Bottom Patches for the Canadian Pacific nearshore: Project and methods overview

Edward Gregr, Kayleigh Gillespie and Joanne Lessard

Fisheries and Oceans Canada
Science Branch, Pacific Region
Pacific Biological Station
3190 Hammond Bay Road
Nanaimo, British Columbia
V9T 6N7

2022

Canadian Technical Report of Fisheries and Aquatic Sciences 3472

Fisheries and Oceans Canada Pêches et Océans Canada

Canada

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques.*

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 456 de cette série ont été publiés à titre de Rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de Rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de Rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Canadian Technical Report of Fisheries and Aquatic Sciences 3472

2022

BOTTOM PATCHES FOR THE CANADIAN PACIFIC NEARSHORE: PROJECT AND METHODS OVERVIEW

by

Edward Gregr[2], Kayleigh Gillespie[1], Joanne Lessard[1]

1. Science Branch
Pacific Region
Ecosystem Science Division
Fisheries and Oceans Canada
Pacific Biological Station
3190 Hammond Bay Road
Nanaimo, British Columbia
V9T 6N7

2. SciTech Environmental Consulting
2136 Napier Street
Vancouver, British Columbia
V5L 2N9

Cat. No. Fs97-6/3472E-PDF ISBN 978-0-660-43118-5 ISSN 1488-5379

Correct citation for this publication:

Gregr, E.J., Gillespie, K. and Lessard, J. 2022. Bottom Patches for the Canadian Pacific nearshore: Project and methods overview. Can. Tech. Rep. Fish. Aquat. Sci. 3472: vii + 36 p.

1 TABLE OF CONTENTS

2 LIST OF TABLES

3 LIST OF FIGURES

4 ABSTRACT

Gregr, E.J., Gillespie, K. and Lessard, J. 2022. Bottom Patches for the Canadian Pacific nearshore: Project and methods overview. Can. Tech. Rep. Fish. Aquat. Sci. 3472: vii + 36 p.

Marine coastal regions are highly productive ecosystems and provide important habitat for many valued species. Here we describe a continuous, object-based (i.e., polygon) prediction of bottom type, a key indicator of habitat in this ecosystem. The model is built with the best available data using an approach that is simple, quantitative, and transparent. The approach is amenable to iterative improvement as data quality and availability improve. To support the use of bottom type in applications such as habitat models we also developed a corresponding confidence layer based on the agreement with, and distance between, available substrate observations. Bottom patches predict areas of similar substrate by extrapolating field observations and samples using assembly rules based primarily on depth and data density. Bottom patches are created semi-automatically from available data sources in three steps: mapping substrate codes from different sources to a common substrate attribute table; validating spatial locations, and extrapolation using Thiessen polygons. We used Python scripts to ensure consistency across different regions, speed up and standardize data transformations, and identify data errors and exceptions. Where no observations or samples were available, we used bottom type predictions from a random forest model. We created bottom patches for the entire Pacific Canadian coast to a depth of 50 m. The patches span 35,000 km of coastline and include 864,531 bottom patches ranging in size from 4 m^2 to just under 30 km^2.

5 RÉSUMÉ

Gregr, E.J., Gillespie, K. and Lessard, J. 2022. Bottom Patches for the Canadian Pacific nearshore: Project and methods overview. Can. Tech. Rep. Fish. Aquat. Sci. 3472: vii + 36 p.

Les régions côtières des océans sont des écosystèmes très productifs et fournissent un habitat important pour plusieurs espèces. Nous avons créé une carte de substrat compréhensive, basée sur des polygones benthiques en utilisant les meilleures données disponibles et une approche simple, quantitative et transparente. L'approche se prête à une amélioration itérative à mesure que la qualité et la disponibilité des données s'améliorent. Nous avons également développé une couche de confidence correspondante basée sur l'accord entre les observations et leur distance pour soutenir les applications écologiques telles que les modèles d'habitat.

Les polygones benthiques sont créés semi-automatiquement à partir d'une variété de sources de données en trois étapes comprenant : convertir les codes de substrat de différentes sources à une table d'attributs de substrat commune, validation des emplacements spatiaux et extrapolation à l'aide des polygones Thiessen. L'utilisation de scripts Python garantit la cohérence de l'application dans différentes régions, accélère et normalise les différents processus de transformation des données et aide à identifier les erreurs et les exceptions de données. Dans les zones où aucune observation ou échantillon n'était disponible, nous avons utilisé des valeurs de type de substrat à partir d'un modèle prédictif. Nous avons créé des polygones benthiques pour l'ensemble de la côte pacifique canadienne depuis la ligne des eaux à marée haute jusqu'à une profondeur de 50 m. Les polygones couvrent l'ensemble des 35 000 km de côtes et comprennent 864 531 polygones benthiques dont la taille varie de 4 m^2 à un peu moins de 30 km^2.

6 INTRODUCTION

Maps representing shallow nearshore habitats have been needed for many years for a variety of marine spatial initiatives and there has been considerable effort in recent years to create spatial layers to inform those initiatives. One of these efforts was the development of the bottom patches (BoPs). The BoPs are polygons describing areas of similar bottom type, delineated by depth and the best available bottom type information, to which other known physical attributes can be attached.

This report describes the evolution of the BoP approach from the prototype methodology described in Gregr et al. (2013) including the automation of the process and its application to the Pacific Canadian coast. This research and development project led to the creation of the 20 m bathymetries (Davies et al., 2019), the first coastwide description of nearshore marine bottom type, and the development of a random forest predictive model of substrate (Gregr et al., 2021). We developed an object-based approach for two reasons. First, because observed heterogeneity is high and model accuracy cannot be guaranteed locally using predictive models. We argue a more accurate map can be achieved by anchoring the model predictions with substrate 'patches' defined using local substrate data and a specified degree of spatial autocorrelation. Second, using patches improves flexibility and processing efficiency over raster-based approaches because polygons can hold a large number of diverse attributes, from multiple spatial scales. They can also maintain links to their data source allowing for local estimates of confidence and can be aggregated or disaggregated based on different needs.

The patches are organized into four regions: Haida Gwaii (HG), North Central Coast (NCC), Queen Charlotte Sound and Strait of Georgia (QCSSOG), and West Coast Vancouver Island (WCVI). We combined QCS and SOG for ease of processing.

6.1 History of BoP development

The work was done over several years (Table 1) and other spatial layers were developed in conjunction with this work (e.g., 20 m bathymetries, fetch, and predicted substrate model). The BoPs were conceptualized by the Nearshore Habitat Working Group, a collection of scientists (including biologists, ecologists, geologists, hydrographers, and oceanographers) from DFO, NRCan, CHS, academia, and the private sector working in the coastal marine environment (Gregr et al. 2013). There was a consensus that groups interested in nearshore/benthic habitat mapping should collaborate on standards for data collection and representation. A key objective for the group was to create a spatial layer that could be used to support province-wide habitat analyses of coastal species. A prototype of the BoPs (Gregr et al. 2013) was funded and developed through the Strait of Georgia Ecosystem Research Initiative.

The work described in this report extended this prototype to the entire coast under DFO's World Class Tanker Safety System program (circa 2013). Following Gregr et al. (2013) we applied the BoP methods regionally to what have since evolved into the five bathymetric regions. The process took four years (2014 – 2017), and required navigating a rapidly evolving data landscape and the processing of large volumes of data not previously used for this purpose. We digitized some CHS field sheets in early years, and processed a large amount of CHS data in 2015, for which little metadata were available. We also produced models of fetch (a proxy for wind-wave exposure) for the entire coastline to support the predictive substrate models, and built the prototypes for the 20 m bathymetries now widely used within DFO.

Table 1: History of Bottom Patch (BoP) development.

When	What	Details
2010-2013	SOG BoPs version 1	Gregr et al. (2013). 3-year effort as part of DFO's Ecosystem Research Initiative. The lack of data in the nearshore was identified by the Nearshore Habitat Working Group which was active during this time. This proof of concept was built manually.
2012-2015	WCVI BoPs	Developed in collaboration with Gregr (2016), a significant part of this effort was devoted to the development of the first 20 m bathymetry in Pacific Canada. These patches were completed May 2015. The shellfish spatialization script was developed in early 2015 to capture data collected on transects defined with both a start and end point, and those with only a starting point.
2015	PRCC BoPs	The prototype for this region (Prince Rupert to Cape Caution (PRCC), later renamed the NCC) was completed July 2015. It was the first version to be fully generated using Python scripts.
2016	HG BoPs version 1	Completed September 2016, this version was built with a prototype 20 m bathymetry built in collaboration with Parks Canada (see Davies et al., 2019 'for details). First use of the random forest substrate model.
2017	NCC BoPs version 2	Completed January 2017 with updated shellfish data.
2017	QCSSOG BoPs	Completed January 2017 with updated shellfish data.
2017	HG BoPS version 2	Completed June 2017 with updated shellfish data.

Most regions were re-processed with updated Canadian Hydrographic Service (CHS) substrate data provided in 2015 as part of the Oil Spill Response Program, and again in early 2017 with updated shellfish observations from Fisheries and Oceans Canada (DFO). We used Python scripts (with ArcGIS) to prepare the data and generate the BoPs (Gregr and Peterman, 2022). This report synthesizes regional working documents into a single document describing the BoPs and how they were produced.

7 METHODS

To maximize the substrate data available we designed our approach to accommodate the diverse substrate data collected for various purposes across the Pacific Canadian coast. To be included in the BoPs, a data set requires only the appropriate look-up table to match the standardized BoP Bottom Type (BType) codes (Table 2) and a spatial reference for each datum. For example, data from Natural Resources Canada (NRCan) contributed to the WCVI region where a unique, shallow water survey had been conducted. The stages of the BoP process (Figure 1) include data standardization, spatialization of transect data, translation of source-specific substrate descriptions to the BType codes, the creation of Thiessen polygons for all points in each data source, and the intersection, attribution, and assembly of the resulting fragments into BoPs. The production of the BoPs also required the creation of depth ribbons, and a background substrate model to assign substrate to areas with no data. Details on each of these steps are below.

The BoP methods were adapted over the course of the project as existing data sets were updated, additional data sets were added, and results were refined. The most significant updates since the 2013 prototype were to the methods for creating the depth ribbons and the predicted substrate model (details below). Other updates included the spatialization of transect data and the development of error correction and automation routines. We maintained consistency in the processing of the BoPs by using a series of Python scripts to support different parts of the process.

We assigned each data set a unique SourceKey to allow each BoP to be traced back to its source data. The data source was also used in the rule base to assign BType and confidence to the BoPs. Other key attributes include BoPID (a numeric key) and DepthCode (text description of the BoP's depth ribbon).

We used the BC Albers projection for this analysis. Earlier prototypes were built with ArcGIS 9.3. BoPs built after 2015 used ArcGIS 10.2. Generally, we removed the Z and M coordinates whenever we transformed a data set - we found these coordinates interfered with some spatial operations.

