

# GEM 511: Advanced GIS for Environmental Management

Paul D. Pickell

2025-12-19



# Contents

<b>Welcome</b>	<b>5</b>
How to use these resources . . . . .	5
How to get involved . . . . .	5
<b>1 Spatial interpolation and visualization of LiDAR Data</b>	<b>7</b>
Lab Overview . . . . .	7
Learning Objectives . . . . .	7
Deliverables . . . . .	8
Data . . . . .	8
Task 1: Preprocess LiDAR data in PDAL . . . . .	8
Task 2: Visualize LiDAR data in QGIS . . . . .	15
Task 3: Prepare LiDAR in ArcGIS Pro . . . . .	19
Task 4: Apply and evaluate spatial interpolation algorithms in ArcGIS Pro . . . . .	23
Task 5: Visualize 3D data in ArcGIS Pro . . . . .	25
Summary . . . . .	27
<b>2 Salmon Stream Network Analysis</b>	<b>29</b>
Lab Overview . . . . .	29
Learning Objectives . . . . .	30
Deliverables . . . . .	30
Data . . . . .	32
Task 1: Create stream segments . . . . .	33
Task 2: Create routes and apply topology . . . . .	36
Task 3: Record stream attributes with linear referencing . . . . .	42
Task 4: Trace the network . . . . .	44
Task 5: Overlay and query the network . . . . .	47
Summary . . . . .	51
<b>3 Terrain Analysis and Riparian Area Management</b>	<b>53</b>
Lab Overview . . . . .	53
Learning Objectives . . . . .	53
Lab Background . . . . .	53

Deliverables . . . . .	54
Data . . . . .	55
Task 1: Prepare the DEM and Map Stream Networks . . . . .	55
Task 2: Extract Stream Characteristics . . . . .	59
Task 3: Mapping Watersheds and Riparian Management Areas . . . . .	64
Summary . . . . .	67
<b>4 Machine Learning with Geospatial Data</b>	<b>69</b>
Lab Overview . . . . .	69
Learning Objectives . . . . .	70
Deliverables . . . . .	71
Data . . . . .	72
Task 1: Exploratory Regression in ArcGIS Pro . . . . .	73
Task 2: Sampling for Training Datasets . . . . .	76
Task 3: Forest-Based Classification . . . . .	79
Summary . . . . .	82
<b>5 Analyzing Green Equity Using Geographically Weighted Regression</b>	<b>83</b>
Lab Overview . . . . .	83
Learning Objectives . . . . .	84
Deliverables . . . . .	84
Data . . . . .	85
Task 1: Prepare census data . . . . .	86
Task 2: Practice geographically weighted regression . . . . .	94
Task 3: Analyze your own census characteristics . . . . .	101
Summary . . . . .	102
<b>6 Web Mapping</b>	<b>103</b>
Lab Overview . . . . .	103
Learning Objectives . . . . .	103
Deliverables . . . . .	104
Data . . . . .	104
Task 1: Adding content to your ArcGIS Online profile . . . . .	104
Task 2: Creating your ArcGIS StoryMap . . . . .	108
Summary . . . . .	114
<b>7 Writing Metadata and Documentation</b>	<b>115</b>
Lab Overview . . . . .	115
Learning Objectives . . . . .	115
Deliverables . . . . .	116
Data . . . . .	116
Task 1: Depositing in the Dataverse . . . . .	117
Summary . . . . .	121

# Welcome

These are the course materials for GEM 511 in the Master of Geomatics for Environmental Management program (MGEM) at the University of British Columbia (UBC). These Open Educational Resources (OER) were developed to foster the Geomatics Community of Practice that is hosted by the Faculty of Forestry at UBC.

These materials are primarily lab assignments that students enrolled in GEM 511 will complete and submit for credit in the program. Note that much of the data referenced are either public datasets or otherwise only available to students enrolled in the course for credit. Deliverables for these assignments are submitted through the UBC learning management system and only students enrolled in the course may submit these assignments for credit.

## How to use these resources

Each “chapter” is a standalone lab assignment designed to be completed over one or two weeks.

Students enrolled in GEM 511 will submit all deliverables through the course management system at UBC for credit and should consult the schedule and deadlines posted there. The casual user can still complete the tutorials step-by-step, but the data that are not already publicly available are not hosted on this website and therefore you will not have access to them.

Unless otherwise noted, all materials are Open Educational Resources (OER) and licensed under a Creative Commons license (CC-BY-SA-4.0). Feel free to share and adapt, just be sure to share with the same license and give credit to the author.

## How to get involved

Because this is an open project, we highly encourage contributions from the community. The content is hosted on our GitHub repository and from there you can open an issue or start a discussion. Feel free to open an issue for any

typos, factual discrepancies, bugs, or topics you want to see. We are always looking for great Canadian case studies to share! You can also fork our GitHub repository to explore the source code and take the content offline.

# Chapter 1

# Spatial interpolation and visualization of LiDAR Data

Written by Paul Pickell

## Lab Overview

The aim of this lab is to use LiDAR data from the University of British Columbia Malcolm Knapp Research Forest (MKRF) to create Digital Elevation Model (DEM) using a variety of spatial interpolation approaches. We will investigate how these methods compare to one another, and explore their strengths and weaknesses. Additionally, we will explore point cloud manipulation and visualization using both QGIS and ArcGIS Pro.

---

## Learning Objectives

- Interpret metadata for a LiDAR acquisition and point cloud
  - Manipulate a LiDAR point cloud with a variety of different tools
  - Generate and evaluate DEMs from a LiDAR point cloud using different terrain spatial interpolation approaches
  - Create maps and 3D visualizations of point clouds and interpolated surfaces
-

## Deliverables

Responses to the questions posed throughout the lab on the course management system (25 points)

Table that shows zonal statistics of the difference rasters across elevation and slope ranges of the binned DEM (10 points)

A panel of maps that symbolizes the DEMs produced and the difference rasters that were produced by the comparison to the binning approach (15 points)

---

## Data

We will be working with LiDAR data collected over the UBC Malcolm Knapp Research Forest (MKRF) in Maple Ridge, British Columbia. These data are publicly available from the Master of Geomatics for Environmental Management (MGEM) Data Store and the instructions for accessing these data are given in the tasks below.

---

### Task 1: Preprocess LiDAR data in PDAL

Point cloud data are large. For example, the point cloud collection that we will be working with contains 1,671,233,402 points! Typically, we should not interact with point cloud data in a desktop environment until we have to. Graphical user interfaces like QGIS and ArcGIS Pro introduce a large amount of computational overhead when working with point cloud data and these software are more suited for visualizing the data rather than processing them. Point cloud data are much more commonly hosted on remote servers nowadays, in cloud-optimized formats, and available for on-demand and query-ready streaming.

In this task, we will explore large LiDAR acquisitions that have been collected at the University of British Columbia (UBC) Malcolm Knapp Research Forest (MKRF). LiDAR collections are typically tiled to reduce the overhead with transacting with individual files. We are only going to process a handful of tiles, but we need to first grab the right tiles for our area of interest (AOI).

**Step 1:** Navigate to the MGEM Data Store and inspect the UBC MKRF 2016 LiDAR collection: [https://206-12-122-94.cloud.computeCanada.ca/UBC\\_MKRF\\_LiDAR\\_2016/](https://206-12-122-94.cloud.computeCanada.ca/UBC_MKRF_LiDAR_2016/)

At the top, you will see some generic metadata for the collection. Below that, you will see a web map showing the tiles. If you click on one of the tiles, it gives the direct download universal resource locator (URL) and file size. If you scroll

down, all the tiles are listed in a table along with a metadata file that provides some more specific information about each individual tile-file.

**Step 2:** Click to open one of the metadata txt files in your browser. This is json-formatted metadata for the associated LiDAR tile. Scrolling through it you will find summary statistics over a number of different dimensions. What attributes describe these LiDAR data?

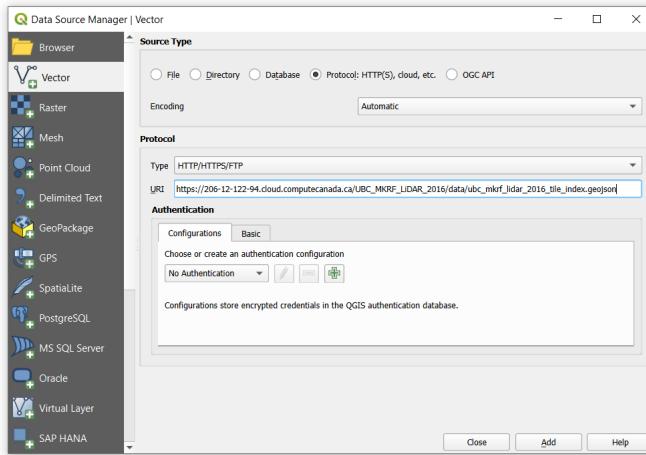
```
{
  "statistic": [
    {
      "average": 532269.9535,
      "count": 206325,
      "maximum": 532299.99,
      "minimum": 532236.53,
      "name": "X",
      "position": 0,
      "stddev": 17.47786941,
      "variance": 305.4759191
    },
    {
      "average": 5455340.249,
      "count": 206325,
      "maximum": 5455399.99,
      "minimum": 5455281.04,
      "name": "Y",
      "position": 1,
      "stddev": 34.89327592,
      "variance": 1217.540705
    },
    {
      "average": 169.5849529,
      "count": 206325,
      "maximum": 199.89,
      "minimum": 139.88,
      "name": "Z",
      "position": 2,
      "stddev": 13.46191763,
      "variance": 181.2232263
    }
  ]
}
```

Suppose that we have some AOI within the research forest that we need to retrieve the LiDAR data from. How could we figure out which tiles we need without downloading all of them from the server? In most cases, you should have a polygon tile index available to assist you with this task. The tile index is another form of metadata, albeit spatial metadata.

**Step 3:** Right-click on the **tile\_index.geojson** file at the top of the file listing and select “Copy Link”. Open QGIS and click the “Open Data Source Manager” button (Layer > Data Source Manager).

On the left of the Data Source Manager dialogue, select “Vector”, toggle on “Protocol: HTTP(S), cloud, etc”, then paste the URL you copied into the “URI” field (uniform resource identifier). “Add” the layer to your map view then “Close” the dialogue and inspect the result.

## 10CHAPTER 1. SPATIAL INTERPOLATION AND VISUALIZATION OF LIDAR DATA



A nice feature about QGIS is that it supports the ability to read any openly-specified geospatial file directly from a remote source. ArcGIS Pro only allows you to read some sources published on compatible remote databases (e.g., ArcGIS enterprise geodatabase or PostgreSQL) or from layers published on ArcGIS Online.

Now that we have spatial tile metadata, we can perform spatial intersection to find the right tiles. Our AOI is going to be the following longitude-latitude bounding box:

Lower left (LL): -122.55275, 49.32325

Upper right (UR): -122.52506, 49.34135

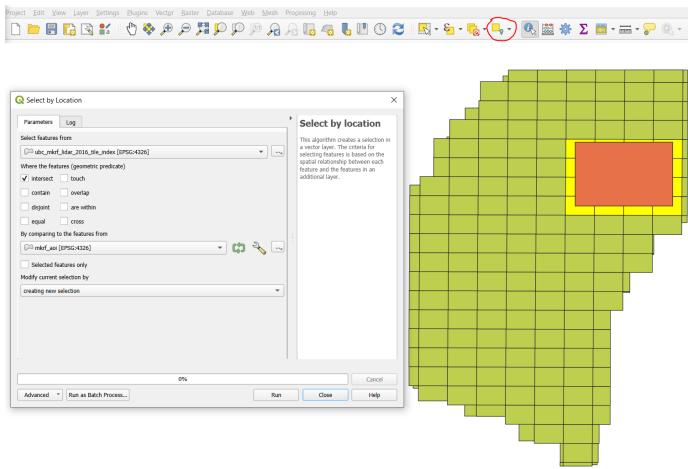
**Step 4:** Open a notepad text editor and convert this bounding box into a geojson polygon feature with the following syntax:

```
{
  "type": "Polygon",
  "coordinates": [
    [
      [LL_longitude, LL_latitude],
      [UR_longitude, LL_latitude],
      [UR_longitude, UR_latitude],
      [LL_longitude, UR_latitude],
      [LL_longitude, LL_latitude]
    ]
  ]
}
```

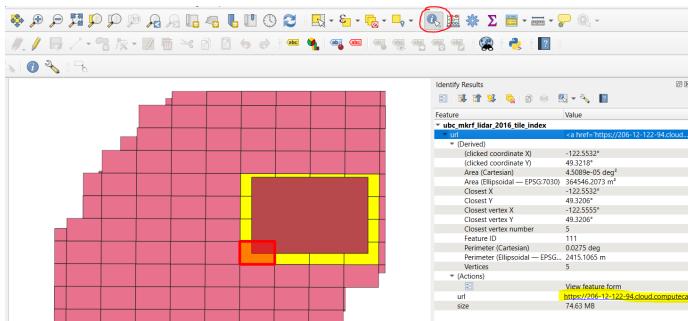
Replace with the correct longitude/latitude values. Note that this creates a square polygon and the fifth coordinate is the same as the first, which topologically encloses the polygon. Save the geojson in your QGIS project folder as

“mkrf\_aoi.geojson” then open the file in QGIS. You should see that our AOI spans 16 total tiles.

**Step 4:** Click the “Select by Location” tool and “Select features from” the `ubc_mkrf_lidar_2016_tile_index` layer by intersecting with the `mkrf_aoi` you just made. Now the intersecting tiles are selected. Turn off `mkrf_aoi` layer so that you can see the selection.



**Step 5:** Click the “Identify Features” tool and click on one of the selected tiles, which will highlight it in red and open the attributes for the polygon. Expand “`ubc_mkrf_lidar_2016_tile_index`”, “`url`”, “(Actions)”, and “`url`”. Click the hyperlinked URL to download the tile to your QGIS project folder. Repeat this step for all 16 tiles.



**Step 6:** Add all of the downloaded tiles to your QGIS map canvas. The default symbology is the classification attribute, but only ground returns have been classified for the research forest.

However, the tiles are all still stored in separate files and some portions are outside our AOI. So next, we are going to filter, merge, and crop the point cloud, but we are going to do this outside of QGIS because it will be faster and more reliable. We are going to use the Point Data Abstraction Library

## 12CHAPTER 1. SPATIAL INTERPOLATION AND VISUALIZATION OF LIDAR DATA

(PDAL) command line utility to perform this processing. You can read more about the extensive PDAL functionality here: <https://pdal.io/> Note that many of the functions we are going to use with PDAL are available as tools through QGIS, but you have less control over the options and parameters in QGIS.

**Step 7:** Open the OSGeo4W shell and navigate to your QGIS project folder where your LiDAR data are located. For example, `cd C:\users\paul\Documents\QGIS\mkrf_lidar`. Type the command `pdal --version` to make sure that PDAL was installed with your current version of QGIS. If a version number is returned in the console window, then continue to the next step, otherwise ask your instructor to help you install PDAL using the OSGeo4W installer.

PDAL can run functions in two ways. First, as subcommands, much like you have used with GDAL in prior labs. For example, `pdal merge [filename1.laz] [filename2.las] output.copc.laz` will use the `merge` subcommand and the file names that follow to merge many different las/laz/copc files together into the named output. Subcommands are generally good for small or incidental tasks like converting a file format or re-projecting data, but if you want to apply a more complex workflow then you should use a pipeline.

Pipelines are JSON-formatted files that give PDAL a set of instructions for reading, processing and writing point cloud data. With pipelines, you can define every step of the processing that you want and you can specify the finest level of detail at every stage. Pipelines are executed linearly, so we read the instructions from top-down. Below is an example of a pipeline that we are going to use, which is described in more detail below:

```
{
  "pipeline": [
    "AQ11.copc.laz",
    "AQ12.copc.laz",
    "AQ13.copc.laz",
    "AQ14.copc.laz",
    "AR11.copc.laz",
    "AR12.copc.laz",
    "AR13.copc.laz",
    "AR14.copc.laz",
    "AS11.copc.laz",
    "AS12.copc.laz",
    "AS13.copc.laz",
    "AS14.copc.laz",
    "AT11.copc.laz",
    "AT12.copc.laz",
    "AT13.copc.laz",
    "AT14.copc.laz",
    {
      "type": "filters.range",
```

```

        "limits":"Classification[2:2]"
    },
    {
        "type":"filters.crop",
        "bounds": "([-122.55275,-122.52506],[49.32325,49.34135])",
        "a_srs":"EPSG:4326"
    },
    {
        "type": "filters.merge"
    },
    {
        "type":"writers.las",
        "filename":"mkrf_aoi_lidar.las"
    }
]
}

```

This pipeline will take all of our input tiles and filter, crop, merge, and write them out to a new file called “mkrf\_aoi\_lidar.las”.

```
{
    "type":"filters.range",
    "limits":"Classification[2:2]"
},
```

This first stage applies a filter stage with a function called range. Basically, this filter is telling PDAL that we only want the points that are classified as ground returns Classification[2:2] where 2:2 indicates the range of values from 2 to 2, which is the ground return classification code. So only ground returns are passed to the next stage:

```
{
    "type":"filters.crop",
    "bounds": "([-122.55275,-122.52506],[49.32325,49.34135])",
    "a_srs":"EPSG:4326"
},
```

This next filter stage crops the ground returns from the previous stage using the bounds of our AOI. We need to specify the spatial reference system **a\_srs** of these coordinates since the LiDAR data are in a projected coordinate system (EPSG:26910). So only ground returns that fall within our AOI are passed to the next stage:

```
{
    "type": "filters.merge"
},
```

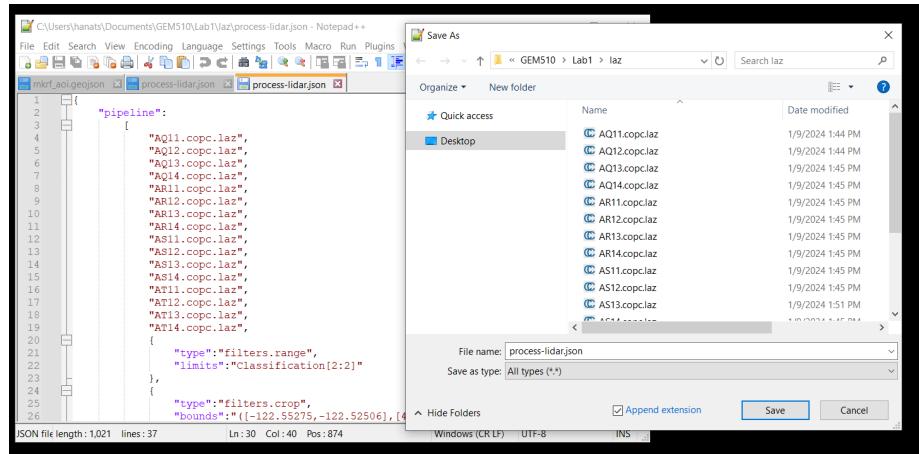
This next filter stage merges all the ground returns in our AOI into a single stream, which is then passed to the last stage that writes it to an output file in

## 14 CHAPTER 1. SPATIAL INTERPOLATION AND VISUALIZATION OF LiDAR DATA

LAS (uncompressed) format:

```
{  
    "type": "writers.las",  
    "filename": "mkrf_aoi_lidar.las"  
}
```

**Step 8:** Copy the contents of the pipeline to a text editor and save the file as “process-lidar.json” in your QGIS project folder. It should be in the same place as the .laz tiles you just downloaded. The output will also be written to this folder.



