2D Seismic Cutline Ecological Recovery Mapping

SUMMARY REPORT

Submitted to:

Jacqueline Howard
BC Energy Regulator
Environmental Management Specialist
2950 Jutland Road
Victoria, BC V8T 5K2

Submitted by:

Caslys Consulting Ltd.
Unit 10 – 6782 Veyaness Road
Saanichton, B.C., V8M 2C2

Contact: **Luanne Richardson**
Tel: (250) 652-9268; Fax: (250) 652-9269
Email: lrichardson@caslys.ca

March 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 BACKGROUND

Northeastern British Columbia is the primary oil and gas producing region within the Province and the landscape includes seismic cutlines of various ages resulting from decades of exploration activities. The BC Geographic [Data] Warehouse has publicly available spatial data that delineates the locations of seismic cutlines within the Province of BC. Although the seismic cutlines are mapped with adequate location accuracy and precision, there is minimal attribute data that defines the age of the cutline, or the current status of ecological recovery within the cutline. Some cutlines have also been superseded by alternate activities such as pipelines, roads, forest harvest, or ecological restoration efforts; while others may have been impacted by fire or experienced natural forms of regenerative shrub or tree canopy recovery.

The BC Energy Regulator (BCER; formerly the BC Oil and Gas Commission) contracted Caslys Consulting Ltd. (Caslys) to enhance the seismic cutline dataset by adding ecological recovery attributes and noting various alternate land uses that can impact recovery. In doing so, this dataset will assist with facilitating the understanding of the characteristics of seismic cutlines; helping to measure the current level of ecological recovery and allowing for improved restoration planning in the future.

To successfully complete this update, we addressed the following objectives:

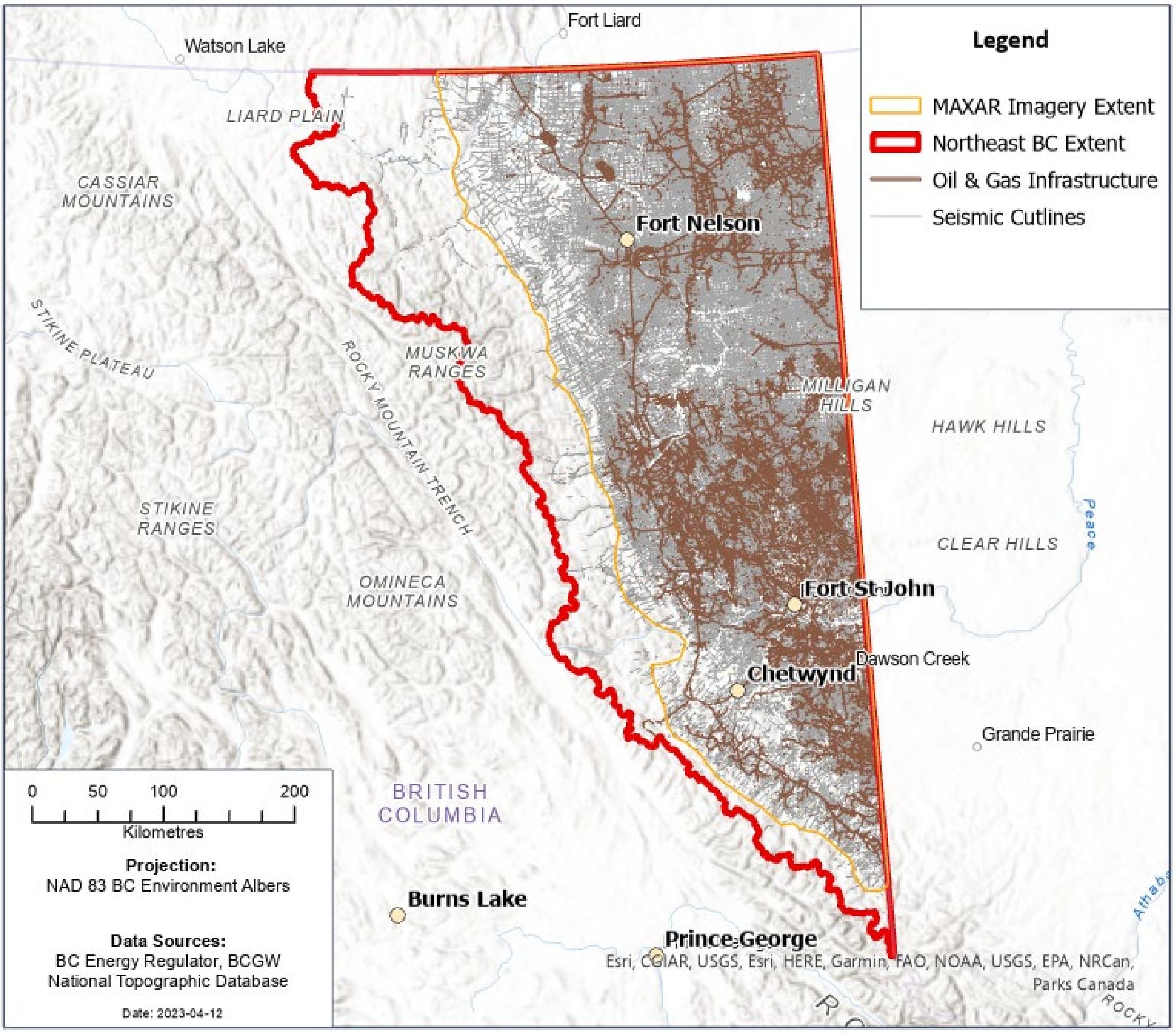
1. Developed a mapping process that allows for the update of seismic cutlines with alternate land use characteristics, cutline width, and ecological recovery attributes;
2. Assessed attributes of seismic cutlines for individual segments of each cutline, generally breaking the lines at locations where land use or adjacent vegetation cover also changes; and
3. Conducted a significant amount of manual interpretation with the best available high-resolution imagery as the basis for evaluating seismic cutline characteristics.

1.1 Study Area

The study area for this project (Figure 1) is defined generally as the foothills and relatively flat regions where oil and gas activity has been active within Northeastern British Columbia. This region is centred on the lands surrounding the Fort Nelson, Fort St. John and Chetwynd communities, and is bounded by borders with Alberta, NWT, and the Rocky Mountain ranges on the west. Caslys worked with imagery that BCER purchased to cover the vast majority of this region. Caslys assisted with the selection of best-available MAXAR high-resolution satellite imagery for the area depicted on the Study Area Map (Figure 1). Caslys was provided with access to this imagery for the duration of the project. Our interpreters also have access to a streaming service that includes similar quality MAXAR data for the cutlines that fall outside of the purchased MAXAR data extent. The streaming data serves the same purpose, but can draw slower, and was therefore not practical as a freely available dataset across the full study area. Notice that only a small portion (less than 1%) of seismic cutlines exist west of the purchased data extent.

More specifically, the study area for this project is defined by the seismic cutline dataset available from the BC Geographic Warehouse (TRIM Miscellaneous Lines) and is focused on seismic cutlines that originate from 2D seismic survey programs. More modern 3D seismic survey programs are out of scope for this project due to their different footprint density and differences in ecological recovery.

Figure 1. Study Area Map



1.2 Source Data

1.2.1 Seismic Cutline Data

The original source of seismic cutlines for this project were derived from the BC Geographic Warehouse (BCGW) TRIM dataset entitled “TRIM Miscellaneous Lines”, comprised of features mapped to the Provincial 1:20,000 scale standards for linear features that are not captured in other map layers such as roads, powerlines, pipelines, etc. This dataset was processed by BCER to split the lines at intersections, as well as at additional breakpoints such as where Vegetation Resource Inventory (VRI) polygons indicate changes in vegetation patterns. This process creates finer details in the data with shorter segments of lines so that ecological recovery can be assessed with an appropriate resolution. This approach splits lines to account for many of the vegetation, forest harvest or forest fire occurrences along a single seismic cutline. Using the files provided by BCER, Caslys completed a number of additional processing steps to refine the input data and finalize a set of over 1.6 million line segments for ecological recovery assessment and attribution. This is covered further in Section 2.2.