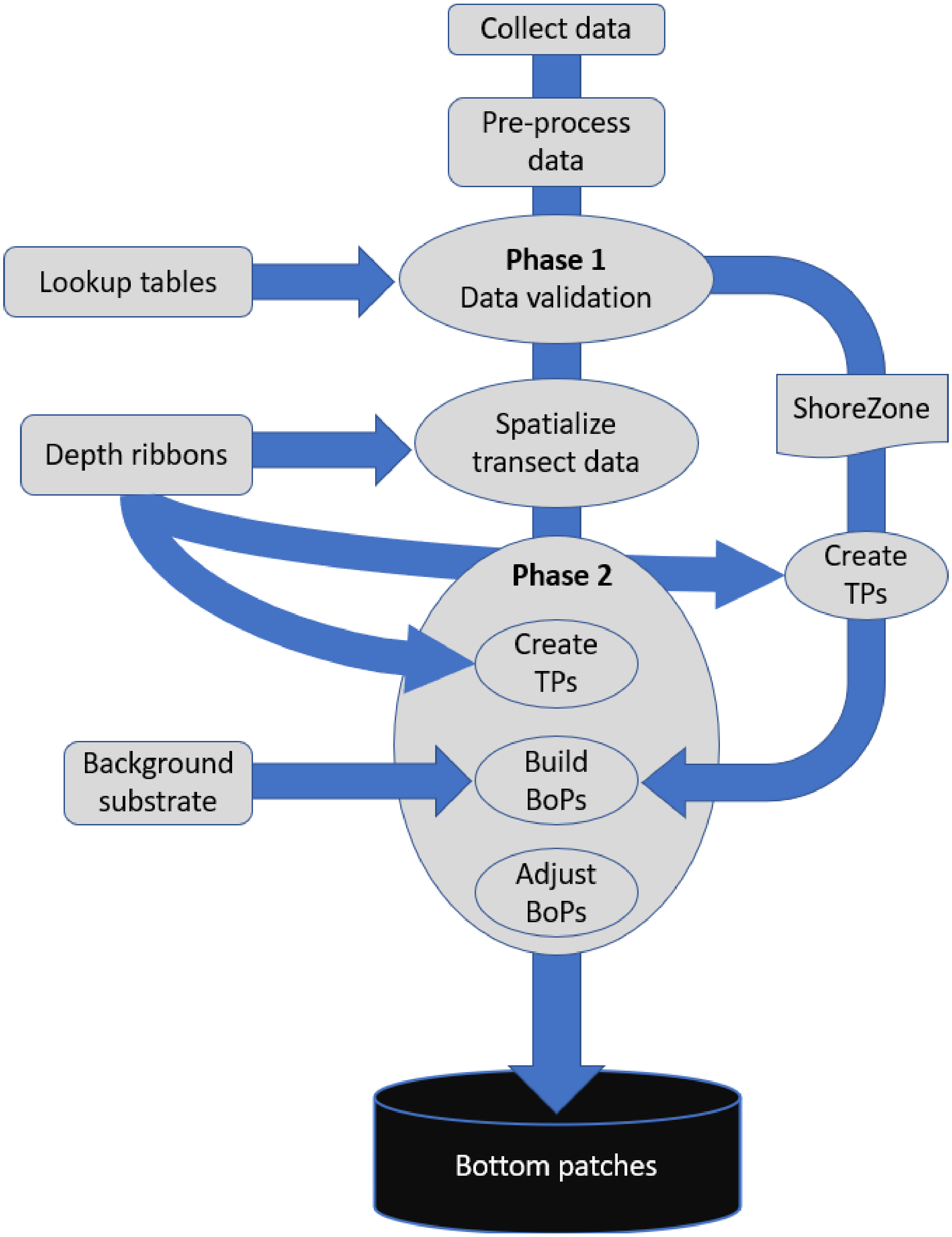


Figure 1: The steps in creating the bottom patches for Canada's Pacific coast. Lookup tables (maintained in an MS Excel configuration file) are used to validate the available data sets after initial pre-processing. Thiessen Polygons (TPs) are created separately for the ShoreZone data and the transect-based data. The results are intersected and re-assembled according to explicit assembly rules to build bottom patches (BoPs).

Table 2: The bottom type (BType) attributes used for the Bottom Patches including description of each class, originally proposed for the Strait of Georgia (Gregr et al. 2013).

Primary and secondary bottom type categories	**Code**	**Bottom type description**
Hard	1	Immobile substrates that support well-developed epibenthic communities, with a low likelihood of infaunal organisms.
Bedrock dominant	1a	Largely (>80%) bedrock, with little relief in terms of boulders or corals. May contain some patches of sand/mud/other.
Boulder dominant	1b	Largely (>80%) dominated by boulders and cobbles; crevices amongst boulders provide habitat complexity; some soft sediment may exist below the boulder-cobble armour layer and support some infauna.
Mixed	2	Mix of hard and soft substrate with a likelihood of both infaunal and epibenthic communities represented
Soft surface, patchy distribution of larger particles	2a	Mix of soft sediments with patchy distribution of larger particles (cobble, boulder) with overall cover <80%. Diverse biota expected with both infaunal and epilitho communities.
Soft surface overlaying hard substrate	2b	Mix of soft sediments distributed over bedrock with patches not to exceed 80% cover. Epibenthic-dominated community expected with potential for some infaunal organisms.
Soft	3	Unconsolidated bottom type with negligible hard components. Very low likelihood of epibenthic organisms.
Sand/shell	3a	Sand or shell dominant (>80%) potentially mixed with larger particles to granules.
Mud	3b	Mud dominant (>80%) potentially mixed with larger particles to granules.

7.1 Depth ribbons

The idea of depth ribbons originated when nearshore bathymetric data were limited to depth polygons describing depth zones on marine charts (i.e., the Murfitt data set – a mosaic of polygons from best resolution digital charts; Joanne Lessard, personal communication). Since depth zones on charts vary with chart resolution, Gregr et al. (2013) combined these zones into a single set of spatially consistent classes - the depth ribbons. Since depth in nearshore ecosystems is a proxy for wave energy and light penetration, the absolute depth at any point has less ecological relevance than the abiotic conditions encountered at the bottom. Thus, the depth ribbon classification emphasises the ecological importance of different nearshore depth zones rather than seeking high resolution depth values. Working with the Nearshore Habitat Working Group we defined five ecologically relevant zones. These depth ribbons included: the intertidal (ITD) and 0-5 m, 5-10 m, 10-20 m, and 20-50 m depth zones (Gregr et al. 2013). The definition of the high water line varied by region. These depth ribbons now underpin the structure of the BoPs and are integral to the spatialization of substrate observations collected on georeferenced transects (i.e., shellfish and herring survey data). They also provide a means of limiting extrapolation to a single, ecologically-based depth range. The depth ribbons are supported by a land (above datum) mask and a deep (50-100 m) depth ribbon to assist with the BoP processing.

Following Gregr et al. (2013), we derived the depth ribbons from prototypes of the 20 m bathymetries built from CHS field sheet data and supplemented with terrestrial elevations from the Canadian Digital Elevation Data raster tiles (Natural Resources Canada, 2013). These regional bathymetries were finalized by Davies et al. (2019).

We created depth ribbons regionally as the high resolution of the bathymetries and the complexity of the Pacific Canadian coast made coastwide processing prohibitive. We built the ribbons manually to account for regional differences in bathymetry and changes in data (see following section) and ArcGIS versions. To facilitate Phase 2 processing (Figure 1) we intersected the ribbons with a 5 km regular grid clipped to the boundaries of each regional study area.

A critical feature of depth ribbons is that they provide continuous, consistent nearshore representation, especially in high slope areas where depths cannot be represented on a 20 x 20 m^2 grid. To capture the depth ribbons in high slope areas we first resampled the 20 m bathymetries to 4 m resolution and then contoured the finer bathymetries to create nearshore depth ribbons effectively capturing these zones even in steep sections where they often occur in between the 20 m bathymetric pixels (Figure 2). Capturing these areas of high slope leads to more contiguous and realistic ribbons and this realism is inherited by the BoPs. We used the simplify option of the Raster-to-Polygon operation for all regions except WCVI where the ribbons were not re-sampled or simplified. It was our experience building ribbons for the earlier WCVI prototype that led to this over-sampling approach used subsequently to create smoother ribbons.

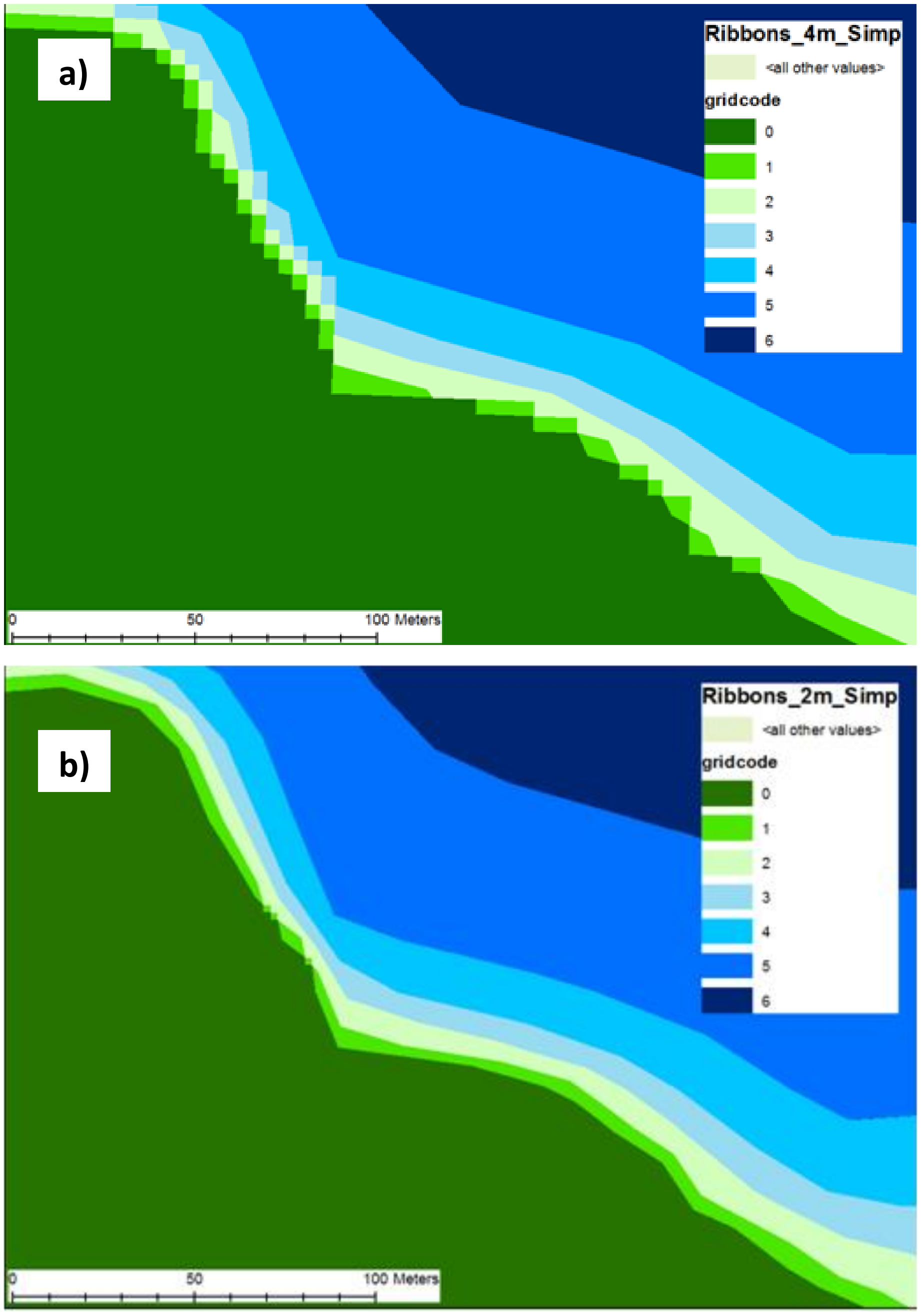


Figure 2: Comparison of simplified depth ribbons derived from a) 4 m and b) 2 m digital elevation models in a test area of high relief. Gridcode values correspond to depth zones (0 = land, 1 = intertidal, 2 = 0-5 m, 3 = 5-10 m, 4 = 10-20 m, 5 = 20-50 m, and 6 = >50m). The ribbons generated at the higher resolution create a more realistic representation of depth zones along this steep coastline. Coarser elevation models (at 20 or 100 m resolution) provide much poorer representations of these steep coastal areas.

7.2 Substrate data pre-processing

We separated the substrate data into Grabs and Observations (Obs) features because the sampling approaches have different biases. The difference is rooted in how and what type of data are typically collected. Grabs are literal point samples - physical samples taken with a grab or core. Grabs can provide detailed information on grain size and sediment composition, but are biased towards soft sediments as a failed grab is typically recorded as simply "Hard". As individual points, they do not provide information on local heterogeneity. Obs are typically collected on transects (e.g., shellfish dive surveys) and do not contact the substrate. They are thus less effective at distinguishing among soft bottom types and can be confounded by veneers (e.g., thin layers of soft substrate over bedrock). However, they do provide a better picture of heterogeneity, and often provide data (e.g., species observations) in addition to substrate.

We developed the substrate classes used in BoPs. Since each source data set may code substrate differently, a common set of classes needed to be defined so that different sources could be merged and used to create the BoPs. The BType is the substrate classification used for the BoPs (Table 2) and was developed based on feedback from the Nearshore Habitat Working Group. Unique processing requirements for each source data set are described in the following sections.

