Land Cover Atlas for Urban and Rural BC
FVRD Summary Report

Prepared for:

Ministry of Forests, Lands,
Natural Resource Operations and Rural Development

July 2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

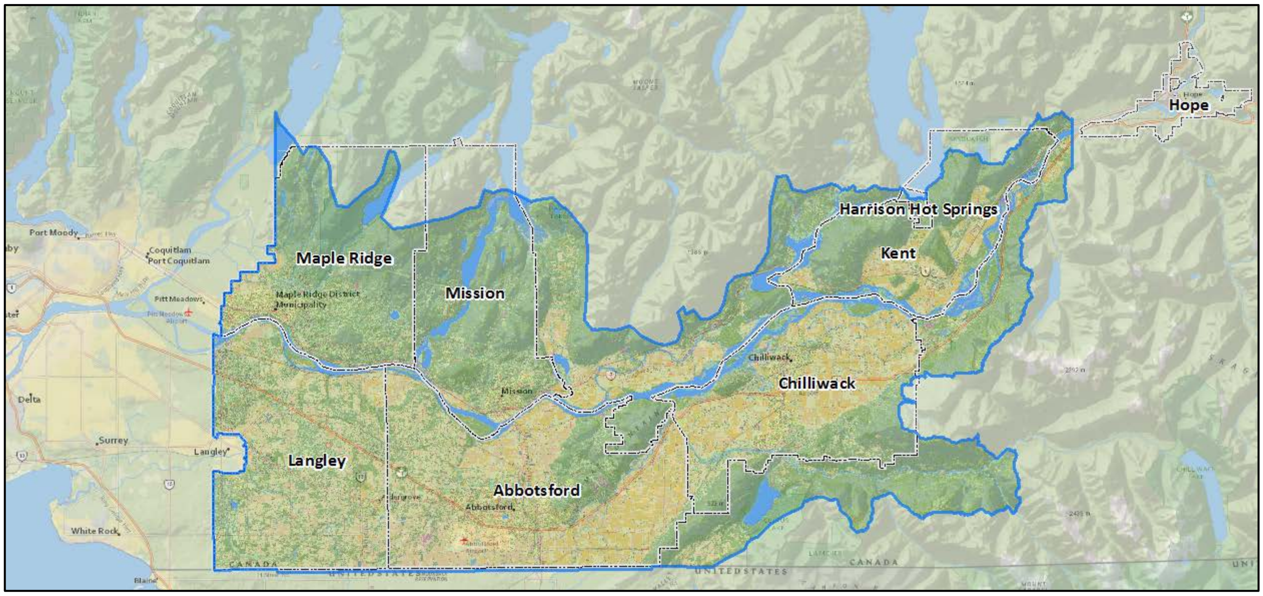
Land cover information is an essential tool for long term land use planning and resource management in the Fraser Valley. This land cover mapping project is intended to develop more comprehensive coverage of the Fraser Valley Regional District by building on the methods developed through pilot projects completed for Still Creek, Clayburn Creek and Chilliwack River watersheds (Caslys 2020). The mapping approach draws from multi-temporal Sentinel-2 satellite imagery and available provincial LiDAR datasets. A broad range of applications can benefit from this data including; carbon balance analysis, quantifying impervious surfaces to aid in water management from an ecological or engineering standpoint, monitoring landscape change and broad municipal or regional planning initiatives.

As satellite imagery becomes more accessible, the potential to generate fine-scale, current land cover products over large areas is increasingly feasible. Sentinel-2 imagery is an ideal data source for large land cover classification projects since the data are freely available, the sensor has a high revisit time, and the 10-metre resolution provides a good balance between classification detail and processing effort. Available LiDAR digital elevation model (DEM) and point cloud data provide a mechanism to enhance the spatial resolution of land cover mapping while also accurately resolving more detailed classes. For this project, over 30 land cover classes are derived from 17 core cover types at 5-metre resolution by leveraging the LiDAR point cloud data. Analysis benefits from the spatial precision of the LiDAR and the ability to explore and utilize Sentinel-2 per pixel spectral data that spans seasonal and yearly variations that help resolve land cover types. Seasonal variation is particularly useful for discerning deciduous and coniferous tree classes as well as agricultural classes. Project methods continue to draw from the random forest classification approach (Brieman 2001) used by Caslys on past watershed mapping pilot studies. Unlike the previous watershed-based pilot studies, the mapping output is increased from 10 metres to 5 metre pixel resolution and the 'Urban Mixed' Land Cover class is removed since cover types are more precisely defined at higher output resolution allowing paved driveways, lawn, shrub and treed areas to be distinguished more precisely.

1.1 STUDY AREA

The study area was defined in a manner to maximize coverage of the Fraser Valley Regional District (FVRD) while constrained by a full scene extent of Sentinel-2 imagery and the extents of the available provincial LiDAR dataset.

Figure 1. Study Area Extent



Sentinel-2 imagery is acquired along a polar orbit that trends in the north or south direction and adjacent orbit paths are collected to the east or west on different dates. The mapping approach can be largely automated through batch processes and classification training for a single date. The scope for this project was defined to maximize extents for a single Sentinel-2 scene across the FVRD. Although most of the FVRD is captured in a single scene, the areas around Hope fall outside the scene boundary. To maximize the size of the mapping area, Langley and portions of Maple Ridge (within Metro Vancouver) were included in the mapping extent. Areas of each municipal area that fall outside of the LiDAR extent are left unmapped. Although out of scope, portions of municipalities not covered by LiDAR could be mapped with a nested classification scheme to identify primary cover types in the interests of having complete coverage of the region.

2.0 SOURCE DATA

To complete the land cover classification, data was sourced from the GeoBC portal or from publicly available data access websites hosted by the municipalities. Data was also sourced from FVRD primarily for electoral areas not covered by municipalities. In some cases Open Street Map data was used to inform layers such as building footprints, airports, and railways. Source data included vector data, LiDAR data, and satellite imagery. The details for each data source are outlined in the following sections.

2.1 Satellite Imagery

The Copernicus Sentinel program is a wide swath, high-resolution satellite mission undertaken by the European Space Agency to monitor vegetation, soil, water cover, inland waterways and coastal areas. The constellation consists of two polar orbiting satellites located in the same sun-synchronous orbit. The return period in the mid-latitudes is 2-3 days, and the Satellites started capturing reliable and standardized format data data in 2016. The Multi Spectral Instrument (MSI) on-board collects 12 bands of varying resolution. The blue, red, green, and near infrared bands have a resolution of 10m, the Red edge 1-4, and SWIR 1, 2 are produced at a resolution of 20m, and the aerosol and water vapour bands have a 60m resolution. In this analysis the blue, green, red, NIR, and SWIR1 bands were used and resampled to 5m resolution to integrate with the higher resolution LiDAR dataset. Key inputs are shaded grey in the Sentinel-2 band list below (Table 1).

Table 1. Sentinel-2 Image Bands

Band	Description	Resolution (m)	Wavelength (nm)
1	Aerosols	60	442.3 – 443.9
2	**Blue**	**10**	**492.1 – 496.6**
3	**Green**	**10**	**559 – 560**
4	**Red**	**10**	**665 – 664.5**
5	Red Edge 1	20	703.8 – 703.9
6	Red Edge 2	20	739.1 – 740.2
7	Red Edge 3	20	779.7 – 782.5
8	**NIR**	**10**	**833 – 835.1**
8A	Red Edge 4	20	864 – 864.8
9	Water Vapour	60	943.2 – 945
11	**SWIR 1**	**20**	**1610.4 – 1613.7**
12	SWIR 2	20	2185.7 – 2202.4

The following Sentinel images listed in Table 2 were acquired, and used in the analysis. The images were selected with two priorities in mind; images spaced evenly throughout the growing season, and a lack of clouds over the study area.