1.2.2 Alternate Use Datasets

When assessing seismic lines for their state of regeneration (or ecological recovery), consideration needs to be given for cutlines that have been converted for other use, or where a superseding disturbance has taken place. This is described as 'Alternate Use' for attribution purposes in this project. In situations where an alternate use is noticed in the image interpretation process, the remediation obligation is likely assumed by the new use or party responsible for the disturbance. A common example of seismic line conversion is road construction, where a seismic line is cut, and before it is remediated, a road is built using the cutline as a right-of-way, and potentially increasing the right-of-way width and/or disturbance footprint. The road is the latest use for the area so the road tenure assumes the remediation obligation. Forest fires are an example of a superseding disturbance because once a wildfire burns through an area, the BC Wildfire Service assumes the silviculture obligation for remediating the whole of the fire area. The remediation of a seismic line within a larger burned area needs to be performed with the rest of the burn area in mind.

To assess alternate use, any potential overlaps were identified through geoprocessing, and final decisions on alternate use codes were performed by interpreters. Identifying potential alternate use overlaps helped to provide additional information in the 'clustering process' (see Section 2.9) to group seismic lines with similar characteristics together. These potential overlaps were determined through a geoprocessing approach that compared the seismic line features against relevant vector data that can be found in the table below.

Table 1. Alternate Use Datasets

Alternate Use Type	Data Source file names *
Roads	WHSE_FOREST_TENURE_FTEN_ROAD_SECTION_LINES_SVW WHSE_MINERAL_TENURE_OG_ROAD_SEGMENT_PERMIT_SP WHSE_BASEMAPPING_GBA_TRANSMISSION_LINES_SP
Recreational Use	WHSE_FOREST_TENURE_FTEN_RECREATION_POLY_SVW WHSE_FOREST_TENURE_FTEN_RECREATION_LINES_SVW
Forestry Cutblocks	VEG_CONSOLIDATED_CUT_BLOCKS_SP
Oil and Gas Infrastructure	OG_SURFACE_LAND_USE_NONGEO_DEC2021 (from BC Energy Regulator)
Hydro Transmission Lines	WHSE_BASEMAPPING_GBA_TRANSMISSION_LINES_SP
Forest Fires	PROT_CURRENT_FIRE_POLYS_SP WHSE_LAND_AND_NATURAL_RESOURCE_PROT_HISTORICAL_FIRE_POLYS_SP WHSE_LAND_AND_NATURAL_RESOURCE_PROT_CURRENT_FIRE_POLYS_SP
Burn Severity	WHSE_FOREST_VEGETATION_VEG_BURN_SEVERITY_SP
Agricultural Land Use	ACI_2021_bc_v2 (Annual Crop Inventory 2021, Canada Open Government Portal)

* All datasets sourced from the BC Geographic Warehouse unless otherwise noted.

Although final alternate use attributes are general assigned by image interpretation, having preliminary alternate use attributes assigned to each cutline segment (based on the features in Table 1) increases the efficiency of manual data capture. For example, an interpreter can be assigned a batch of data that corresponds to segments that all parallel mapped road features, or all fall within mapped recent cutblocks. As the user progresses through the records, they can focus on the simple decision of identifying if the seismic cutline segment actually falls along a road or falls within a cutblock, and rapidly assign the attribute codes as suitable.

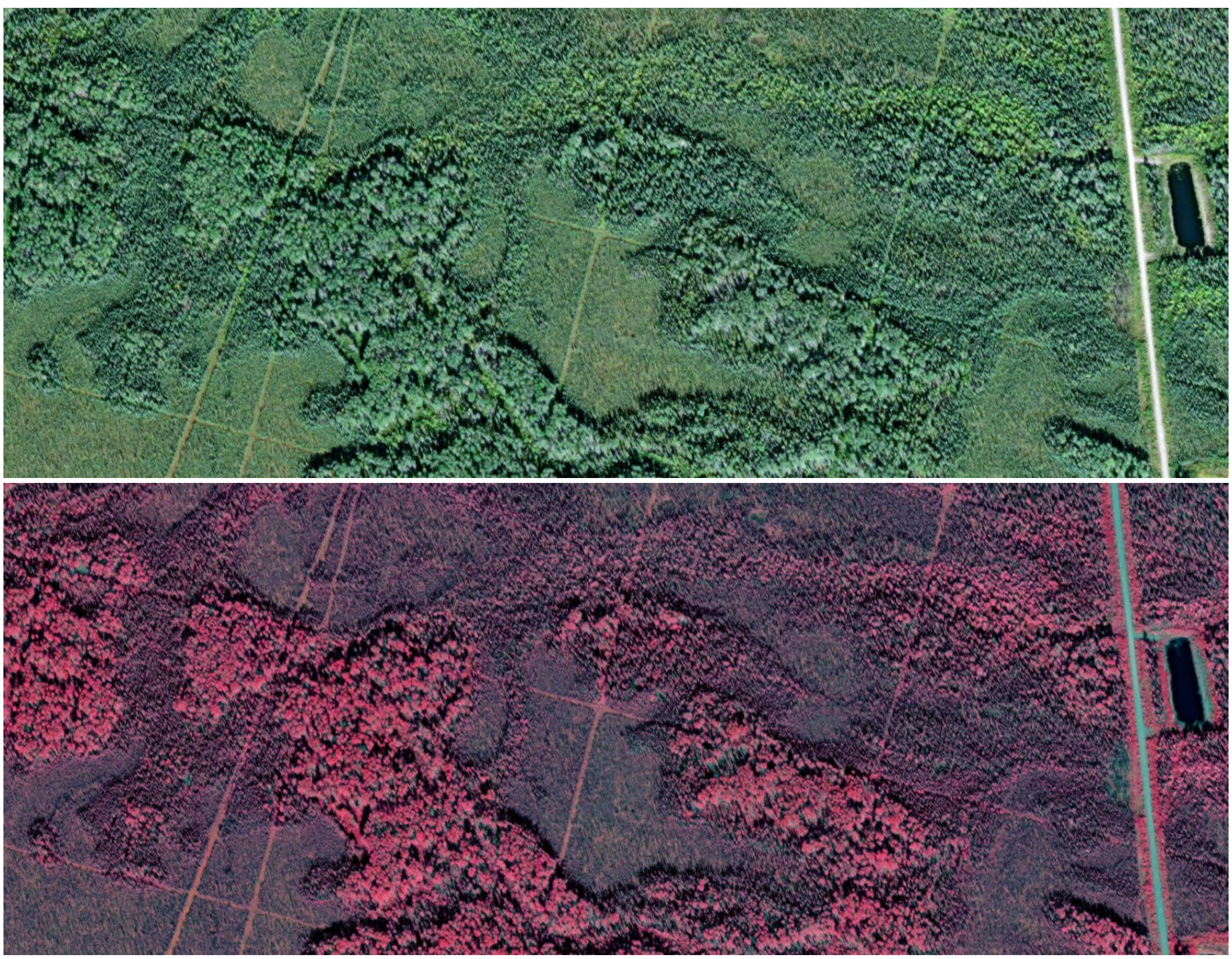
Alternate use fields can also be used to inform final attributes is a small portion of the study area where visual inspection is difficult due to cloud or haze in the available imagery, or in rare cases where the cutline is not easily discerned for other reasons.

1.2.3 High-Resolution Satellite Imagery

The project authority provided Caslys with access to the best available image tiles as supplied by a third-party vendor. MAXAR operates a constellation of satellite sensors that provide regular updates to imagery across the globe from various satellites that generally have 30-to-50-centimetre resolution over the past decade and slightly coarser imagery dating back to 2001. MAXAR offers what is called a Vivid Mosaic that includes the best available images from their archive for any given location. This mosaic is updated quarterly and balances quality based on achieving the best resolution, fewest clouds, and most recent date for the images that make up the mosaic. The dataset was current as of Dec 2022 and included less than 1% cloud cover with about three quarters of the study area captured within the last 5 years and less than 4% of the imagery originates from imagery captured more than 10 years ago. This is the best available data product from a single vendor for the study area. The next best alternative vendor product covered less than a third of the study area. The data in this MAXAR product is commonly found on websites such as Google Maps or Esri World Imagery streaming image services.