7.2.1 Observations

Obs data come primarily from DFO dive surveys conducted for shellfish and herring stock assessment. These data sets required work to create spatial points suitable for BoP processing. The number of observations by bathymetric region (Table 3) shows the relative abundance of these data. After processing, the herring and shellfish data were merged into an Obs feature class along with any other local data sets. The necessary fields required by the Phase 2 script include fcode, SourceKey, BType1, and BType2.

The ShoreZone (SZ) coastline imagery (Coastal Oceans, 2022) provides an additional source of observations which capture a variety of biophysical attributes (see Lerner and Gregr, 2018). These data are treated separately from the Obs dive data because they were collected for the purpose of documenting and classifying geomorphic and biological features. The mapped products are line segments with an associated database which include a wide range of data types.

Table 3: Summary of coastwide substrate observations showing total records in each original source data and the proportion successfully processed into bottom patches based on data completeness, by region.

Data Sources		NCC	HG	WCVI	QCSSOG	Total
Shellfish	Original	32,700	7,842	7,200	18,911	68,097*
	Processed	32,304	7,620	6,541	18,778	65,243
Herring	Original	3,474	1,140	2,076	2,671	9,361
	Processed	3,069	1,079	1,543	2,573	8,262
ShoreZone	Original	27,510	7,894	12,705	22,838	74,580
	Processed	23,414	7,882	12,022	22,812	66,130

**For Shellfish data, the regional numbers do not sum to the total because the WCVI region was never rerun with the updated 2016 shellfish points and is thus currently missing these additional points (n = 1444).*

7.2.1.1 Shellfish survey data

DFO collects substrate data as part of shellfish dive survey transects which can extend from the intertidal to depths greater than 20 m. We included data from abalone, sea cucumber, sea urchin and geoduck stock assessment surveys from 1992 to 2015. Earlier shellfish surveys had different substrate classifications for the various species surveyed. The shellfish protocol was subsequently standardized across species to improve data comparability. Shellfish data are collected on quadrats spaced along a transect. Quadrats are not geolocated, rather their location is informed by the recorded dive depth and the distance along the transect (where both transect endpoints are recorded). We therefore aggregated the quadrat data according to their tide-corrected sample depth to create a unique shellfish record for each depth ribbon. Each record contains one or two coordinates defining its transect, the depth ribbon it belongs to, a derived BType value, a summary of the quadrat data used to derive the BType, and a link to the source survey.

Earlier prototypes of the BoPs used data extracted from the DFO shellfish database with MSAccess queries. Starting in Spring 2015, DFO provided updated coastwide shellfish data as an ArcGIS shape file. The Spring 2015 data contained 57,486 aggregated (to depth ribbon) samples and an updated attribute table. The coastwide data were updated again in June 2016 with additional survey data providing 68,097 samples. BoPs created after 2015 (NCC, HG, and QCSSOG) used the updated shellfish data. The update included a revised method to summarize the quadrat data (Appendix A2), and the provision of the shellfish records as a stack of points located at the start of the associated transect. The updated shellfish attributes also included a numeric DepthCat field that was offset by 1 from the fcode used in Phase 2 processing. We therefore needed to add an fcode field to the shellfish table. We also filled empty values in the assigned BType2 with '0' to so that the BType lookup coded properly.

Earlier surveys recorded only the transect origin, while more recent transects include both a start and an end point. We developed spatialization methods for both the pre-processing of quadrats into shellfish records, and the subsequent spatialization of these records according to available transect position.

For each region, records were extracted, re-projected, and examined for spatial accuracy. We used a Python script to spatialize each stack of substrate records according to their depth and transect location. Transect point stacks found on land within 300 m of the coastline were moved seaward to the nearest corresponding depth ribbon. If transect start and end points were unique, we intersected the transect with the depth ribbons and placed each point in the centre of the depth ribbon corresponding to its depth. For transects where only the start location was recorded, we distributed the points according to the shortest distance across the depth ribbons, assuming that would be the orientation of the transect. This assumption was both the most parsimonious and produced the fewest possible edge cases. This spatialization routine allowed a single transect to produce up to five points (one for each depth ribbon).

7.2.1.2 Herring survey data

Substrate is collected as part of DFO herring spawn transect surveys. The herring spawn sampling program was designed to support stock assessment and focuses on estimating total seasonal egg production. The herring program used both permanent dive transects planned a priori for survey design, and transects drawn in the field (termed historic transects) based on local observations. The field transects and their data were entered into the herring database a posteriori, and are known to be incomplete (Kristen Daniel, DFO, personal communication). To mobilize the herring substrate data set, considerable processing was required to extract and link the field and spatial transect data, correct the quadrat depths for tide height, and position them in space; much of the work was done as part of a herring data recovery project (Gregr, 2010).

BoP prototypes developed for the SoG (Gregr et al. 2013) and WCVI regions (Gregr, 2016) used herring substrate observations (termed Stations in the herring database) from a summary table created by Gregr (2010). Starting in December 2014 we transitioned to using Python scripts to prepare the herring survey field data and combine them with valid transect locations into a new feature class. We preprocessed the field data by applying a summarizing script (Appendix A2) to the entire coastwide herring data set. We processed 116,170 substrate observations and found 14,250 of those had substrate records with valid transect codes, bottom type and percent cover. These data formed the basis of an updated BType cross-walk table (Table A2). For processing, the fields BType1, BType2, and BType3 in the herring source data were renamed to ensure unique names in the Herring cross-walk table (Table A2).

Using this set of coastwide substrate records we spatialized the data for each region using valid transects from the herring surveys. A total of 9361 valid transects (8200 permanent and 1161 field) were available across Pacific Canada. For each region, we used a spatialization script (Appendix A2) to intersect the transects with the depth ribbons to create a new ArcGIS point feature class at the midpoint of each resulting line segment. These segments were then populated from the substrate records by joining on transect code and summarizing the available substrate data for each transect-depth zone intersection (because there were usually multiple observations from a transect in a depth zone). We added attributes to show the dominant observations within each aggregation including the total stations on the transect-depth zone intersection, the number of unique bottom types, the proportion of the most frequently occurring bottom type, and the three most common bottom types. We finalized the herring data by applying the Phase 1 script to validate the contents of the necessary fields and merged the results into the Obs feature class for Phase 2 processing.

7.2.1.3 ShoreZone

The SZ data are based on line features of the Pacific Canadian coastline. We processed the data regionally following Gregr et al. (2013). With the Nearshore Habitat Working Group we assessed the reliability of the SZ data and concluded that as aerial observations, the SZ data provided reliable information on bottom type up to 5 m depth. These data have evolved considerably since their original production in the 1980s and 1990s (see Lerner and Gregr 2018). The data were originally digitized on the terrestrial BC TRIM high water line (HWL) which has known discrepancies with the CHS HWL. As part of intermittent data improvements, DFO began translating the SZ data to a CHS HWL. However, this effort was complicated by the continued evolution of the CHS HWL coastline, and intermittent updates to the SZ data through 2015. Updated SZ data were provided by Coastal and Ocean Resources Incorporated (CORI) in 2015. These data were used in all the final regional BoPs, and were documented and archived as the version of record for DFO Science (Lerner and Gregr, 2018). Of the 74,580 line segments contained in the database, 89% had the exposure and physical form codes needed to inform bottom type (Table 3, Table A3).

We processed the SZ data regionally in two steps. We first passed the cleaned and projected polyline data through Phase 1 data validation (see section below) where all unnecessary fields were removed[1], We then passed the results to a customized Python script (Appendix A2) that transformed the linework into Thiessen polygons. Lastly we created the SourceKey field and populated it with a code reflecting the source of the data (e.g., “SZ_CORI_Mar2015”).

We found that SZ data were missing for some smaller features represented on the CHS HWL coastline (and thus the 20 m bathymetries) because the feature was not represented in the older TRIM coastline, or because of misalignment between the TRIM and CHS coastlines (see Lerner and Gregr 2018).

7.2.2 Grabs

Grab data were sourced primarily from the CHS. Grabs are also available from NRCan, but most of these are deeper than our study area.

The CHS grab data (Table 4) are a large collection of bottom type samples collected by hydrographers in the field. Prior to the digital acoustic surveys conducted today, hydrographers

---

[1] Retained fields included the key PHY_IDENT, the main ShoreZone identifiers BC_CLASS and EXP_CLASS, and the cross-walked BTypes BType1 and BType2.

recorded depth and bottom type on mylar field sheets. These field sheet data served as the basis of marine charts for decades. In 2014, DFO provided bottom quality data from field sheets for the entire BC coast as part of the World Class Tanker Safety System program. We used these data to develop the first regional BoPs for WCVI and PRCC (see Table 1). For the WCVI, we also included data digitized from CHS field sheets and a rare, shallow water survey by NRCan (see Region-Specific methods for WCVI).

In 2016, CHS provided a more comprehensive ArcGIS-compatible geodatabase of all available bottom type information from marine charts and field sheets as points. No metadata was provided but we identified two sources: the raw, unprocessed grabs data from field sheets; and post-processed versions of these data as annotations of bottom type features displayed on digital charts. These post-processed data are called the S57 data, and are part of the standardized structure used in the CHS charting system. We reviewed the extensive data provided by CHS and elected to use the bottom type data from the field sheet data as the definitive source of these data.

We included the Rock, Kelp, and Marsh annotations from the S57 chart data as proxies for bottom type. We coded kelp and marsh features as hard and soft bottom, respectively, and added coastal rocks to the hard samples. We did not use other S57 substrate records as visual inspection confirmed they were generalizations of the field sheet data and thus redundant.

Working with the aggregated data, we found local areas where points were duplicated but offset, perhaps due to projection issues. We removed these regionally by hand, first using a tolerance of 4 m and then with a tolerance of 20 m to address a second set of duplicates with a regular pattern < 20 m from the correct points (we assumed the field sheet data we collected prior to 2015 to be accurate). While more duplicates were apparent in some locations with a separation of about 50 m distant, these were not removed because the process would have removed many non-duplicate points closer than this threshold, and removing them manually would have been prohibitive. Finally, we recovered those Grab points with correct depth values that occurred on land within 500 m of the ITD and placed them in the nearest part of the ITD depth ribbon (see Appendix A2).

Table 4: Summary of main sources of CHS grab sample data used for each region.

Data source	NCC	HG	WCVI	QCSSOG	Total
CHS 2014	--	--	14,463	--	14,463
CHS digitized	--	--	3,696	--	3,696
NRCan digitized	--	--	1,282	--	1,282
CHS 2016	23,496	13,567	--	25,897	62,960
CHS S57	25,880	4,566	--	12,312	2,758
S57 annotations	--	--	--	5,696	5,696

7.3 Background substrate

Since substrate sampling tends to be patchy (at least in the nearshore), large areas of the coast have little or no substrate data. To fill in these data deserts, we used predictive models to define the background substrate layer to use when no substrate data were near. Prior to 2016, different models were used to estimate the background values.

For the WCVI we used a simple model based on bathymetric roughness which showed a reasonable (0.41) correlation with the BType of a sub-sample of substrate data. For the NCC this approach showed a much lower correlation (0.18) so we used the BC Marine Ecoregion Classification (Zacharias et al., 1998).

These simple background models provided a working representation of bottom type while we developed a random forest (RF) classification model. The prototype RF model was developed in 2015 (Haggarty, 2015) and we used it first in the HG prototype. Random forest substrate models for the other regions were produced shortly thereafter and were used in all subsequent BoPs revisions. The models (including the most recent – Gregr et al., 2021) predict Rock, Mixed, Sand, and Mud (BTypes 1, 2, 3a, and 3b respectively).

The RF substrate models can show considerable variability at local scales (e.g., Figure 3). We therefore smoothed all models using a circular majority filter. This removed the pixel-level noise evident in the raw model output while preserving potentially realistic variability. The size of the filter influences the resolution of the resulting BoPs, particularly when the background is the sole determinant of BType (see Figure 4). We found it useful to adjust the radius size for the different regions, as the filter distance producing the best balance of signal to noise varied by region. We defined best as preserving perceived 'real' features while reducing the scattering of values across pixels. We chose a 60 m filter for the NCC and HG regions, and found a filter of 80 m performed better in QCSSOG. The WCVI BoPs were not updated with the random forest substrate model to maintain consistency with the analyses in Gregr (2016).