**Step 9:** Return to the OSGeo4W shell and run the pipeline using the following command: `pdal pipeline process-lidar.json`. It may take several minutes for this step to complete, but in the end you should have a file called “mkrf\_aoi\_lidar.las” in your QGIS project folder. Drag it into QGIS and inspect it. The processed file should be ~90000KB when finished. If you get an error at this step make sure you have downloaded the correct tiles and the filenames match what is being loaded into the pipeline. Once it completes, you can get the summary statistics of the points using the `pdal info` subcommand.

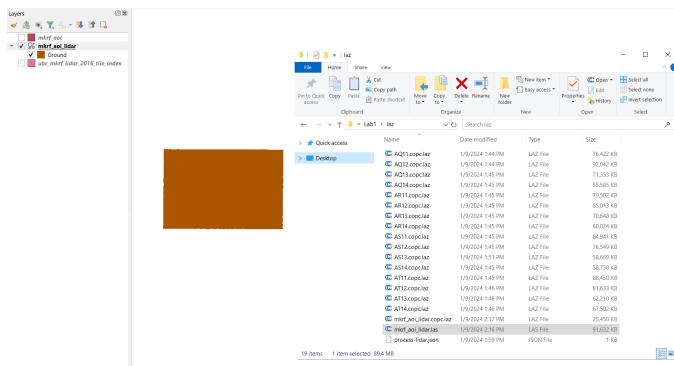
**Q1. What is the measured range of elevation by LiDAR within the AOI? (1 point)** **Step 10:** Finally, we will create a thinned point cloud to compare different interpolation methods in Task 3. In the OSGeo4w shell run the following command:

```
pdal translate mkrf_aoi_lidar.las thinned.las sample --filters.sample.radius=5
```

This will thin the point cloud by selecting a random point within a sphere with 5 m radius. The output will be called “thinned.las” and will be written to the current working directory.

## Task 2: Visualize LiDAR data in QGIS

The default symbology of LiDAR data in QGIS will be the classification, but we only have ground returns in our file, so everything will appear brown. Try symbolizing some of the other attributes like Intensity, ReturnNumber, and ScanAngleRank.

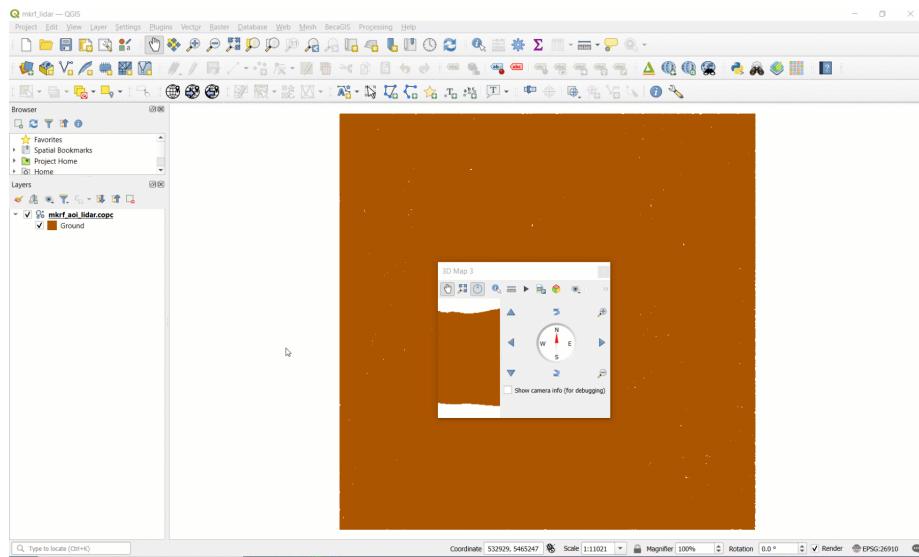


**Step 1:** Right-click the `mkrf_aoi_lidar` layer, open “Properties”, select “Symbology” from the left, and then select “Attribute by Ramp” from the very top drop-down menu. Then choose your attribute and apply your symbology parameters. Make some notes of your observations for these attributes that you can reference in your report.

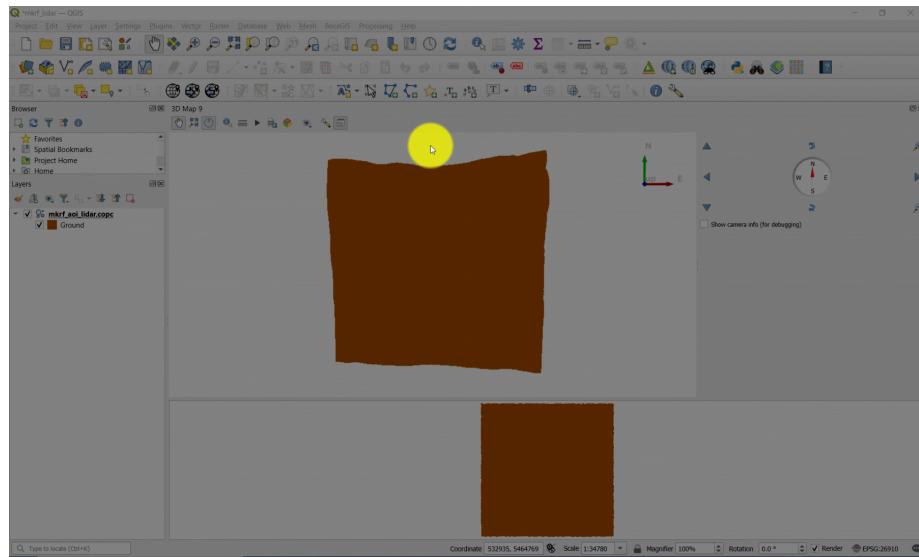
2D is a very boring way to visualize point cloud data, so let’s create a 3D map view in QGIS.

**Step 2:** From the top menu, select “View”, then “3D map views”, and click “New 3D Map View”. A small window will appear. Click and drag it to the top of your map canvas to dock it and make it larger.

## 16 CHAPTER 1. SPATIAL INTERPOLATION AND VISUALIZATION OF LIDAR DATA



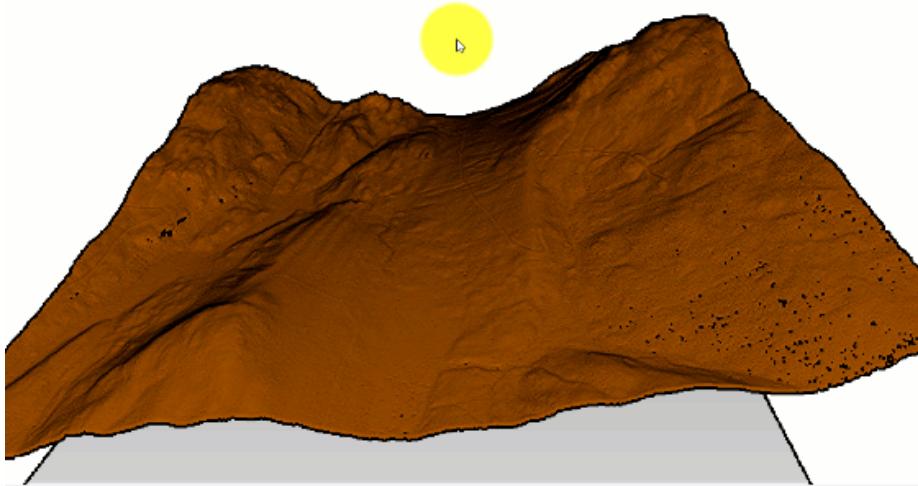
**Step 3:** From the 3D map view menu bar, click the tool icon and toggle on “Show Eye Dome Lighting”. Then, holding the SHIFT key, left-click and drag your cursor from top-to-bottom then release both buttons. This will give you an oblique shaded 3D perspective of the terrain.



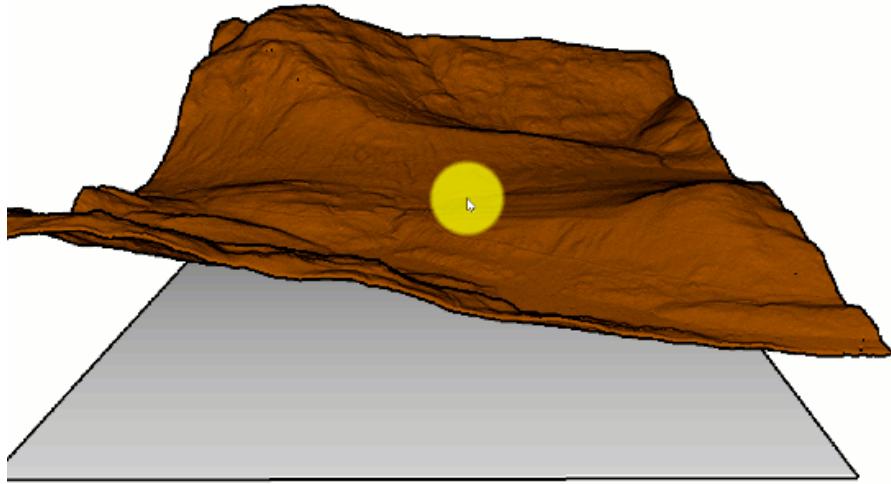
Navigating 3D data can be challenging using a 2D input device like a mouse, so we need to use different keyboard keys to control how we want to change our view:

**Hold SHIFT: Orbit camera around fixed position**

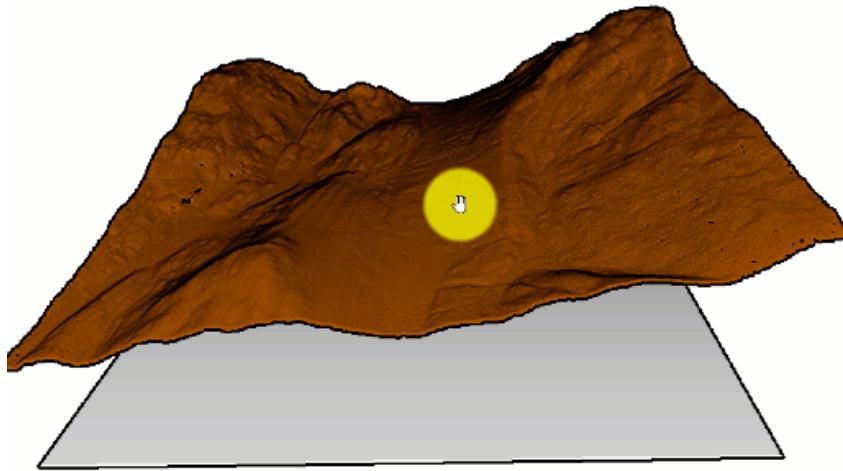
If you want to pan around a fixed position, then hold the SHIFT key and drag your cursor. The point cloud will rotate in the opposite direction. As you can see from the animation below, dragging your cursor in a circular pattern rotates the point cloud around a stationary imaginary point. If you continuously drag your cursor to the right, you will rotate around the fixed point counter-clockwise.

**Hold CTRL: Maintain camera position and change camera angle**

If you want to pan your camera angle up, down, left or right, then hold the CTRL key and drag your cursor in the direction you want to look. This only changes the camera angle, not the camera elevation or position.

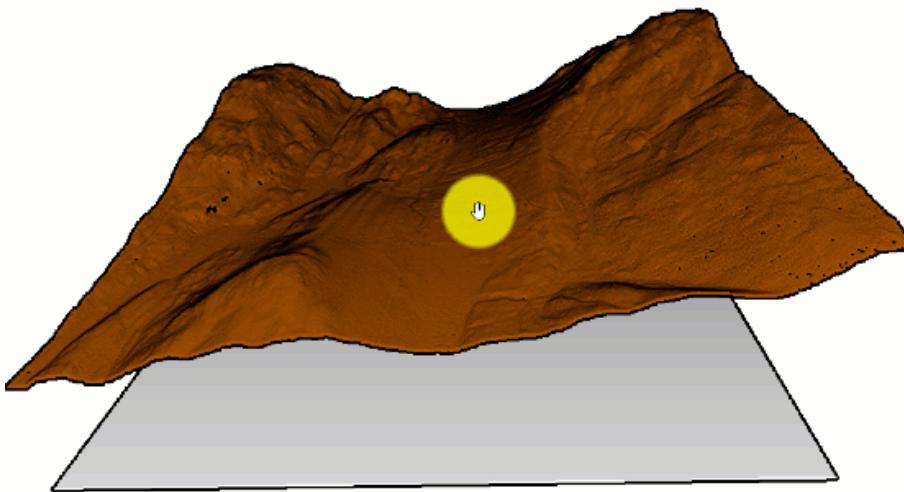
**Hold ALT: Move camera position on the X-Y plane**

If you find yourself wanting to move around, then hold the **ALT** key and drag your cursor in the direction that you want to move the camera. Think of this as sliding along an X-Y plane that is fixed at the camera elevation and angle.

**Scroll: Move camera position to/from the cursor location**

If you want to move towards some feature, then simply point your cursor at it and scroll without clicking. This movement is akin to traversing a ray that

connects your current camera position with a look direction angle (relative to X-Y-Z). Scrolling therefore changes the camera position and elevation. Scrolling down has the effect of “zooming in” while scrolling up has the effect of “zooming out”. Be careful, though, because scrolling will always follow the current position of the cursor, so if you scroll down on one position and then move your cursor and scroll up, your camera position will not be where it started.



**Step 4:** Practice navigating in 3D and observing the different attributes across the ground surface. Make some notes of your observations for your report. There is a button on the top of the 3D map view to save an image of any given view.

---

### Task 3: Prepare LiDAR in ArcGIS Pro

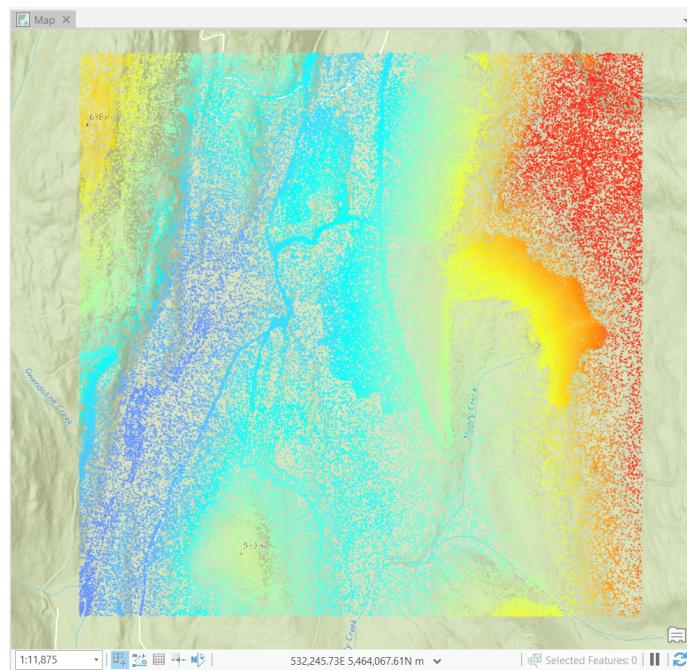
For this task, we will switch to ArcGIS Pro and practice manipulating the LiDAR point cloud to derive a terrain surface using different algorithms. ArcGIS Pro has several tools that we can use to view and analyse LiDAR point clouds. In order to view the dataset, we need to import it as a LAS Dataset. Note that QGIS will open LAS or LAZ files and immediately convert them to COPC format (.copc.laz), but ArcGIS Pro does not currently support reading COPC, so we must use the LAS file. If you want to reproduce this lab with other LiDAR data in ArcGIS Pro, you can convert LAZ to LAS using the “Convert LAS” tool.

**Step 1:** Open a new ArcGIS Pro map project and Search for the “Create LAS Dataset” tool. Specify the “mkrf\_aoi\_lidar.las” file as your input file. Make sure to name the output LAS Dataset and specify the correct coordinate system. Ensure that “Create PRJ for LAS files” is set to “All LAS Files”. Toggle on

## 20CHAPTER 1. SPATIAL INTERPOLATION AND VISUALIZATION OF LIDAR DATA

“Compute Statistics” and run the tool. This will produce a LAS Dataset file (.lasd). **Repeat this step for the thinned las dataset.**

**Step 2:** We can now add our LAS Datasets to the map. Depending on the zoom extent, you may only see the red bounding box of the LAS Dataset file; this is not an error, you just need to zoom in to see the actual points. Alternatively, you can open the dataset in a Local Scene, although due to the size of the point cloud this might cause some lag. Compare the full and thinned point clouds.

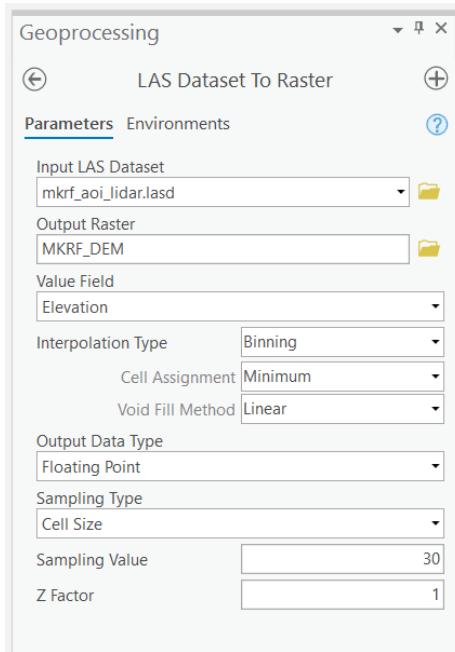


**Step 3:** Another way to explore the dataset is to view the properties. Right-click the LAS Dataset and open the “Properties”. Here we can see some statistics of the point cloud, such as information regarding the Classification Codes and Return Number. You can get more detailed metadata from a Catalog view. From the top ribbon, select the View tab and then click “Catalog View”. Navigate to where you saved the LAS Dataset, right-click it, and open the “Properties” again.