Figure 2 provides a sample of the imagery used, which was available for use with true colour (RGB) and false-colour infrared images. The false-colour infrared imagery is commonly used by photo interpreters to better discern differences in vegetation characteristics as it maintains a higher degree of variability that cannot be seen in the visible spectrum.

Figure 2. Sample MAXAR Imagery (True Colour and False-Colour Infrared)



The 2022 MAXAR mosaic is made up of the best available imagery at that point in time, weighting how recent the image data is over most other characteristics other than cloud cover. In some cases, the sun angle can be very low, or the look angle from the satellite to the ground can be somewhat side-looking. Both of these image characteristics can hinder our ability to clearly see within the cutline in treed areas as shown in Figure 3 and Figure 4.

Figure 3. Shadow

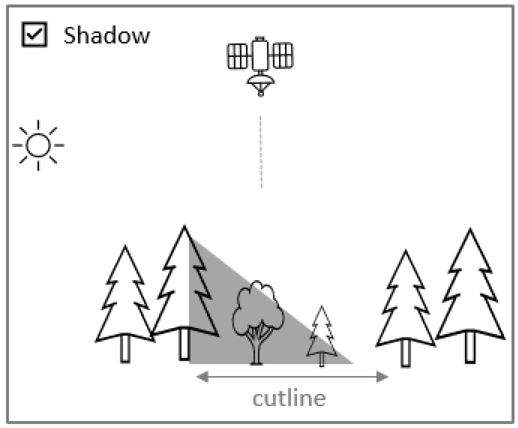
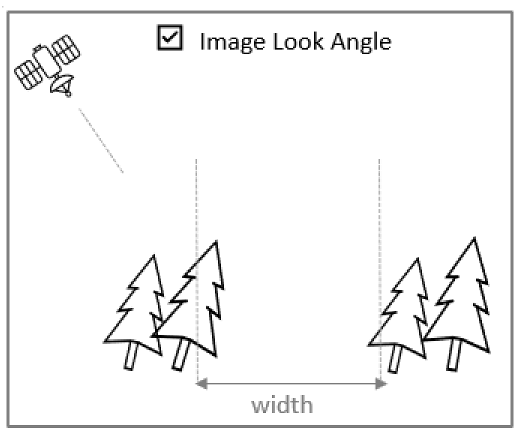


Figure 4. Image Look Angle



As a result of these issues, MAXAR provided additional imagery in the form of the 2019 Vivid Mosaic product. In some locations, this provided Caslys with a second image to review in cases where shadow or other problems limited our ability to see lower vegetation within the cutline clearly. Figure 5 shows one example of how a second image date can help with interpretation for ecological recovery assessments. The most recent image in the 2022 Vivid mosaic was captured on September 4th, 2019 (upper image); while the most recent image in the 2019 Vivid mosaic was captured on July 10th, 2018 (lower image). Although the more recent image (upper) image contains more contrast in vegetation types, it is evident that low sun angle is casting more shadows across some cutlines. The 2018 (lower) image has less shadow, but still allows a clearer view of more of the ground in certain situations.

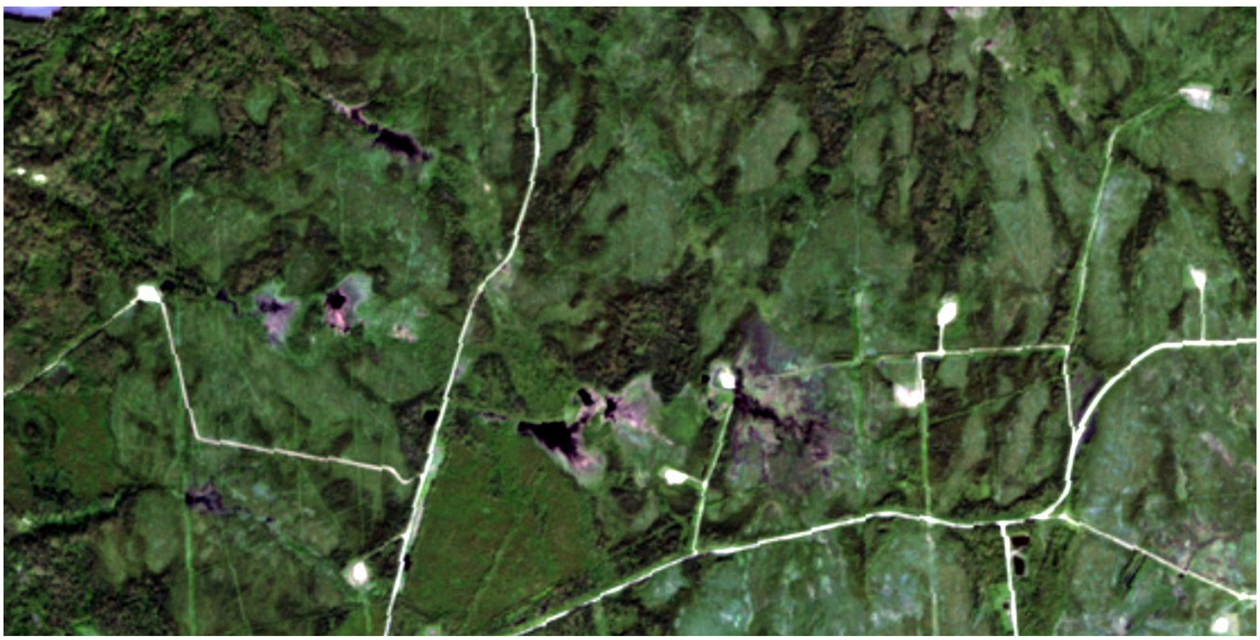
Figure 5. Sample Imagery with Different Acquisition Dates

1.2.4 Sentinel-2 Imagery

Caslys developed a 2022 Sentinel-2 satellite image mosaic with 10-metre resolution across the full study area. A sample of this image is presented in Figure 6. Sentinel-2 Imagery was used to assist with three key project tasks :

- This dataset was used primarily to split longer seismic cutlines into separate, shorter, segments in places where the vegetation transitions from one cover type to another. This helps ensure that the final cutline segments being assessed do not include longer lines characterized with different classes of ecological recovery.
- Image characteristics are extracted from both the high-resolution image mosaic and the Sentinel-2 mosaic to assign spectral characteristics to each line segment. These values help group cutlines into similar clusters that are then used to create batches of work (macro clusters) for image interpreters, and also to help locate line segments that have very similar characteristics to other line segments in the general vicinity (micro clusters). The terms micro and macro clusters are described in Section 2.9.
- Sentinel-2 imagery also provides a recent look at cutlines to assist in attribution in cases where cloud cover may be present in the MAXAR mosaic. This is a minor occurrence but could be applicable where cloud blocks a view of different alternate uses that can effectively be interpreted in the coarser imagery.

Figure 6. Sample Sentinel-2 Imagery



2.0 MAPPING METHODS

The methods applied for attributing seismic cutlines with ecological recovery are presented here with sections that outline each specific step. As an overview, the following key steps have been followed:

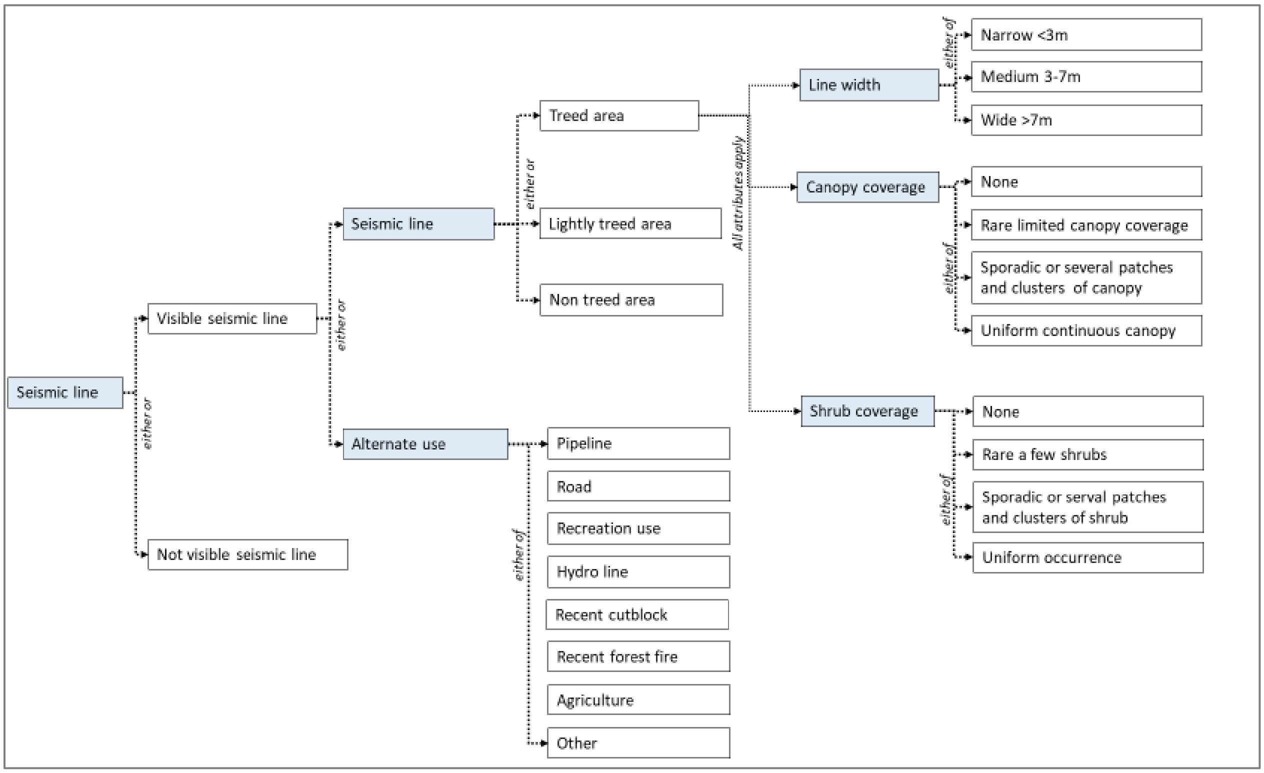
1) Seismic lines received from the BCGW have preliminary splitting and a series of attributes applied that assist with the initial step of breaking lines into reasonable (and shorter) segments where vegetation and land use patterns are expected to differ. Caslys applied additional breaks (described in Section 2.2 to refine the data further so that ecological recovery assessments apply to segments that best match landscape level variability in land use or tree / shrub cover patterns. These segments have unique Segment IDs. In addition to these fine level cutline segments, a spatially related version of cutlines was built that has Line IDs and grouped together parallel segments that connect to form what is considered the full cutline. Section 2.4 outlines the technique to build these lines and the purpose they serve.
2) Using these Segments and Lines, Caslys worked with BCER to develop a comprehensive classification structure so that attributes clearly described ecological recovery within the cutlines, and also identified alternate land uses and cutline widths in a manner that best supports restoration planning and other anticipated uses of this dataset. Section 2.1 highlights the mapping schema or class framework that was used by our team of image interpreters to assess each cutline segment.
3) Section 2.5 highlights the ancillary GIS mapping layers that were used to pre-populate and flag cutline segments regarding the likely occurrences of alternate uses that may exist. These occur where the cutline overlap mapped features such as roads, cutblocks, forest fires, and other land use layers may supersede the previously developed cutline footprint.
4) We then built high-resolution image mosaic datasets (Section 2.6) from the best available imagery supplied by MAXAR, through a data sharing agreement with BCER, and conducted manual interpretation of segment characteristics through a custom GIS workspace that streamlined the data attribution process so that hundreds of thousands of segments could be assessed in a consistently accurate and timely manner (Section 2.8), with quality controls enforced (Section 2.10).
5) During the data capture process, we also extracted characteristics of the available imagery (Section 2.7) and assigned those metrics to each cutline segment. These and other GIS-based metrics allowed us to locate macro clusters of segments (Section 2.9), that are somewhat similar in nature, that act as batches of work assigned to each interpreter. These work batches helped streamline manual efforts and create opportunities to reduce some manual effort when consistent attributes were encountered. Micro Clusters or spectral twins were also found, which enabled us to split the overall segment dataset into two groups – records that must be reviewed manually to determine ecological recovery attributes, and records that could be assigned attributes based on precise similarities to other manually attributed records.

Work was continuously reviewed in a statistically significant sampling approach and feedback was presented back to the group of interpreters to continuously improve calls and increase consistency across the full dataset. The data capture environment also contained automated checks to ensure that users entered valid attributes and an accuracy assessment summarized the quality of the overall desktop mapping approach.

2.1 Cutline Attribution Classification Strategy

Seismic cutlines were assigned attributes through this project using an interpretation guide or classification tree process (Figure 7) that included the presence or absence of a visible cutline, followed by the presence or absence of trees adjacent to the cutline. In other words, we assessed if the line was visible or not, and then since we were assessing forest cutlines, we noted areas where trees were not present adjacent to a mapped cutline and, therefore, we expected that no trees were removed at the time that seismic surveys were completed. Following these simple assessments, we reviewed the imagery in more detail to assess 'alternate uses' on or around the cutline. In the absence of any of these situations, we assessed the width of the cutlines and the level of tree and shrub coverage within the line. This decision framework is illustrated below.

Figure 7. Seismic Cutline Attribute Classes



Interpretation calls were structured to provide as much detail as possible about the condition of the cutline in terms of ecological recovery. '*Alternate Uses'* were assigned as appropriate. It is common for both pipelines and roads to be present along a former cutline; in these cases, the more pronounced or substantial disturbance is recorded. Seismic lines may be noted as '*Not Visible'* for a variety of reasons including:

- No substantial vegetation was present to see a cutline disturbance. There is some overlap with the '*Non-treed'* class which often shows some signs of the seismic survey;
- A cutline segment may cross different land uses but not include any visible tree cutting (e.g., the line segment is partially within an agricultural field and partially in an adjacent non-treed wetland; in which case it would not be appropriate to class as agricultural alone); or
- In some cases, there is full continuous tree cover within the area under the line segment that defines the cutline locations. It can be difficult to assess if the cutline was never completed or if trees have fully recovered from and older cut. In either case they may be classed as not visible.

Some seismic lines occur in areas where tree cover is naturally sparse. The non-treed class covers areas where open sparse canopy covers less than 10% of the landscape. The lightly-treed class covers areas where tree naturally cover between 10-25% of the ground – often in open bogs or fens. These classes generally identify areas where cutlines are less visible and may not see the same benefits from active restoration effort as expected in more densely treed areas.

Alternate uses are applied when the presence of that land use appears to have occurred after the cutline (or seismic survey) was completed. These include the development of other oil and gas infrastructure (often coded as pipelines but can include well pads and other features), roads, recreational access (less substantial than roads), power transmission lines, recent forest harvest cutblocks, forest fires, or agricultural land uses. Other alternate uses have been assigned in limited cases for situations where rural residential or industrial land uses occur. Although rare, (and not a land use), we also apply cloud cover coding in this manner where the cutline is not visible due to cloud in the available imagery.

In the absence of all situations described above, we then assess a cutline for its width into one of three classes:

- *Less than 3 m* wide
- *3 to 7 m wide*
- *Greater than 7 m wide*

Widths are assigned based on the expected width of the cut at the time of the seismic survey. Width can sometimes be masked by partial tree canopy recovery along the sides of the line, but the width assessment is typically influenced by review of other areas along the line where recovery has been slower and can be more easily be determined due to different vegetation types. Although cutline widths can occasionally differ along a single cutline, this appears to be a rare occurrence.

Along with width, tree canopy cover is assessed, and shrub cover is assessed where tree canopy does not fully obscure the understory. For trees and shrubs, the same classification structure is assigned as follows:

- *None (0)* – No visible coverage of trees or shrubs respectively.
- *Rare (2)* – Limited and/or very sparse coverage.
- *Sporadic or Patchy (5)* – Coverage that can include gaps and occasional clusters of trees or shrubs or denser coverage that is not full and includes some open gaps where recovery has not taken place.
- *Uniform* (9) – Coverage that is largely continuous with full canopy for tree or shrubs respectively. Note that tree species is not differentiated adjacent to or within the line, and therefore it is possible to encounter different species within a cutline with full canopy coverage.