We found the small NoData holes created after applying the focal filter can be pushed out of the study area with multiple passes of the filter. However, Phase 2 solves this problem by merging any holes in the final polygon coverage with the dominant neighbouring polygon.

Finally, we converted the filtered raster to polygons using the Smooth option. We added a SourceKey field, and BType1 and BType2 codes. We defined the SourceKey by combining a prefix of ‘RMSM’ with a suffix (e.g., ‘_60m’) noting the size of the majority filter applied.

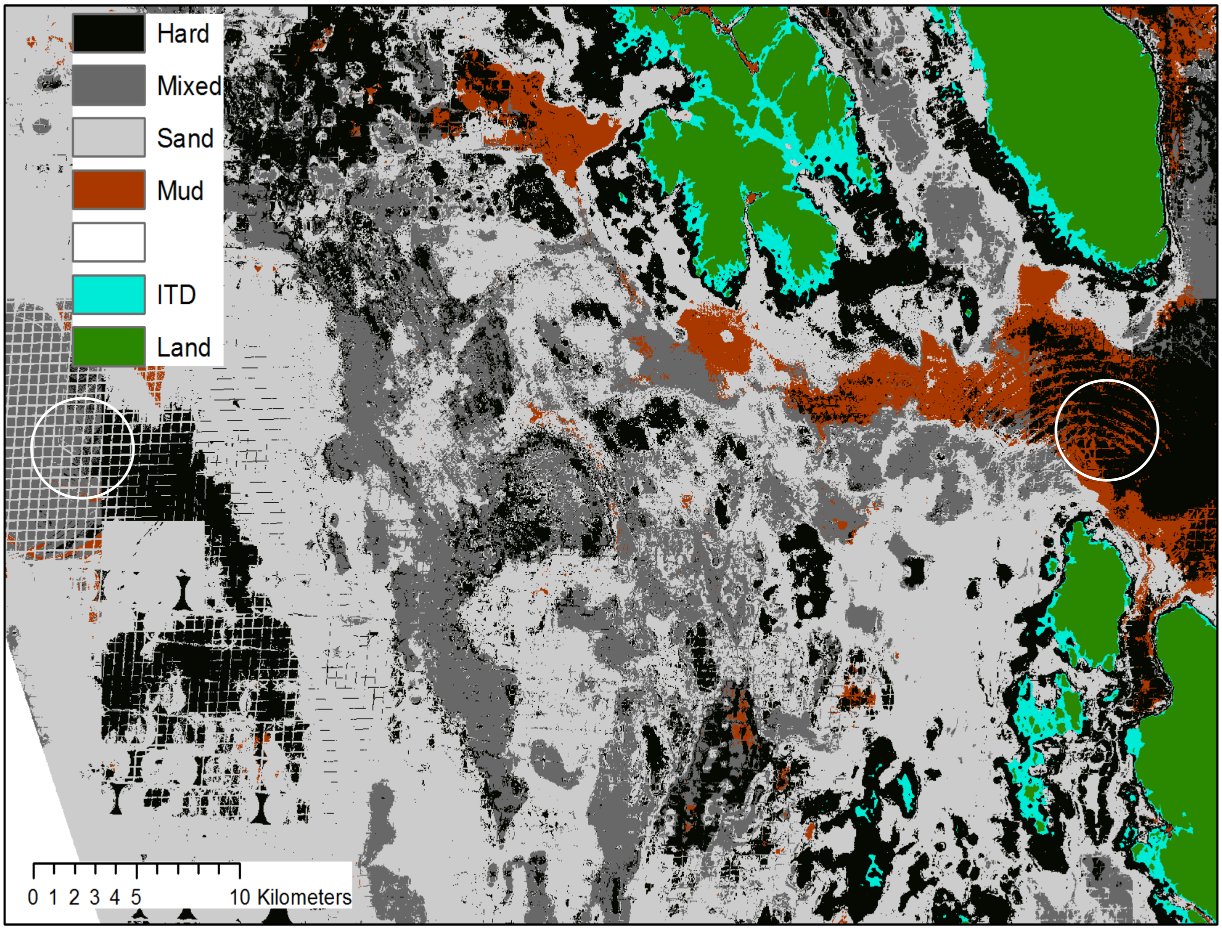


Figure 3: Artefacts in the random forest substrate model. Lower resolution artefacts are visible on the left and right of the figure (circled); higher frequency noise is evident throughout as a kind of fuzz from individual or small groups of pixels. This example is from the most recent substrate model (Gregr et al. 2021). These data are not included in the current bottom patches.

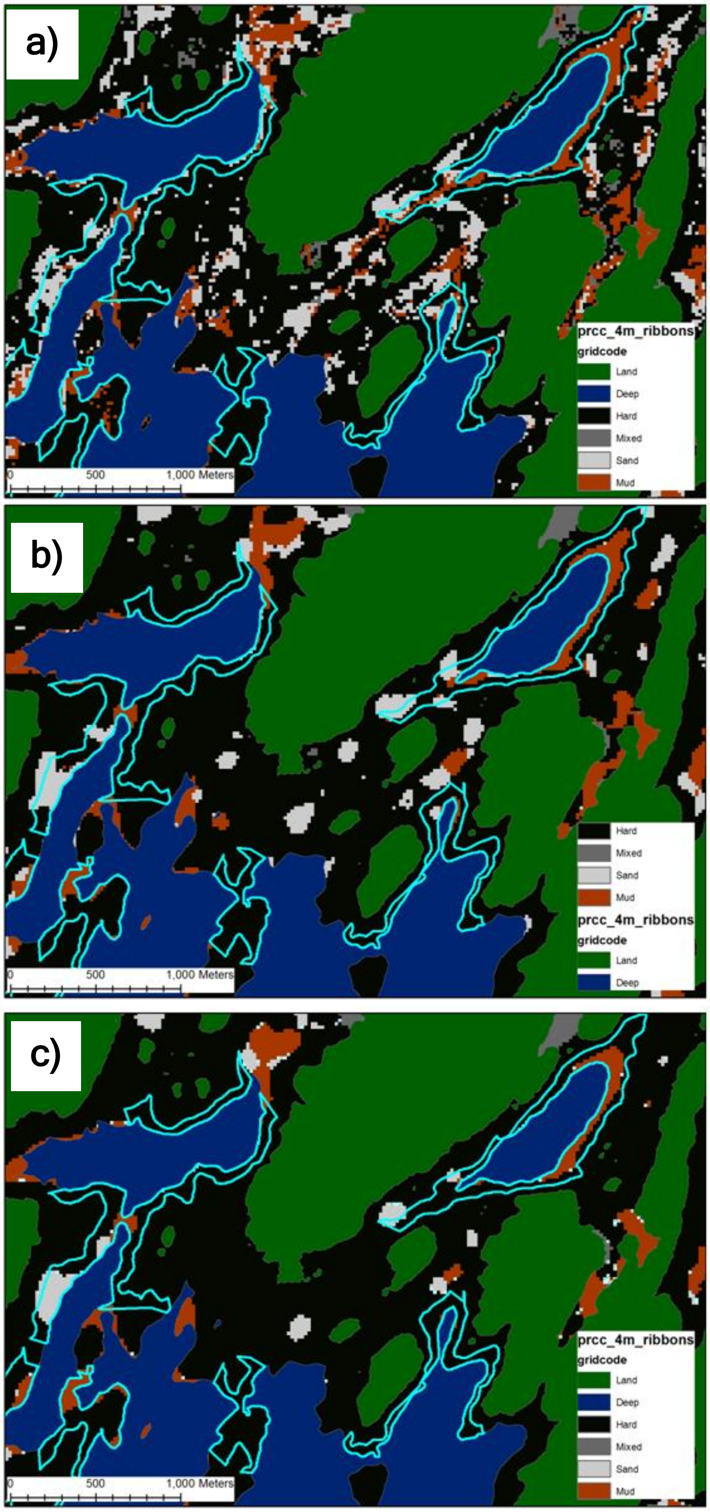


Figure 4: Comparison of random forest substrate model results with a) no smoothing, b) a circular majority filter with a 60 m radius, and c) a majority filter with a 100 m radius. The figure illustrates how increased filtering (from a to c) removes pixel-level noise, but also real heterogeneity. This highlights the trade-off between noise and resolution when filtering the modelled substrate.

7.4 BoP Creation

7.4.1 Phase 1 - Data validation

After pre-processing, we used a validation script to examine each data source for consistency, including field names and completeness of individual records, and to cross-walk the bottom type data from each data source to the BoP BTypes (Table 2). The processing is divided into data cleaning, attribute validation, and clipping stages, with each stage using settings contained in an MS Excel configuration file (Appendix A4).

7.4.2 Phase 2 - Bottom patch production

Phase 2 brings the different data sets together into the BoPs using Thiessen polygons. This process creates Thiessen polygons for the validated Obs and Grabs collections, and combines these with the SZ polygons created during pre-processing. The BoP production script includes 18 steps to create, re-assemble, trim, and validate the final patches. Each step is detailed in Gregr and Peterman (2022) and the process is summarized below.

The Thiessen polygons from each data source are first clipped to the land and 50-100 m masks. They are then intersected, producing a large collection of polygon fragments, with each fragment informed by one or more data sources. The fragments are then assigned BType and Confidence, and refined according to a set of decision rules to limit extrapolation.

7.4.3 Assigning BType and confidence

We used a cascading series of decision rules (Appendix A3) to assign BType and confidence values to each polygon fragment based on distance from, and agreement with, the source substrate data. The confidence values describe the relative accuracy associated with the BType assigned. We assigned BType and confidence values according to the following characteristics of each fragment resulting from the intersection of the source data (in this case, Obs, Grabs, and SZ):

1) The number of data sources informing the polygon
2) The level of agreement of the different sources
3) The distance of the nearest source data point
4) The depth ribbon of the polygon
5) The assumed reliability of the original point data.

The algorithm begins by checking the number of data sources that overlap each polygon fragment. If there is more than one source, the level of agreement between the different source data sets is examined, and BType is assigned based on the dominant substrate type. If there is no agreement among multiple data sources, then the BType of the closest source point within 100 m of the polygon is assigned. If there is no point within 100 m, then the SZ-derived BType is assigned if the fragment is in the two shallowest depth ribbons, and the Obs BType is used for fragments in the deeper ribbons, making the assumption Obs are more reliable than Grabs.

Each BType assignment is paired with a categorical confidence score reflecting the relative confidence in the bottom type assigned. This confidence is based on the characteristics listed above and reflects the disagreement between the BTypes of the data sources that influence each fragment (fragments are typically influenced by multiple data sources, often with different BTypes). Our guiding principle in designing the BType and confidence assignment rules was that coherent fragments (those where multiple sources agreed) and those with a higher spatial density (smaller polygons imply higher resolution sampling), have higher accuracy.

7.4.4 Managing BoP size

The fragments resulting from the intersection of Thiessen polygons are variable in size and shape, reflecting the patchiness and overlap of the source data. To finalize the BoPs, the fragments are merged with their neighbours with the same BType, and any small slivers resulting from the intersections are absorbed by the largest adjacent polygon. We applied a minimum patch size (set to 4 m^2 for all regions).

The influence of an individual substrate datum is a key consideration in the BoP algorithm. Phase 2 provides parameters to limit the distance substrate can be inferred from each datum. This includes a user-specified extrapolation distance, and a polygon reduction ratio. The maximum influence distance limits the influence of a datum to the specified distance. We used 500 m for all final BoP layers. The polygon reduction ratio reduces the number of long narrow BoPs caused by few points occurring on narrow depth ribbons. We used a value of 2 for all final models, meaning a BoP could not be more than twice as long as wide. BoPs exceeding this ratio were recursively cut in half retaining only the half containing the source point until the ratio of length to width was no larger than the reduction ratio. The polygon reduction ratio occurs before the inclusion of the background substrate and as such, the polygons informed by the background substrate can be quite large, especially in deeper depth ribbons. Other environmental variables (e.g., fetch) could be used to refine these larger polygons.

Finally, the background substrate model is used to assign BType to BoPs in areas with no substrate data.

