of hard samples in the analysis likely introduces some bias, but the effect on the model results is unknown.

Positional Accuracy: The layer has a nominal horizontal accuracy of 100 m

Attribute Accuracy: The data was produced from a robust set of bottom type sampling programs. Through the validation component (natural forest modeling), we can be fairly confident in the average accuracy of the attributes. Local accuracy will be highest in nearshore regions where bottom type sampling was highest, and more uncertain in areas of low sampling density. Accuracy is compromised more generally by the inherent, often high, variability of the ocean bottom which decreases with depth.

Logical Consistency: The predictions are based on the relationship between geophysical characteristics and observations of bottom type from a variety of sources. The methods were consistently applied across the study area.

Completeness: These data are complete for the study area

Absence Data: None. The predicted coverage is comprehensive.

Uncertainties: As a model output, the agreement between the predicted and observed substrate will vary across the study area in an unpredictable manner. Ad hoc experiments done during model development suggest the model can be quite sensitive to the input data. Areas with high spatial variability (i.e., 'speckling') may have higher uncertainty than more consistent areas.

Use Restrictions: Data is restricted to DFO Science for spatial analysis only

Change History:

Date of Change	Description of Change
2018-09-03	Initial creation of dataset record
2019-01-25	Update with new data Previous version of the 100m substrate layer used the SciTech bathymetry and several derivatives as predictors for the random forest model. The current version uses a resampled version of the 80m bathymetry created by NOAA and available here: https://data.noaa.gov/metaview/page?xml=NOAA/NESDIS/NGDC/MGG/DEM/iso/xml/4956.xml&view=getDataView&header=none Bathymetry and its derivatives used as predictor layers now use the NOAA data.
2019-03-01	Update with new observation data found here: \\dcbcpbsna01a\Spatial_Datasets\Substrate\KSwan_TemporaryFolder\All_BoP_Grabs_Obs_2019.gdb\Obs_HG These are dive survey points around Haida Gwaii that were added to the input points for generating the model.

Date of Change	Description of Change
2019-11-19	Modify spatial reference to EPSG: 3005. The existing spatial reference had the same parameters, but wasn't registered as the BC Albers projection (3005).
2020-01-31	1. Completed Metadata 2. Added summary statistics file 3. Added obs shapefile
2020-04-02	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-07-21	updated attribute fields
2020-09-16	updated attribute fields
2020-09-16	updated attribute fields
2020-09-16	updated attribute fields
2020-09-16	updated attribute fields
2020-09-16	updated attribute fields
2020-09-16	updated attribute fields
2020-09-16	test of update res order 2
2020-09-23	updated attribute fields
2020-09-23	updated attribute fields

Date of Change	Description of Change
2020-09-23	updated attribute fields
2020-09-23	updated attribute fields
2020-09-23	updated attribute fields
2020-09-23	updated attribute fields
2020-09-23	updated attribute fields
2020-10-15	removed obs shapefile of source point data
2020-10-15	included gitlab code release https://gitlab.com/dfo-msea/environmental-layers/random-forest-substrate/-/releases/v1.0
2020-12-21	DEEP Sub Data: RANGER PACKAGE UPDATE: https://github.com/ejgregr/substrate_model
2021-04-27	DEEP Sub Data: Updated name of GeoTIFF to include Ranger in name.

Species Data:

Code and Name	Age Data	Obs Type

References

Reference: Du Preez, C. 2015. A new arc–chord ratio (ACR) rugosity index for quantifying three-dimensional landscape structural complexity. Landscape Ecology 30:181-192.

Reference: Masson, D., and I. Fine (2012), Modeling seasonal to interannual ocean variability of coastal British Columbia, J. Geophys. Res., 117, C10019, doi:10.1029/2012JC008151

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

Reference: https://cran.r-project.org/web/packages/ranger/ranger.pdf

Collaboration: SciTech Consulting

Other Information: The data were validated as part of a 2018 metadata effort by the MSEA group. The following aspects of the data were reviewed: 1) Spatial reference (EPSG:3005), 2) visual comparison to validate locations are mapped correctly, 3) Evaluated coastlines to determine shoreline overlap, 4) Checking for data gaps, 5) evaluating for null values. Origins of the regional data were also validated to ensure no overlap. This data has not yet been released. Confidentiality and permissions will be decided upon prior to release.

Confidentiality: Not Protected