2.2 Building 2D Cutline Segments

As described in Section 1.2.1, the cutline segments used for this project originated from the Provincial 'TRIM Miscellaneous Lines' layer available from the BC Geographic Warehouse and split at intersections and various land cover breaks based largely on the Vegetation Resource Inventory polygons by the BC Energy Regulator. Once received, Caslys completed the following additional steps to finalize the cutline segments that were used for this project:

1. Dissolving very short line segments that have been generated from previous geoprocessing steps. Caslys selected all lines shorter than 15m and merged them into the adjacent line that was most closely parallel to the short feature by taking advantage of the orientation (azimuth) attribute that was assigned to each line segment.
2. Caslys also selected all line segments longer than 500 metres and split them at locations where vegetation appears to be changing based on the NDVI characteristics of 2022 Sentinel-2 Satellite imagery. A small number of seismic cutline segments remain longer than 500 metres; only in cases where there does not appear to be any substantial vegetation change across their span.

3. Caslys also updated the line orientation attribute on this final seismic cutline segment dataset. Orientation refers to the general azimuth of the line segment which denotes the average direction that the line travels and can be used to assist in grouping cutlines that may be similar to each other.
4. Caslys also added unique segment ID attributes to allow for tracking each individual segment as the project progressed.

2.3 Flagging 3D Cutline Segments

The TRIM Miscellaneous Lines map layer includes predominantly seismic cutlines originating from 2D seismic surveys, and for the most part, does not include a substantial portion of more recent 3D seismic grid surveys. The scope of this project was to focus on 2D seismic lines which tend to be wider and generally older than their 3D counterparts. Although most 3D grids are not mapped in this data, a selection of 3D seismic survey grids are present in the source data and have been flagged by Caslys to allow them to be filtered or selected separately in future analysis. This approach allowed us to avoid manual effort assessing these lines for ecological recovery, but does leave them in the dataset, and through our semi-automated mapping process, allows a portion of the 3D lines to be attributed. The lines were flagged manually based on their layout in densely spaced grids. Lines within a 3D grid that appear to have been also part of past 2D surveys remain un-flagged in the final dataset field *ASSESSMENT_3D*.

2.4 Building Cutline Lines (groups of segments)

In general, Caslys assessed ecological recovery along shorter *segments* of each seismic cutline. One exception to this rule is the first pass assessment of seismic line width. In most cases, seismic cutlines appear to be a uniform width along their full length. Caslys applied a geoprocessing algorithm that looked at line segment connectivity and generally parallel orientation of *segments* to group segments that belong to a common *line* and pieced them together into groups of segments or longer *lines*.

These consolidated lines were used to perform a first pass evaluation of the data to evaluate cutline widths. As well, these lines were applied to certain micro clustering routines that seek similarities of ecological recovery for different segments along a common line (see Section 2.9).

Identifying these lines of connected segments is an important component of the mapping approach, which unfortunately, could not be defined by the geometry and unique line IDs of the source datasets. Instead, the follow logic was applied.

Modifying the line connectivity geometry was complex due to the size of the dataset and the amount of overlapping line features. Challenges arose in cases where the linework turned a 90-degree angle but was digitized as a single cutline. Therefore, we had to split lines in various locations and then dissolve them back into lines where the orientation of each segment was relatively similar. This task was completed using a pairwise function and generated approximately 140,000 individual lines from the over 1.6 million cutline segments. This created line features covering multiple connected segments, but breaks lines at significant turns, gaps or intersections with other lines. Updating line orientation angles at steps during this task was key to dissolving features to connect common lines, while avoiding connections at intersections and sharp turns. This allowed original segments (with unique Segment IDs) to be grouped and processed based on their relationship to longer lines having unique Line IDs.

2.5 Pre-Processing Vector Datasets

Mapping alternate use classes for cutlines can be improved upon by first flagging cutlines that likely belong to certain alternate use categories. In many cases, spatial data already exists to indicate the locations of features such as roads, power transmissions lines, pipelines, well sites, agricultural land, forest fires, cutblocks, etc. We assign flags to each segment to note the common association between these mapped features (See Table 1 and Section 1.2.2) and each cutline segment.

For example, a cutline that passes through a cutblock polygon is likely going to be assigned the cutblock alternate use attribute code. Although we only flag the data as such to put many records into clusters (or batches of work) for our interpreters, the final call is made by the interpreter to validate the finding of this pre-processing GIS analysis. Similarly, a cutline that runs in close proximity and parallel to a mapped road, has likely been superseded by the road right-of-way.

The following geoprocessing steps were completed to identify and flag these associated land uses:

- Lines that fall within polygons for forest fires, cutblocks, or agricultural areas were assigned flags for corresponding alternate uses. This is calculated as a line in polygon selection where greater than 90% of the cutline falls inside the polygon.
- All time sensitive data was compared, where appropriate, to best assess the order that development took place. For example, cutblocks and fires that had date fields were compared against lines that had the cutline dates (ADMIN_YEAR / ASRC_YEAR) populated. Any cutline without a year was assumed (with advice from BCER) to be older and assessed for overlap without time filtering. Through this processing approach, a cutline that was developed in 2005 would not be flagged with an alternate use if inside a cutblock from 1990.
- In the case of forest fires, burn severity data was queried to only include polygons with moderate to severe classes.
- In the case of oil and gas infrastructure, such as pipeline a right-of-way or well site, the polygonal reference dataset was used to overlay with cutlines in a similar manner. A cutline was considered coincident with the oil and gas infrastructure if 90% of the line fell within the polygon.
- In the case of agricultural lands, Federal Government 2021 Annual Crop Inventory was used to define simple polygons based on the aggregate of classes that were associated with rangeland or cropland (120, 122, 131, 136, 146, 153, 162, and 195), and then generalized to remove smaller polygons outside the key agricultural areas found in Northeast BC. This dataset was reviewed against the satellite imagery to ensure that the process performed in a way that was fit-for-purpose.
- Road features included source files with centrelines and some features as polygons. Centreline features were buffered to represent the full road right-of-way and then assigned as coincident features with alternate use when cutlines were at least 90% within the road polygons.
- In all cases described above, there is a significant number of features to process, and the volume of data can exceed the capacity of Esri geoprocessing functions and computer memory capacity. As a result, all functions were completed by tiling the study area inputs and processing each tile independently. Some overlap was included between tiles to remove any data gaps and results were merged back together following the geoprocessing steps.

2.6 High-Resolution Image Mosaic Dataset

The imagery data provided by MAXAR was 2.68 Terabytes in size and divided into 9,828 image tiles. As such, this dataset required structuring before it could be effectively loaded into Esri ArcGIS Pro software. The most efficient way to do this was to create a *Mosaic Dataset*. It is important to note that this is different from simply mosaicking the data together. A basic mosaic creates a new output raster that is a combination of all the input rasters. This would take multiple terabytes of additional space and have significant performance problems when trying to load the imagery. A *Mosaic Dataset* differs from a basic mosaic by only referencing the location of the source imagery. This enables rapid data refresh rates with multiple users and requires minimal additional disk space.

2.7 Clustering and Image Data Metrics per Segment

An important component of the project approach is to cluster seismic line segments into groups that have similar characteristics. In doing do, we leverage the ability to assign similar batches of work to interpreters, seek opportunities to reduce manual interpretation when those batches appear to be consistently one ecological recovery class, and locate very similar segments that can have ecological recovery attributes assigned through automated means based on their similarity to other features that have been assessed manually. To accomplish this clustering of segments, it is required to assign spectral characteristics from the imagery under each segment to the segment itself. This is achieved using both MAXAR and Sentinel-2 imagery for a wide variety of image characteristics such as NDVI and image texture values This approach is completed by buffering the cutline segments to build polygons and assess zonal statistics from the image pixels that fall under those polygons (and subsequently join back to the line features by unique Segment ID).