7.5 Region-specific methods

7.5.1 North Central Coast (NCC)

The ribbons for this region were updated with the 20 m bathymetry produced by DFO in August 2015. We used a 2 m bathymetry interpolated from the 20 m raster using bilinear interpolation to improve the ribbon resolution in high relief areas. We set the HWL elevation at 4 m for this region based on the distribution of raster elevations at the vertices of the most recent HWL feature (see Davies et al. 2019 for details on CHS HWL).

We used the same DFO herring transects as the PRCC region prototype for the final NCC BoPs. Two changes to the herring source data were necessary to successfully pass them through Phase 1 processing. First, we found that longer attribute names in the transect file had been clipped because of the limit on the length of attribute names in ESRI shape files. We cross-checked the clipped names with the herring lookup sheets in the MS Excel configuration file and renamed the internal herring BType fields so that the Phase 1 script could create and populate the BType1 and BType2 fields. After passing through the Phase 1 script, we spatialized them using the updated NCC depth ribbons.

Local Obs data from towed video arrays were provided by Hakai Research Institute and the Central Coast Indigenous Resource Alliance (CCIRA). The Hakai data were local to Calvert Island, while the CCIRA data were more broadly distributed in the central NCC. The towed video data had high spatial resolution so we sub-sampled it to a resolution of 1 m. We then checked for null values and errors, reducing the number of records from 27,893 to 7,591 valid records on the depth ribbons. We merged the results with the Obs dive data prior to Phase 2 processing. The data had enough detail to assign both primary and secondary BTypes (Table 2).

The Hakai data included ROV surveys from 2012 and 2014 which were extracted from Excel workbooks with Python into a single point feature class with 31,567 points. Of these a final set of 27,899 points had valid geomorphic and attribute codes and were imported directly into ArcGIS and merged with the Obs dive data. The 2012 codes differed somewhat from those standardised in 2014 necessitating separate look-up tables. The 2012 data also contained shorter tows more broadly distributed than the 2014 surveys.

After Phase 1 processing, we created the SourceKey field manually and a simple meaningful name was assigned to all records within each data set. The data were then merged, retaining the BType1, BType2, and SourceKey fields. We added a DepthCode field to the merged file to identify the depth ribbon using a spatial join. The resulting feature class was used in Phase 2.

The Grabs prepared for the PRCC prototype were used in the final NCC version. The data were re-processed using the Phase 1 scripts to ensure all codes were up to date and the data were correctly structured. We added several new codes to the CHS grabs attribute table (Table A4).

The final version of the NCC Grabs included the S57 data with duplicates removed (see Substrate data pre-processing, above). We passed the data through the Phase 1 script and added the SourceKey field manually.

7.5.2 Haida Gwaii (HG)

We set the HWL elevation to 4.9 m for the final version of the HG depth ribbons based on the distribution of raster elevations found at the vertices of the most recent CHS HWL feature (Davies et al. 2019).

We selected data from the full DFO Shellfish database using the HG boundary file used for the ribbon production. We spatialized the shellfish data using the latest stand-alone script (Appendix A2).

We created regional subsets of the CHS Grabs and S57 data and passed them through the Phase 1 script. Points on land within 500 m of the ITD depth ribbon were recovered. We used a spatial join to assign the appropriate DepthCode to each point.

7.5.3 West Coast Vancouver Island (WCVI)

We used this region to prototype the automation of the BoP production process. Many of the data cleaning and validation activities implemented in the validation script originated during the development of the WCVI BoPs, as did approaches to the spatialization of observations collected on transects. Differences from other regions include unsmoothed depth ribbons and a missing BoPID.

We set the HWL elevation to 5 m for the depth ribbons in this region based on the distribution of raster elevations found at the vertices of the most recent CHS HWL feature (Davies et al. 2019). During production we found that smoothing the ribbons created gaps in areas of high variability. The ribbons were therefore rebuilt without smoothing using raster to polygon conversion after resampling the 20 m bathymetry to 2 m with interpolation. While at close inspection the resulting ribbons and BoPs appear jagged, the gaps that appeared between smoothed ribbons were avoided. The desire to avoid these jagged features led to the development of the re-sampling approach used to build the depth ribbons for the other regions.

The Obs used this region are from earlier versions of the herring, shellfish and SZ databases, and have not been rerun with data provided in 2015 and 2016. The Grabs for this region are also based on earlier (2014) substrate data compiled by CHS, and on CHS and NRCan data digitized to fill in gaps in the WCVI substrate coverage (Table 4). We removed any areas of overlap between the digitized sheets and the 2014 substrate data.

While NRCan data are typically collected deeper than our study area, the WCVI area is an exception as a pair of NRCan surveys provide good coverage of the nearshore. Originally digitised as part of the WCVI prototype, these data (Bornhold and Barrie, 1991) have since been compiled into the NRCan Expedition Database (Natural Resources Canada, 2015). A total of 1,282 points were digitized containing 3 classes: Gravel, Sand, and Mud.

7.5.4 Queen Charlotte Strait / Strait of Georgia (QCSSOG)

Depth ribbons for the QCS and SOG regions were created separately from the 20 m bathymetries and combined for the BoP analysis. The QCS ribbons were based on bathymetry provided in November 2015, while the SOG bathymetry was last updated in October 2016. We set the HWL elevation to 3.25 m based on the distribution of raster elevations at the vertices of the most recent HWL feature (Davies et al. 2019).

The updated shellfish data provided by DFO in June 2016 contained almost 1,000 more valid points within the QCSSOG study area than the earlier version. We created the final BoPs using the revised 2017 workflow and the stand-alone shellfish script (Appendix A2).

Regional (QCS and SOG) subsets of the CHS Grabs and S57 data were created and merged for this analysis. The combined file was passed through the Phase 1 script and BType values assigned from a combined Grabs and S57 lookup table. Points within 500 m of the ITD depth ribbon were recovered.

8 RESULTS

The BoP values are best viewed by coding the polygons by BType1 and BType2 (e.g., Figure 5a). In addition to clearly showing bottom type, this view allows confidence to be inferred from the size of the polygons (smaller polygons indicate higher data density and thus greater confidence). The Confidence attribute (Figure 5b) explicitly shows the level of agreement between the different source data sets and allows quick reference to areas dependent on predicted background substrate and thus data deficient.

8.1 Bottom patch overview

Comparing the distribution of the BoP classes across regions gives information on the regional differences in bottom type, while the relative influence of the different data sources on the BoPs gives a sense of source data distribution, bias, and regional coverage. These results are summarized below.

The relative contribution of the different data types (Table 5) shows that the SZ data influenced about half of all the BoPs, while Grabs influenced between 19 and 25% depending on region. The influence of these data were fairly consistent across regions. Obs contributed less than 15% to any region, and were especially poorly represented in HG where they informed only 4% of the BoP polygons. Between 27 and 45% of the BoPs were influenced by the background layer and about 20% were influenced by multiple data sources (the amount over 100% in the proportion total).

A total of 864,531 BoPs were defined for the entire Pacific Canada coastal zone (Table 6). The NCC accounted for about half of these, while HG, with its smaller coastal zone, contributed just over 10%. The primary BTypes (Hard, Mixed, Soft) were distributed most evenly in the NCC, while the WCVI showed the greatest unevenness (Table 6).

Regionally, the final 431,639 NCC BoPs cover over 6,700 km^2, with most BoPs being influenced by SZ and Background, and the Obs and Grabs data influencing almost equal proportions (Table 5).

The QCSSOG region produced 235,754 BoPs covering over 3,900 km^2. The BoPs in this region were most influenced by SZ (60%) and the least by Background (27%) (Table 5).

There are 86,825 BoPs in HG spanning 10,807 km^2. HG is the only region that was more influenced by Background (45%) than SZ (39%). This region also had the highest proportion of BoPs influences by Grabs (25%).

On the WCVI, we defined a total of 110,313 polygons covering an area of over 4300 km^2 (Table 6). BoPs in this region followed the common pattern of most being influenced by SZ (47%) and Background (37%), with less influence coming from Grabs (22%) and Obs (12%) (Table 5).

The mixed category comprises the largest number of BoPs (44%) with about equal proportions soft and hard (27% and 28% respectively). HG and NCC are similar in having fairly equal substrate proportions (Table 6). The proportion of BTypes on WCVI is highest for mixed (44%) with little soft (17%) (Table 6).

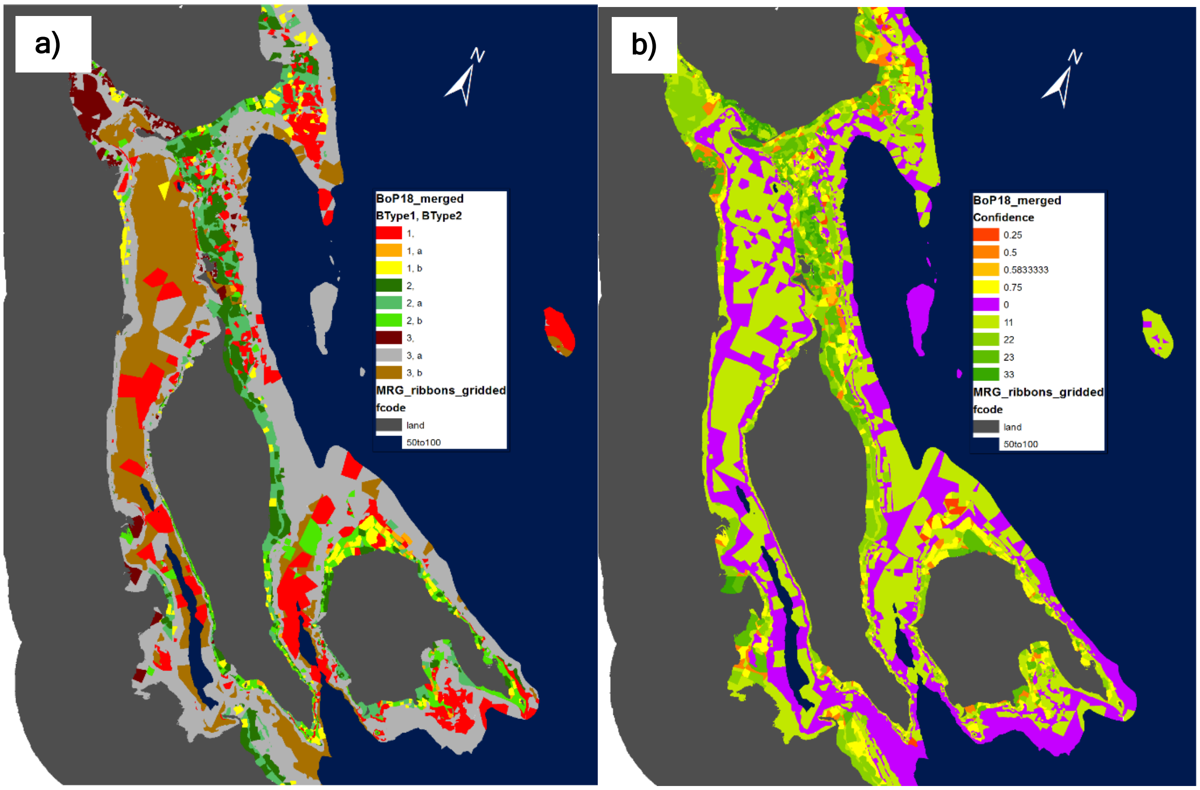


Figure 5: a) Bottom patches (BoPs) coded for BType1 and BType2, and b) BoP confidence surface in Baynes Sound. Land is dark grey and deep water is dark blue. This highly sampled area shows how BoPs can resolve into a credible representation of bottom type with adequate sampling. The confidence surface shows areas of high agreement across data sets, as well as where the background (modelled) substrate data were used.

Table 5: The number and proportion of bottom patches (BoPs) influenced by the major source data sets. Influence means the data source contributed to a BoPs shape and assigned bottom type. Many BoPs were influenced by more than one data source.