**Q2. What differences do you observe between the thinned and non-thinned datasets in terms of number of points, average point spacing, and the measured range of elevation? (3 points)** **Step 4:** Search for the “LAS Dataset to Raster” tool, and use “mkrf\_aoi\_lidar.las” (not thinned) as the input. Since we are interested in creating a terrain surface model, we want to use the “Binning” “Interpolation Type”, and make sure that we use the “Minimum” (i.e., the lowest “Elevation”) points in each “Cell Assignment” of the output raster. “Sampling Value” refers to the resolution of the raster that

we are creating, in other words, the spacing of our raster cell samples. Set the “Sampling Value” to **1** (meters), name the output raster “MKRF\_DEM” and run the tool.

**Q3. How might changing the parameters of this tool (“Interpolation Type”, “Cell Assignment”, and “Sampling Value”) affect the statistical qualities of the output DEM? (6 points)**



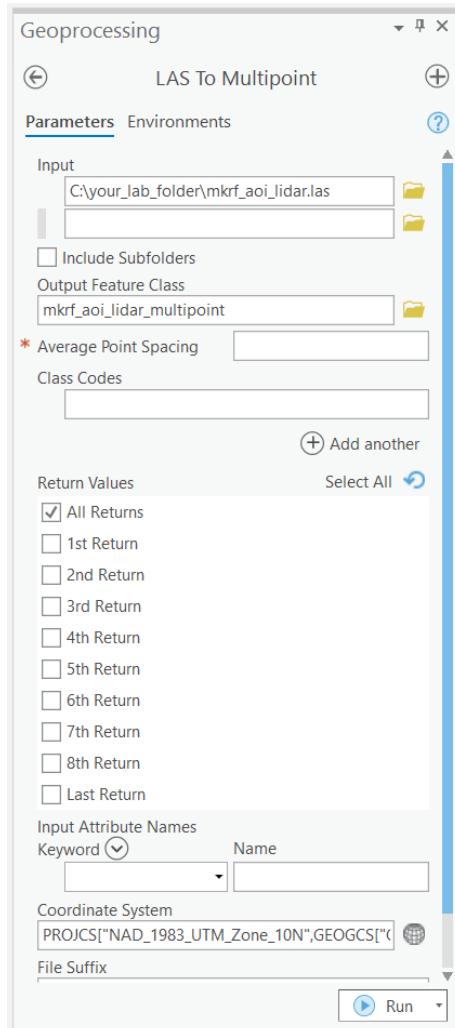
The tool above is the most straight-forward way to create a DEM from a LiDAR point cloud in ArcGIS Pro. Because this DEM was fit using all available lidar points we will consider it to be our “ground-truth” dataset.

In the following steps you will create a 2D point dataset using the **thinned LiDAR point cloud**. We will use this to test and compare three interpolation methods in the next task.

Note, for future reference you can use these methods to spatially interpolate all other kinds of attributes that are not terrain!

**Step 5:** Search for the “LAS to Multipoint” tool. Input your thinned LAS Dataset. You will notice that there is a box asking for the “Average Point Spacing”. Enter the value that you recorded from Step 3. Set the correct coordinate system and then run the tool.

## 22CHAPTER 1. SPATIAL INTERPOLATION AND VISUALIZATION OF LIDAR DATA



**Step 6:** Search for the “Multipart to Singlepart” tool. The multipoint feature class from the previous step is the input then run the tool. ArcGIS Pro will attempt to draw the symbols for every single point in the multipart and singlepart feature classes, so keep them toggled off.

**Step 7:** Now we need to add the Z value as an attribute. Search for the “Add Z Information” tool and use the singlepart features as the input. The only option available should be “Spot Z” and make sure it is toggled then run the tool. This may take several minutes.

We now have a 2D point feature class that we can use to compare different spatial interpolation algorithms!

## Task 4: Apply and evaluate spatial interpolation algorithms in ArcGIS Pro

Since we now have points representing height values, we can use the raster interpolation toolset to experiment with three different interpolation methods: Natural Neighbor, Inverse Distance Weighting (IDW), and Spline. You can read more about the tools we will be using from the ArcGIS Pro documentation.

Search for each of the interpolation tools listed above individually (for IDW, the tool is called “IDW”) and explore their parameters. Natural Neighbor can be found within the “LAS Dataset to Raster” tool, change the “Interpolation Type” to “Triangulation”, and “Interpolation Method” to “Natural Neighbor”.

**Step 1:** Create three rasters for each interpolation method: Natural Neighbor, IDW, and Spline. We suggest writing these rasters out to GeoTiffs so that you can also view them in QGIS, if you want. For each tool, keep the default settings, but make sure that you generate the rasters at 1 m resolution. Note that the Spline tool will take several minutes to run. If you have trouble running this step try generating the rasters at a slightly coarser resolution (5 m).

Take some time to inspect each raster, and look at their similarities and differences (make sure that you are using a common symbology when comparing). We can compare the differences between our interpolated rasters with the binned DEM that we created in the last task.

**Step 2:** In order to compare the rasters, we will make a raster of the differences between the interpolated surface and the binned DEM. You will use the “Raster Calculator” tool to do this. Subtract each interpolated surface from **MKRF\_DEM**. Name each difference raster according to the interpolation method: “spline\_diff”, “nn\_diff”, and “idw\_diff”. We suggest writing these rasters out to GeoTiffs so that you can also view them in QGIS, if you want.

Next, we are going to quantitatively evaluate the differences and calculate statistics over different elevation and slope classes to examine how the interpolation methods perform.

**Step 3:** Change the Symbology of the **MKRF\_DEM** to Classify and set the number of Classes to 3 with the method of your choice. Use the “Reclassify (Spatial Analyst)” tool to reclassify the binned **MKRF\_DEM** into “high”, “medium”, and “low” areas using this classification scheme. (If you symbolize first, then open the Reclassify tool it should populate the start and end fields with the symbolized values.)

**Step 4:** Use the “Slope (Spatial Analyst)” tool to create a raster called “mkrf\_slope” using degrees. Then, use the “Reclassify” tool to reclassify the degree slope values into three classes using the same method.

We will use these reclassified areas as zones to calculate statistics about the difference between the interpolated surfaces and the binned DEM.

## 24CHAPTER 1. SPATIAL INTERPOLATION AND VISUALIZATION OF LIDAR DATA

**Step 5:** Search for the “Zonal Statistics as Table (Spatial Analyst)” tool. The “Input Raster or Feature Zone Data” are the reclassified zone rasters you just made. “Zone Field” should be “Value”. The “Input Value Raster” is the raster that we want to summarize over the zones, which are all of the difference rasters we calculated in the earlier step. Run this tool for both each combination of the two topography zonations and the three difference rasters for the different interpolation methods. You should finish with six tables:

- idw\_diff\_elev
- idw\_diff\_slope
- nn\_diff\_elev
- nn\_diff\_slope
- spline\_diff\_elev
- spline\_diff\_slope

**Q4.** Summarize and describe how the interpolation methods differ from the binned DEM across elevation and slope classes. Which zones/methods are most similar or dissimilar? (10 points)

**Table that shows zonal statistics of the difference rasters across elevation and slope ranges of the binned DEM (10 points)** **Step 6:** Create a final map with the following elements:

- Each of the interpolated surfaces and difference rasters (6 total)
  - The interpolated surfaces should be symbolized using **Shaded Relief** symbology
- Summary statistics showing the mean and standard deviation across slope and elevation zones
- North arrow and appropriate labels for each map panel

**A panel of maps that symbolizes the DEMs produced and the difference rasters that were produced by the comparison to the binning approach (15 points)** Note that you can add standalone tables to your Map Layout by first clicking on the table in the Contents pane then Insert > Table Frame. Click and drag to set the extent of the table on the map, (similar to how you add a Map Frame).

**Q5.** Which spatial interpolation algorithm would you recommend the research forest use? Defend and justify your choice. (5 points)

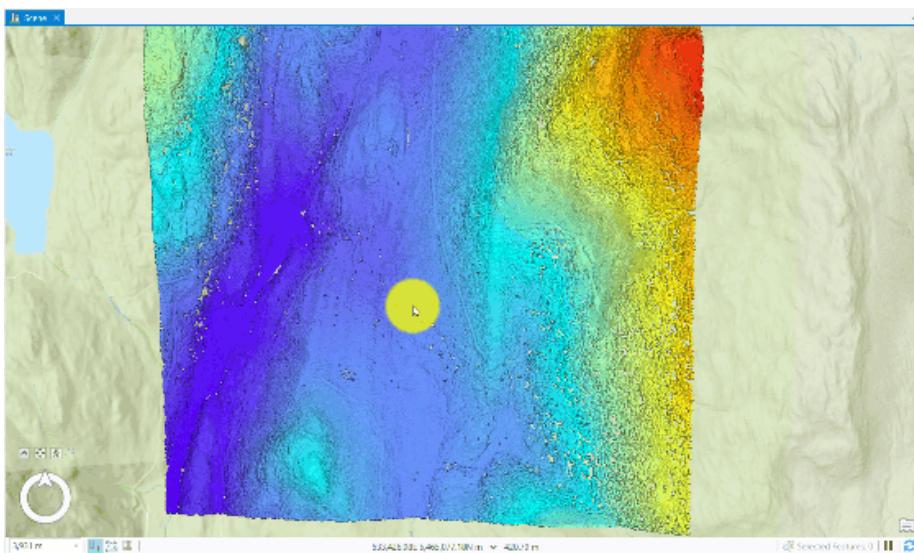
---

## Task 5: Visualize 3D data in ArcGIS Pro

In this last brief task are some instructions for visualizing the point cloud and interpolated DEM surfaces in ArcGIS Pro. You can produce and export some 3D scenes that may help to illustrate your observations in your report.

**Step 1:** In ArcGIS Pro, click the “Insert” tab, click and expand “New Map”, and select “New Local Scene”.

**Step 2:** Add the “mkrf\_aoi\_lidar.las” file to the scene. Navigating an ArcGIS Pro scene is nearly the same as it is in QGIS. Though, instead of using keys on the keyboard to toggle between the different camera axes to manipulate, ArcGIS Pro provides an on-screen navigator in the bottom left. In this way, you can manipulate an ArcGIS Pro scene using just your mouse.



By default, ArcGIS Pro uses its own “WorldElevation3D/Terrain3D” source for the “Ground” in the Contents Pane. You can change the ground source to any DEM or really any other raster that you want to use the values of to create a “ground” surface. We are going to use this feature to explore the differences between the surface models we derived and the actual point cloud.

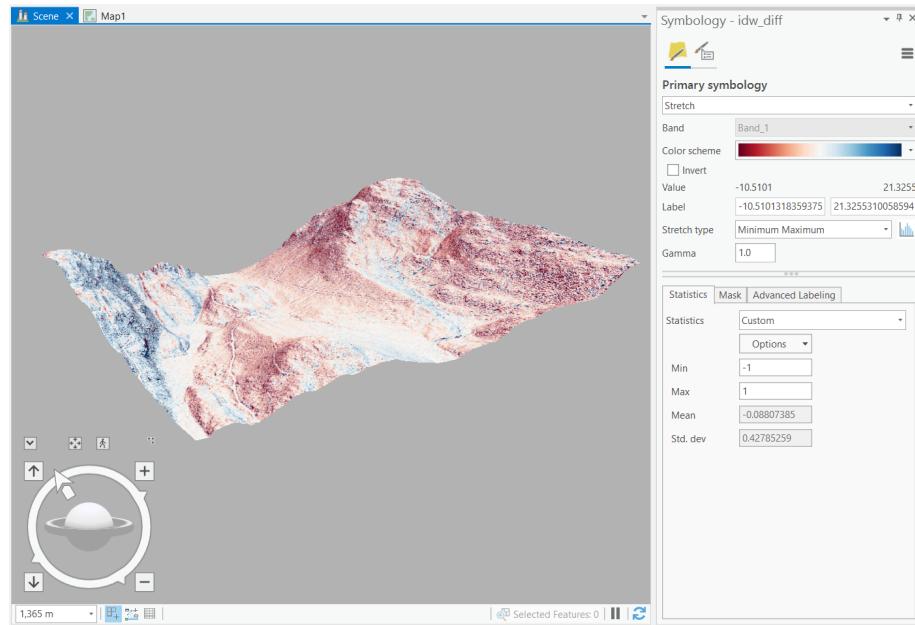
**Step 3:** Add the three interpolated DEM rasters to the scene and also the difference rasters. Toggle everything off in the scene. Drag the three DEM rasters under “Ground”. Leave the three difference rasters under “2D Layers”.

**Step 4:** For each difference raster, toggle it on and change the symbology to “Stretch”, using a diverging “Color Scheme” and a “Minimum Maximum” “Stretch Type”. Below in the “Statistics” tab, click “Dataset” and then select “Custom” from the drop-down menu. Modify the “Minimum” and “Maximum” values so that they are the same magnitude, but one is positive and the other is

negative. You select the appropriate values based on the range of values that you observe in the difference raster and you will need to test different values to get your desired effect. For example, if the range of difference values is -10.51 and 21.32, then you might try setting the “Minimum” to -10 and the “Maximum” to 10. If you want the colors to be more apparent, then incrementally reduce the magnitude of your values: try -7, 7 next; then -5, 5; etc. until you reach your desired visualization. It is important to ensure that the midpoint is zero when using a diverging color scheme.

**Step 5:** Toggle on each pair of DEM (under “Ground”) and difference raster (under “2D Layers”). The ground source adds elevation to the difference raster that you symbolized, but does not impact or change the point cloud because it is inherently 3D data. Since we are using our own custom ground source, we can actually pan and observe a cross-section of the point cloud sitting on, above, or below the ground source and then observe local deviations between the point cloud and the derived DEM surface.

The white background might make it difficult to observe the differences. You can change the background color of the scene by left-clicking on “Scene” in the Contents Pane, open the “Properties”, and then select the background color. We suggest using a light grey. There are other options for illumination that you can also modify in the “Properties”, if you want. The image below shows an example of what the IDW might look like:



## Summary

We covered a lot of ground in this lab about LiDAR processing with GIS software. One thing that should be apparent is that the vast majority of your time will be spent preparing LiDAR data for analysis. Point clouds are dense. Advances in scanning technology are creating new challenges for managing, accessing, organizing, and manipulating these dense data. Remember, we always want to work with the least amount of LiDAR data that our task requires. For spatial interpolation of terrain, this means only working with the necessary ground returns in our AOI. By filtering the tiles and points to only those we needed for the interpolation task, we reduced our data overhead by over 650%! You also practiced evaluating different spatial interpolation algorithms, which can have a large impact on subsequent analyses with those surfaces. When choosing an interpolation algorithm, you must always consider how dense your data are (i.e., how long will it take for me to produce this surface), what properties you want the output surface to have (i.e., statistical intervals, reproducing the true ranges, etc.), and what the surface will be used for (i.g., hydrology mapping, suitability analysis, etc.). Finally, you also practiced visualizing 3D data to augment and enhance your interpretations of these 2D interpolated surfaces.

Return to the **Deliverables** section to check off everything you need to submit for credit in the course management system.



## Chapter 2

# Salmon Stream Network Analysis

Written by Paul Pickell

### Lab Overview

An important application of GIS for watershed management is characterizing the physical components of streams, rivers, and water bodies. In this lab, you will learn how to use the Hydrology, Network Analyst, and Linear Referencing toolsets in ArcGIS Pro to extract hydrologic information from a digital elevation model (DEM) in order to model a network of streams.

In British Columbia, salmon are an important species for economic, cultural, and ecological reasons. Thus, there is extensive research and management efforts to better understand and sustainably manage salmon populations. There are five Pacific salmon species, Chinook, Coho, Chum, Pink, and Sockeye salmon. Salmon are diadromous species, spending their lifecycle in both freshwater and the ocean. We are particularly interested in habitat for salmon spawning. Salmon return to freshwater habitats to spawn and there are various requirements for salmon spawning that differ by species:

- Stream order
- Proximity to a lake
- Upwelling/downwelling
- Water quality
- Water depth
- Water velocity

- Substrate material
- Slope gradient
- Land cover

For this lab, you will identify the preferred habitat for salmon spawning based on proxy measures and how it corresponds to our stream network that we have derived from the DEM. This lab will walk you through how to measure stream order and slope. For your final report, you will also be expected to analyze an additional habitat requirement for salmon spawning. (Note: if you use some information that is not on the aforementioned list, please describe how it is a habitat requirement for salmon spawning in your report.)

In this lab, we will focus on Vancouver Island for a few reasons. First, islands make it possible to calculate flow accumulation exactly, which is important for creating a (nearly) topologically correct stream network. Second, Vancouver Island is large, which will enable us to look at natural sources of topological errors. Third, many species of Pacific salmon return to highly developed Vancouver Island to spawn and often come into conflict with human development.

---

## Learning Objectives

- Derive hydrology layers from a DEM
  - Model dams on streams as barriers by applying network topology
  - Evaluate a capability model of salmon habitat along stream segments using linear referencing
  - Trace a stream network and calculate reachable upstream salmon habitat
- 

## Deliverables

Lab report with the following specification:

6 pages maximum PDF including figures, tables and references (5 points). Single-spaced, 12-point Times New Roman font (1 point). All figures and tables should be properly captioned (1 point).

Results should address the following questions and requirements (25 points):

- **You are only expected to comment on a single conservation unit.**
- Map showing the reach of your streams from the ocean with dams as barriers

- How many dams/ocean outlets are in your conservation area?
- How much of the stream network is accessible and inaccessible?
- How well do the streams you created line up with the mapped streams in the ArcGIS basemap?
- Zoom to Lake Cowichan. What is going on here? What could you do to automatically fix this for all large lakes?
- Report any other metrics that you calculated that might be useful for making your arguments in the Discussion section.
- Comment on both small and large scale observations that you made throughout the process. Small scale observations would be investigating the behavior/structure of individual routes, while large scale observations would be the overall differences between total salmon habitat (all streams) and habitat that can actually be reached unimpeded from the ocean.