Table 2. Sentinel-2 Imagery used in classification

ID	Date	Tile	Satellite
L2A_T10UEV_A019644_20190327T191016	2019-03-27	T10UEV	2A
L2A_T10UEV_A010807_20190401T191440	2019-04-01	T10UEV	2A
L2A_T10UEV_A011665_20190531T191656	2019-05-31	T10UEV	2B

L2A_T10UEV_A012380_20190720T191856	2019-07-20	T10UEV	2B
L2A_T10UEV_A021503_20190804T192036	2019-08-04	T10UEV	2A
L2A_T10UEV_A021646_20190814T191644	2019-08-14	T10UEV	2A
L2A_T10UEV_A022790_20191102T191543	2019-11-02	T10UEV	2A

2.2 LiDAR Data

Light detection and ranging (LiDAR) is an active remote sensing technique whereby laser light is transmitted towards a feature of interest. The amount of time that light takes to reflect back to the sensor is used to measure the distance of that feature from the sensor. The resulting dataset is known as a point cloud and from it numerous other datasets can be produced for various applications. From April 2016 to July 2017, LiDAR data was captured for the Fraser valley by the National Disaster Mitigation Program (NDMP). The resulting point clouds had an average point density of 12 pp/m^2 and were processed into a digital elevation model (DEM) and a digital surface model (DSM) which is defined by the first return from each LiDAR pulse. Subtracting the DEM from the DSM also provides a canopy height model (CHM) which defines vegetation and building heights.

2.3 Ancillary Image Data

Although not used directly for mapping tasks, various air photo or online high-resolution satellite images were referenced to interpret cover types when defining training sites for the image classification process. This included colour 10cm orthophotos that were collected along-side the LiDAR dataset as well as any municipal web mapping sites with orthophotos or publicly available sub-metre resolution satellite imagery. These datasets also provide the validation for accuracy assessment tasks.

2.4 Ancillary Vector Data

For the project study area, detailed hydrology datasets and land use zoning datasets were downloaded from municipal open data portals or acquired by the project authority from selected municipalities without self-serve data portals. Other vector datasets including the digital road atlas (DRA), the freshwater atlas, and regional/ municipal boundaries were downloaded from the BC Data Catalogue operated by DataBC. The Open Street Map (OSM) initiative supplied the most suitable rail and airport datasets. FVRD provided a hydrology and land use zoning dataset to fill gaps in some municipal or electoral areas. Building footprints were extracted from the LiDAR point cloud and cross referenced against a similar approach available from NRCan. Table 3 on the following page highlights specific data sources as compiled into a single coverage of the study area for each data theme.

Table 3. Compiled Vector Data Sources

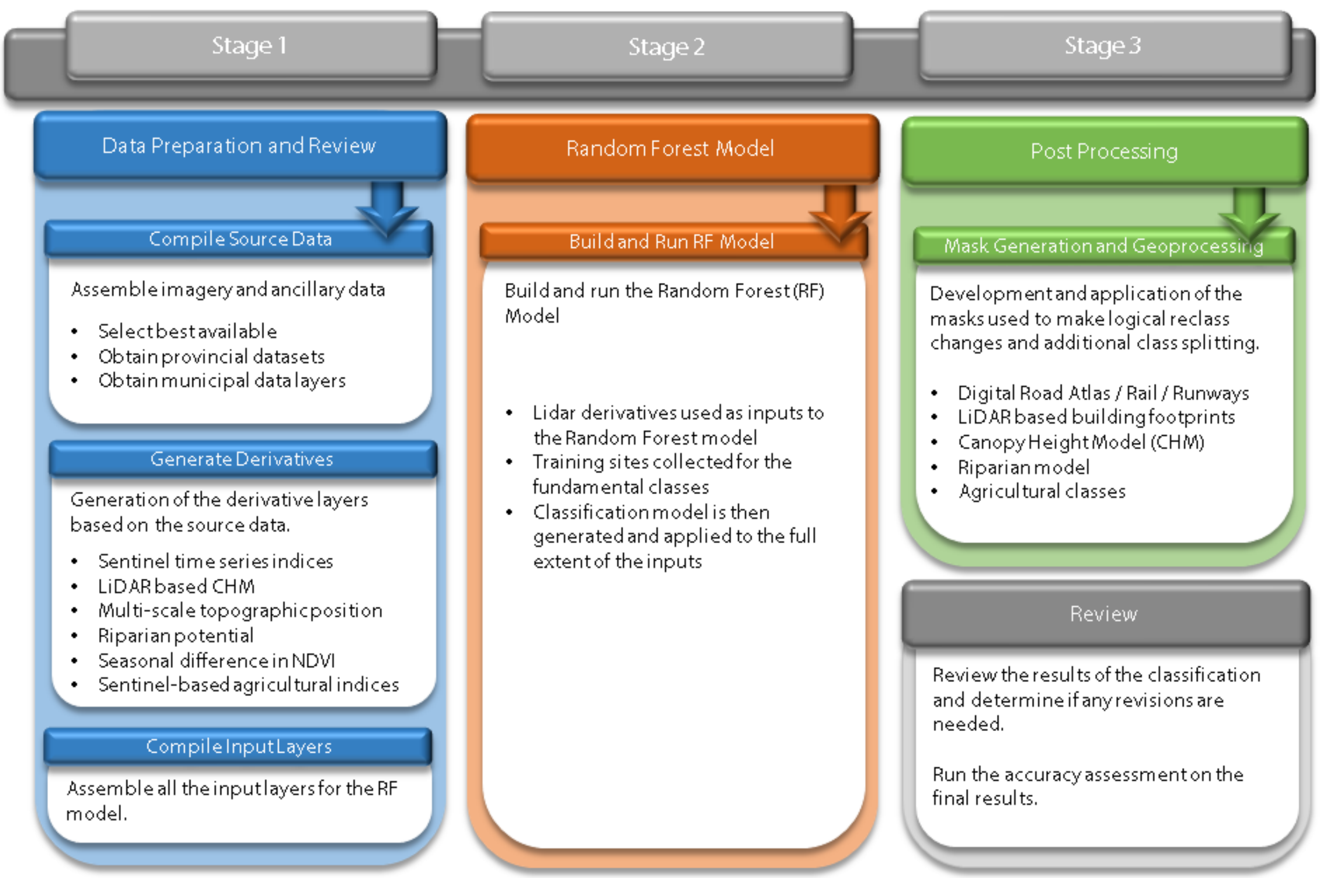
	Data Themes						
Area	**Hydrography**	**Roads**	**Rail**	**Airport Runways**	**Zoning**	**Forest Cutblocks**	**Building Footprints**
Langley	Municipal	DRA	OSM	OSM	Municipal	GeoBC	LiDAR /NRCan
Maple Ridge (south portion)	Municipal	DRA	OSM	OSM	Municipal	GeoBC	LiDAR /NRCan
Maple Ridge (north portion)	OSM	DRA	OSM	OSM	Municipal	GeoBC	LiDAR /NRCan
Abbotsford	Municipal	DRA	OSM	OSM	Municipal	GeoBC	LiDAR /NRCan
Chilliwack	Municipal	DRA	OSM	OSM	Municipal	GeoBC	LiDAR /NRCan
Mission (south portion)	Municipal	DRA	OSM	OSM	Municipal	GeoBC	LiDAR /NRCan
Mission (north portion)	OSM	DRA	OSM	OSM	Municipal	GeoBC	LiDAR /NRCan
Kent	OSM	DRA	OSM	OSM	FVRD	GeoBC	LiDAR /NRCan
Harrison H.S.	OSM	DRA	OSM	OSM	FVRD	GeoBC	LiDAR /NRCan
Electoral Areas	FVRD /OSM	DRA	OSM	OSM	FVRD	GeoBC	LiDAR /NRCan

3.0 METHODS

To generate a land cover raster for the full study area, a random forest classification approach was used. Random forest modelling is a machine learning algorithm that iteratively classifies an image using a random selection of predictor variables (Brieman 2001). This iterative process generates hundreds of unique decision trees, to evaluate each predictor variable and assign the best fit land cover class. The approach compensates for any model weaknesses present in the individual trees (or input variables); thereby, producing better predictive results. Additionally, it is data-driven approach that is flexible enough to accommodate the variety of predictor variables, that are available from Sentinel-2 and LiDAR derived inputs and easily scales to process the full study area to output a land cover classification and associated class probabilities which defines the stronger and less critical input variables. The random forest approach is driven by the collection of training data for each class.