NCC		
Source	**BoPs influenced**	**Proportion**
ShoreZone	215,056	0.50
Background	160,132	0.37
Observations	59,443	0.14
Grabs	82,528	0.19
Total	517,159	1.20
HG		
Source	**BoPs influenced**	**Proportion**
ShoreZone	33,429	0.39
Background	38,784	0.45
Observations	3,289	0.04
Grabs	21,520	0.25
Total	97,022	1.13
QCSSOG		
Source	**BoPs influenced**	**Proportion**
ShoreZone	142,115	0.60
Background	64,019	0.27
Observations	32,792	0.14
Grabs	49,293	0.21
Total	288,219	1.22
WCVI		
Source	**BoPs influenced**	**Proportion**
ShoreZone	51,579	0.47
Background	40,455	0.37
Observations	12,782	0.12
Grabs	24,656	0.22
Total	129,472	1.18

Table 6: The proportion of the dominant bottom types (BTypes), their mean size, and proportion of total area covered by region.

NCC			
BType1	**Proportion**	**Mean size (m^2)**	**Area (km^2)**
1	0.385	15,447	0.379
2	0.340	12,491	0.270
3	0.274	20,107	0.351
Total count:	431,639	**Total area:**	6786.23
QCSSOG			
BType1	**Proportion**	**Mean size (m^2)**	**Area (km^2)**
1	0.295	15,346	0.270
2	0.437	8,444	0.221
3	0.268	31,767	0.509
Total count:	235,754	**Total area:**	3943.29
HG			
BType1	**Proportion**	**Mean size (m^2)**	**Area (km^2)**
1	0.371	35,083	0.105
2	0.244	66,914	0.131
3	0.385	247,018	0.764
Total count:	86,825	**Total area:**	10806.79
WCVI			
BType1	**Proportion**	**Mean size (m^2)**	**Area (km^2)**
1	0.388	30,643	0.298
2	0.441	37,869	0.419
3	0.171	65,774	0.283
Total count:	110,313	**Total area:**	4396.09
Grand Total :	864,531		

9 DISCUSSION

The BoPs were conceived before any well-resolved bathymetry or coastal substrate maps were available. As such, they filled an important data gap in the development of habitat suitability models for coastal species, while also providing the impetus for the development of high resolution bathymetric and substrate layers.

The BoPs have been used to support several coastal models to date. Gregr et al. (2018) found that the BoPs led to kelp habitat suitability models with comparable performance to those built with a random forest model of substrate when evaluated using independent data. This suggests that any increase in precision of the classification models over the BoPs may be within the uncertainty of the overall model. Robinson et al. (2021) found the BoPs contributed significantly to the habitat of sandlance under a variety of model frameworks. They also showed better agreement than random forest models (e.g., Gregr et al. 2021) to available, independent observations of substrate (Robinson, personal communication).

Recognizing the need to capture processes from different resolutions, Misiuk et al. (2018) showed that the best-performing models relied on predictors derived from a range of resolutions. Similarly, Porskamp et al. (2018) showed how sediment model performance varied with predictor resolution, and Gregr et al. (2021) further showed a correlation between predictor resolution and variable performance with depth. This emphasizes the difficulty of using a single resolution, gridded framing to capture processes operating at different spatial scales.

An object-based framing (like BoPs) allows a variety of potential attributes to be assigned. Importantly, such attributes could be derived from different resolutions, providing a solid conceptual framing for considering processes from multiple scales. Object-based approaches are also increasingly being used in image analysis as remote sensing resolutions increase (Lightfoot et al., 2020). Algorithmically combining pixels into polygons avoids the salt and pepper appearance of pixel-based methods, and can improve accuracy (Lightfoot et al., 2020). Polygons can also be a more conceptually accessible organization of classes, providing potentially useful features related to shape and neighbour relationships, and reduce computational effort (Mitchell et al., 2018). Segmentation approaches (where pixels are grouped into polygons algorithmically) include kriging (Bostock et al., 2019), and rule-based approaches, of which the BoPs presented here are one example.

A further advantage of the object-based approach is the ability to immediately convey the data density and quality in the underlying source data sets (e.g., Figure 6). Patches of high data density (e.g., Figure 6a) provide insight in the true regional heterogeneity, a useful measure for assessing model fit. Data quality (Figure 5b) is inferred both from the size of the polygons, and the level of (dis)agreement between independent data sets (e.g., Figure 6a).

One challenge to the BoP process is that the intersection of different layers leave slivers and thin wedges that when combined are unlikely to accurately reflect local bottom types (e.g., Figure 6b). However, it is not clear whether such discontinuities are worse than model interpolations that, while providing aesthetic boundaries, do not contain information on their local accuracy. A comparison of dive survey and SZ data suggests that local high resolution surveys could be used to assess the sufficiency of local sampling density (Figure 6a).

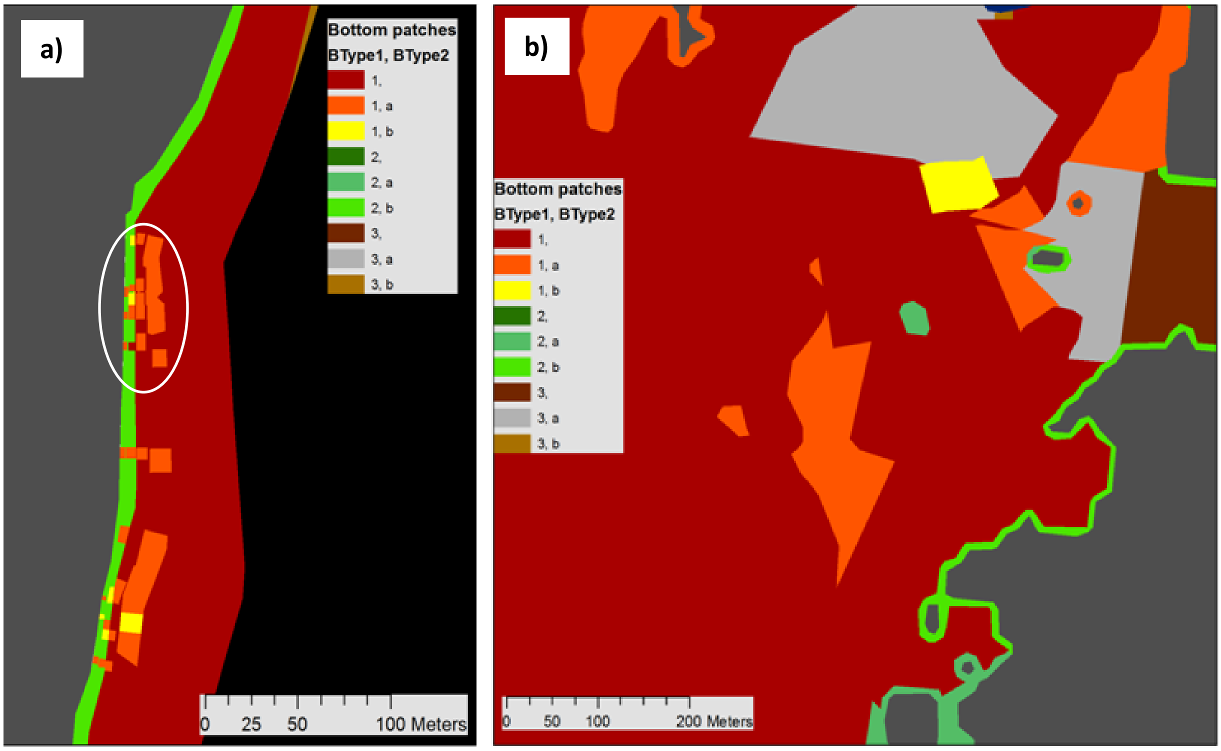


Figure 6: Two types of unusual features resulting from the BoP algorithm settings and performance. Dark grey is land, black is deep (> 100 m). Panel a) shows the effect of combining data from different scales and the potential for disagreement between dive survey and ShoreZone data (circled example). Panel b) shows the small fragments and thin wedges resulting from the intersection, deletion, filling, and merging of polygons to form the final BoPs.

9.1 Next steps

Available data and modelling methods have continued to evolve since the BoPs were completed in 2017. However, the challenge of single integrated representation of substrate for the Pacific Canada shelf remains. Integrating predictions from recently completed substrate models (Gregr et al. 2021) with updated BoP methods could now address this notable data gap using a leading edge, multi-scale approach. The updated BoPs would also provide important insights into the accuracy of the predictive models.

Refinements to the BoPs methods should include adding a categorical exposure layer to partition the patches more accurately around headlands. One challenge with an object-based approach is determining the definition of the most basic unit, in terms of size and attributes. A consultation with the broader nearshore modelling community could help determine this limit, while also providing feedback and potential improvements to the methods. The effectiveness of the Thiessen polygon approach could also be compared to other potential segmentation methods, which may alleviate the challenge of slivers, and oddly shaped polygons. Any update to the BoPs should also include a review of data available from various sources beyond those used herein. An update of the WCVI BoPs is also overdue.

Bathymetric artefacts remain in the source data underlying the 20 m bathymetries. Some localized smoothing (in areas where survey transects are visible) and an accurate (i.e., variable) HWL from CHS would allow the sea and land side elevations to be locally interpolated, significantly improving the 20 m bathymetries and their derivatives.

10 CONCLUSIONS

The BoPs represent an object-based classification of the coastal zone in Pacific Canada that combines the accuracy of substrate point data with the generality of a substrate classification model applied to those same model outputs. The process for creating the BoPs is reproducible, and can be updated with additional data or refined methods as necessary. The corresponding confidence layer also provides information on the local accuracy of the assigned substrate values. The benefits of an object-based approach include the ability to assign multiple attributes to each BoP, and to integrate values from different resolutions. Given that predictive models are sensitive to the non-stationarity of substrate both across depths and regions, the ability to include values across scales makes the BoPs a leading approach to creating a comprehensive, shelf-to-slope substrate map.

11 ACKNOWLEDGEMENTS

Allison Thomson digitized the NRCan data and Caitlin Millar the CHS field sheet data at the University of British Columbia, in the Institute For Resources, Environment, and Sustainability. Pete Wills with the Canadian Hydrographic Service was a strong and generous supporter of this work. Michael Peterman wrote the Python scripts. Andrew Day and Kevin Head from West Coast Aquatic provided critical stakeholder engagement and support during the development of the BoPs for the West Coast of Vancouver Island. Careful reviews by Ashley Park, Cliff Robinson, and Beatrice Proudfoot greatly improved the clarity of this report. Thank you all very much.

12 REFERENCES

BORNHOLD, B. D. & BARRIE, J. V. 1991. Surficial sediments on the western Canadian continental shelf. *Continental Shelf Research,* 11**,** 685-699.

BOSTOCK, H., JENKINS, C., MACKAY, K., CARTER, L., NODDER, S., ORPIN, A., PALLENTIN, A. & WYSOCZANSKI, R. 2019. Distribution of surficial sediments in the ocean around New Zealand/Aotearoa. Part B: continental shelf. *New Zealand Journal of Geology and Geophysics,* 62**,** 24-45.

COASTAL OCEANS. 2022. *ShoreZone of the British Columbia coastline* [Online]. Available: https://mcori.maps.arcgis.com/apps/Viewer/index.html?appid=c76377500f814914ad90149f229d4d66 [Accessed April 2022].

DAVIES, S. C., GREGR, E. J., LESSARD, J., BARTIER, P. & WILLS, P. 2019. Coastal digital elevation models integrating ocean bathymetry and land topography for marine ecological analyses in Pacific Canadian waters. *In:* CANADA, F. A. O. (ed.).

GREGR, E. J. 2010. Herring Survey Substrate Recovery Project.

GREGR, E. J. 2016. *Sea otters, kelp forests, and ecosystem services: modelling habitats, uncertainties, and trade-offs.* PhD, University of British Columbia.

GREGR, E. J., HAGGARTY, D. R., DAVIES, S. C., FIELDS, C. & LESSARD, J. 2021. Comprehensive marine substrate classification applied to Canada's Pacific shelf. *PLoS One,* 16**,** e0259156.

GREGR, E. J., PALACIOS, D. M., THOMPSON, A. & CHAN, K. M. 2018. Why less complexity produces better forecasts: An independent data evaluation of kelp habitat models. *Ecography*.

GREGR, E. J. & PETERMAN, M. 2022. Python Scripts for Nearshore Bottom Classification: Documentation and User Guide. Available: https://github.com/ejgregr/bop_scripts_2017 [Accessed May 2022].

HAGGARTY, D. R. 2015. *An evaluation of the effectiveness of rockfish conservation areas in British Columbia, Canada.* Doctor of Philosophy, University of British Columbia.