Discussion should address the following questions and requirements (20 points):

- Why are reach and connectivity important for salmon conservation?
  - What are some approaches to modelling reach within networks? Connectivity?
  - Consider how you might model the other salmon habitat requirements laid out at the beginning of the lab. Give brief examples for how you would model at least three of the criteria that we did not consider in the lab (i.e., do not discuss stream order, stream gradient). Assume that you have access to any typical data source (e.g., LiDAR, optical satellite, RADAR, land inventory, gauge stations, weather stations, etc).
  - How would you expect the stream network to change if you used a raster with a cell size of 5 m or 100 m?
  - How might additional topological errors and unidentified/unfilled sinks in the DEM impact your analysis?
  - Discuss any limitations to the analysis based on your observations and suggest how the modelling process might be improved further.
  - What are your final recommendations to the province for conservation of spawning salmon habitat in Vancouver Island streams?
  - Reference any peer reviewed sources as needed.
-

## Data

All data for this lab are accessible via the UBC PostgreSQL server. Instructions for connecting to the server are given in prior labs. We will be using data from the `salmon` database. To save time for this lab, the data have already been assembled and pre-processed for you, but below are some details about how that was achieved.

### `vancouver_island_dem`

This was the general process applied to create the DEM in case you want to reproduce it or apply the same method for another project:

1. Download all the tiles for Vancouver Island from the B.C. Government [here](#)
2. Mosaic the rasters together using the “Mosaic to new raster” tool
3. Buffer a polygon shapefile of Vancouver Island by 25 m
4. Convert the result of step 3 to a raster with cell size 25 m and project to NAD 1983 BC Environment Albers. This is a mask that extends just beyond the shore of Vancouver Island to account for mapping errors.
5. Extract the DEM from step 2 by the mask in step 4 to produce a DEM for only Vancouver Island that excludes all other nearby islands:



The DEM above was filtered several times for sinks. Generally, you can use the “Fill” tool to achieve this. However, Vancouver Island has many naturally-

occurring geomorphic features that can cause real sinks (as opposed to random errors). In order to create a depressionless DEM, the DEM was iteratively put through the following process:

1. Calculate flow direction
2. Calculate sinks
3. Calculate watersheds using the flow direction raster and the sinks raster as pour points
4. Calculate minimum elevation of the sinks using the “Zonal statistics” tool
5. Calculate maximum elevation of the watershed for the sink using the “Zonal fill” tool
6. Subtract result of step 4 from result of step 5 for the sinks mapped in step 2. This is the depth of each sink.
7. Add the sink depth to the DEM to fill the sink
8. Repeat steps 1-7 several times until you remove all sinks or achieve a stopping criterion

The original DEM contained >3,000 sinks. After iterating the above 12 times, there were fewer than 300 sinks. You can subtract the DEM from the filled DEM to observe where fill was added. Notice any patterns for where sinks are located on the island?

#### **dams**

The dams were acquired from the BC Government here. These data were clipped to Vancouver Island and represent dam structures as polyline features.

#### **Salmon conservation units (CU)**

Conservation units for Chinook, Chum, and Coho were acquired from Oceans and Fisheries Canada here. These polygon areas were mapped considering ecology, life history, and molecular genetics (Waples et al 2001).

---

### **Task 1: Create stream segments**

**Step 1:** Before starting the lab, you will need to download all of the data from the **salmon** database on the UBC PostgreSQL server. Currently, ArcGIS Pro cannot read Post GIS raster datasets so we will first open the Vancouver Island DEM in QGIS and Export it as a GeoTiff, which can be opened in ArcGIS Pro.

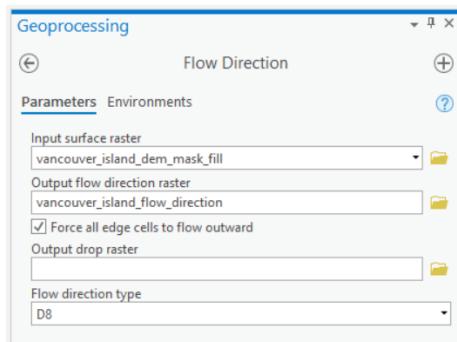
Open QGIS and connect to the **salmon** PostgreSQL server. Expand the **salmon** database in the Browser pane and right-click on “vancouver\_island\_dem” > Add Layer to Project.

Right click on the DEM layer in the Layers pane > Export > Save as. Save the file as a GeoTiff in your Lab 2 project folder. Leave the default settings. Close QGIS.

**Step 2:** In ArcGIS Pro, add the Vancouver Island DEM to your map project. Clip the DEM to the “vancouver\_island\_boundary” using the clip raster tool, toggle on “use input features for clipping geometry”.

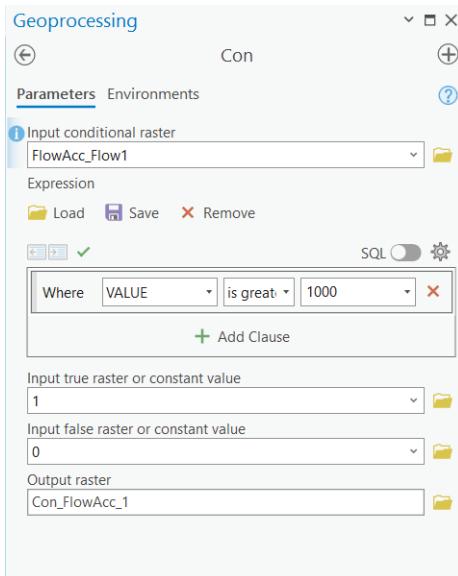
You should save all of your outputs from the tools in this lab directly into your ArcGIS Pro project geodatabase. The reasoning for this is because we will only be working in ArcGIS Pro for this lab and also because the ArcGIS Pro topology tools only work with data inside an ArcGIS Pro geodatabase.

Calculate the flow direction of the filled DEM, using the “Flow Direction” tool. The flow direction tool determines which raster cells flow into other cells based on the elevation difference the immediate neighborhood. Keep the defaults of the tool. Ensure that “Force all edge cells to flow outward” is toggled on and leave all other parameters with the default value. The cell values in the output raster represent the steepest down slope direction of each pixel.



**Step 3:** Use the output from the “Flow Direction” tool as an input into the “Flow Accumulation” tool. Again, use the defaults within the tool, but select integer as the output data type because the output of this tool is a count. This tool will create a raster that represents the number of pixels that will accumulate into any given pixel. You may need to zoom into the output to be able to see these streams. Play around with the symbology to visualize the stream network. (Hint: try using classify and various methods within symbology).

**Step 4:** To define streams, you need to set a threshold for the flow accumulation. Open “Con”(Image Analyst Tools). This tool allows you to set a true/false condition and assign values to a new raster based on whether the condition is met. Set the expression so that a pixel that has a count equal to or more than 1000 will be assigned a value of 1. For “Input true raster or constant value”, use value 1. For “Input false raster or constant value”, use value 0.

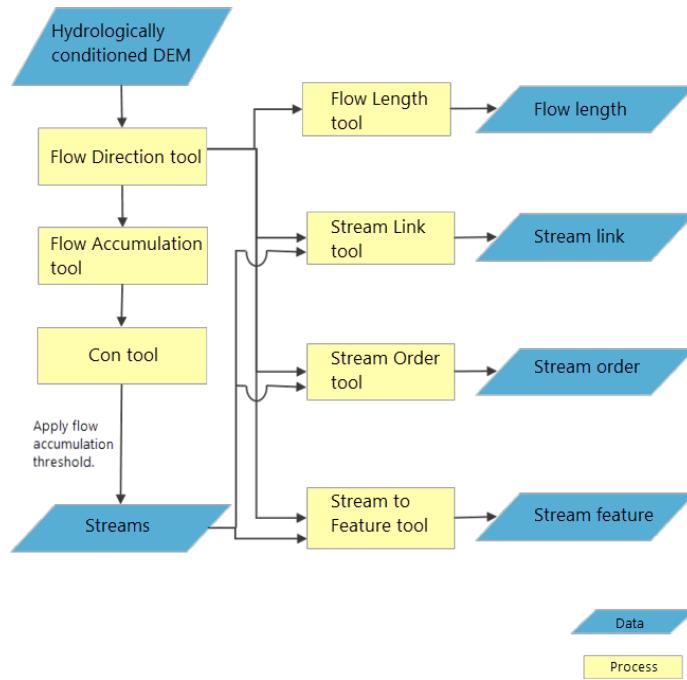


**Step 5:** Next, you will create a grid of stream links, which assigns unique values to each stream segment, so that the user can see how the streams are connected. Open the “Stream Link” tool and use the output from the “Con” tool and the flow direction raster here.

**Step 6:** You will also use the “Stream Order” tool using the same inputs as in step 4 and the default parameters. The “Stream Order” tool assigns a numeric order to segments of the raster, essentially classifying branches of the stream network. The lowest order streams do not have any other streams that flow into it, while higher order streams can have multiple levels of tributaries that flow into it. This tool will take some time to run on the whole island. Take a moment to stretch and hydrate!

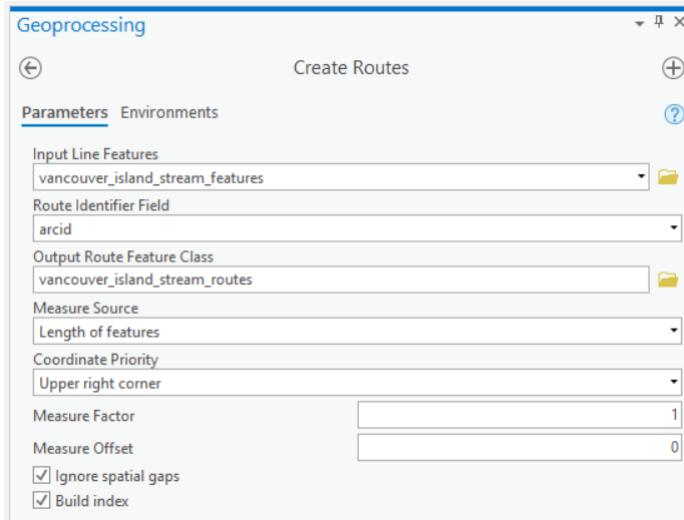
**Step 7:** You will also convert your raster into polylines using the stream links raster and flow direction raster as inputs for the “Stream to Feature” tool. Toggle off “Simplify polylines”.

Below is a useful flowchart of the process of creating a stream network from a DEM (ArcGIS Pro).



## Task 2: Create routes and apply topology

At this point, you have only created polyline features representing streams. Next will convert those segments to routes and apply topology for our streams, as well as incorporate dams into the network.

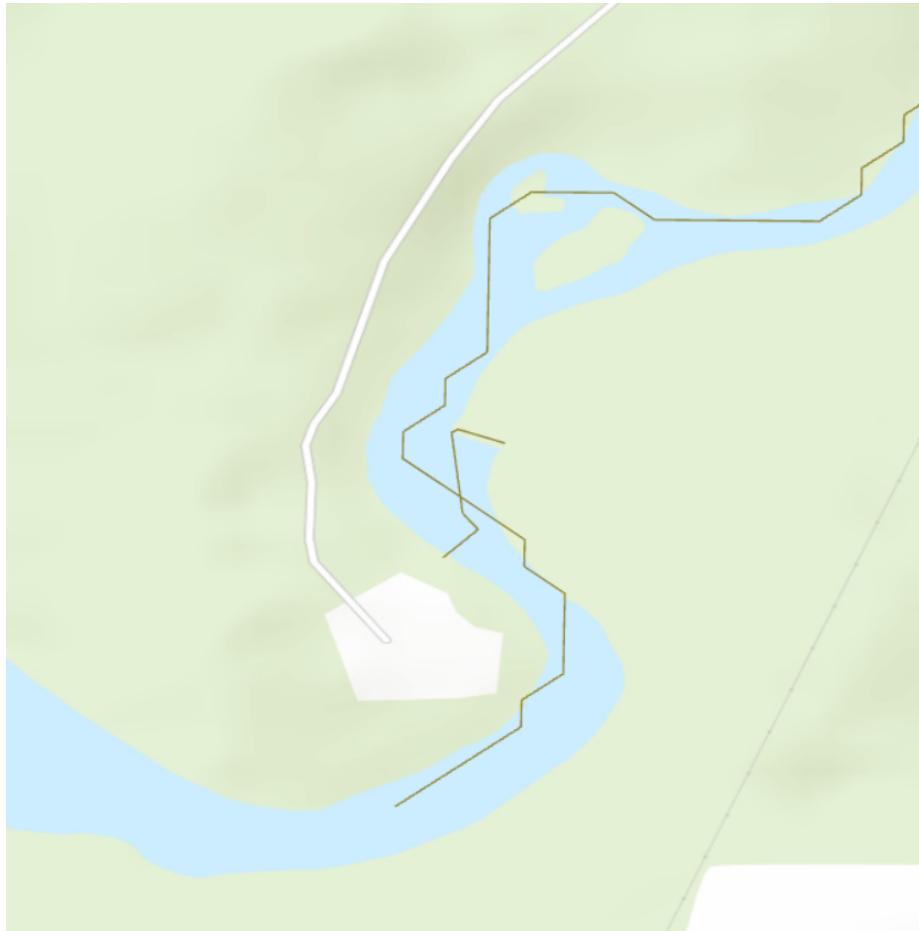


**Step 1:** Create a blank feature dataset and name it “stream\_routes\_dams”. To do this, right-click the geodatabase in the Catalog pane and select New Feature Dataset. For the projection, use NAD 1983 BC Environment Albers (EPSG: 3005).

**Step 2:** Convert your stream features into routes using the “Create routes” tool. Set the Route ID field to “arcid”. The coordinate Priority parameter can be set to any corner. When you go to save your output, navigate into your feature dataset (it “feels” like entering another folder from the Catalog view), and save the output as “vancouver\_island\_stream\_routes”.

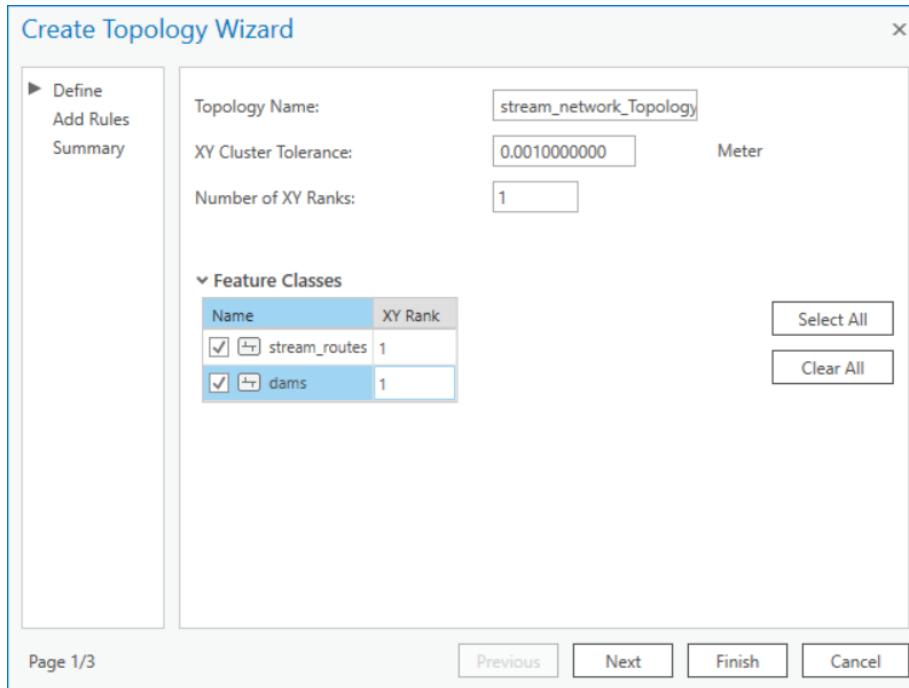
**Step 3:** Right-click on the feature dataset, select “Import”, and then navigate to the **dams** layer on the UBC PostgreSQL server and import this layer into your feature dataset.

The dams are modelled as polylines. We need to identify where these dams intersect the stream routes. We want to return a single point for that location so that we can properly model the dams in our network, which act as barriers to salmon as they travel through the stream network. We will use topology to achieve this.

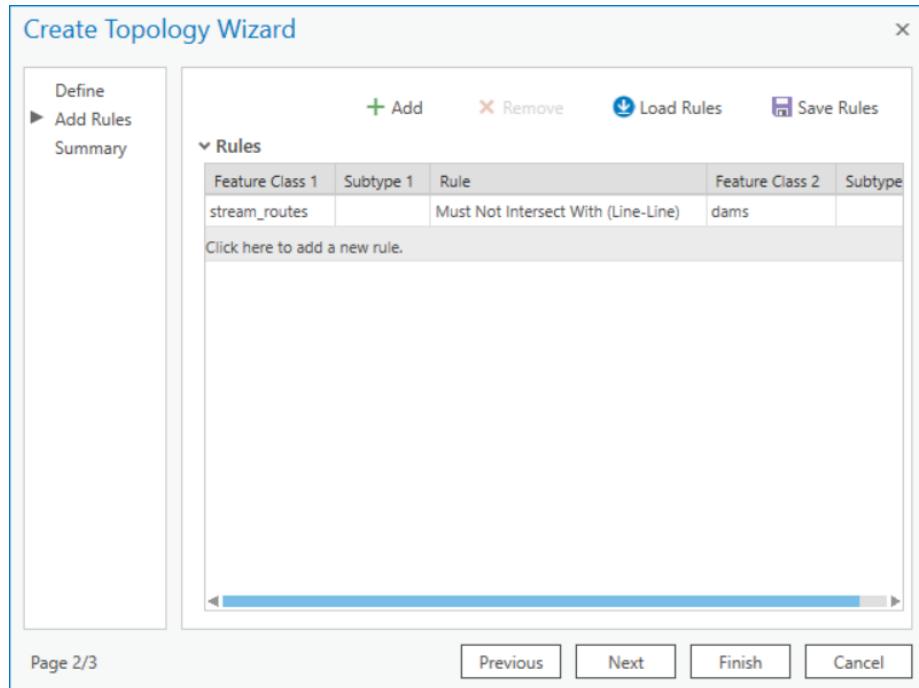


**Step 4:** Right-click the feature dataset and select “New” and “Topology”. This will open the Create Topology Wizard dialogue box.

**Step 5:** On the first page of the wizard, toggle on the dams and stream routes feature classes.

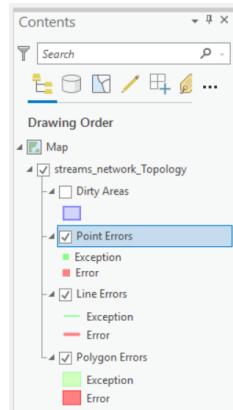


**Step 6:** On the second page of the wizard, add a new rule so that the stream polylines must not intersect with the dam polylines.



Once done, click “Finish”. Refresh the feature dataset in the Catalog pane to see the new Topology.