At the highest level, the classification process consists of three stages: data preparation and review, random forest model generation, and classification post-processing (Figure 2). Each stage is made up of multiple steps that are described in detail in the following sections. This classification framework can be applied in the identical manner for other study areas as required with additional (localized) training data.

Figure 2. Geoprocessing analysis workflow



3.1 Satellite Image Preparation

3.1.1 Image selection

Cloud-free Sentinel images were collected for seven dates in 2019: March, April, May, July, August (x2), and November. June, September and October imagery was not included in the analysis, as no cloud-free images were available in those months. These images form a multi-temporal 'stack' that can be used to infer seasonal differences (as illustrated in Figure 3) and act as the foundation for the classification.

Figure 3. Seasonal Variation in Sentinel-2 Imagery

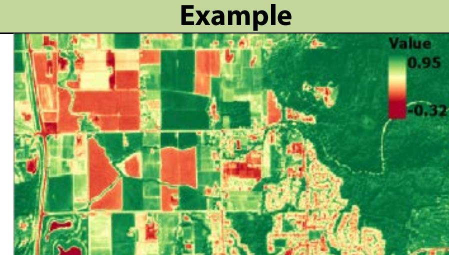
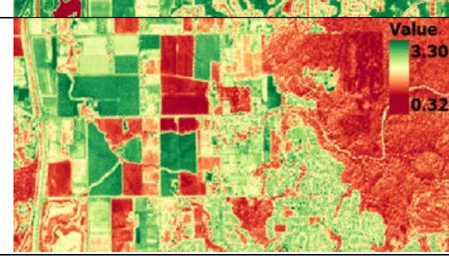
3.1.2 Image processing

The images in the stack are clipped to the boundary and the pixels co-aligned (or snapped) to ensure that pixel location was consistent throughout the analysis. All images were also topographically corrected to minimize the impact of shadows caused by terrain and resampled with a bilinear algorithm to 5-metre pixel resolution. Although resampling from the native 10-metre resolution to 5-metre pixels does not create more detail, it allows the satellite imagery to be stacked and analysed in conjunction with higher resolution LiDAR data.

3.1.3 Generation of derivative layers

From the image stack, two indices are generated: Normalized Difference Vegetation Index (NDVI) and a Red-Green Index as illustrated in Table 4. Using the NDVI values from each of the seven time periods, the NDVI Stack is summarized for cumulative statistics such as Maximum NDVI, Minimum NDVI, NDVI Standard Deviation, and Mean NDVI values for each pixel location. Different land cover types tend to have different value ranges for each of these multi-temporal statistics.

Table 4. Indices generated from Sentinel 2 imagery

Name	Formula	Example
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - Red}{NIR + Red}$	
Red Green Index (RGI)	$Red/Green$	

3.2 LiDAR Data Preparation

3.2.1 Generation of Terrain Derivatives

From the LiDAR data, four terrain related products were generated:

- digital elevation model (DEM);
- digital surface model (DSM) derived from first return values from each LiDAR pulse;
- canopy height model (CHM) which includes vegetation and building heights; and
- classified building raster that defines building footprints

The CHM was calculated by subtracting the DEM from the DSM. These derivatives were produced at a 2.5m resolution to retain data precision and then final derivatives were resampled to 5 metres and snapped to align with the satellite image data as inputs to the initial Random Forest model

3.2.2 Other Lidar Derivatives

The LiDAR intensity, and number of returns per laser pulse were also used as inputs to the Random Forest model, LiDAR return intensity is tied to each xyz point in the point cloud and values can be used to better infer the cover types which disperse and/or return different levels of intensity. The count of returns per pulse from the LiDAR platform also provides a useful statistic that can assist with separation of various cover types.

LiDAR is also processed to identify the extents of building rooftops, and since parallax is not a significant source of error, this raster dataset provides an accurate delineation of building footprints. This dataset was compared to an existing LiDAR derived building footprint file and available high resolution imagery to ensure that the algorithm functions accurately.

3.3 Ancillary Data Preparation

Ancillary data were used in the classification process to facilitate the separation of land cover classes that have similar spectral signatures and/or LiDAR derived parameters. (For example, 1m tall shrubs in a natural riparian area can be confused with 1m tall crops in an agricultural area – and land use zoning information can help separate the two classes.) Each of the ancillary datasets presented below was clipped to the appropriate study area and then used as a raster overlay with classification results to modify (re-assign) the classes as appropriate.

3.3.1 Agricultural Zoning Mask

The land use zoning data was assigned into agricultural or non-agricultural areas at the parcel level from the source data. In some cases it is evident that portions of a parcel are being actively worked, while other portions of a parcel are more natural or undeveloped. This includes parcels that extend into the Fraser River or into forested areas. In these cases, the parcels were split and assigned the appropriate class. The purpose of the zoning mask is to resolve any confusion between natural and agricultural herbaceous vegetation remaining in the final classification.

3.3.2 Road Mask

The road mask was generated using the provincial digital road atlas data. The road surface attribute defines the paved versus loose road surface types which can also classified tree cover classes to identify roads that are covered or partially covered by tree canopy. ‘Unknown’ and ‘Overgrown’ road surface types were not incuded in the road mask. Any road line feature intersecting the 5 metre grid cell was assigned into the paved or loose road class to later overlay and update the classification.

3.3.3 Building Footprint Mask

The building footprint mask was generated from the CHM and the LiDAR point cloud. Building heights were aggregated into three height categories: 0 - 3 metres (not a building); 3 - 15 metres (building); and greater than 15 metres (tall building). The purpose of the building mask was to resolve any remaining confusion in the classification between buildings and other anthropogenic features and to inform the building classes, as a result any pixel that fell under a building was classified as building.

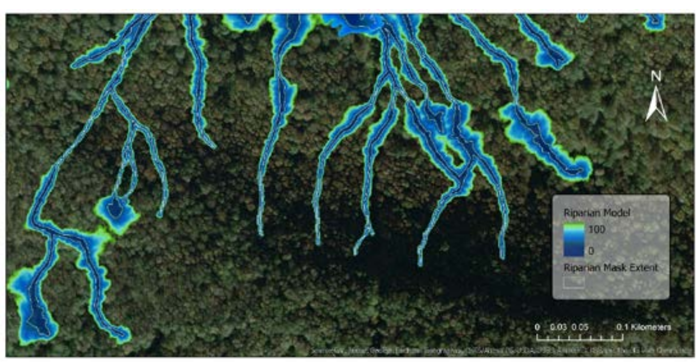
3.3.4 Canopy Height Mask

The canopy height mask was generated from the CHM layer. The vegetation heights were reclassified into five height categories: 0 - 2 metres; 2 - 5 metres; 5 - 10 metres; 10 - 15 metres; and greater than 15 metres. The purpose of the canopy height mask was to split apart the natural vegetation classes into a variety of height classes reflecting different vegetation types, and areas of young or regenerating vegetation.

3.3.5 Riparian Potential Mask

A terrain-based model based on the LiDAR derived DEM and freshwater atlas features was run to help identify riparian areas. The model uses slope values to determine the likelihood that water would move through the landscape from mapped hydrographic features such as wetlands, streams, rivers and lakes (Figure 4). This approach generates a much more realistic estimate of riparian habitat than the application of a constant setback distance: in steep terrain riparian corridors are narrow whereas in flatter terrain riparian areas are wider. In the case of ditches, which are designed and engineered to be constrained to a fixed channel, a fixed buffer of 5 metres is applied to account for the approximate wetted channel width plus embankments and the approximate distance shrub or tree roots may extend to reach wetter soils. The threshold distances used in the riparian model are determined largely through comparison to a sample of locations where more lush vegetation is evident in high resolution imagery adjacent to mapped streams and lakes. The riparian mask is later used to modify land cover classes and assign a riparian modifier to shrub and treed cover types.