Each line was buffered by 15 meters on either side of the centerline for MAXAR processing, and 50 meters either side of the centerline for Sentinel-2 processing. The 15m buffer used for the MAXAR imagery was chosen to account for the full width of the cutline and the positional accuracy (i.e., error) of the cutlines relative to the image pixels where cutlines are visible. This distance is increased for the Sentinel-2 data because the pixels that capture the cutlines are larger, and the positional accuracy of the dataset is slightly poorer. The buffer distance could not simply be the width of the widest possible cutline but had to consider the potential offset caused by the accuracies of the source data. 15 meters was chosen because it provided a good balance of ensuring the pixels in the seismic lines were captured without introducing so much of the surrounding landscape to dilute the spectral characteristics within the cutline. The Sentinel-2 imagery was included to provide a spectral representation of the surrounding landscape. As a result, we generally found cutlines in deciduous forests (for example) to have common characteristics, and those in coniferous forests to be grouped separately as well.

Once the buffers were created for each segment, the zonal statistics were calculated for each buffer. The Sentinel-2 statistic was an average Normalized Difference Vegetation Index (NDVI). The MAXAR data has significantly more statistics, including spectral band averages, an average NDVI, and multiple texture models. These spectral traits of each cutline segment are fundamental to the clustering procedures described in Section 2.9.

2.8 ArcGIS Pro Manual Data Capture Environment

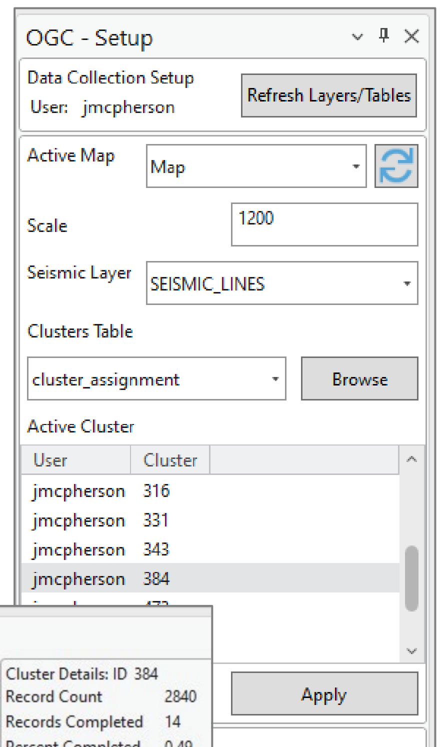
Caslys developed a custom data entry interface to allow for the following data capture requirements:

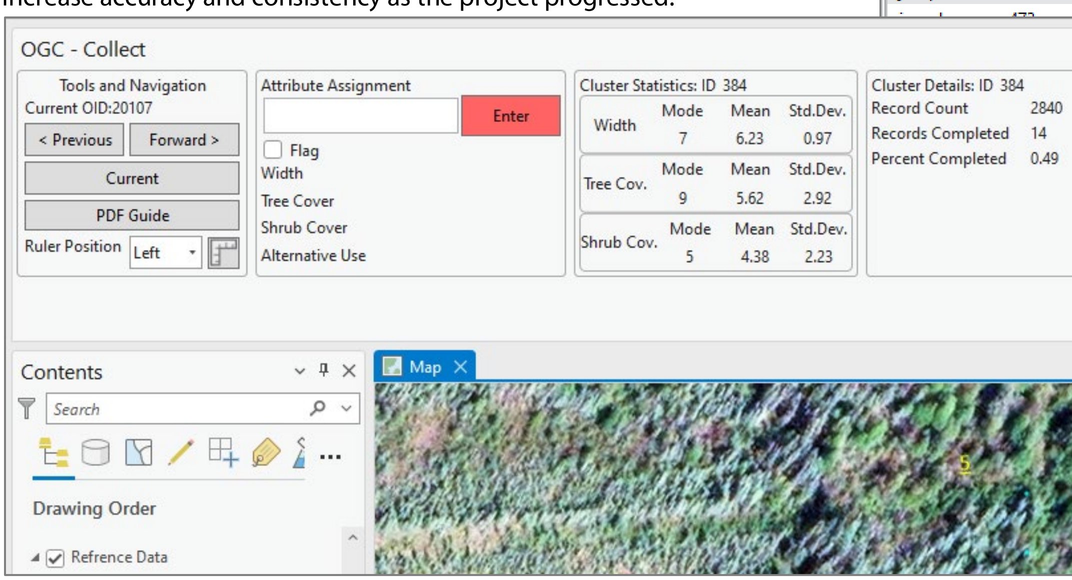
1. Assigned batches of work to individual image interpreters and manage the progress of attribute calls as they progressed through the work batch.
2. Automatically and randomly panned to the next segment in a work batch to save time scanning for segments to attribute, and to provide our team with results that are randomly completed. This

randomized approach allowed for certain decisions to be made about the similarity between features in a work batch and reduced the amount of manual work in cases where a very high level of similarity existed. For example, we may define a macro cluster (work batch) based on various spectral characteristics, line orientation, and presence of an overlapping alternate use dataset. As the interpreter progressed and found that over 90% of all calls were identical, our data administrator sometimes chose to assign a common call to all records in the batch and move on to a new cluster.

3. Added measuring scale bars with 3-metre and 7-metre diameters that drew adjacent to the cutlines to quickly allow for width assessment. This included labels on the map for image dates from both the 2022 MAXAR mosaic and the 2019 MAXAR mosaic so that a user could quickly toggle both images to make the best interpretation calls.
4. Included functions to conduct random QA/QC of ecological recovery calls and provide feedback to all interpreters to increase consistency across the full team.
5. Included rapid entry keyboard codes that required minimal clicks to record attributes; thereby increasing the number of calls that could be made per minute. We also only allowed entries to be saved if they conformed to a set database schema to avoid erroneous entries.

The seismic line recovery toolbox was developed within the Esri ArcGIS Pro environment and included a *Setup Dialog* (shown right) where the interpreter could select from batches of work (clusters) that had been assigned to them and defined the default zoom scale each time the map panned to a new location. Once setup, the interpreter used the *Collect Dialog* (shown below) and simply selected the current feature and reviewed the imagery to enter the cutline attribute codes for Alternative Uses or Width, Tree and Shrub cover, recorded as 3-coded digits. On average, our team members were able to attribute approximately 4-5 cutline segments per minute. Behind the tools, each entry was automatically committed to the spatial database with time stamp, username, and all required attribute ecological recovery fields were populated. The toolbox could be toggled into QA/QC mode where senior interpreters could randomly review a portion of an individual's work and record pass/fail and correct attributes that were shared with the team to increase accuracy and consistency as the project progressed.





Each night, the data captured by our entire team was automatically consolidated into the merged attribute database and statistics about team performance and accuracy were tracked by the senior data administrator. In doing so, we reviewed effort, provided regular team feedback through cross-training sessions, and identified challenges or opportunities to enhance data capture protocols. Through this detailed work, we are familiar with the nature of seismic cutlines in Northeastern BC, the types of landscapes where ecological recovery appears to be occurring or not, and where alternate uses have superseded the disturbance caused by prior seismic surveys.

2.9 Clustering Approaches to Develop work Batches

Through this project our team manually reviewed a significant (>500,000) number of cutline segments. The batches of work each interpreter reviewed were defined by geographic, line orientation, image spectral and texture characteristics, as well as the presence or absence of various alternate use datasets.

Several thousand of these batches or Macro Clusters were developed. Macro Clusters are defined as groups of line segments that for various reasons tend to be somewhat similar. They may have similarities that include a high percentage of common alternate use codes being applied, or they may have similar ecological recovery characteristics. In many cases, they may simply have similar ecological settings but have different levels of recovery. The intent of Macro Clusters was to align common content into batches of work that have potential to speed data capture and reduce some manual effort. Manual effort was saved as each interpreter progressed through their batch of work. If a substantial proportion of the batch was completed and the ecological recovery attribute calls were highly consistent, the data capture software flagged the Macro Cluster to the data administrator and requested that the interpreter begin a new batch of work. The data administrator then reviewed the batch, and if deemed appropriate, assigned a common set of attributes to all remaining records in the Macro Cluster. To demonstrate this by example: if an interpreter completed 250 records of Macro Cluster batch and found that over 90% (or over 225) of those records had been assigned a common attribute (for example, Alternate Use = Road), then the remaining 500 records in that Macro Cluster could be assigned an extrapolated value of Road without the need to review each record manually. Note that manual and extrapolated values are reported individually in the accuracy assessment stage of the project to define the accuracy losses associated with this process.