**Step 7:** Validate the topology by right-clicking the topology in the feature dataset and clicking “Validate”.



**Step 8:** You can click and drag the topology into the map to view any errors. We expect there to be errors because we know that the dams should intersect with our stream routes. You should see red squares all over the island. In the Contents Pane, right-click the “Point Errors” layer, select “Data”, and select “Export Features”. Save the points to your feature dataset

as “stream\_dam\_junctions”.

At this point, take a moment to appreciate where we are in this problem. We have modelled streams as routes and we now know where dams intersect with those routes. Salmon are diadromous species, they swim from the saltwater ocean up the freshwater streams to spawn. Therefore, we need one more piece of information in order to model how salmon swim upstream. We need to know where the streams meet the ocean.

**Step 9:** Take the Vancouver Island boundary polygon provided and perform a negative buffer, called a nibble. Use -25 meters for the buffer distance. We are doing a negative buffer because we want to ensure that this edge intersects with the streams that terminate at the ocean. Remember that the streams were generated by an imperfect and imprecise DEM, therefore this process will ensure that every stream that reaches the ocean will intersect this boundary.

**Step 10:** Convert the nibbled Vancouver Island boundary to a polyline using the “Polygon to line” tool. Be sure to toggle off “Identify and store polygon neighboring information”. **Save the result to your feature dataset with the stream links and dams.**

Next, we will use the same topology process to identify where the island boundary intersects the stream routes, representing where salmon enter the stream network from the ocean.

**Step 11:** Right-click the network topology in the Catalog and select “Properties”. This will open the topology wizard dialogue again. On the left, select “Feature Class” and toggle on the nibbled Vancouver island boundary that you created in step 10 and the stream routes. Toggle everything else off. On the left, select “Rules”. Highlight the previous rule and click the “Remove” button at the top. Then, add a rule so that the stream routes must not intersect with the nibbled island boundary. Refer to steps 5 and 6 above if you need clarification.

**Step 12:** Again, validate the topology, add the topology to the map view, and export the Point Errors to a point feature class in your feature dataset. Name it “stream\_ocean\_junctions”. Refer to steps 7 and 8 if you need clarification.



---

### Task 3: Record stream attributes with linear referencing

Now that we have stream routes, we can calculate some information that relate to salmon habitat. Water quality is important for salmon spawning habitat. Steeper gradients (slope) may be more susceptible to erosion that could be one proxy for lower water quality. In this task, you will practice extracting information about the streams and storing that information in the routes using linear referencing.

**Important:** Some of these operations can be computationally intensive because we will be dealing with up to  $10^5$  features for the entire island.

**Step 1:** Calculate the slope of the filled DEM using “Percent rise” as the output measurement and “Planar” as the Method.

**Step 2:** Now use the “Extract by Mask” tool to extract the slope raster by the mask of the stream links raster you created back in Task 1. This will produce a raster with slope values only where our streams exist.

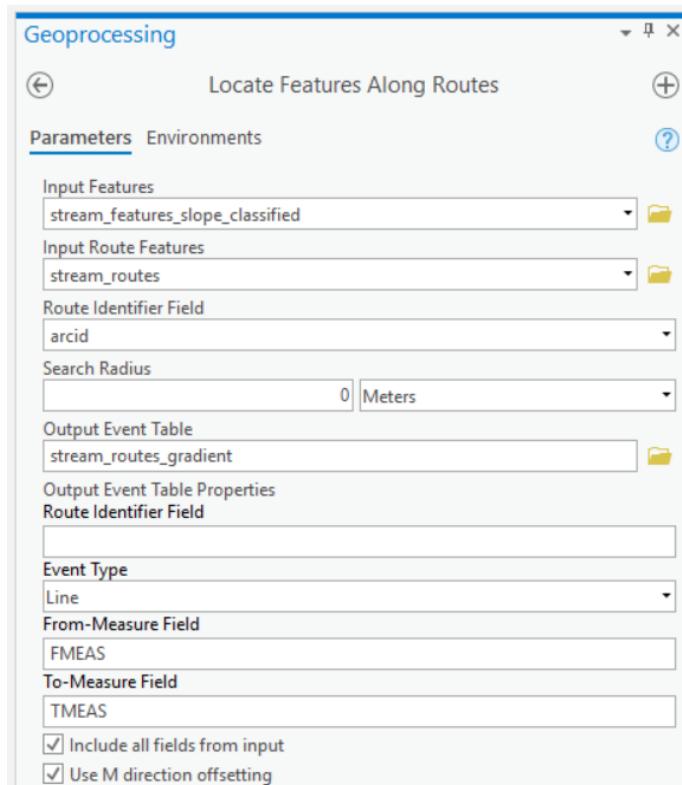
**Step 3:** In British Columbia, stream gradients  $< 20\%$  are generally considered to be fish-bearing. Therefore, use the “Reclassify” tool to classify slopes  $< 20\%$  to a value of 1 and all other slopes to a value of 0 for the stream raster you created in the last step.

**Step 4:** Use the “Stream to Feature” tool to convert the classified slope raster to line features. Parameterize this tool as follows:

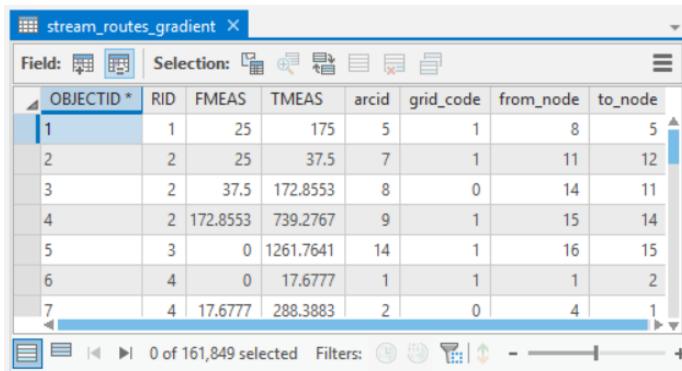
- Input stream raster is the classified slope raster you just made in step 3
- Input flow direction raster is the original flow direction raster you made in Task 1
- Name the output “stream\_features\_slope\_classified”
- **Uncheck** “Simplify polygons”

Now you have another feature class of streams that should spatially match your stream routes, but these new polylines contain the gradient information. Now we will overlay this information with the routes. Open the attribute table of this layer, the field “grid\_code” stores the value of the reclassified slope raster for each stream segment.

**Step 5:** Use the “Locate Features Along Routes” tool to overlay the gradient information with the routes.



This will produce a route event table that should look something like this:



The screenshot shows the ArcGIS attribute table for the 'stream\_routes\_gradient' layer. The table has eight columns: OBJECTID\*, RID, FMEAS, TMEAS, arcid, grid\_code, from\_node, and to\_node. The first few rows of data are:

OBJECTID*	RID	FMEAS	TMEAS	arcid	grid_code	from_node	to_node
1	1	25	175	5	1	8	5
2	2	25	37.5	7	1	11	12
3	2	37.5	172.8553	8	0	14	11
4	2	172.8553	739.2767	9	1	15	14
5	3	0	1261.7641	14	1	16	15
6	4	0	17.6777	1	1	1	2
7	4	17.6777	288.3883	2	0	4	1

At the bottom of the table, it says "0 of 161,849 selected".

`grid_code` in the event table represents the classified stream gradient (0 is 20%, 1 is <20%). Note that you have more records in this table than you have routes, because many routes are being segmented based on the gradient. Thus, you see repeating route id's (RID).

Add a Short field to the event table called “GRADIENT” and calculate this field as `!grid_code!` to transfer the `grid_code` values to a more meaningful field name.

**Step 6:** Repeat steps 4 and 5 above for the stream order raster that you generated in Task 1. Ensure that the `stream_order` raster is used as the “Input stream raster” and “Simplify polygons” is toggled off. In the end, you should have a second event table where the `grid_code` (values between 1 and 6) indicates the stream order for each route. Again, add a Short field to the event table called “SORDER” and calculate this field as `!grid_code!` to transfer the `grid_code` values to a more meaningful field name.

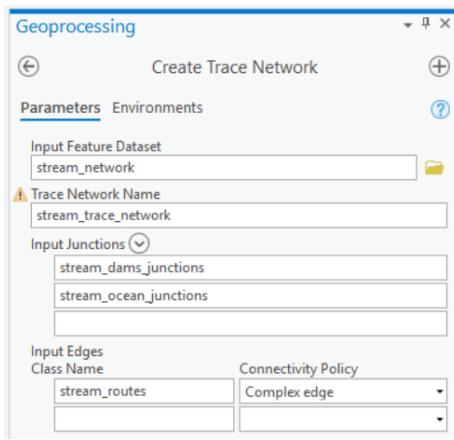
## Task 4: Trace the network

Now we can create a network that will allow us to trace the routes from the ocean to the dams. The goal of this task is to identify routes through the stream network that flow to the ocean and are unimpeded by dams. The following link provides more information about trace networks in ArcGIS Pro: <https://pro.arcgis.com/en/pro-app/latest/help/data/trace-network/what-is-a-trace-network-.htm>

The following steps will use the “`stream_dams_junction`”, “`stream_ocean_junctions`” and the “`vancouver_island_stream_routes`” layers. Before running the next steps save your ArcGIS Project. Ensure you have cleared any selected features from the map.

**Step 1:** Open the “Create Trace Network” tool:

- The Input Feature Dataset should point to your feature dataset in your geodatabase
- Trace Network Name can be named “stream\_trace\_network”
- Input Junctions should contain both the “stream\_dams\_junction” and “stream\_ocean\_junctions” from the Topology geodatabase.
- Input Edges should contain the clipped stream routes and set the Connectivity Policy to “Complex edge”



**Step 2:** Run the “Enable Network Topology” tool on the trace network you just created.

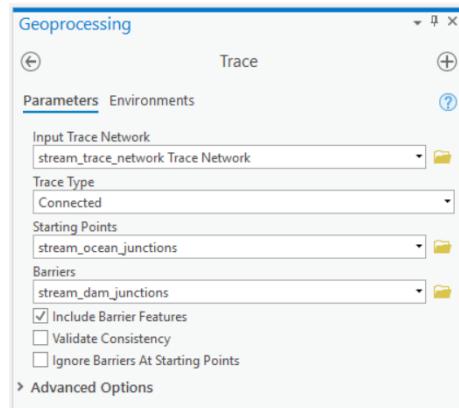
**Step 3:** Run the “Validate Network Topology” tool on the trace network you just created. You might receive a warning: “A dirty area is not present within the specific extent.” Ignore it.

**Step 4:** Open the “Trace” tool. This tool traces a path along a network that meets specified criteria. It also supports modelling barriers, so that the tracing terminates when a barrier feature is reached. Here is how we will parameterize this tool:

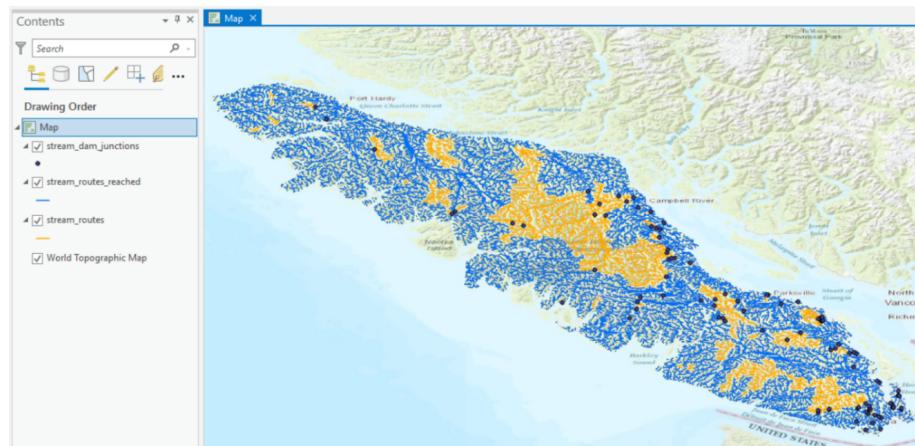
- Input Trace Network is the trace network that you created earlier
- Trace Type is “Connected”
- Starting Points are the “stream\_ocean\_junctions” (make sure this is the layer stored in the Topology geodatabase)
- Barriers are the “stream\_dam\_junctions” (make sure this is the layer stored in the Topology geodatabase)
- Toggle on “Include Barrier Features”
- Toggle off “Validate Consistency”
- Toggle off “Ignore Barriers at Starting Points”

- Leave all Advanced Options as defaults

After the tool has finished running, toggle off the stream trace network and toggle on the stream routes layer (“vancouver\_island\_stream\_routes”). You should see some of the routes are now highlighted on your map.



**Step 5:** Add a Short field to the stream routes attribute table called “REACHED”. Calculate this field with a value of “1” for the routes selected. Invert the selection and calculate the field again for the selected features with a value of “0”. Finally, clear the selection and symbolize the routes based on the “REACHED” field. Congratulations! You have now modeled which route segments can be reached from the ocean and are not impeded by dams. Your stream\_routes should look something like this (blue = unimpeded streams, yellow = stream reaches that are impeded by a dam):



**Important:** If your map looks significantly different from above, like most of the streams are impeded, then try these fixes:

- Zoom in closely to the ocean and dam junctions to ensure they are overlapping with the stream routes.

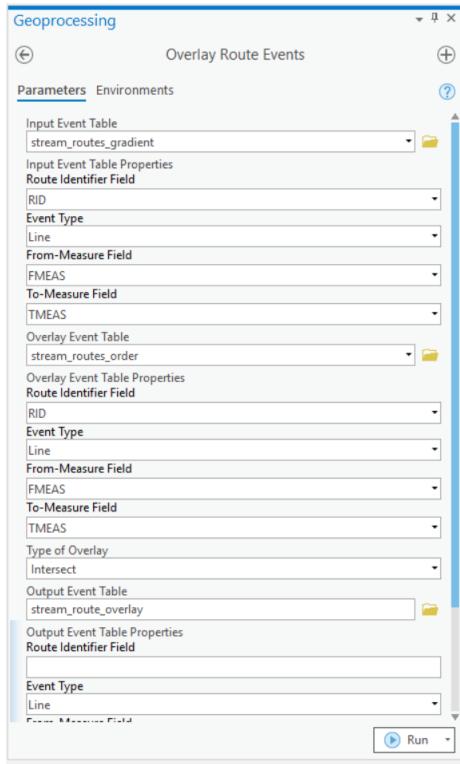
- Apply a new topology rule that dam/ocean junctions must be coincident with the stream routes.
  - If there are any errors from above, then use the “Snap” tool to force the ocean and dam junctions to. The input features are the features that you want to move (snap), in this case the ocean and dam junction points. The type should be set to “Edge” and usually a distance of 25-30 meters is sufficient, but check the output
  - If you are still having trouble, try deleting the Trace Network from the feature dataset and rebuilding it (“Create Trace Network” + “Enable Network Topology” + “Validate Network Topology”).
  - If you are unable to delete the trace network, your geodatabase may be corrupt. Try creating a new geodatabase and import or re-create the important feature classes. Then, re-create the feature dataset and rebuild the trace network.
- 

## Task 5: Overlay and query the network

Now you will practice doing an overlay with your route event tables to transfer all the attributes into a common linear referencing system. This will allow you to then query the network for stream segments that meet some basic conditions for salmon habitat.

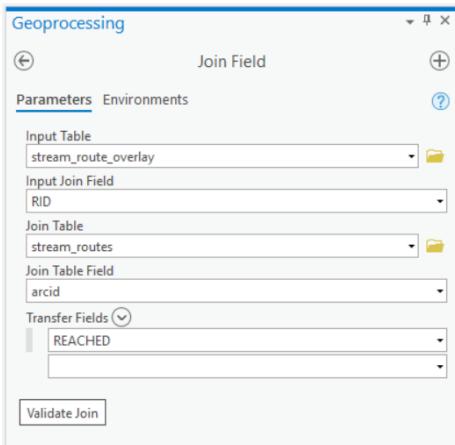
**Step 1:** Parameterize the “Overlay Route Events” tool like below:

- Input Event Table is your classified gradient event table from Task 3
- Overlay Event Table is the stream order event table from Task 3
- From-Measure Field is FMEAS
- To-Measure Field is TMEAS for both input event tables and the output event table
- Type of Overlay is Intersect
- Name the Output Event Table “stream\_route\_overlay”
- Leave the Route Identifier Field empty for the output, but ensure it is set to RID for both the input tables



This will produce a new output event table that contains the intersection of stream gradient and stream order along the routes. As well, this event table should contain which routes can be reached from the ocean before running into a dam. Keep in mind that this is just a table, not a feature class. This is known as dynamic segmentation. Next we will add one last field, visualize the everything and do a simple query.

**Step 2:** Join the field “REACHED” from the original routes to the overlay event table you just created using the “Join Field” tool. The stream\_routes layer should be the layer with the REACHED field:



You should now have a table that has all the fields we need: SORDER, GRADIENT, and REACHED. (Note that some fields have been hidden from view in the table below.)