Figure 4. Riparian model example



3.3.6 Agricultural Type Mask

Using the Sentinel-2 spectral values throughout the year an agricultural class mask was produced to separate cultivated, perennial, and herbaceous agriculture. Through this model, cultivated fields are defined as those that have a very high standard deviation in the seasonal NDVI values from March, Through summer to November. This is common as cultivated fields include green crops covers (high NDVI) and tilling or harvesting periods with bare soils (low NDVI). Perennial crops such as blueberries are isolated though less dramatic seasonal NDVI ranges and taller vegetation heights (>50cm). Cranberry perennial crops were initially identified spectrally via a dark red spectral signature in spring or fall but then delineated manually (to reduce noise or class confusion) based on manual interpretation of the fields. All other agriculturally zoned areas with lower vegetation heights (<50 cm) and suitable NDVI values throughout spring, summer and fall were assigned to the herbaceous class which includes pasture and hay crops. Each of these classes are used as a mask to later assign specific agricultural types to the vegetated portions of agricultural lands where the vegetation is less than 2.5 metres in height. This threshold height was defined as the tallest height for corn fields in the study area. Tree fruit crops are not separated from other tree classes.

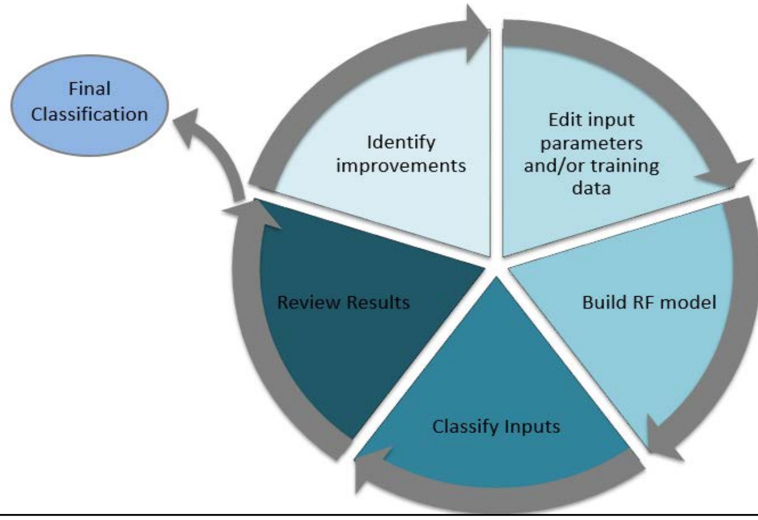
3.3.7 Tree Type Mask

Using the Sentinel-2 spectral values throughout the year a mask was produced to separate deciduous from coniferous trees. The March image provides a leaf-off NDVI value for each pixel location, while the early July image acts as the leaf-on NDVI value. A subtraction of leaf-on NDVI minus leaf-off NDVI results in high values for deciduous dominant canopy and low values for coniferous dominant canopy. This analysis is only completed in areas where the canopy height model (CHM) identifies trees as greater than 3 metres in height. (i.e., deciduous and coniferous shrubs are not differentiated.)

3.4 Random Forest Classification

The random forest classification model produces the initial version of the classification. It involves selecting training sites per class, assembling important LiDAR derivative rasters, and iteratively fitting the random forest models to optimize the initial classification results. For this project, the random forest model was built to identify five classes: non-vegetated (bare ground, gravel, pavement etc.), buildings, low vegetation, medium vegetation, and tall vegetation. These five classes can be discerned quite accurately in terms of spatial precision and minimal confusion between classes. Figure 5 illustrates the workflow used to generate the random forest classification.

Figure 5. Random forest workflow



3.4.1 Training Site Selection

The first step required for the random forest classification is the generation of a training dataset. The quality of the training dataset is important to the quality of the classification results, as it is the building block for the predictive model. Training sites were defined to be a representative sample of each class and span the variability of each class across the full study area extent. Training sites are validated through visual interpretation of Sentinel-2, LiDAR data as well as available high resolution imagery. Training sites were delineated as polygons and were collected for each class. An iterative process of collecting training sites and running trials of the classification provides feedback to guide additional training data requirements and refine specific class separations. Table 5 summarizes the number of training sites collected for each class in terms of the count of pixels with the training polygons.

Table 5. Training sites by class for each study area.

Class	Pixel Count
Non-Vegetated Ground (Bare, Gravel, Paved etc.)	533
Low Vegetation	2,518
Medium Vegetation	565
Tall Vegetation	5,137

3.4.2 Predictor Variables

All predictor variables for the random forest model were derived from LiDAR data derivatives to leverage the higher spatial resolution of LiDAR data and to limit class confusion that results from the broad spectral variability of optical imagery. The Sentinel-2 imagery remains as a key next step in refining the longer list of land cover classes. The seven input parameters (i.e., model predictor variables) used are: building classification, DEM, CHM, intensity max, intensity min, intensity average, and returns per point (from each laser pulse).

3.4.3 Random Forest Model Generation

The random forest classification was performed using the 'randomForest' package (Liaw and Wiener 2002), supported by R statistical software (R Core Team 2019). The goal of the random forest model was to classify the study area into the initial 5 land cover classes based on the predictor variables.

The random forest model has two analysis parameters that can be adjusted to improve model fit and accuracy: the number of predictor variables considered by each decision tree, and the number of decision trees generated for the random forest. Through iterations of training data collection, the selected model parameters were determined to be highly effective with strong classification separation accuracy. Within the R statistical software, each of the random forest models considered the five variables at each split and generated 400 trees.

3.4.4 Model Evaluation

Fitting the random forest models was an iterative process, where a model was built and the results were evaluated both visually and quantitatively. Visually, the classification was compared to the source imagery to ensure that landcover classes were being captured accurately. Quantitatively, the confusion matrix generated from the training data provide a strong indicator of model performance.

The model also outputs two measures of variable importance that can be used to evaluate and improve model fit: mean decrease accuracy and the mean decrease Gini index. The mean decrease in accuracy represents the number of observations that are incorrectly classified by removing the parameter in question from the model. The mean decrease in Gini is the average of a parameters total increase in node impurity if removed from the model. Node impurity represents how well the decision tree has split the data. An impure node has not split the data adequately (as confusion between classes persists), and therefore may require another decision 'branch'. These metric inform the mapping team as to where additional training may be required and which input variables are most effective.

3.5 Post-Processing

A series of post-classification processing steps were applied to resolve any remaining confusion in the random forest classification and to split classes into the full range of additional land cover types. Following this step, which leverages the multi-temporal Sentinel-2 image stack, a total of 34 land cover classes were defined. The post-processing workflow involves applying the masks generated from the ancillary datasets according to a set of conditional rules that follow this logic: In cases of current class 'A' falling under value 'B' of the mask the resulting land cover type should be converted to class 'C'. For example, if herbaceous vegetation fell within the boundaries of the agricultural zoning mask, then it became agricultural herbaceous vegetation. The masks were applied through a series of raster algebra and reclassification steps in the following order:

- Road mask: Defines loose and paved classes as well as identifies where roads are tree covered.
- Building mask: Defines the building footprints.
- Agricultural zoning mask: Separates vegetation into agricultural and non-agricultural land cover classes.
- Canopy height mask: Assigns height classes that define building heights and vegetation heights which influence all vegetated classes.
- Riparian potential mask: Adds the riparian potential modifier to herbaceous, shrub and all tree classes.
- Agriculture type mask: Splits agricultural land into Annual, Perennial, or Herbaceous classes.
- Tree type mask: Adds the deciduous or coniferous modifier to all tree cover classes.
- Log boom / storage mask: applies a manually delineated mask to covert bare areas in water to the log boom storage class.
- Maintained grass mask: applies to large patch areas where low vegetation is continually green from March to November 2019.
- Greenhouse mask: Greenhouses were delineated manually using satellite image interpretation and polygon digitizing. Buildings under this mask were converted to the greenhouse class. An unsupervised classification of buildings was performed to highlight buildings that exhibit the spectral signatures associated with greenhouses. These signatures help to locate greenhouses but they are not reliable enough to use directly in an automated manner.