LERNER, J. & GREGR, E. J. 2018. Description of the British Columbia ShoreZone Mapping System. Available: https://www.researchgate.net/publication/360345821_Description_of_the_British_Columbia_ShoreZone_Mapping_System [Accessed May 2022].

LIGHTFOOT, P., SCOTT, C. & FITZSIMMONS, C. 2020. Using object-based image analysis with multi-temporal aerial imagery and LiDAR to detect change in temperate intertidal habitats. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 30**,** 514-531.

MISIUK, B., LECOURS, V. & BELL, T. 2018. A multiscale approach to mapping seabed sediments. *PLoS One,* 13**,** e0193647.

MITCHELL, P. J., DOWNIE, A.-L. & DIESING, M. 2018. How good is my map? A tool for semi-automated thematic mapping and spatially explicit confidence assessment. *Environmental Modelling & Software,* 108**,** 111-122.

NATURAL RESOURCES CANADA 2013. Canadian Digital Elevation Model Product Specifications. *In:* CANADA, N. R. (ed.).

NATURAL RESOURCES CANADA 2015. Expedition Database. *In:* CANADA, N. R. (ed.).

PORSKAMP, P., RATTRAY, A., YOUNG, M. & IERODIACONOU, D. 2018. Multiscale and hierarchical classification for benthic habitat mapping. *Geosciences,* 8**,** 119.

ROBINSON, C. L., PROUDFOOT, B., ROOPER, C. N. & BERTRAM, D. F. 2021. Comparison of spatial distribution models to predict subtidal burying habitat of the forage fish Ammodytes personatus in the Strait of Georgia, British Columbia, Canada. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 31**,** 2855-2869.

ZACHARIAS, M. A., HOWES, D. E., HARPER, J. R. & WAINWRIGHT, P. 1998. The British Columbia marine ecosystem classification: Rationale, development and verification. *Coastal Management,* 26**,** 105-124.

13 APPENDICES

13.1 A1. Attribute lookup tables

The Look up tables from the MS Excel spreadsheet (cross-walk to user document)

Table A1: Bottom type (BType) classification of shellfish survey substrate data (updated from Gregr et al. 2013).

BType	Primary substrate	Secondary substrate
1a	1	All values
	2	0, 1, 5, 6, 7, 8, 9, 10, 11, null
	0	1, 2
1b	2	3, 4
	3	All values
	0	3
	4	1, 2, 3, 5, 6, null
	5, 6, 7, 8, 10, 11	1, 2, 3
2a	4	0, 7, 8, 9, 10, 11
	5	0, 4 6, 7, 8, 9, 10, 11, null
	6	0, 4, 5, 7, 8, 9, 10, 11, null
	11	4, 5, 6
2b	0	4, 5, 6, 11
	7, 8, 10	4, 5, 6
	11	0, 7, 8, 9, 10, null
3a	0	7, 8, 10
	7	0, 8, 9, 10, 11, null
	8	0, 7, 9, 10, 11, null
	10	0, 7, 8, 9, 11, null
3b	0	9, null
	9	All values

Shellfish codes: 1. bedrock smooth, 2. bedrock crevices, 3. boulders, 4. cobble, 5. gravel, 6. pea gravel, 7. sand, 8. shell (old code), 9. mud, 10. crushed shell, 11. whole shell, 0. wood debris.

Table A2: Lookup table for Herring observations

This lookup table is to translate the three reported dominant substrate types into BType classes. We designed the table to be adaptable by allowing the definition of dominant type to be configured here. The table accommodates all observed combinations of Rock, Boulders, Cobbles, Pebbles, Sand, Shell, and Mud to be coded to the most appropriate BType class. Used by the *Herring_summarize.py* Python script and applied once to the available coastwide herring dataset.

BType1DomOperator	BType1Dominant	BTypeUniOperator	BTypeUnique	BType1orig	BType2orig	BType3orig	BType1	BType2	Herring substrate
>=	0.8	>=	1	R			1	a	Rock
>=	0.8	>=	1	B			1	b	Boulders
>=	0.8	>=	1	C			1	b	Cobbles
>=	0.8	>=	1	P			2	b	Pebbles
>=	0.8	>=	1	S			3	a	Sand
>=	0.8	>=	1	SH			3	a	Shell
>=	0.8	>=	1	M			3	b	Mud
<	0.8	==	2	R,B,C,P	R,B,C,P		1	b	Mix of hard substrates with none dominant
<	0.8	==	2	R,B,C,P	S,SH,M,P		2	a	Hard dominant, with some soft patches
<	0.8	==	2	S,SH,M	R,B,C,P		2	b	Soft dominant with hard patches
<	0.8	==	2	S,SH	S,SH		3	a	Sand/shell dominated soft
<	0.8	==	2	S,SH	M		3	a	Sand/shell dominated soft
<	0.8	==	2	M	S,SH		3	b	Mud dominated soft
>=	0.5	>=	3	R,B,C,P	R,B,C,P	R,B,C,P	1	b	Mix of hard substrates with none dominant
>=	0.5	>=	3	R,B,C,P	R,B,C,P	S,SH,M,P	2	a	Hard dominant, with some soft patches
>=	0.5	>=	3	R,B,C,P	S,SH,M,P	R,B,C,P	2		Mixed, neither dominant
>=	0.5	>=	3	R,B,C,P	S,SH,M,P	S,SH,M,P	2		Mixed, neither dominant
>=	0.5	>=	3	S,SH,M	R,B,C,P	R,B,C,P	2		Mixed, neither dominant
>=	0.5	>=	3	S,SH,M	R,B,C,P	S,SH,M,P	2		Mixed, neither dominant
>=	0.5	>=	3	S,SH,M	S,SH,M	R,B,C,P	2	b	Soft dominant with hard patches
>=	0.5	>=	3	S,SH	S,SH	S,SH,M,P	3	a	Sand
>=	0.5	>=	3	M	M	S,SH,M,P	3	b	Mud
>=	0.5	>=	3	S,SH,M	S,SH,M	S,SH,M,P	3		Mix of soft substrates with none dominant
<	0.5	>=	3	R,B,C,P	R,B,C,P	R,B,C,P	1	b	Mix of hard substrates with none dominant
<	0.5	>=	3	R,B,C,P	R,B,C,P	S,SH,M,P	2	a	Hard dominant, with some soft patches
<	0.5	>=	3	R,B,C,P	S,SH,M	S,SH,M	2		Mixed, neither dominant
<	0.5	>=	3	R,B,C,P	S,SH,M	R,B,C,P	2	a	
<	0.5	>=	3	S,SH,M	R,B,C,P	R,B,C,P	2		Mixed, neither dominant
<	0.5	>=	3	S,SH,M	S,SH,M	R,B,C,P	2	b	Soft dominant with hard patches
<	0.5	>=	3	S,SH,M	R,B,C,P	S,SH,M	2	b	
<	0.5	>=	3	S,SH	S,SH	S,SH,M,P	3	a	Sand
<	0.5	>=	3	M	M	S,SH,M,P	3	b	Mud
<	0.5	>=	3	S,SH,M	S,SH,M	S,SH,M,P	3		Mix of soft substrates with none dominant

Table A3: Lookup table for ShoreZone observations

We combined the ShoreZone exposure class (EXP_CLASS) with the coastal class (BC_CLASS) following Gregr et al. (2013).

EXP_CLASS	**BC_CLASS**	**BType1**	**BType2**	**Description**
VE,E	1,2,3,4,5,6,7,8,9,10,33	1	a	Rock Ramp, wide
VE,E	11,12,13,14,15,16,17,18,19,20	2	b	Rock Platform, wide
VE,E	21,22,23,32	1	b	Rock Cliff
VE,E	24,25,26,34	2	a	Rock Ramp, narrow
VE,E	27,28,30	3	a	Rock Platform, narrow.
SE	1,2,3,4,5	1	a	Ramp with gravel beach, wide
SE	6,7,8,9,10,21,22,23,24,25,26,32,34,35	2	a	Platform with gravel beach, wide
SE	11,12,13,14,15,16,17,18,19,20,33	2	b	Cliff with gravel beach
SE	27,28,29,30,31	3	a	Ramp with gravel beach
SP	1,2,3,4,5,11,12,13,14,15,33	2	b	Platform with gravel beach
SP	6,7,8,9,10,21,22,23,24,25,26,32,34,35	2	a	Ramp w gravel & sand beach, wide
SP	16,17,18,19,20,27,28,30	3	a	Platform w gravel & sand beach, wide
SP	29,30,31	3	b	Cliff with gravel/sand beach
P	1,2,3,4,5,11,12,13,14,15	2	b	Ramp with gravel/sand beach
P	6,7,8,9,10,21,22,23,34,35	2	a	Platform with gravel/sand beach
P	16,17,18,19,20,24,25,26,27,28,30,31,32,33	3	a	Ramp with sand beach, wide
P	29,30,31	3	b	Platform with sand beach, wide
VP	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15	2	b	Cliff with sand beach
VP	16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,35	3	b	Ramp with sand beach, narrow
VP	34	2	a	Platform with sand beach, narrow

Table A4: Lookup table for Canadian Hydrographic Service (CHS) grabs

BType1	BType2	CHS Feature code	CHS Description
1		BQHD	hard
1		BQRC	rock
1		DLRA	rock awash
1		DLRK	rock below datum
1		DLRKREP	reported
1		DLSF	intertidal rock
1	b	DLBE	boulder
1	b	BQBO	boulder
1	b	BQBS	boulders and sand
1	b	BQBG	boulder gravel
1	b	BQSN	shingles
1	b	BQSS	stone
2	a	BQCA	coarse
2	a	BQCO	cobble
2	a	BQSS	stone or cobbles
2	b	BQPB	pebble
2	b	BQGR	gravel
2	b	BQGB	gravel and boulders
3	a	BQGS	gravel sand
3	a	BQSO	sand
3	a	BQSD	sand
3	a	BQSN	sand
3	a	BQSP	sand
3	a	BQSG	sand gravel
3	a	BQSH	shell
3	a	BQSM	sand mud
3	a	BQWS	weed sand
3	b	BQCY	clay
3	b	BQFN	fine
3	b	BQFS	fines and sand
3	b	BQMD	mud
3	b	BQMG	mud gravel
3	b	BQMS	mud sand
3	b	BQOZ	ooze
3	b	BQRE	Red
3	b	BQBR	Brown (colors are assumed to be mud)
3	b	BQGN	Green
3	b	BQGY	Grey
3	b	BQBL	Black
3	b	BQYW	Yellow
3		BQWD	Weed (sand/mud not discernable)

Table A5: Lookup table for Canadian Hydrographic Service (CHS) S57 observations

BType1	BType2	CHS Feature code	CHS Description
1		NFKE	Kelp
3	b	NFMS	Mudflats
1	a	CLLWRL1R	Rock ledge
1	a	CLLWRL2R	Rock ledge
1	a	CLLWRL6R	Rock ledge
1	b	DLRA	Rock awash
1	b	LDRA	Rock awash
1	b	DLRK	Rock
1	b	DLSF	Intertidal (sinking) rock

Table A6: Lookup table for Natural Resources Canada (NRCan) grabs (updated from Gregr et al. 2013).