OBJECTID *	RID *	FMEAS	TMEAS	SORDER	GRADIENT	REACHED
1	1	25	175	1	1	1
2	2	25	37.5	2	1	1
3	2	37.5	172.8553	2	0	1
4	2	172.8553	739.2767	2	1	1
5	3	0	1261.7641	1	1	1
6	4	0	17.6777	1	1	0
7	4	17.6777	288.3883	1	0	0
8	4	288.3883	1295.8631	1	1	0
9	5	0	220.2007	4	4	4

**Step 3:** Use the “Make Route Event Layer” tool to convert the output event table to a feature class. Parameterize the tool as follows:

- Input route features will be the routes you originally created in Task 1
- Route identifier field is “arcid”
- Input event table is the “stream\_route\_overlay” table that you created from step 1
- From-measure field is “FMEAS”

- To-measure field is “TMEAS”
- Layer name or Table view is “stream\_route\_overlay\_layer”

**Step 4:** Query the event layer from step 2 for stream segments with the following requirements:

- Can be reached from the ocean
- Is 1<sup>st</sup> or 2<sup>nd</sup> order stream
- Has a gradient not steeper than 20%

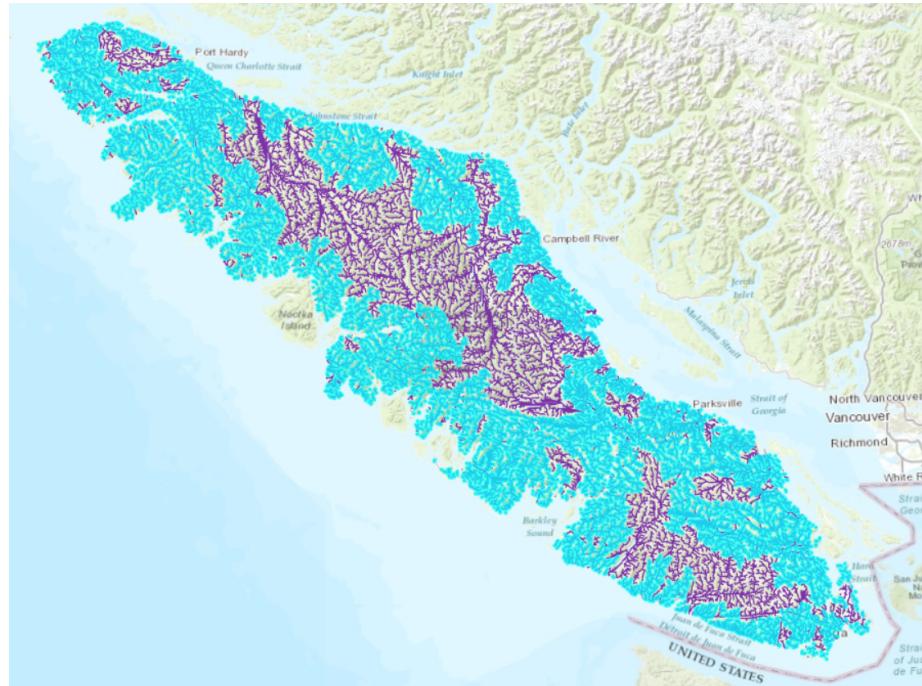
Export the selected features to a new feature class called “accessible\_salmon\_habitat”.

**Step 5:** Repeat the query using the following requirements:

- Cannot be reached from the ocean
- Is 1<sup>st</sup> or 2<sup>nd</sup> order stream
- Has a gradient not steeper than 20%

Export the selected features to a new feature class called “inaccessible\_salmon\_habitat”.

If everything goes *swimmingly*, then the selected stream segments that meet the habitat criteria and can be reached from the ocean will look something like this:



For your final report, you will clip these layers to one of the salmon conservation units.

---

## Summary

Would you have guessed that you could do all that analysis from a single DEM? That is the power of applying focal functions to the raster (flow direction, flow accumulation, slope) and deriving a network with topology. In this lab, we were principally concerned with events and attributes along the stream segments. Although we did not explicitly calculate connectivity measures, the connectivity of the network was implied through the modelling approach using flow direction from the filled DEM. If you attempted to perform this same analysis without the aid of a DEM, then you would need to take additional precautions to ensure that your polyline features have topological connectivity and valid junctions. Still, our approach using a DEM is prone to artifacts from the raster. Stream segments are limited to eight turning angles and the resolution of the DEM can impact stream attributes.

Return to the **Deliverables** section to check off everything you need to submit for credit in the course management system.



## **Chapter 3**

# **Terrain Analysis and Riparian Area Management**

Written by Hana Travers-Smith

### **Lab Overview**

In this lab you will take on the role of a forest manager with the following task: Define reserve (no harvest) areas within a topographically complex riparian forest using current “best management” practices. To do this you will use a Digital Elevation Model and high-resolution basemaps. The deliverable for this assignment will be a written report where you will assess the application of DEMs for forest management in riparian regions.

---

### **Learning Objectives**

- Characterize riparian areas using hydrological toolsets
  - Execute tasks using Python scripting
  - Evaluate harvestable areas near riparian management areas
- 

### **Lab Background**

Riparian areas are found adjacent streams, lakes and wetlands. In many areas they contain highly productive and commercially valuable timber. However,

## 54CHAPTER 3. TERRAIN ANALYSIS AND RIPARIAN AREA MANAGEMENT

vegetation in riparian zones also has important ecological function by stabilizing stream banks, regulating water temperature, providing nutrient inputs and reducing runoff into water bodies. In British Columbia, Riparian Management Areas (RMAs) are defined around sensitive streams and rivers that forbid or limit forest harvest activities within them. The width of these buffer zones depend on stream class, which is defined by channel width and fish-bearing potential. Thus, it is critical for forest managers to accurately map and measure stream networks.

Airborne Laser Scanning (ALS) can be used to derive detailed maps of terrain and vegetation structure to aid in riparian forest management. Previous work has been done using ALS to map and extract stream characteristics over large areas and assign stream classes (Tompalski et al., 2017). You will use similar methods to classify streams and define RMAs over a small area in the Nahmint valley, British Columbia.

Throughout this lab stream *order* refers to the Strahler or Shreve ordering system seen in your previous lab, and stream *class* refers to the BC government stream classification used to determine riparian management areas.

More background information on riparian forest management in British Columbia can be found here: <https://www2.gov.bc.ca/gov/content/industry/forestry/managing-our-forest-resources/silviculture/silvicultural-systems/silviculture-guidebooks/riparian-management-area-guidebook>

---

## Deliverables

Responses to the questions posed throughout the lab on the course management system (25 points)

Table of summary statistics of number of stream links, average stream length, width and percent gradient grouped by stream order (10 points)

A map including the following layers and requirements: (15 points)

- Watersheds, contour lines
  - Stream networks symbolized by stream class
  - Inset map showing example stream network and Riparian Management Areas
  - Text showing the total area in the Reserve and Management Zone and the length of the stream network
-

## Data

The data for this lab are a Digital Elevation Model (DEM) of the Nahmint watershed region in British Columbia and is accessible via the UBC PostgreSQL server. Instructions for connecting to the server are given in the tasks below. We will be using data from the `nahmint` database.

---

### Task 1: Prepare the DEM and Map Stream Networks

**Step 1:** The following steps should largely be review from the previous terrain modelling assignment! First, you will need to open QGIS and connect to the `nahmint` database on the UBC PostgreSQL server and export the `Nahmint_DEM` layer as a GeoTiff to work with in ArcGIS Pro. **Set the NoData values as -9999 to 0 and the CRS as EPSG:3005.**

Open ArcGIS Pro and import the `Nahmint_DEM` GeoTiff you just created. We will do a few pre-processing steps before generating a stream network.

Use the **Fill** tool to remove any sinks from the DEM.

Sinks are small imperfections in the DEM that create areas where water cannot flow out of. If sinks are not eliminated, water flow can get trapped within these depressions, leading to unrealistic pooling of water and incorrect delineation of watershed boundaries.

Navigate to Analysis > Tools > Fill (Spatial Analyst).

- Input Surface Raster: **Nahmint\_DEM**
- Output Surface Raster: **Nahmint\_fill**
- Z limit: **leave blank**

The Z-limit represents the minimum depth of sinks that will be filled. For example, if it is set to 10 m then only sinks deeper than 10 m will be filled. For now leave this field blank, this will fill all sinks in the data.

**Step 2:** Use the “Flow Direction” tool to calculate the direction of water flow across the landscape. Use the following parameters for the tool:

- Input surface raster: **Nahmint\_fill**
- Output flow direction raster: **Nahmint\_FlowDir**
- Flow Direction Type: D8
- Leave the rest blank/unchecked

There are three flow modelling algorithms, but we will use the simplest: D8. In this model water will flow from one cell to its steepest downslope neighbour. The cell will then be assigned a value based on which of its 8 neighbours water will flow into.

## 56CHAPTER 3. TERRAIN ANALYSIS AND RIPARIAN AREA MANAGEMENT

Click Run. You should now have something like the image below:

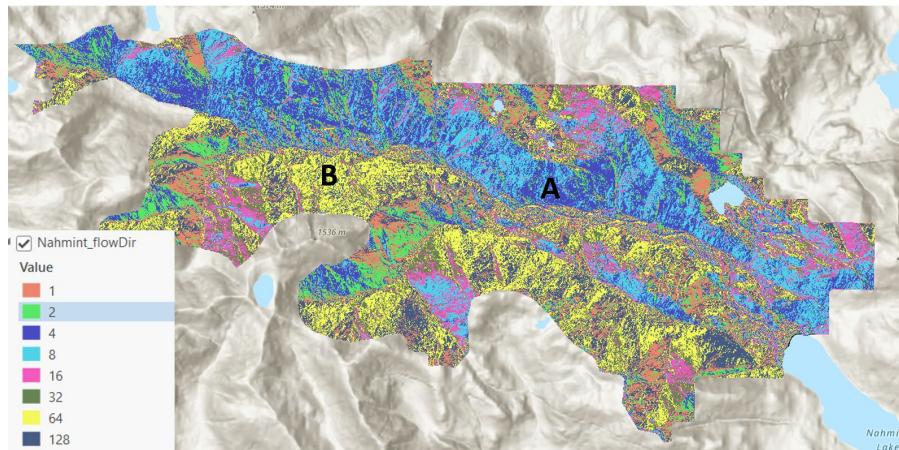


Figure 3.1: Flow Direction.

What direction is water flowing at points A and B? You will need to interpret this raster in your Results and Discussion. Use the ArcGIS Pro documentation to understand the output if necessary.

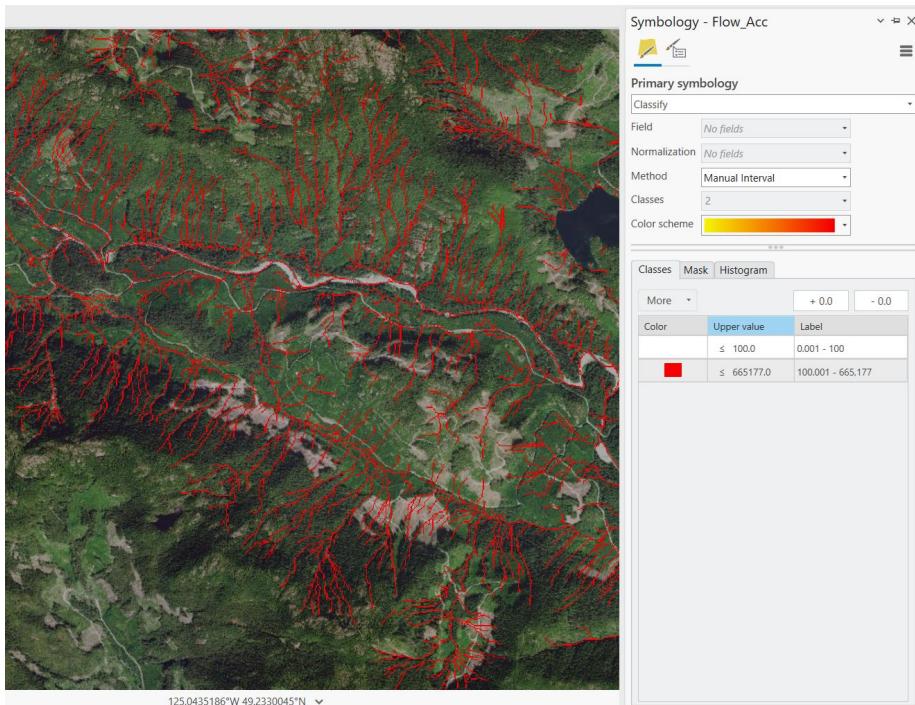
**Step 3:** We will use the flow direction raster calculated in the last step to calculate flow accumulation, which counts the total number of cells that will flow into a given cell. For example, a cell located at the bottom of the hill will have high flow accumulation and a cell at the top of a hill will not have any flow accumulation. Open the “Flow Accumulation” tool and parameterize it as follows:

- Input flow direction raster: **Nahmint\_FlowDir**
- Output flow accumulation raster: **FlowAcc**
- Output data type: **Integer**
- Input flow direction type: **D8**

Leave all other fields blank then run the tool.

**Step 4:** Next, we will create a raster-based stream network using a threshold in the flow accumulation raster. For example, if the threshold is 100, then only cells with flow accumulation greater than 100 will be counted as a stream. Cells with flow accumulation less than 100 will be set to a background value of 0.

To see how different thresholds impact stream identification, change the symbology of the flow accumulation raster and use the **Manual Interval** symbology to set two classes. See the example below for a stream network with a flow accumulation threshold of 100 (cells with flow accumulation < 100 are set to no color). Note that in the example below the threshold is likely too small, and results in many small streams that are not evident in the ArcGIS basemap.

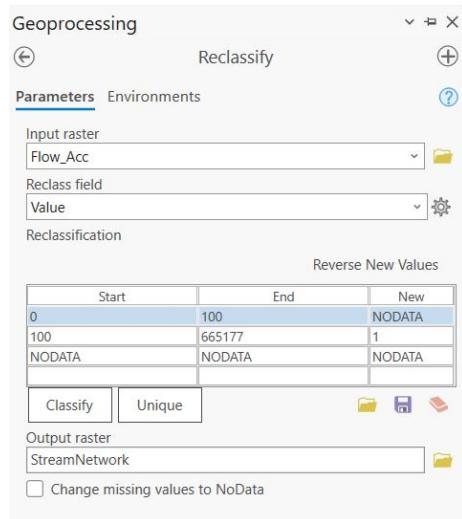


Compare your stream network to the streams visible in the ArcGIS satellite basemap and the filled DEM (use the shaded relief symbology for a better visualization). Experiment with different flow accumulation thresholds. What other features could be misclassified as a stream using this method? Find a threshold that captures a realistic stream network visible in the basemap and DEM, without too many misclassified features.

**Q1. Comment on the accuracy of stream networks compared to basemaps. Discuss potential sources of error and how they could be improved. (5 points)** Once you have selected a threshold, open the “Reclassify (Spatial Analyst Tools)” tool.

Use the threshold you have selected as the Start and End values. Set the cells representing streams to a “New value” of 1 and all other cells to NODATA. Save the new raster as “Stream\_reclass”.

## 58CHAPTER 3. TERRAIN ANALYSIS AND RIPARIAN AREA MANAGEMENT



**Q2.** Report the flow accumulation threshold value that you used to create Stream\_reclass. (1 point) Zoom into the reclassified flow accumulation raster to see your stream network.

**Step 5:** Next, we will assign a stream order to each segment in the stream network. Stream ordering is a system to classify streams based on the number of tributaries (smaller streams) that flow into it. First order streams have no tributaries and higher order streams may have multiple levels of tributaries flowing into it.

Open the “Stream Order” tool and parameterize it as follows:

- Input stream raster: **Stream\_reclass**
- Input flow direction raster: **Nahmint\_FlowDir**
- Method of stream ordering: **Strahler**
- Output raster: **StreamOrder**

**Q3.** Compare and contrast the Strahler and Shreve methods for stream ordering. (5 points) Zoom in to see the **StreamOrder** output. The value of each pixel corresponds to the order of each stream segment. Now re-run the tool using the Shreve method and compare the two outputs (You will compare these two outputs in the Discussion section of your report).

**Step 6:** Use the “Stream to Feature (Spatial Analyst Tools)” tool to create polyline features representing the ordered stream network. Parameterize the tool as follows:

- Input stream raster: **StreamOrder**
- Input flow direction raster: **Nahmint\_FlowDir**
- Output polyline features: **StreamNetwork\_polyline**

- Simplify polylines: **Checked**

Open the polyline attribute table. The “grid\_code” field corresponds to the stream order of the segment. The tool also generates attributes describing the start and end points of each segment and the segment length.

**Step 7:** Finally, we will assign a unique numeric ID to each stream segment using the “Stream Link” tool. Parameterize the tool as follows:

- Input stream raster: **Stream\_reclass**
- Input flow direction raster: **Nahmint\_FlowDir**
- Output raster: **StreamLinks**

Zoom in and examine the output.

---

## Task 2: Extract Stream Characteristics

In this task we will calculate stream gradient and estimate width for each segment in the network. These metrics will be used to assign a stream class and define the width of the riparian management area.

Percent gradient is a measure of the streams “steepness” and is calculated as:

$$\frac{EC}{SL} * 100$$

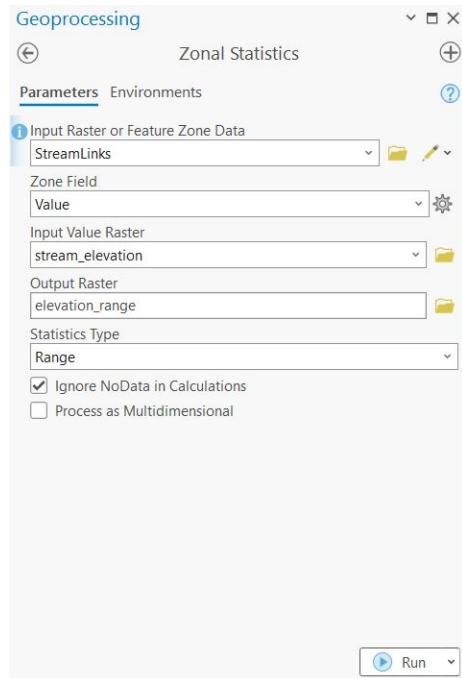
*EC* is elevation change *SL* is stream length

Streams with a gradient <20% are considered “fish bearing” in the absence of other field data.

**Step 1:** First, use the “Raster Calculator” tool to multiply the classified flow accumulation layer **Stream\_reclass** and the **Nahmint\_DEM** layer. This will give elevation within streams. Name the output **streams\_elevation**.

**Step 2:** Open the “Zonal Statistics” tool. This tool calculates summary statistics in one raster layer within “zones” or groups defined in another layer. This is a very useful tool for many applications! We will use the unique IDs from the **StreamLinks** layer as “zones” to calculate the change in elevation within each stream segment.

## 60CHAPTER 3. TERRAIN ANALYSIS AND RIPARIAN AREA MANAGEMENT



The resulting raster shows the gain in elevation within each stream segment.

**Step 3:** We will convert the elevation change raster to polygon using the “Raster to Polygon” tool. (You may have to first convert the raster from floating point to integer format using the Int Tool). Save the output of “Raster to Polygon” as **elevation\_change\_polygon** and check **Simplify Polygons**. This step will allow us to perform a spatial join with the vectorized stream network. Open the attribute table and note that some attribute names are the same as the **Stream-Network\_polyline**. The attribute “grid\_code” represents the elevation gain within each stream segment.

You can change the field names in both layers so you don’t get confused after joining the two features.

**Step 4:** Open the “Spatial Join” tool and join the **StreamNetwork\_polyline** with the **gradient\_polygon**. Parameterize the tool as follows:

- Target Features: StreamNetwork\_polyline
- Join Features: elevation\_change\_polygon
- Output Feature Class: StreamNetwork\_join
- Join operation: Join one to one
- Keep all target features: Checked
- Match Option: Largest Overlap
- Search Radius: 3 Meters

**Step 5:** Open the attribute table of the **StreamNetwork\_join** feature class.

If you have duplicated attribute names, the attributes are ordered first with the Target Feature and the Join Features second. You may rename and delete unnecessary attributes. You will need to keep the fields relating to stream order, change in elevation and the length of the polyline features. (There are two Shape\_Length attributes from the join so figure out where they come from.)