4.0 RESULTS

Results were aggregated to 17 more generalized groupings to produce a stratified classification product (Table 6) (Note: Class 18 – Perennial snow cover was not present in this study area). The results presented in the following sections summarize the classification results for the detailed classes. General classes can be applied for certain end-user applications and any class aggregation would see an increase to classification accuracy although classes defined by heights from LiDAR are highly accurate and therefore do not increase or decrease overall accuracy in a significant manner. A breakdown of the detailed class descriptions and associated raster pixel values are available in Appendix A.

Table 6. Land cover class schema

General Class Name	Detailed Class Name	Class Description
Building	Building	Building, less than 15m tall
	Building - tall >15m	Building, greater than 15m tall
	Greenhouse	Greenhouse
Rail	Rail	Railway
Road	Road - Paved	Road, paved surface
	Road - Paved (under tree canopy)	Road, paved, under tree canopy
	Road - Loose	Road, loose (gravel or other unsealed surfaces)
	Road - Loose (under tree canopy)	Road, loose, under tree canopy
Pavement	Pavement	Non-road pavement: road shoulders, sidewalks, parking lots, driveways, farm vehicle areas or industrial yards
Bare Ground	Gravel	Gravel: gavel pit, gravel parking lot
	Bare Ground	Bare ground: bare ground, new land clearing
Agriculture – Annual Crop	Agriculture – Annual Crop	Agriculture, annual crops (likely tilled)
Agriculture – Grass / Herb	Agriculture – Grass / Herb	Agriculture, grass and herbaceous vegetation, often pasture (not tilled)
Agriculture - Perennial	Agriculture - Perennial	Agriculture, perennial, often berry crops
Grass/Herb	Grass / Herb	Grass and herbaceous vegetation outside of agricultural and riparian areas
Grass / Herb - Riparian	Grass / Herb - Riparian	Grass and herbaceous vegetation within the modeled riparian area, often fringing wetlands
Maintained Grass	Maintained Grass	Grass that is continually green (May to November 2019): playing field, golf course
Shrub / Sapling (2-5m)	Shrub / Sapling (2-5m)	Shrub and sapling, 2 to 5m in height
Shrub / Sapling (2-5m) - Riparian	Shrub / Sapling (2-5m) - Riparian	Shrub and sapling, 2 to 5m in height, within the modeled riparian area

continued from previous page

General Class Name	Detailed Class Name	Class Description
Deciduous Tree	Deciduous Tree Short (5-10m)	Deciduous tree, 5 to 10m in height
	Deciduous Tree Medium (10-15m)	Deciduous tree, 10 to 15m in height
	Deciduous Tree Tall (>15m)	Deciduous tree, greater than 15m in height
Coniferous Tree	Coniferous Tree Short (5-10m)	Coniferous tree, 5 to 10m in height
	Coniferous Tree Medium (10-15m)	Coniferous tree, 10 to 15m in height
	Coniferous Tree Tall (>15m)	Coniferous tree, greater than 15m in height
Tree Riparian	Deciduous Tree Short (5-10m) - Riparian	Deciduous tree, 5 to 10m in height, within the modeled riparian area
	Deciduous Tree Medium (10-15m) - Riparian	Deciduous tree, 10 to 15m in height, within the modeled riparian area
	Deciduous Tree Tall (>15m) - Riparian	Deciduous tree, greater than 15m in height, within the modeled riparian area
	Coniferous Tree Short (5-10m) - Riparian	Coniferous tree, 5 to 10m in height, within the modeled riparian area
	Coniferous Tree Medium (10-15m) - Riparian	Coniferous tree, 5 to 10m in height, within the modeled riparian area
	Coniferous Tree Tall (>15m) - Riparian	Coniferous tree, greater than 15m in height, within the modeled riparian area
Water	Water	Water, visible from above with 10m resolution imagery, most streams are obscured by vegetation at this observational scale
Perennial Snow	Perennial Snow*	(* Not present in this study area)

4.1 Land Cover Classification Results Summary

The random forest classification and subsequent masking techniques leveraged Sentinel-2 imagery to successfully classify the desired land cover types. Subsequent vector data masking integrated improved the separation of classes to provide reliable class enhancements. Overall accuracy remains high for the number of classes in this model and can useful for a variety if end-user decision support tasks.

4.1.1 Classification Results

The images below provide an overview and samples of full detail within the study area. Refer to the digital deliverables of this project to review or analyse the full dataset. Although different vector source files were conflated to build hydrography and zoning masks, the merged product remained consistent and results in a consistent product across all mapped municipalities. The study area contains the full diversity of land cover types, most of which are present in the sample area illustrated near Abbotsford (Figure 6). These land cover types are also presented in an aggregated set of general classes for the same area (Figure 7). Appendix B provides a summary of the coverage per class within each municipal area. This table includes the total mapped area per municipality to assist in calculating the proportion of the municipality that has been mapped (since portions of some municipalities were not mapped due to partial LiDAR coverage).

