Basic Information

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

Title: Shallow substrate model (20m) of the Pacific Canadian coast

Dataset ID: substrate20m

Status: Ongoing

Quality Control: Completed

Summary: The shallow substrate bottom type model was created to support near shore habitat modelling. Data sources include both available observations of bottom type and environmental predictor layers including oceanographic layers, fetch, and bathymetry and its derivatives. Using weighted random forest classification from the ranger R package, the relationship between observed bottom type and predictor layers can be determined, allowing bottom type to be classified across the study areas. The predicted raster files are classified as follows: 1: Rock 2: Mixed 3: Sand 4: Mud

The categorical substrate model domains are restricted to the extent of the input bathymetry layers (see data sources) which is 5 km from the 50 m depth contour.

Maintainer Email: Dana.Haggarty@dfo-mpo.gc.ca

Cite this data as: Dr. Dana Haggarty and Edward Gregr, Joanne Lessard, Cole Fields, Sarah Davies, 2020. Shallow substrate model (20m) of the Pacific Canadian coast. Published Jan 31 2020. Data distributor: Dr. Dana Haggarty, Marine Spatial Ecology Section, Fisheries and Oceans Canada, Nanaimo, BC.

Start Date: 1984-01-01

End Date: 2019-12-31

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

Program Manager:

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: 200 - 401 Burrard Street, Vancouver, British Columbia, V6C 3S4, Canada
Phone: 250-756-7386

General

General metadata compatible with the Canada Open Data metadata standard.

Topic Category: Oceans

Date Completed: 2015-01-01

Date Published: 2020-01-31

Update Frequency: Irregular

Dataset Level: Dataset

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

Science

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

Science Keywords: rugosity, sediment, substrate

Theme: Classification

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. These points were subdivided into 5 regions (North Central Coast, Haida Gwaii, Strait of Georgia, Queen Charlotte Strait, and West Coast Vancouver Island). 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 each 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.

The bathymetry and its derivatives 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, curvature, and rugosity. 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. As input to the random forest model, the original non-smoothed bathymetry was used. Ocean energy layers were also included – mean bottom ocean currents and mean tidal speed on the bottom. The SOG region uses the Salish Sea NEMO model (Allen) as source data for the ocean energy layers. The other 4 regions are sourced the the BC ROMS model (Masson). 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 20 m resolution rasters. See the scripts section for a link to the ROMS data processing. The source data for the NEMO model had higher resolution in the Strait of Georgia and so it was decided that this model would be used for the SOG region. They were first interpolated to a 40 m cell resolution (because of computational limitations) and then resampled to 20 m. The NEMO Tidal and Circulation layers were smoothed using Focal Statistics in ArcGIS with a 13 cell neighbourhood. Fetch was also included as a predictor. Fetch points were interpolated to the extent of the raster stack (specific methods and exact source data are unclear). DFO 20 m bathymetry layers (that included terrestrial elevation data) were used as the bathymetry source. 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 (bathymetric derivative; Arc-Chord rugosity)
9: Circulation
10: Tidal
11: Fetch

Data Sources

Source: https://www.gis-hub.ca/dataset/coastline-fetch-bc
Source: https://www.gis-hub.ca/dataset/bathy_20m
Source: https://www.gis-hub.ca/dataset/substrate-preds-qcs-20m
Source: https://www.gis-hub.ca/dataset/substrate-preds-hg-20m
Source: https://www.gis-hub.ca/dataset/substrate-preds-sog-20m
Source: https://www.gis-hub.ca/dataset/substrate-preds-ncc-20m
Source: https://www.gis-hub.ca/dataset/substrate-preds-wcvi-20m

Scripts or Software Routines:

Data layers preparation/creation: https://gitlab.com/dfo-msea/environmental-layers/bathy-derivatives

Random Forest model: https://gitlab.com/dfo-msea/environmental-layers/random-forest-substrate/-/releases/v1.1
ROMS data processing: https://gitlab.com/dfo-msea/environmental-layers/bc-roms-climatologies
NEMO data processing: https://gitlab.com/dfo-msea/environmental-layers/salishsea-nemo

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 20 m

Attribute Accuracy: The data was produced from a robust set of bottom type sampling programs. Through the validation component (natural forest modelling), we can be fairly confident in the accuracy of the attributes. Local accuracy will be best 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 areas described.

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.

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

Change History:

Date of Change	Description of Change
2020-12-21	Substrate - 20 m data package: R PACKAGE: Changed to ranger package and code repo (https://github.com/ejgregr/substrate_model). Ranger supports case weighting.

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: Soontiens, N., Allen, S., Latornell, D., Le Souef, K., Machuca, I., Paquin, J.-P., Lu, Y., Thompson, K., Korabel, V., 2016. Storm surges in the Strait of Georgia simulated with a regional model. Atmosphere-Ocean 54 1-21. https://dx.doi.org/10.1080/07055900.2015.1108899

Collaboration: SciTech Consulting

Confidentiality: Not Protected