In addition, Micro Clusters were assigned into very small groups for line segments found along a common cutline, or within the local area that had the same image capture date. Micro Clusters are pairs or small groups of segments that have extremely similar spectral characteristics (essentially twins), as defined by those same parameters defined in Section 2.7. Upon review of these Micro Cluster (or twin) groups, we could see through manual review that each segment in the group would receive the same ecological recovery attributes in most cases. The realities of this project meant that not all segments could be manually reviewed in a timely or cost-effective manner; Micro Clusters provided a means to review one member of this group and automatically assign the same attributes to other segments that have the same characteristics. For this reason, all twins within the Micro Clusters were assigned into two categories: One or more member was assigned as a *Progenitor* (a member that had attributes assigned through manual interpretation); other *Child* records were assigned attributes of the *Progenitor* from the group. The following paragraphs define this process in more detail.

The primary way manually interpreted results were extrapolated to other records was through Micro Clusters of cutline segments. These small clusters were created through a process of spatial twinning. This twinning was performed using two primary data sources: the spectral statistics generated for each segment that was described in Section 2.7 above; and the Vegetation Resource Inventory (VRI) land cover information. This data was compiled to create an index that could be compared across seismic line segments to evaluate how similar they were. This also created the ability to tune the degree or level of twinning that would be leveraged for extrapolation. The level of constraints applied to the index values balanced the size of the groupings versus the potential of including incorrect records where twins do not have the same ecological recovery characteristics; and therefore, would be extrapolated with an introduced error.

As the constraints applied to the twinning process increased, the matches between records becomes stronger, but the size of the group is also reduced. Due to the nature of this dataset, with a significant number of records that cannot all be reviewed manually, Micro Clusters and twins have provided an effective means of reducing the manual effort, while balancing quality and accuracy of attribute calls.

It is not accurate to locate twins across the full extent of the project area because image acquisition dates and landscape characteristics vary greatly from one area to another. Instead, we assigned two tiers of Micro Clusters. The first seeks to locate twins from segments along a common cutline. The second finds twins in the general vicinity (along other cutlines) that falls within the same general area and exists on the portions of the high-resolution mosaic that were captured on the same date – having the same sun angle, atmospheric conditions, and satellite look angle.

The first tier of spatial twinning was performed on the seismic line level. As described is Section 2.4, seismic lines segments were processed to create longer line features of segments that share connectivity. By performing spatial twinning along a common line feature, segments were less likely to have a spectral and VRI variance in their attributes. This allowed the constraints of the twinning to be expanded, while still maintaining strong matches.

The second tier of spatial twinning matched segment twins across different lines, while maintaining common characteristics from common MAXAR image dates as defined in the metadata polygons that accompanied the satellite imagery files. The image metadata strips provided some geographic restrictions on the pool of seismic line segments, as well as helped to reduce any spectral variation that could happen between imagery acquisitions. This resulted in a much larger pool of potential twins and potential error so only very strong matches were considered during the area twinning. The result was an increase to the total seismic line segments captured by the twinning process, while maintaining the confidence between matches.

In both tiers of twinning, manual effort was taken to review all twins in some groups to define the threshold, or twin levels where twinning procedure still produced records with common ecological recovery attributes.

Once the micro clusters were created from the twin groupings, each cluster had a seismic line segment assigned as the *Progenitor,* and the rest assigned as *Child* records. By manually evaluating the Progenitors records, the resulting classification could be extrapolated to the Child records in that same Micro Cluster. This set of Progenitor records was the primary set of seismic line segments that were manually assessed.

Despite the effectiveness of the tiered twinning approach, there were still seismic line segments that could not be matched to a twin. These segments could have been included if the twinning constraints where relaxed, but this would compromise the quality of the results. To avoid this, the remaining seismic line segments without twins were flagged and manually assessed.

2.10 Quality Control Measures

Quality control was an integral part of the interpreter training and on-going validations. Systems were developed to have senior staff members review a percentage of other interpreters' work to correct any errors and provide feedback and direction to the team of interpreters. These quality checks were also performed between senior staff to ensure there was a common 'less-biased' understanding on how segments should be attributed.

These quality checks were performed very frequently during the early stages of the project as our team was gaining familiarity with the landscape and seismic line attributes. Although review continued throughout the project timeline, the percentage of records reviewed could be reduced as confidence in results from each team member was assessed. The QC process involved a senior interpreter randomly assessing several hundred segments attributed by a staff member and either approving the code or rejecting it and providing a correct code. These records were then reviewed with the interpreter one-on-one and discussions took place with the full group to help correct any errors and increase consistency across the team.

Staff meetings with the entire interpretation team were held approximately bi-weekly to review any common challenges, discuss any changes to the interpretation codes discussed with the client, and reinforce a common understanding of segment attribution. These meetings also provided an opportunity to discuss edge cases and develop a collective approach for how less common situations should be attributed.

Once the final attribution of seismic lines was completed, a quality control measure was implemented to review all early records collected by every staff member. The initial QC process did involve an intensive review of staff attribution when they first started but it did not involve looking at every record they attributed. This final step reviewed all records attributed in the early learning stage of the project to ensure that biases and errors made early in the project were fully re-assessed.

2.11 Final Data Consolidation

Once all required segments had been attributed, final steps could be taken to extrapolate that data across all the remaining segments and generate the final geodatabase. This included combining line widths calls, using QA/QC data where applicable, rather than the initially interpreted attributes, and extrapolating the progenitor data across both macro and micro clusters.

Once the width assessment was complete on the line features, it was then joined back to the master seismic line segments layer using the unique line ID. This connected all the segments that shared a line to the width attribute given to that line. This was then used to populate the final width attribute.

We implemented improvements that resulted from the QA/QC process to ensure that extrapolated records were derived from the updated calls.

When performing the extrapolation, consideration needed to be given to the order the micro and macro clusters were applied. Because of the nature of the statistics used in the macro clustering and the twinning process used in the micro clustering, the extrapolation took place in reverse order of the cluster assignment, starting with macro cluster, then micro clusters using area twinning, and finally micro clusters using line twinning. This ensured that calls made through the micro cluster along a single line were given the highest priority, micro clusters in the same vicinity were given the second highest priority, and macro cluster approaches were given the lowest priority since they tended to include a lower level of reliability.

It is important to note that segments identified as 3D seismic grid lines were included in the clustering for extrapolation but not part of the manually assessed progenitors (as they were considered out of scope for manual efforts on this project) or manually assessed non-twinned data. This resulted in 3D flagged attributions being fragmented, some containing no attributes, and others containing full or partial attributes. These features are included in the final deliverable but can be filtered out as they are not assessed fully through this project.

Macro cluster extrapolation was performed to assign attributes to null features in clusters that reached statistical homogeneity when being assessed. GIS summary statistics functions were used to generate the modal values for each macro cluster that did not reach 100% completion. These values where then joined back to the segment database using the macro cluster unique identifier and all null values were assigned the model value or the appropriate alternate use or ecological recovery attributes respectively.

The micro clustering extrapolation was performed next with the area-based twinning followed by the line-based twinning. Summary statistics were generated for each cluster ID and the representative code from the progenitor was applied to the rest of the records in the area and line micro clusters respectively.

The result of the extrapolation is a final dataset that contains either a width and cover call or an alternate use code for every record, excluding features identified as 3D. A final pass of this dataset was performed to ensure all the attributes had the correct formatting. Some examples were ensuring all forest fire calls used 'ff' rather than 'FF' or 'Ff', as well as removing any line widths assigned to features with no cover information.

2.12 Accuracy Assessment Methods

With the final dataset created and all extrapolation records completed, an accuracy assessment was performed from over 10,000 randomly selected records to evaluate the effectiveness of both the manual interpretation and the extrapolated records. This assessment was done using a stratified random sample of 40% manually assessed records and 60% extrapolated records to eliminate biases associated with fields that we populated without direct manual review. The accuracy assessment was performed by our experienced senior staff members.

The results (Figure 8) show an overall accuracy of 83.6%, with 8.0% being classified as minor errors and 8.4% classified as more significant errors. Minor errors were circumstances where the class was incorrect, but close to being considered correct; an example of this would be a line being classified as 502 when it is an active pipeline or hydro right of way. The provided code is correctly describing the right of way but is still incorrect in this example because pipeline alternate use is the more correct classification. Other errors were defined by situations where the coded information is clearly not accurate. This is most commonly due to interpreter error or extrapolation error.