Figure 6. Detailed Classes - Abbotsford Area Sample

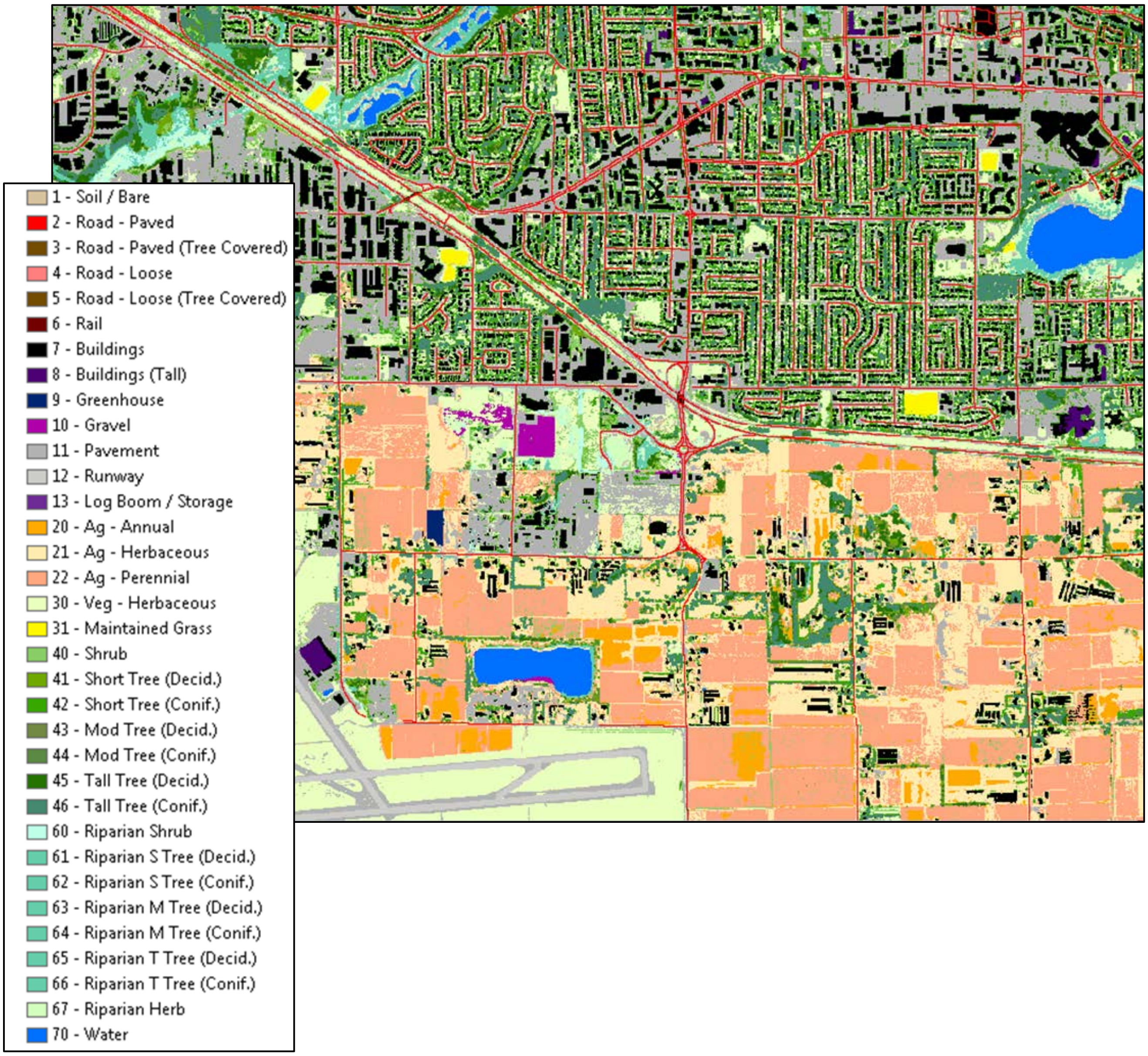


Figure 7. General Classes - Abbotsford Area Sample

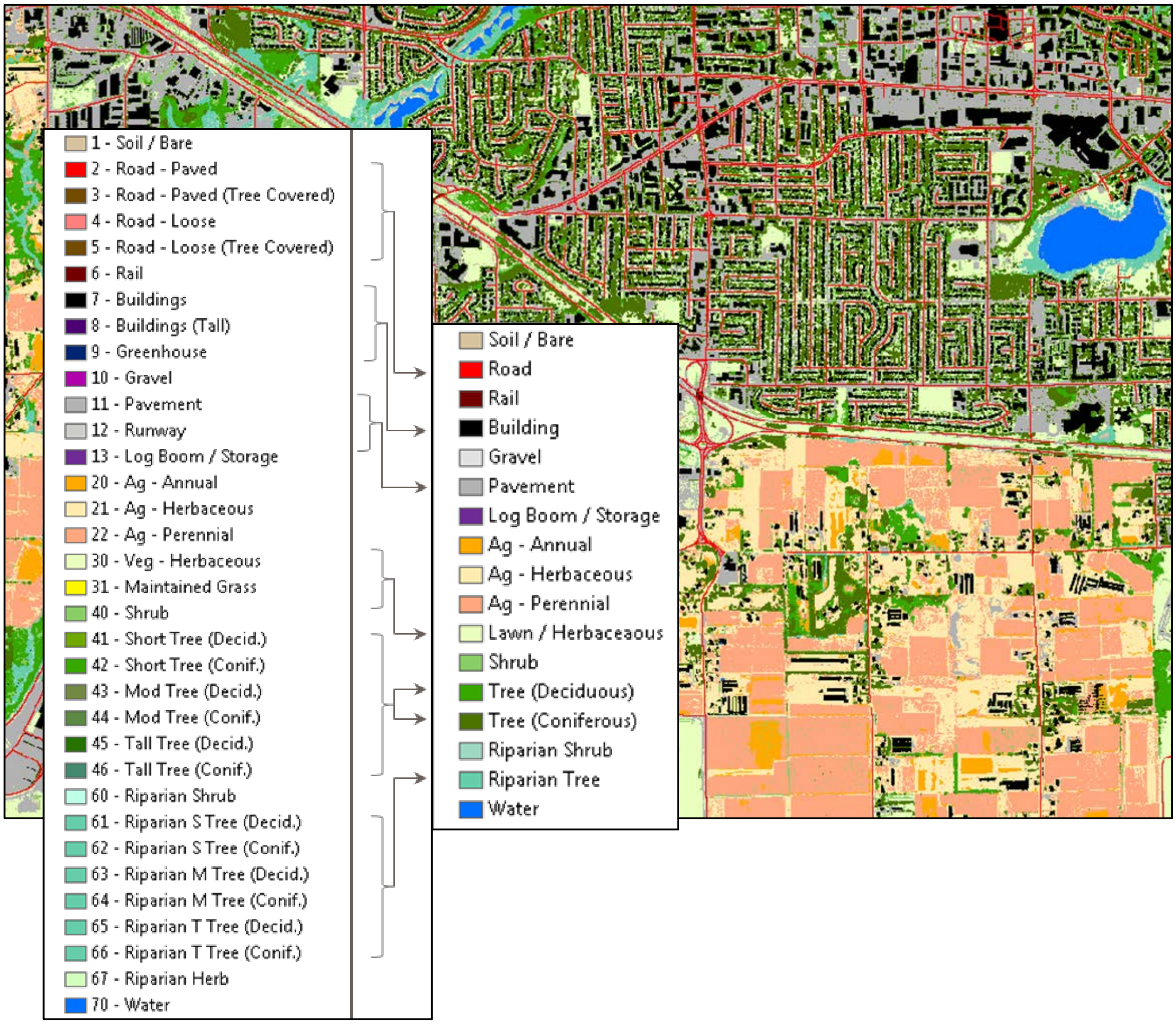
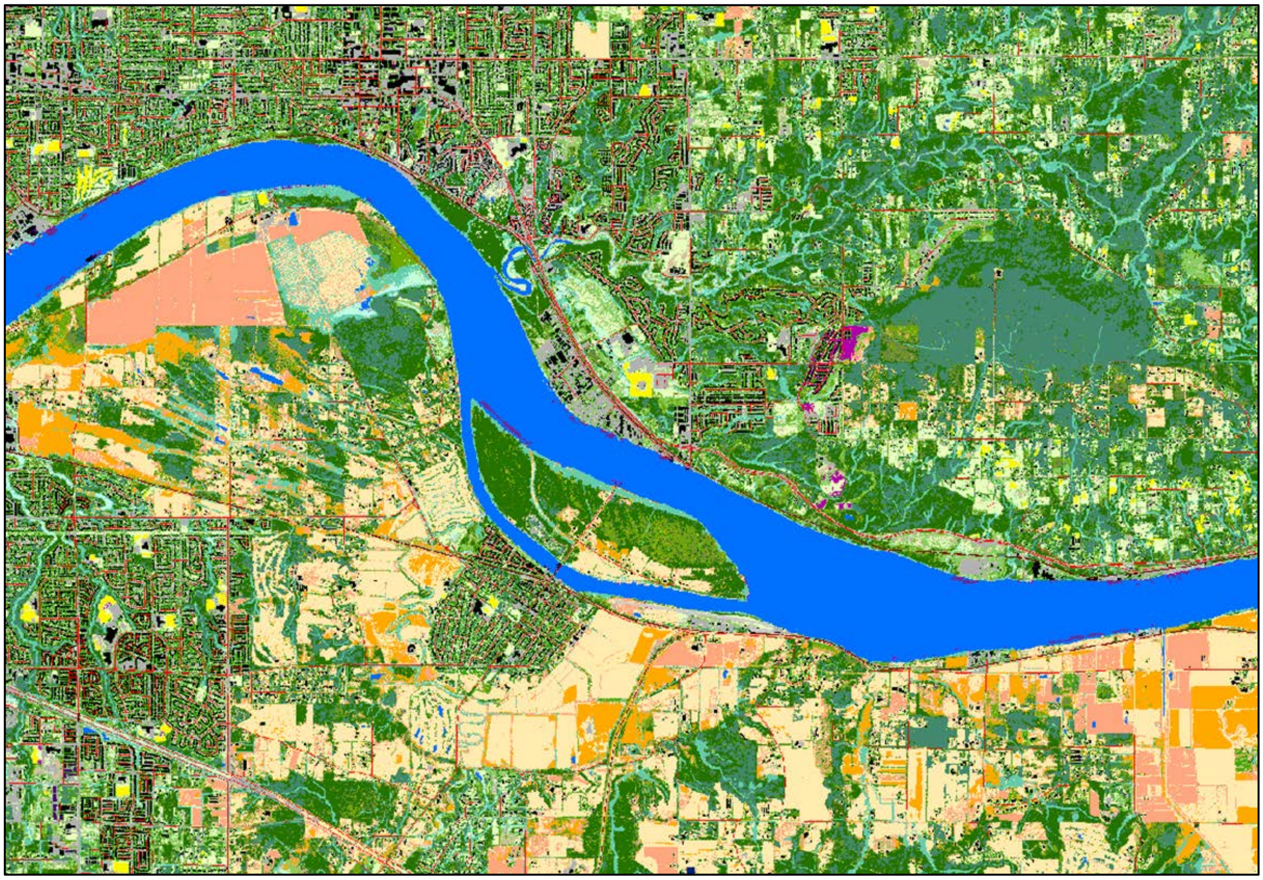


Figure 8. Detailed Classes - Langley / Mission Area Sample



When compared to previous project 10 metre results, the 5 metre output (Figure 9) illustrates how urban and residential areas are well mapped with various land cover types instead of a blended 'urban mix' class that previously was too coarse to discern the detail of smaller features.