BType1	BType2	NRCan sediment code	Description
1	a	R	Bedrock
1	b	B	Boulders
2	b	G	Gravel
2	b	mG	Mud-gravel
2	b	msG	Mud-sand-gravel
2	b	sG	Sand-gravel
3	a	gmS	Gravel-mud-sand
3	a	gS	Gravel-sand
3	a	mS	Mud-sand
3	a	S	Sand
3	a	smG	Sand-mud-gravel
3	b	gM	Gravel-mud
3	b	gsM	Gravel-sand-mud
3	B	M	Mud
3	B	sM	Sand-mud

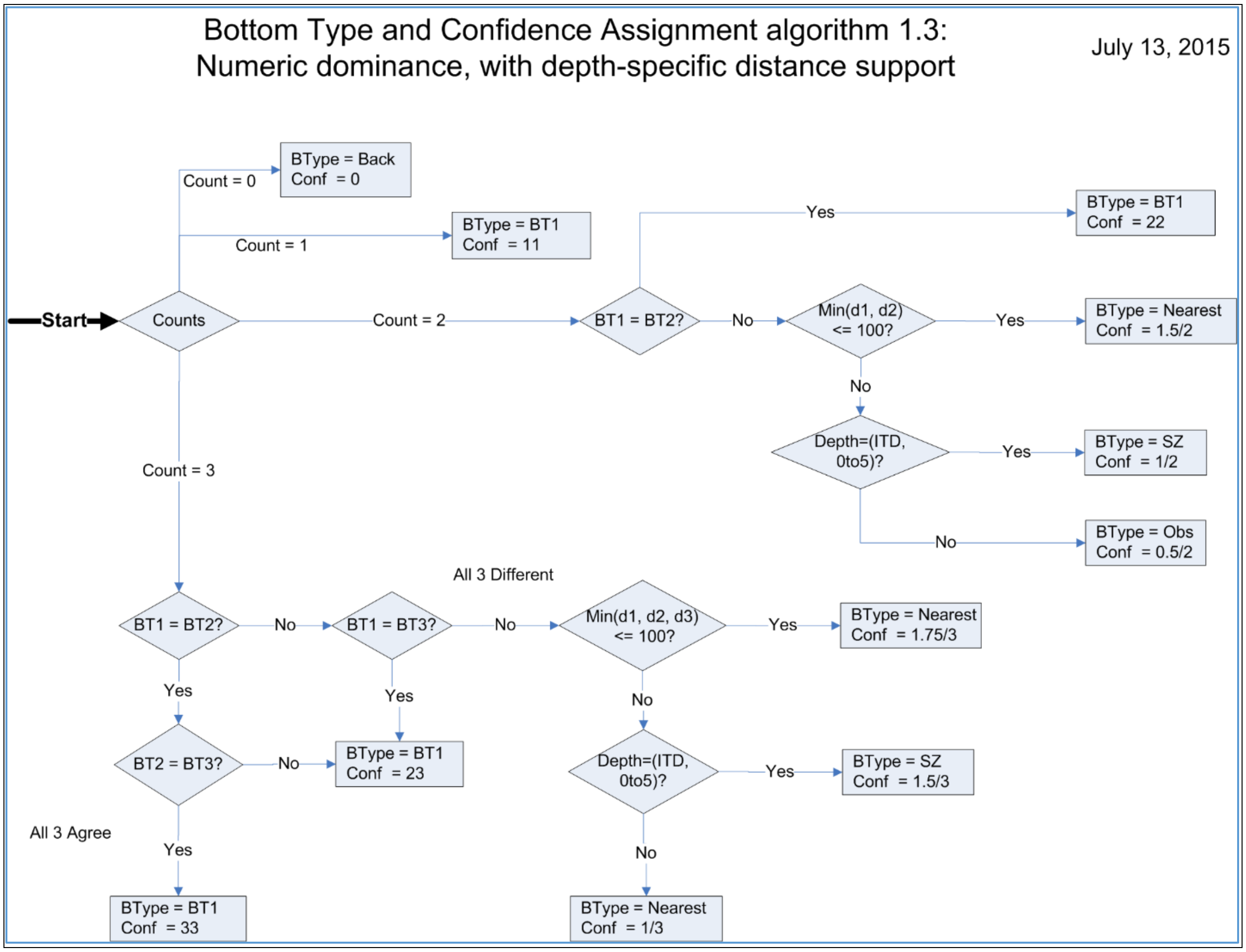
13.2 A2. Description of scripts used in the Bottom Patch (BoP) production and data preparation (from Gregr and Peterman 2022).

Script Name	Description
Herring_Summarize.py	Single use script. Run on the entire Herring database. Path to Access file hard-coded in script. Results are placed in the same Access file and named *Herring_Date_Time.* Takes data compiled into MSAccess table and organises the recorded field samples by transect and depth ribbon, defining StationCount, BTypeUnique, BTypeDominant, and the top 3 BTypes recorded for the group. This summary table is then joined back to the transects. The summary data are used by BType_Phase 1 processing to assign BoP BTypes using an updated (Aug 2015) lookup table. This improves on and refines the earlier BType assignment done using SQL queries in the Access file. Invalid BTypes are removed, leading to some data loss.
Do_shorezone.py	Single use script. Run on the ShoreZone data by region, after passing the full database through the Phase 1 script for BType assignment and regional feature selection. Steps implemented in the **doShoreZone_v10x.py** script: 1. Create point list using the *Feature_Vertices_to_Points* tool in the ArcGIS Data Management toolbox. 2. Create Thiessen polygons (TPs) from this full set of points, retaining ALL attributes 3. *Dissolve* the vertex-based TPs into Z unit TPs using on the PHYIDENT field. 4. *Join* the dissolved geometry with the undissolved TPs to recover the attributes (only the first match is joined which is ok because the PHYIDENT pieces are duplicates). 5. Remove TPs with no data (e.g., missing CoastalClass or Exposure fields) and repair geometry. 6. Merge to retain only the relevant attributes and ensure they conform to the format specified for the WorldClass processing script (e.g., Rename PhyIdent to SourceKey).

	7. Clip the cleaned SZ polygons to the ITD and 0to5 depth ribbons, thereby breaking up the SZ TPs into fragments, each of which is a potential BoP. Convert the fragments to SinglePart, allowing for correct selection. Repair geometry. 8. Add DepthCode field (Text, 12) to each file; populate with "ITD" or "0to5" as appropriate.
BType_Phase1_v1.4.py	We used this script to examine and validate the data contained in each source data set. This script reviews projections, validates field names, and confirms the existence and content of the necessary fields. Best practice was to pass all the source data through Phase 1 to ensure downstream processing steps did not fail. Using parameters in an Excel spreadsheet (*BTypeLookup.xls*), translates specified source data sets into BType compatible attributes. Script processing is controlled using parameters set in a separate sheet in the MS Excel workbook. During the final round of revisions to the methods, the standardization of the shellfish observations was moved to the pre-processing/merged with the spatialization script to simplify the processing of these data(see below).
herring_spatialize.txt	The Python code to spatialize the herring data runs best inside a Python window within an open ArcGIS session. This code is included as an appendix in Gregr and Peterman (2021). Memory management seemed to be an issue when run as Python code from a DOS command window.
shMasterController1.7.py shellsupport.py	Disaggregates the shellfish point data according to depth ribbon.
BType_Phase2_v1.42.py	All source feature classes are defined at the top of the script.

13.3 A3. Bottom patch and confidence decision tree

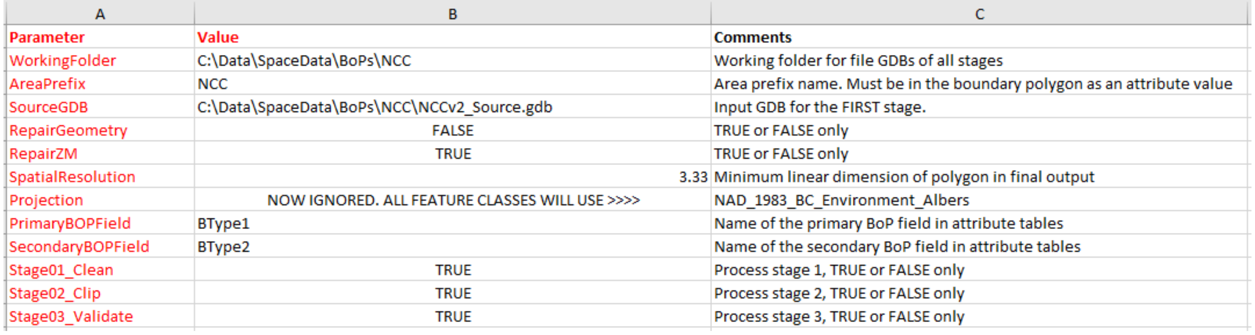
The assignment of bottom type attributes and confidence is based on the numerical dominance of the bottom types of the source data, and is conditioned with several rules related to depth and distance.



13.4 A4. Example configuration work sheets for bottom patch processing

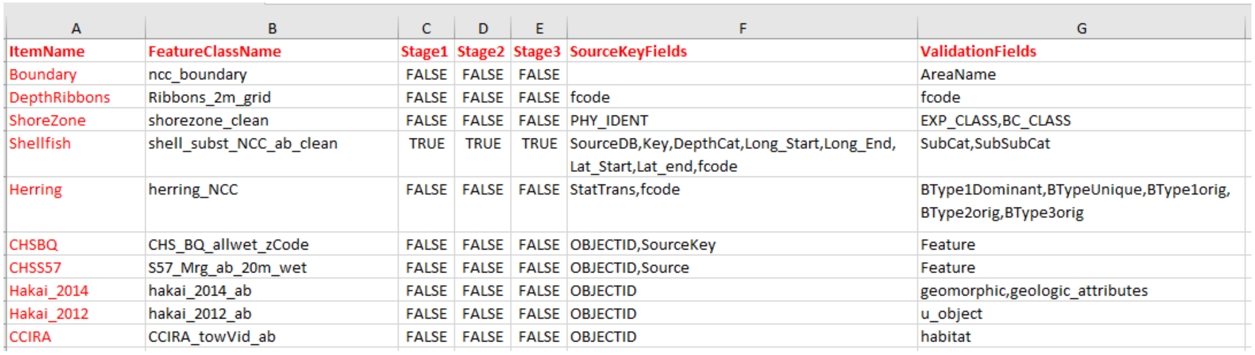
The parameters used to control the bottom patch scripts are maintained in an MS Excel configuration file, which also includes the substrate data lookup tables (Appendix A1). The parameters identify a) data sources and selected processing tasks, and b) the steps to be applied to each data source.

a) the model parameter sheet showing the configuration parameters for the Python scripts.



A	B	C
Parameter	**Value**	**Comments**
WorkingFolder	C:\Data\SpaceData\BoPs\NCC	Working folder for file GDBs of all stages
AreaPrefix	NCC	Area prefix name. Must be in the boundary polygon as an attribute value
SourceGDB	C:\Data\SpaceData\BoPs\NCC\NCCv2_Source.gdb	Input GDB for the FIRST stage.
RepairGeometry	FALSE	TRUE or FALSE only
RepairZM	TRUE	TRUE or FALSE only
SpatialResolution	3.33	Minimum linear dimension of polygon in final output
Projection	NOW IGNORED. ALL FEATURE CLASSES WILL USE >>>>	NAD_1983_BC_Environment_Albers
PrimaryBOPField	BType1	Name of the primary BoP field in attribute tables
SecondaryBOPField	BType2	Name of the secondary BoP field in attribute tables
Stage01_Clean	TRUE	Process stage 1, TRUE or FALSE only
Stage02_Clip	TRUE	Process stage 2, TRUE or FALSE only
Stage03_Validate	TRUE	Process stage 3, TRUE or FALSE only

b) the data sheet showing the source feature classes and the stage of processing to apply.



A	B	C	D	E	F	G
ItemName	**FeatureClassName**	**Stage1**	**Stage2**	**Stage3**	**SourceKeyFields**	**ValidationFields**
Boundary	ncc_boundary	FALSE	FALSE	FALSE		AreaName
DepthRibbons	Ribbons_2m_grid	FALSE	FALSE	FALSE	fcode	fcode
ShoreZone	shorezone_clean	FALSE	FALSE	FALSE	PHY_IDENT	EXP_CLASS,BC_CLASS
Shellfish	shell_subst_NCC_ab_clean	TRUE	TRUE	TRUE	SourceDB,Key,DepthCat,Long_Start,Long_End, Lat_Start,Lat_end,fcode	SubCat,SubSubCat
Herring	herring_NCC	FALSE	FALSE	FALSE	StatTrans,fcode	BType1Dominant,BTypeUnique,BType1orig, BType2orig,BType3orig
CHSBQ	CHS_BQ_allwet_zCode	FALSE	FALSE	FALSE	OBJECTID,SourceKey	Feature
CHSS57	S57_Mrg_ab_20m_wet	FALSE	FALSE	FALSE	OBJECTID,Source	Feature
Hakai_2014	hakai_2014_ab	FALSE	FALSE	FALSE	OBJECTID	geomorphic,geologic_attributes
Hakai_2012	hakai_2012_ab	FALSE	FALSE	FALSE	OBJECTID	u_object
CCIRA	CCIRA_towVid_ab	FALSE	FALSE	FALSE	OBJECTID	habitat