Create a new field called “PercentGradient” and use the “Calculate Field” tool to calculate the gradient for each stream segment using the change in elevation and the length of the stream segment. Streams with gradient < 20% will be considered as “fish bearing”.

**Q4. How many streams have fish bearing potential? What is the total length of these streams? (2 points)** **Step 6:** Next, we will use the ESRI basemaps to measure stream width for a sample of streams stratified by order. We will use the average stream width per order to approximate the width for all streams in the study area. (For example, if the average width of streams in order 1 is 30 m, all streams in order 1 will be assigned that width).

Stream width is defined using the entire channel width, not just the part with water currently flowing through it. Use the examples below to guide you with your mapping. You should be able to see channels for streams in orders 2-4, but order 1 streams will likely not be visible. We will assign these streams an average width of 1 m. Note that there are more sophisticated methods to calculate stream width from a DEM!



S4 Stream



S3 Stream



S2 Stream (it is barely visible)

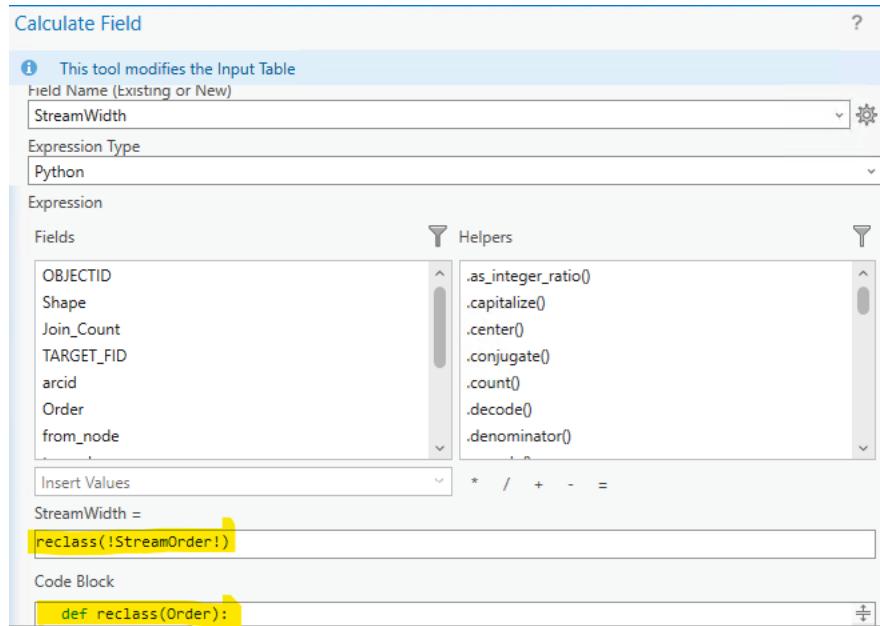
Measure 10-15 stream widths per order using the “Measure” tool on your toolbar and record these measurements in an excel document. When you are satisfied with your measurements calculate average width of each stream order. Use the shaded relief DEM and the stream network polyline to help you find some of the smaller order 2 streams.

**Step 7:** In the **StreamNetwork\_join** layer, create a new field called “StreamWidth”. We will use an **ifelse** statement to conditionally assign

values to the stream width attribute depending on stream order. Select the “StreamWidth” column in the attribute table and click “Calculate”. In the “Calculate Field” tool, change the “Expression Type” to “Python”.

To use Python in the tool we need to populate two fields. First, the top “Expression” field, which should read `StreamWidth =`. This is where we can call a function and define the inputs to the function. Here we have called the function `reclass` with the “StreamOrder” attribute as the input.

Next, in the “Code Block” we need to define the `reclass` function, so that we can use it as part of an expression. Here we will use an `if else` statement to set the values of stream width based on stream order.



Modify the following code and paste it into the “Code Block” (delete the square brackets before running). Verify the expression using the check button > Apply.

```
def reclass(Order):
    if (Order == 1):
        return [average stream width here]
    elif (Order == 2):
        return [average stream width here]
    elif (Order == 3):
        return [average stream width here]
    else:
        return [average stream width here]
```

If you get an “Invalid Field” error it is because the variable in the “Expression Field” (between the two !) does not match a field in the attribute table. So

## 64CHAPTER 3. TERRAIN ANALYSIS AND RIPARIAN AREA MANAGEMENT

make sure to match the name of the appropriate field. You can use the pre-generated Field names to generate your expression to make sure this step works, just double-click on one to add it to your expression.

**Table of summary statistics of number of stream links, average stream length, width and percent gradient grouped by stream order. (10 points)**

**Q5. How did stream characteristics change across stream order? Is this what you expected? (5 points)**

---

### Task 3: Mapping Watersheds and Riparian Management Areas

**Step 1:** Use the following table to assign a stream class (S1-S6) based on average channel width and fish bearing potential. Once a stream class is assigned, the Reserve Zone defines the region around the stream where harvest is prohibited and the Management Zone defines the region where limited harvest is permitted. Note that not all stream classes may be present in the dataset. (We do not have the S1 large rivers class in the study area so no need to include it.)

Riparian class	Average channel width (m)	Reserve zone width (m)	Management zone width (m)	Total RMA width (m)
S1 large rivers	$\geq 100$	0	100	100
S1 (except large rivers)	$>20$	50	20	70
S2	$>5 \leq 20$	30	20	50
S3	$1.5 \leq 5$	20	20	40
S4	$<1.5$	0	30	30
S5	$>3$	0	30	30
S6	$\leq 3$	0	20	20

 Fish stream or community watershed

 Not fish stream and not in community watershed

In the **StreamNetwork\_join** layer, create a new attribute called “Reserve\_BufferDist” and set the “Number Format” to “Numeric”.

Select the “Reserve\_BufferDist” field in the attribute table and open the “Calculate Field” tool. Change the “Expression Type” to “Python”.

This time we will combine the **if else** syntax from the previous task with conditional statements using stream gradient and width to assign buffer distances for

each stream class. The example below defines the Reserve Zone buffer distance of 0 m for the S1 class using the stream gradient and stream width attributes.

The Reserve Zones and Management Zones are typically measured from the edge of the stream channel, however our streams are represented as lines down the centre of the stream channel. **Think of a way to modify the buffer distances given in the table to reflect the approximate width of the stream.**

Your code should look something like the following. Make sure the highlighted attribute names in the Expression field match the names in your table, depending on how you named your columns they may look slightly different than this screenshot:

Input Table  
streamNetwork

Field Name (Existing or New)  
Reserve\_bufferDist

Expression Type  
Python

Expression

Fields

- OBJECTID
- Shape
- Join\_Count
- TARGET\_FID
- arcid
- Order
- from\_node

Helpers

- .as\_integer\_ratio()
- .capitalize()
- .center()
- .conjugate()
- .count()
- .decode()
- .denominator()

Insert Values

Reserve\_bufferDist =  
myCalc(!gradient!, !StreamWidth!)

Code Block

```
def myCalc (gradient, width):
    if (gradient < 20)and(width >= 100):
        return 0
```

Use the following template in the “Code Block” to calculate the buffer distance for the remaining Reserve Zones (S1-S6). For classes with a width of 0 set the buffer distance to 0.

```
def myCalc (gradient, width):
    if (gradient < 20)and(width >= 100):
        return 0
    elif (gradient ...)and(width > ... and width < ...):
        ...
    else:
        return ...
```

## 66CHAPTER 3. TERRAIN ANALYSIS AND RIPARIAN AREA MANAGEMENT

Repeat the same process for the Management Zone widths.

**Step 2:** Open the “Buffer” tool and parameterize it as follows:

- Input Features: **StreamNetwork\_join**
- Output Feature Class: **Reserve\_buffer**
- Distance: Use drop-down menu to change to from Linear Unit to **Field**
  - **Reserve\_BufferDist**
- Side Type: Full
- End type: Round
- Dissolve: Dissolve all output features into a single feature

Inspect the output.

Repeat the Buffer step for the Management Zone widths. That is, first create a new field that captures the Management Zone widths in the table above. Then calculate a new field that represents the total of the stream + reserve + management buffer and create a new buffer that uses this value.

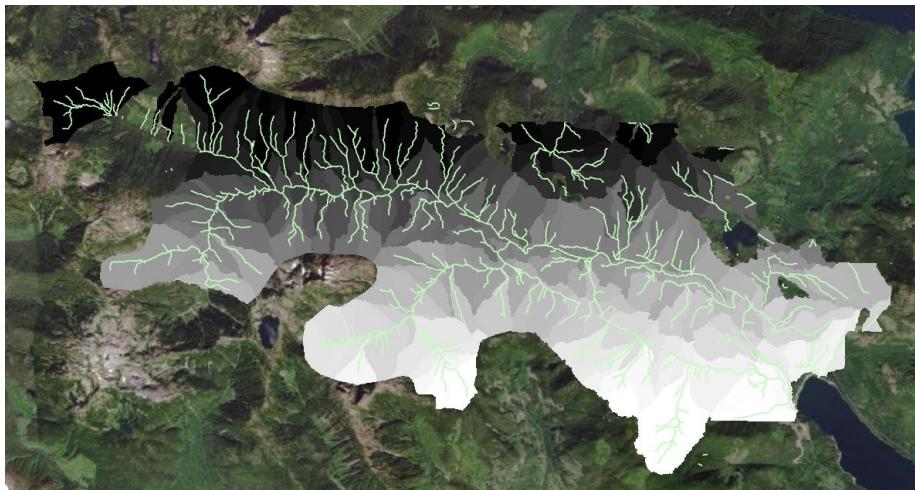
**3.0.0.0.1 Q6. How much forest area is retained within the Riparian Reserve and Management Zones? (2 points)** **Step 3:** Next, we will map watersheds. A watershed is an area of land where all the water that falls or flows into it converges to a common outlet, such as a river, lake, or ocean. It is bounded by a drainage divide, which is a boundary that separates water flowing into different watersheds.

The watershed tool uses flow direction and stream links to delineate watershed boundaries. The watershed boundaries will be defined such that water flows into each of the stream links.

Open the “Watershed” tool and parameterize it as follows:

- Input D8 flow direction raster: **Nahmint\_FlowDir**
- Input raster or feature pour point data: **StreamLinks**
- Output raster: **Nahmint\_watersheds**

The output will be a new raster where the cell values correspond to each unique watershed catchment. You should have something like the following (do not worry if it is not exactly the same as the screenshot):



**Step 4:** Finally, we will create contour lines to help interpret the DEM. Open the “Contour” tool and parameterize it as follows:

- Raster: **Nahmint\_DEM**
- Output features class: **Nahmint\_contour**
- Contour interval: 100
- Base contour: [Set to the lowest elevation in the study area]
- Z factor: 1
- Contour type: Contour

**3.0.0.0.2 Q7.** Use the watershed map, contour lines and flow direction layers to assess how runoff from harvesting may impact stream networks. Reference specific areas/examples from the maps you created. Do you think the mapped RMAs provide adequate protection to streams in this area? Why or why not? (5 points)

---

## Summary

The aim of this lab was to practice solving a real-world forest management problem using raster and vector data. You have hopefully gained a thorough understanding of the tools available for modeling hydrology in ArcGIS Pro. In addition, the basemaps in ArcGIS Pro can be a valuable source of high-resolution satellite imagery. Be sure to stand up, stretch and drink some water!

Return to the **Deliverables** section to check off everything you need to submit for credit in the course management system.



## **Chapter 4**

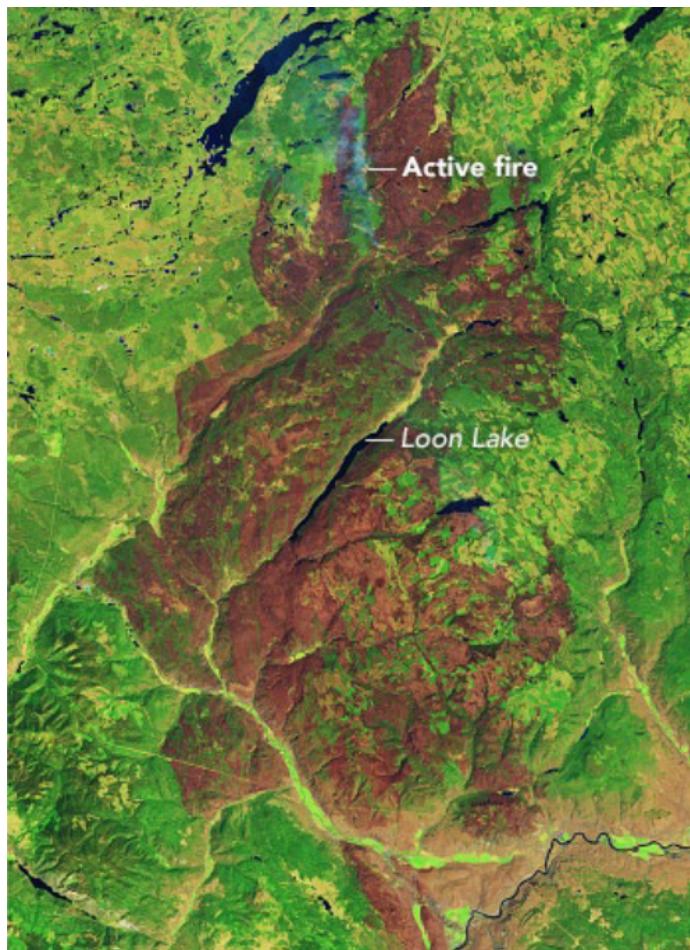
# **Machine Learning with Geospatial Data**

Written by Claire Armour

### **Lab Overview**

In this lab, we will be using simple machine learning tools in ArcGIS Pro to predict two variables related to fuel in a post-burn scenario. In Task 1, you will check for spatial autocorrelation in our data and select a subset of input variables which will be most useful in predicting the outcome variables. In Task 2, you will create three different sample datasets using three types of sampling. In Task 3, you will be using a Random Forest classifier to create prediction surfaces for the two variables.

Our area of interest is a large rectangular area in southern British Columbia containing the burn boundary from the 2017 Elephant Hill wildfire. The fire began on July 6th, 2017 and burned 191,865 ha in the south-central Interior region of BC, just northeast of Cache Creek. It was the largest fire by area in the record-breaking 2017 BC wildfire season and was likely started by smoking materials (matches, cigarettes, etc.).



---

## Learning Objectives

- Compare and contrast sampling designs on machine learning predictions
  - Train, test, and predict a continuous variable (crown closure) and a categorical variable (fuel type) using the random forest algorithm
  - Evaluate models by interpreting model statistics like  $R^2$ , AIC, VIF, and confusion matrices
-

## Deliverables

Lab report with the following specification:

6 pages maximum PDF including figures, tables and references (3 points). Single-spaced, 12-point Times New Roman font (1 point). All figures and tables should be properly captioned (1 point).

Results should address the following questions and requirements (25 points):

- How many fuel types are there in the VRI dataset? Which fuel type has the fewest number of points?
- What are the results of the spatial autocorrelation (SA) tool and what do they tell us? Why do we test CROWN\_CLOS for spatial autocorrelation but not FUEL\_TYPE?
- Report the diagnostic results of the Exploratory Regression ( $R^2$ , AIC, VIF, variable significance) **for three of best models**. What model had the highest  $R^2$ ? Based on this, what variables did you choose to include in your final model? Use the statistical terms in the results to justify your choice.
- Select the best model for FUEL\_TYPE and CROWN\_CLOS and briefly justify your choice. Map one of the FUEL\_TYPE rasters and one of the CROWN\_CLOS rasters. Be sure to specify in a caption or title what sampling design / parameter settings you are displaying in each.

Discussion should address the following questions and requirements (20 points):

- For the fuel type model discuss the performance of each model. Were there large differences in model performance among the three sampling strategies? What are the benefits and drawbacks of using each sampling design, both in terms of statistical power and the limits of field data acquisition?
- For the crown closure model discuss the impact of changing the Random Forests parameters on the model performance. What parameters should be optimized and why?
- We used the Exploratory Regression tool to select our input variables. Discuss why you think these variables were important in predicting crown closure. We used the same variables for both the crown closure and fuel type models, what variables might be better for predicting fuel type that we didn't include?
- What are the advantages and limitations of using remote sensing data to predict crown closure and fuel types? What about climate data? Terrain data? Why should we use these data for this problem?
- Reference to any peer reviewed sources as needed.

---

## Data

All data for this lab are accessible via the UBC PostgreSQL server. Instructions for connecting to the server are given in the tasks below. We will be using data from the `elephanthill` database.

The data for this lab comes from several different sources, including the Landsat archive and Landsat-derived indices, the ClimateNA program, the Vegetation Resources Inventory (VRI) from the BC Ministry of Forests, and the ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) sensor on the Terra satellite. The VRI is a polygon dataset with hundreds of attributes describing harvest, disturbance, species, volume, etc., for all of BC. The individual rasters and their descriptions are listed on below.

Variable Name	Formula or Description	Raster Name
Elevation	Elevation above vertical datum (meters) Source: Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 3	ASTER_DEM
Aspect	Azimuth direction, measured clockwise from North (degrees) Source: ASTER_DEM	ASTER_Aspect
Slope	Vertical angle from horizontal datum (degrees) Source: ASTER_DEM	ASTER_Slope_Deg
Topographic Radiation Solar Aspect Index (TRASP)	Assigns a value of zero to the north-northeast direction: $TRASP = \frac{1 - \cos(\frac{\pi}{180}(\alpha - 30))}{2}$ Source: ASTER DEM	ASTER_TRASP
Mean Annual Precipitation (MAP)	Millimeters Source: ClimateNA	CNA_MAP
Mean Annual Temperature (MAT)	Degrees Celsius Source: ClimateNA	CNA_MAT
Precipitation as Snow (PAS)	Millimeters Source: ClimateNA	CNA_PAS
Summer Heat Moisture Index (SHM)	$\frac{\text{mean warmest month temperature}}{\frac{\text{May to September precipitation}}{1000}}$ Source: ClimateNA	CNA_SHMz
Temperature Difference (TD)	$\text{mean warmest month} - \text{mean coldest month}$ Source: ClimateNA	CNA_TD

Variable Name	Formula or Description	Raster Name
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{NIR - Red}{NIR + Red}$ Source: Landsat	Landsat_NDVI
Normalized Difference Burn Ratio (NBR)	$NBR = \frac{NIR - SWIR}{NIR + SWIR}$ Source: Landsat	Landsat_NBR
Tasseled Cap Brightness (TCB), Wetness (TCW), Greenness (TCG)	$TCB = Band1 * coeff1$ $TCW = Band1 * coeff1$ $TCG = Band1 * coeff1$ Source: Landsat	Landsat_TCB Landsat_TCW Landsat_TCG

The data have been processed ahead of time so that you may focus more on machine learning and sampling design. If you wanted to replicate this workflow in another area of your work, here are the steps we used:

1. Obtained the ClimateNA, ASTER, VRI, and Landsat data appropriate to the study area/time period.
2. Projected all the data to NAD 1983 BC Environmental Albers geographic coordinate system and clipped it to the study area boundary.
3. Calculated Normalized Difference Vegetation Index (NDVI), Normalized Burn Ratio (NBR), and Tasseled Cap Wetness (TCW), Greenness (TCG), and Brightness (TCB) using the appropriate bands from the Landsat image.
4. Normalized all the rasters (including the indices) using the standard formula where x is a pixel value, mu is the mean pixel value of the raster, and omega is the standard deviation of the raster.
5. Cleaned the VRI data by removing all unneeded fields and records with <Null> values for our variables of interest.
6. Used the “Extract Multi Values to Points” tool to extract the values of all our rasters to our VRI point feature class (similar to a Spatial Join). Checked for <Null> or values again and removed.