Figure 9. Full Resolution - Langley Residential Sample

Figure 10. Full Resolution - Chilliwack Sample

4.2 Accuracy Assessment

To complete an accuracy assessment, a stratified random sample of reference locations was generated for each class. A total of 540 comparisons were made between the mapped class and the interpreted ground feature based on high resolution orthophotos. A larger percentage of samples were taken for classes that tend to perform poorly to ensure that the accuracy assessment was not biased by over sampling stronger, easily classified cover types. The number of samples for each height class of tree is slightly lower because the precision of the LiDAR provides a very reliable separation of tree heights and our ability to assess tree height is no better than the LiDAR separation. The random sampling approach ensures statistical validity of the approach while also distributing the samples across the full study area to account for any localized variation. The overall user accuracy is assessed at 93.2% which is a much stronger result for a classification with over 30 cover types when compared to an approach that uses satellite imagery alone. The error matrix can be referenced in Appendix C to explore individual class accuracies. The most significant errors include:

- Differentiation with the various agricultural classes;
- Differentiation between some semi-permeable gravel surfaces and impermeable paved / concrete areas; and
- Some localized errors due to spatial mis-alignment or incorrect attributes for vector inputs such as road features.

5.0 DISCUSSION

The efforts from this project have resulted in an efficient and well defined procedure to map land cover at a 5 m resolution that is replicable where LiDAR data is available. The procedure could be modified to produce fewer (nested) classes with a predominantly Sentinel-2 image classification where LiDAR is not available to provide complete coverage of regional areas. The random forest classification models successfully identified land covers with strong performance for a variety of end-user applications.

5.1 Data Caveats

When using these classified products there are a few data caveats that should be noted. These data limitations originate in the Sentinel and LiDAR source data, as well as, from the manually collected training samples.

5.1.1 Sentinel Imagery

This multi-temporal Sentinel based classification requires that cloud free images are collected several times throughout the year for each study area and that the dates when each of the images were collected are spread out throughout the growing season. Cloud cover is an obvious detriment to the mapping process but a lack of cloud-free images in certain dates could have the following impacts on results:

- Less effective separation of deciduous and coniferous classes if imagery is not available for leaf-off and leaf-on dates.
- Less effective separation of agricultural classes if imagery is not available for several dates across the growing season to allow for mapping periods of full growth and periods of tilling as well as periods of relative dry and relative wet ground.

Batch processing of the imagery and classification training data allows for mapping of areas as wide (in the east-west direction) as 290 km depending on cloud cover and the maximum width of the image swath. In some cases mapping can be completed for areas that extend further than this in the north-south direction along the path of the satellite orbit where the image acquisition dates and data collection settings remain similar. For study areas that extend beyond a single image swath, additional training areas must be collected to account for variations in dates and spectral characteristics.

5.1.2 LiDAR Data

As seen in the model results, the LiDAR derived data is an important input parameter for the classification. However, the LiDAR data was not collected at the same time as the Sentinel imagery leading to discrepancies between the vegetation and building heights in the CHM and the spectral values in the imagery. This impacts the classification of features that have changed between Sentinel data capture date in 2019 and the LiDAR data capture date in 2016. This issue is minor over a large study area but could be more problematic in areas with a high degree of development over this timeframe.

During this project it became evident that some small areas have LiDAR data flaws that prevented the use of elevation values or intensity values. Some processing considerations and reclassification was required in these areas to account for errors due to these data gaps.

Additionally, only the area with LiDAR data can be mapped with the same classes and precision. If there is a desire to map areas not covered by LiDAR to achieve complete coverage of a municipality or regional district, it would be possible by developing a nested set of classes that could be mapped to align well for various end-uses. This product would be based on Sentinel-2 and ancillary vector data alone.

5.1.3 Ancillary Vector Data

The classification implements a masking stage to integrate ancillary vector data into the process. As a result, the accuracy, precision, and coverage of the ancillary datasets have an impact on the final classification accuracy. For example, the data used to inform the riparian classification are dependant on the location of water features and the agricultural classes are dependent upon accurate parcel level zoning data. As such, it is important that the best possible data is being acquired to inform the masking stage.

6.0 REFERENCES

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Caslys Consultig Ltd. (2020). Lower Mainland Land Cover Classifcation – Summary Report, provided to BC MFLNRORD

Liaw, A. & Wiener, M. (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

R Core Team (2020). R: A language and environment for statistical computing

APPENDIX A

Raster Value	Class Name	Class Description
7	**Building**	Building, less than 15m tall
8	**Building - tall >15m**	Building, greater than 15m tall
9	**Greenhouse**	Greenhouse
2	**Road - Paved**	Road, paved surface
3	**Road - Paved (under tree canopy)**	Road, paved, under tree canopy *determined by reference to the image classification*
4	**Road - Loose**	Road, loose surface
5	**Road - Loose (under tree canopy)**	Road, loose, under tree canopy, *determined by reference to the image classification*
6	**Rail**	Railway
11	**Pavement**	Non-road pavement: parking lots, large driveways, farm vehicle yards
10	**Gravel**	Gravel: gravel pit, gravel parking lot
1	**Bare Ground**	Bare ground: bare ground, new land clearing
20	**Agriculture – Annual Crop**	Agriculture, annual crops
21	**Agriculture – Grass / Herb**	Agriculture, grass and herbaceous vegetation, often pasture
22	**Agriculture - Perennial**	Agriculture, perennial, often berry crop
30	**Grass / Herb**	Grass and herbaceous vegetation outside of agricultural and riparian areas
67	**Grass / Herb - Riparian**	Grass and herbaceous vegetation within the modeled riparian area, often fringing wetlands
31	**Maintained Grass**	Grass that is continually green (May to October 2019): playing field, golf course
40	**Shrub / Sapling (2-5m)**	Shrub and sapling, 2 to 5m in height
60	**Shrub / Sapling (2-5m) - Riparian**	Shrub and sapling, 2 to 5m in height, within the modeled riparian area
41	**Deciduous Tree Short (5-10m)**	Deciduous tree, 5 to 10m in height
43	**Deciduous Tree Medium (10-15m)**	Deciduous tree, 10 to 15m in height
45	**Deciduous Tree Tall (>15m)**	Deciduous tree, greater than 15m in height
61	**Deciduous Tree Short (5-10m) - Riparian**	Deciduous tree, 5 to 10m in height, within the modeled riparian area
63	**Deciduous Tree Medium (10-15m) - Riparian**	Deciduous tree, 10 to 15m in height, within the modeled riparian area

65	**Deciduous Tree Tall (>15m) - Riparian**	Deciduous tree, greater than 15m in height, within the modeled riparian area
42	**Coniferous Tree Short (5-10m)**	Coniferous tree, 5 to 10m in height
44	**Coniferous Tree Medium (10-15m)**	Coniferous tree, 10 to 15m in height
46	**Coniferous Tree Tall (>15m)**	Coniferous tree, greater than 15m in height
62	**Coniferous Tree Short (5-10m) - Riparian**	Coniferous tree, 5 to 10m in height, within the modeled riparian area
64	**Coniferous Tree Medium (10-15m) - Riparian**	Coniferous tree, 5 to 10m in height, within the modeled riparian area
66	**Coniferous Tree Tall (>15m) - Riparian**	Coniferous tree, greater than 15m in height, within the modeled riparian area
70	**Water**	Water, visible from above with 10m resolution imagery, most streams are obscured by vegetation at this observational scale