The accuracy assessment also found that manually assessed records averaged 88.8% correct and extrapolated records averaged 80.1% correct. This falls in line with expectations where extrapolation will introduce more error than a manual evaluation, but not in a manner that we expect will greatly impact restoration planning and other decision support tasks. The split between minor errors and more significant errors is approximately 50/50 independent of the manual or extrapolated attribution process.

The following chart (Figure 8) shows the accuracy assessment breakdown by classification. Where widths and ecological recovery values are applied, these values have been grouped based on widths, and *Low* versus *High* vegetation recovery codes. Low is a group of codes that generally have minimal tree or shrub recovery; while

'High' is a group of codes that generally have more substantial tree and shrub cover within the cutline. 3, 5 and 7 refers to the cutline width class codes below:

- *Less than 3 m* wide (coded as 3)
- *3 to 7 m wide (coded at 5)*
- *Greater than 7 m wide (coded as 7)*

Figure 8. Accuracy Assessment Breakdown

When performing the accuracy assessment, interpreters were instructed to approve records where the call was reasonable, even if it may not have been the precise call they would have provided. This approach accounted for slight differences of opinion or biases, while still identifying minor and more significant errors effectively.

3.0 RESULTS

The primary result of this project is the attributed SEISMIC_LINES feature class within the supplied Esri format geodatabase that accompanies this report. Section 3.1 highlights the structure and fields contained within this dataset; Sections 3.2 and 3.3 identify some additional learnings that can be considered as this data is reviewed and used to plan restoration efforts for cutlines in Northeastern BC.

3.1 Data Schema

Table 2 presents the field and their purpose within the attributed SEISMIC_LINES derived from this project.

Table 2. Data Attribute Descriptions

Field Name	Description
OBJECTID	The feature identifier assigned by ArcGIS
Shape	This is an ArcGIS generated field indicating the type of geometry
SegmentID	This is a unique identifier used to connect features across datasets when processing
ASSESSMENT_3D	The field is identifying if a record was flagged as a 3D seismic feature. A "Y" in this field indicated that it is a 3D feature. It may be fully or martially attributed
Raw Attributes	This is the full attribute code provided to the record. It will either be a numeric code indicating width, tree cover, and shrub cover, or a letter code indicating alternate use
Alt. Use	This field contains the alternate use letter code. If null, then there is not alternate use identified
Width	This is the width class of the line feature (3,5, or 7)
Tree cover	This is the tree cover class for the line feature (0,2,5, or 9)
Shrub cover	This is the shrub cover class for the line feature (0,2,5, or 9)
Shape_Length	This is an ArcGIS generated field indicating the geometry length in meters

3.2 Image Date

Section 1.2.3 mentioned the availability of 2022 version MAXAR mosaic and the additional 2019 version MAXAR mosaic provided to allow for a second view in areas where satellite image characteristics are less ideal. Throughout the project, our team was able to assess the added value of the older dataset, which included different imagery for about 25% of the study area. This older imagery sometimes had less shadow, which can decrease the amount of shadow in some cutlines. In these areas where different (older) images were present in the 2019 version mosaic, we found that only ~15% of our calls (based on a sample) were found to benefit from the older mosaic product. In most of these cases, the benefit was not significant, but still an improvement to see more clear pixels within the cutline. Overall, this dataset does not greatly improve the overall quality of the final dataset, but it may assist in further review. The 2022 version mosaic is resampled to 30cm resolution and includes RGB and Infrared channels, while the 2019 version mosaic was created in an era when 50cm resolution was the standard and no infrared channel has been included.

You can refer to the image strip metadata shapefiles that are included with the imagery files, as delivered by the image vendor, as these polygons can be used to understand the precise image acquisition dates for all areas of the mosaic, and also include the sun angle and look angle of each component image. This dataset can be used to label the mapping dates and other characteristics when reviewing the imagery in software such as Esri ArcMap or ArcGIS Pro.

3.3 Lessons Learned and Recommendations

This section highlights a series of lessons learned pertaining to the mapping approach and recommendations to mitigate impacts of these data issues when working with the data for future planning initiatives. These concepts are presented as limitations of the supplied dataset and recommendations for use of the data.

3.3.1 Limitations:

- Image dates: Although the image vendor has the best available archive of high-resolution data for this region of British Columbia, shadows are an inherent issue when reviewing data this far north of the equator where sun elevations are lower. There is specifically an issue in forested areas where cutlines trend in the east-west direction. This issue is magnified in imagery that is acquired when days are shorter and sun angles are lower. The 2019 imagery provides some improvement in certain cases, but also takes significant additional effort to review multiple images, when in many cases the data does not alter the initial call that an interpreter would make. Also, you need to keep in mind that this off-the-shelf image product provides the best value for this data purchase and is designed to include more recent and higher resolution data in many areas. It is our opinion, following this effort, that options to explore custom image purchases would increase image costs substantially and not result in a proportional improvement to cutline attribute accuracy. This being said, BCER may wish to explore additional options or custom processing samples and options with image vendors.
- Ground Truthing and Quality Assessments: The results of this project are defined through desktop analysis based on the available imagery datasets. This project is not an assessment of seismic recovery based on local ground-based field assessments. As a result, confidence in this dataset should reflect this approach. Actual ecological recovery is impacted by our ability to see vegetation in areas where imagery is less than ideal and is also restricted to interpretation at a scale of approximately 1:2,000 (or 1:1,500 in some cases where native imagery files within the mosaic are better than 40 cm resolution). Although, 1:2,000 scale is a highly accurate mapping scale, it is not a substitute for detailed assessment in the field. Satellite-based assessment does however minimize the environmental impacts associated with gaining ground access to cutlines. This data product should not be used in a one-to-one comparison with someone standing within a seismic line to assess ecological recovery. It is a landscape level planning tool for building more detailed remediation strategies on a project-by-project level.
- The accuracy assessment is based on the same desktop review of imagery and does not benefit from field verification. Errors were only identified where differences of opinion were considered to have a potential influence on the restoration planning and anticipated uses of this dataset.

3.3.2 Recommendations:

- The accuracy assessment shows a slight improvement of manually attributed records over the results of extrapolated records. Given a significant amount of additional effort to review those records would provide small gains in the final results; however, we do not consider that this would be the best use of resources to make a small improvement in accuracy.
- A more precise method of mapping ecological recovery within cutlines would involve the use of even higher-resolution (<5cm) imagery and/or LiDAR to more clearly identify details in the cutlines that are shaded in the imagery. In theory, higher-resolution aerial imagery could have increased pixel depth and you could see reductions of the effects of shadow, just like a person on the ground can see clearly when

in the shade. More likely, LiDAR data (which is more likely to be collected in future) could resolve issues related to image quality and would be a proven method to increase mapping quality and precision. LiDAR point-cloud data can be used to map ecological recovery without manual interpretation, and instead rely on automated classification techniques that are proven to be accurate. As this path forward is restrictive due to costs and data volumes for the entire study area, it still would provide a suitable method for evaluating the accuracy of this dataset more precisely. A small pilot study of LiDAR data could be acquired across dispersed locations in the study area. High-resolution stereo imagery can also be used for point-cloud generation as another option with similar benefits to LiDAR.

- Machine learning algorithms provide a modern technique to leverage computer processing to map ecological parameters from imagery. Machine learning algorithms require a substantial amount of training data to generate valid results. The manually assessed results of this project are a suitably large dataset to be applied to a machine learning mapping approach and could be worth exploring to improve upon the extrapolation methods used in this project or if there is interest to revisit this analysis on an ongoing basis. This project dataset likely represents the most comprehensive cutline recovery dataset in Canada at this time.
- In the event that similar ecological recovery assessments are completed for 3D seismic grids, it is recommended that those initiatives assess cutlines in such a manner (and with a similar schema and approach) that it can be integrated with the 2D results. It may be possible to adjust or group some classes for the purposes of 3D seismic lines that tend to be thinner and could be mapped quicker with a more simplified attribute classification strategy that nests within this 2D class structure.