APPENDIX B

Land Cover Area Summary by Municipality

Area Description	Electoral Areas hectares	Abbotsford hectares	Chilliwack hectares	Maple Ridge hectares	Kent hectares	Mission hectares	Langley hectares	Harrison Hot Springs hectares	Total Mapped Area hectares	% Area	
Soil / Bare	628.7	240.1	358.6	275.8	126.0	189.4	366.9	7.3	2192.6	1.0%	1.0%
Road - Paved	222.6	602.9	369.4	255.4	83.8	174.4	504.4	9.7	2222.4	1.0%	1.4%
Road - Paved (TreeCover)	47.3	51.0	34.8	43.8	12.5	36.0	61.6	0.8	287.7	0.1%	
Road - Loose	64.3	17.6	35.1	13.0	50.6	18.2	13.2	0.0	212.0	0.1%	
Road - Loose (TreeCover)	98.9	9.1	17.6	6.4	40.5	14.6	3.5	0.0	190.7	0.1%	
Rail	34.3	56.2	26.1	20.1	29.5	20.2	31.0	0.0	217.4	0.1%	
Building	194.5	1407.4	1006.5	617.8	109.2	315.7	1379.0	19.7	5049.8	2.3%	2.4%
Building Tall	0.3	11.1	5.0	3.4	0.1	0.9	7.8	0.4	29.0	0.0%	
Greenhouse	0.5	109.8	42.2	8.8	0.0	0.0	93.3	0.0	254.6	0.1%	
Gravel (Semi-Permiable)	242.9	188.3	126.5	55.3	119.6	57.8	33.0	0.9	824.1	0.4%	3.1%
Pavement (Impervious)	436.0	1538.3	1173.4	827.3	161.7	460.8	1293.3	29.4	5920.0	2.7%	
Runway	0.0	37.9	9.5	0.0	0.0	0.0	22.1	0.0	69.5	0.0%	
Log Boom / Storage	1.4	6.9	0.0	13.0	0.8	23.5	8.2	0.0	53.8	0.0%	
Ag - Annual	1514.4	5739.2	4787.2	195.2	1764.0	183.4	1397.8	0.0	15581.3	7.1%	25.9%
Ag - Herbaceous	3268.8	8751.1	5127.0	462.2	1812.2	303.1	8341.6	0.0	28065.8	12.8%	
Ag - Perennial	1223.5	5718.6	1978.9	376.4	541.3	150.9	3099.7	0.1	13089.4	6.0%	
Veg - Herbaceous	754.3	1141.1	774.4	1026.8	231.2	821.2	1058.5	14.4	5821.9	2.7%	7.7%
Maintained Grass	80.1	97.5	162.4	78.0	33.3	67.5	62.3	0.0	581.2	0.3%	
Shrub	2949.1	1222.1	1289.6	1507.1	735.8	1405.7	1247.9	25.6	10382.8	4.7%	
Short Tree (Decid)	2466.6	772.5	680.8	590.0	712.4	551.0	941.3	19.0	6733.5	3.1%	46.1%
Short Tree (Conif)	1587.1	1244.4	835.9	1153.7	495.1	1250.4	1562.8	30.8	8160.3	3.7%	
Mod Tree (Decid)	1375.6	561.8	469.0	408.3	434.5	358.6	701.4	12.7	4321.9	2.0%	
Mod Tree (Conif)	1295.6	424.1	285.0	667.2	384.5	841.4	570.7	15.2	4483.6	2.0%	
Tall Tree (Decid)	9574.3	3895.7	2907.5	2059.9	2840.1	2106.9	3620.7	70.2	27075.3	12.4%	
Tall Tree (Conif)	17806.0	2274.4	2713.0	10662.2	5834.4	8218.6	2598.9	229.8	50337.4	23.0%	
Riparian Shrub	466.4	203.1	291.5	135.3	124.5	259.2	253.4	6.8	1740.2	0.8%	0.8%
Riparian S Tree (Decid)	512.5	257.6	252.0	105.9	207.0	133.1	304.1	4.2	1776.3	0.8%	4.5%
Riparian S Tree (Conif)	108.6	99.6	99.1	65.1	32.1	92.3	190.6	1.8	689.2	0.3%	
Riparian M Tree (Decid)	200.5	131.2	151.1	59.8	97.2	61.6	145.6	3.2	850.3	0.4%	
Riparian M Tree (Conif)	64.5	46.0	36.0	47.0	14.9	65.1	78.3	1.0	352.7	0.2%	
Riparian T Tree (Decid)	744.9	661.8	598.3	332.2	297.3	202.6	621.8	7.3	3466.2	1.6%	
Riparian T Tree (Conif)	450.8	241.2	151.9	602.4	87.2	491.3	282.1	4.6	2311.5	1.1%	
Riparian Herb	64.2	75.3	89.5	43.6	28.8	55.6	80.1	1.3	438.6	0.2%	
Water	5629.5	1064.6	1806.6	1239.2	2069.9	2753.7	803.7	33.8	15401.0	7.0%	7.0%
Totals	**54,109**	**38,899**	**28,691**	**23,958**	**19,512**	**21,685**	**31,780**	**550**	**219,184**	100%	

APPENDIX C

Accuracy Assessment – Error Matrix

Red flagged errors are due to precision errors in source vector datasets

Mapped vs. Actual Classes	Soil / Bare	Road - Paved	Road - Paved (TC)	Road - Loose	Road - Loose (TC)	Rail	Building	Building Tall	Greenhouse	Gravel (Semi-Permiable)	Pavement (Impervious)	Runway	Log Boom / Storage	Ag - Annual	Ag - Herbaceous	Ag - Perennial	Veg - Herbaceous	Maintained Grass	Shrub	Short Tree (Decid)	Short Tree (Conif)	Mod Tree (Conif)	Mod Tree (Conif)	Tall Tree (Decid)	Tall Tree (Conif)	Riparian Shrub	Riparian S Tree (Decid)	Riparian S Tree (Conif)	Riparian M Tree (Decid)	Riparian M Tree (Conif)	Riparian T Tree (Decid)	Riparian T Tree (Conif)	Riparian Herb	Water	Total	Stratified User Accuracy
	1	2	3	4	5	6	7	8	9	10	11	12	13	20	21	22	30	31	40	41	42	43	44	45	46	60	61	62	63	64	65	66	67	70		
Soil / Bare	18									1																									19	94.7%
Road - Paved		27		1																															28	96.4%
Road - Paved (TC)		2	27																																29	93.1%
Road - Loose			1	28																															29	96.6%
Road - Loose (TC)				1	14																														15	93.3%
Rail						10																													10	100.0%
Building							29																												29	100.0%
Building Tall								8																											8	100.0%
Greenhouse							1		12																										13	92.3%
Gravel (Semi-Permiable)		1	1							25	4																								31	80.6%
Pavement (Impervious)										3	24																								27	88.9%
Runway												9																							9	100.0%
Log Boom / Storage													9																						9	100.0%
Ag - Annual														26		1																			27	96.3%
Ag - Herbaceous	1														12	4	2																		19	63.2%
Ag - Perennial	1													3	1	20																			25	80.0%
Veg - Herbaceous	1																25				1														27	92.6%
Maintained Grass																		11																	11	100.0%
Shrub					2														24																26	92.3%
Short Tree (Decid)			1																	10								2							13	76.9%
Short Tree (Conif)																					12							1							13	92.3%
Mod Tree (Decid)																			1			12													13	92.3%
Mod Tree (Conif)																							9												9	100.0%
Tall Tree (Decid)																								8											8	100.0%
Tall Tree (Conif)																									8										8	100.0%
Riparian Shrub														1												9									10	90.0%
Riparian S Tree (Decid)																											11	2							13	84.6%
Riparian S Tree (Conif)																												4							4	100.0%
Riparian M Tree (Decid)																													7						7	100.0%
Riparian M Tree (Conif)																														8					8	100.0%
Riparian T Tree (Decid)																															8	2			10	80.0%
Riparian T Tree (Conif)																																9			9	100.0%
Riparian Herb	1																																10		11	90.9%
Water																																		13	13	100.0%
Number Sampled	22	30	30	30	16	10	30	8	12	29	28	9	9	30	13	25	27	11	25	10	13	12	9	8	8	9	11	9	7	8	8	11	10	13	540	**93.2%**