## Task 1: Exploratory Regression in ArcGIS Pro

**Step 1:** Start ArcGIS Pro and create a new project with a map template.

**Step 2:** Connect to the UBC PostgreSQL database and import all of the data from the `elephanthill` database into your project geodatabase. This should include 14 rasters, the Vegetation Resource Inventory (VRI) point feature class, and the Elephant Hill (EH) boundary. Open the attribute table for the VRI and

see what fields are there. We will not be using the EH boundary for any analysis but it is interesting to see where the fire took place. It truly is a megafire!

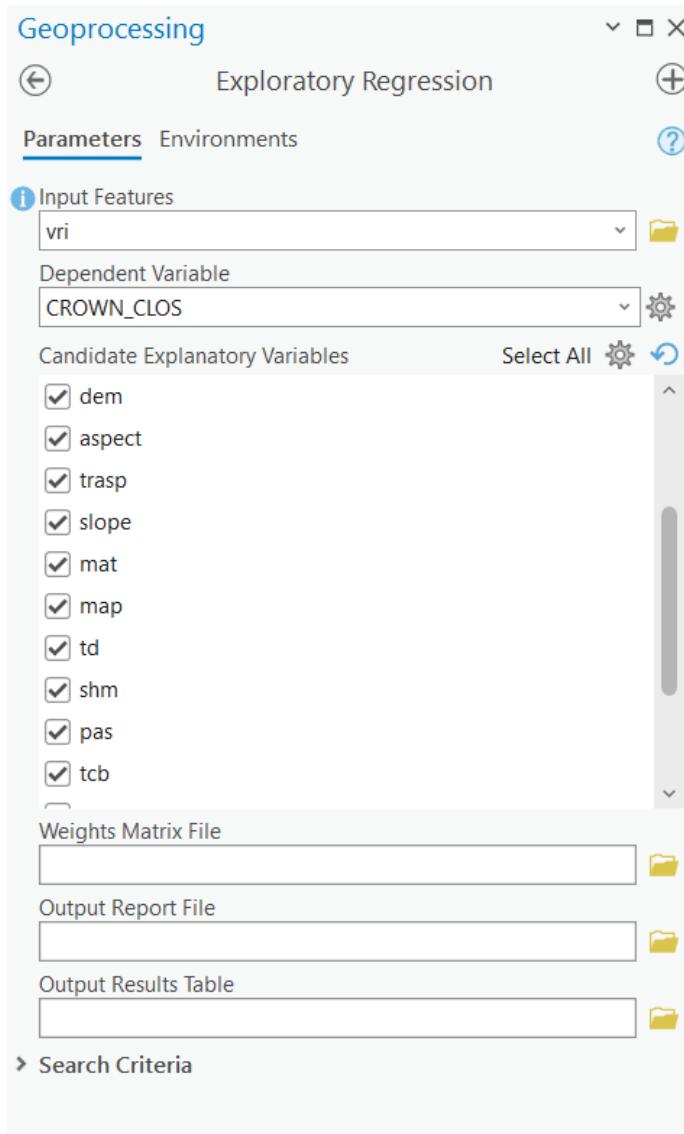
**Step 3:** Set the “Primary symbology” on the VRI feature class to “Graduated Colors” and choose “CROWN\_CLOS” as the field to display. Take a look at the results and then set “FUEL\_TYPE” as the field to display. These are the two variables we will be predicting – crown closure (also known as canopy cover) and fuel type. CROWN\_CLOS is continuous and FUEL\_TYPE is categorical. You will notice there are quite a few VRI points (over 84,000), so save your laptop the extra rendering work and turn off this layer unless you are actively looking at it.

**Step 4:** Open the “Spatial Autocorrelation (Global Moran’s I)” tool. Use the VRI feature class as the input and set the input field to “CROWN\_CLOS”. Change the “Conceptualization of Spatial Relationships” parameter to “K Nearest Neighbours” and set the “Number of Neighbors” to “20”. Check the box that says “Generate Report”. Leave the rest of the tool parameters with the default values and run the tool. Observe the results by viewing the generated report (found in your default folder/geodatabase) or by selecting “View Details”.

**HINT: Due to the constraint of completing this lab in ArcGIS Pro, we are unable adjust our sampling parameters to account for spatial autocorrelation. Fortunately, this is a learning exercise, so you will get to discuss this conundrum in your report!**

Next, we are going to use Exploratory Regression to determine which of our rasters will be most effective in predicting our variables of interest.

**Step 5:** Open the “Exploratory Regression” tool and set the VRI feature class as your “Input Features”. Select “CROWN\_CLOS” as your “Dependent Variable”. Remember that as part of the pre-processing, all raster data were extracted at the VRI points. Select the fields corresponding to the rasters in the VRI feature class as your “Candidate Explanatory Variables”. Leave out the original Landsat bands. Save the “Output Report File” as “regression\_results.txt”. This will create a text file in your default folder with the results of the tool, which can also be viewed under “View Details”. Leave the rest of the tool parameters as the default values and run the tool.



**Step 6:** Select the equation you think will work best for both “CROWN\_CLOS” and “FUEL\_TYPE”. Remember to look at all the different measures of model “goodness” ( $R^2$ , AIC, VIF, multicollinearity, etc.). Our variable rasters fall into three categories: terrain, spectral, and climate. Our final equation should include at least one variable from each category. Write down the variables from your chosen equation somewhere as you will have to re-enter them later.

## Task 2: Sampling for Training Datasets

As noted earlier, the VRI feature class is enormous, and it is an outlier regarding the typical number of field samples we would have to work with. Training on 80,000+ points would also require an equally vast amount of computing power, so we are going to cull this dataset into a more manageable 500 points. This also gives us the opportunity to create and compare sampling designs: random sampling, stratified random sampling, and fixed interval point sampling.

**Step 1:** First, we will make a randomly sampled dataset. Open the tool “Subset Features” and set the VRI feature class as the “Input features”. Name your “Output training feature class” as “random\_sample” and leave the “Output test feature class” blank (this option creates a second feature class for testing but we do not need that). Change the “Subset size units” to “ABSOLUTE\_VALUE”, as we want a specific number of points and not a percentage of the dataset. Set the “Size of training feature subset” to 500 and run the tool. View the output on your map.

**Step 2:** Next, we want to make a stratified randomly sampled dataset. There are four major categories of fuel types in our dataset and their distribution is highly uneven. This design will ensure the rarest class is as equally represented as the most abundant class in our sample. As there are four categories (fire, non-vegetated, vegetated non-forest, and forest), we will sample 125 points from each. The list below tells you what fuel type codes correspond to each category:

- Fire – “N-fire”
- Non-vegetated – “N”
- Vegetated non-forest – all fuel types starting with “S” or “O” (as in octopus).
- Forest – all fuel types starting with “M”, “C”, or “D”.

Open the attribute table for the VRI dataset, right-click on the “FUEL\_TYPE” field, and select “Summarize”. Set the “Field” to “FUEL\_TYPE” and the “Statistic Type” to “Unique”. Leave “Case Field” as is and select “OK”. Inspect the output statistics table in your Contents Pane.

**Step 3:** Using the “Select by Attribute” tool, select all the features in the VRI feature class that have a “FUEL\_TYPE” of “N-fire”. Right-click the feature class, select “Selection”, then select “Make Layer from Selected Features”.

**Step 4:** Using the “Subset Features” tool again, set your parameters to subset 125 points from the layer you created in the previous step. Name the output “fire\_sub\_sample” and check the attribute table to make sure all of the records in this feature class have a fuel type of “N-fire”.

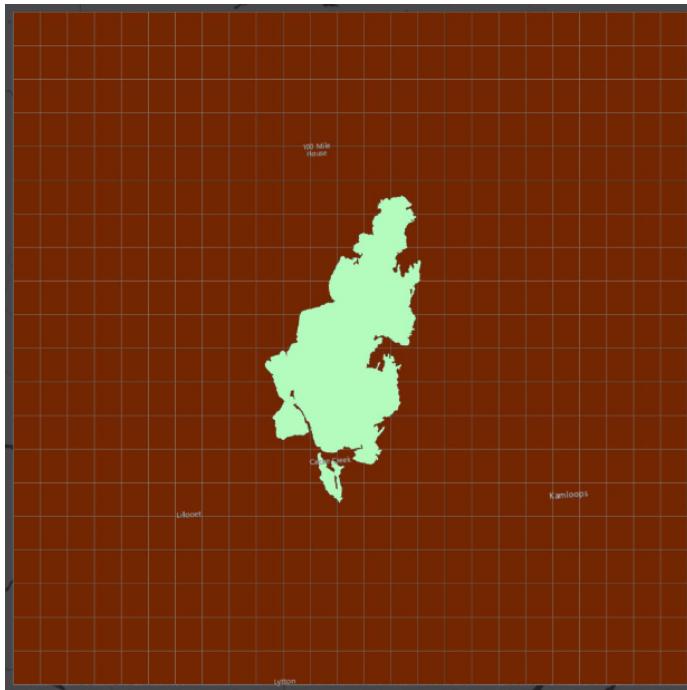
**Step 5:** Clear your selection. Right-click “VRI”, select “Selection”, then select “Clear Selection” and repeat steps 3 and 4 for the remaining three categories. Give the outputs logical names and clear your selection/delete the selection layer between each.

**HINT:** Use the “begins with” operator to select all fields with a certain letter rather than “is”, but beware that both “N” and “N-fire” begin with the same letter.

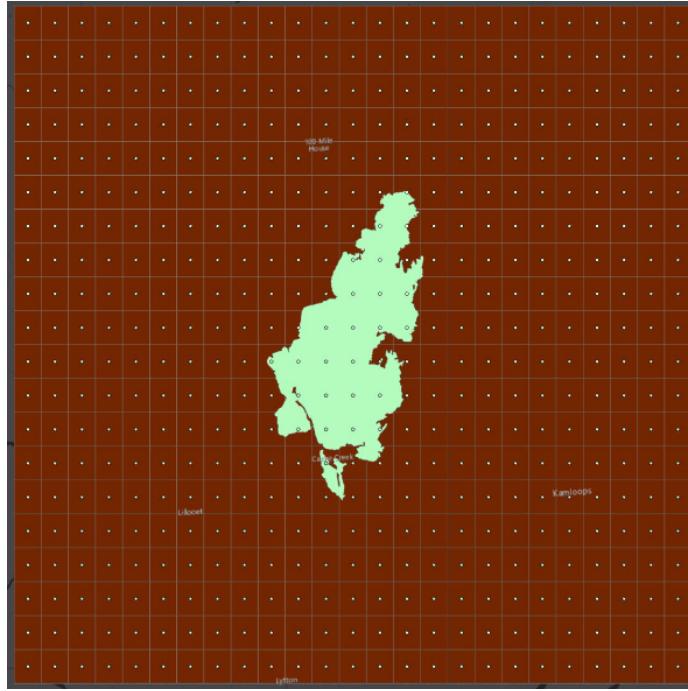
Double-check that your sub-samples have the correct number and fuel types in them.

**Step 6:** Using the “Merge” tool, combine your four sub-sample data sets and call the output “stratified\_sample”. This sample should have 500 points.

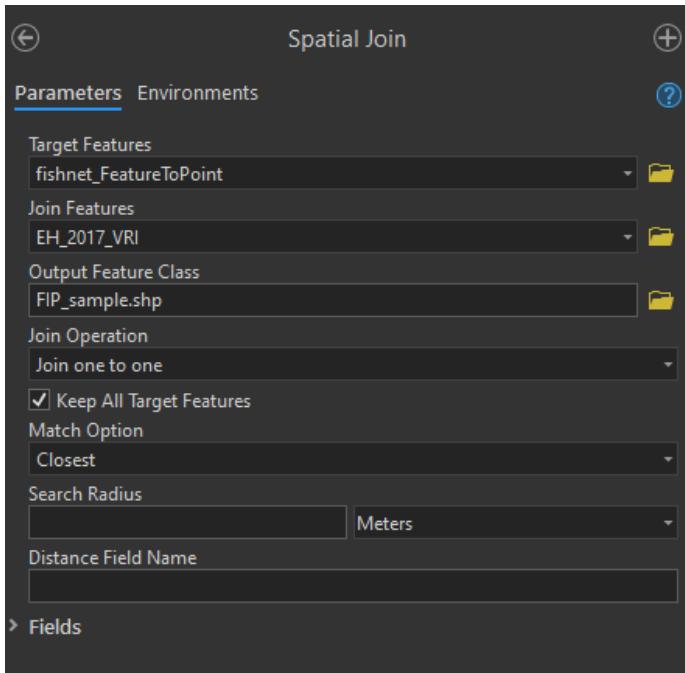
**Step 7:** To make the fixed interval point (FIP) sampled dataset, we will create a regularly spaced grid from which to sample our points by creating a fishnet of polygons, converting those polygons into centroid points, and then sampling the closest VRI point to our fishnet points. Open the “Create Fishnet” tool and name your “Output Feature Class” “fishnet”. Under “Template Extent” click “Browse” and select the vri shapefile. The X and Y Extents and Origin Coordinates should automatically populate. Set the “Number of Rows” to 20 and the “Number of Columns” to 25. This divides the extent into 20x25 (500!) identical segments. Change the “Geometry” to “Polygon” and run the tool. Your output feature class should look like the grid below.



**Step 8:** The tool will also create a point layer called “fishnet\_label” with points in the center of each polygon (shown below).



**Step 9:** We have a nice lattice of points across our study area, and we have a point feature class to sample from, but we need to put them together. Open the “Spatial Join” tool and set “fishnet\_label” (points not the polygons!) as the “Target Features” and your VRI point feature class as the “Join Features”. Set “Closest” as the “Match Option”. This will find the closest VRI point to the fishnet and join it to the table. The “Join Operation” should be “Join one to one” and your “Search Radius” can be left blank. Name your “Output Feature Class” “FIP\_sample” and run the tool.



You now have three sampling sets to try your machine learning with!

---

### Task 3: Forest-Based Classification

Our last task will be using the “Forest-based Classification and Regression” tool to run our machine learning. ESRI provides a helpful overview of the tool here.

We will be running it a total of nine times (three with FUEL\_TYPE, six with CROWN\_CLOS) and it can be slow, so please be patient and leave yourself lots of time.

**Step 1:** Open the “Forest-based and Boosted Classification and Regression” tool. When you search for the tool, check that it is from the “Spatial Statistics” toolbox and NOT the “GeoAnalytics Desktop” toolbox as they have different inputs. Parameterize the tool as follows:

- **Prediction Type:** Predict to Raster
- **Input Training Features:** random\_sample
- **Variable to Predict:** FUEL\_TYPE
- **Treat Variable as Categorical:** check this box (our variable is categorical)
- **Explanatory Training Rasters:** drag in your variable rasters that you selected from Exploratory Regression in Task 1. The screenshot below

shows one potential raster but you should have a maximum of 5.

- **Output Prediction Surface:** random\_output\_FT (CC for CROWN\_CLOS, FT for FUEL\_TYPE)
- **Match Explanatory Rasters:** they should already match, but double check the prediction rasters are matched to the correct field in the training data.
- **Additional Outputs > Output Classification Performance Table (Confusion Matrix):** random\_matrix\_FT
- **Validation Options > Training Data Excluded for Validation:** set to 30%
- **Validation Options > Number of Runs for Validation:** set to 2

Leave all other parameters as the default value or blank. Run the tool. You may get a warning about low numbers of certain fuel types – ignore it (for now...). When the tool finishes, click “View Details” and copy/paste the info under “Messages” into a text file. This is where you can find the regression statistics ( $R^2$ , RMSE, etc) for the training and validation data for your report.

**Step 2:** Run the tool again, but this time use “stratified\_sample” as your “Input Training Features”. Change the name of your outputs accordingly, but keep everything else the same. When this is finished, run the tool a third time with “FIP\_sample” as the input. Remember to keep your regression results in a text file. If you do forget, just go to “History” under the “Analysis” tab and hover over the tool.

You should now have three Random Forests outputs for fuel type from the three sampling strategies: - random\_output\_FT - stratified\_output\_FT - FIP\_output\_FT

**HINT: Under the “Analysis” tab, view your tool “History”.** Double-click the “Forest-based Classification and Regression” tool and it will pop up in your geoprocessing pane with the same inputs and settings you used for that particular run – this way, you can just edit the inputs that are changing between runs rather than having to reenter the settings each time.

**Step 3:** For CROWN\_CLOS, instead of changing the sampling design, we are going to experiment with different parameter settings on the tool for a total of **6 runs** (three different parameters, two settings each).

Open the “Forest-based Boosted Classification and Regression” tool again and set these inputs which will act as our control:

- **Prediction Type:** Predict to Raster
- **Input Training Features:** random\_sample
- **Variable to Predict:** CROWN\_CLOS
- **Treat Variable as Categorical:** Uncheck this box (our variable is continuous interval)
- **Explanatory Training Rasters:** same as FUEL\_TYPE

- **Output Prediction Surface:** control\_output\_CC
- **Match Explanatory Rasters:** same as FUEL\_TYPE
- **Advanced Forest Options > Number of Trees:** set to 50
- **Validation Options > Training Data Excluded for Validation:** set to 30%

Leave all other parameters as the default value or blank. Run the tool. When it is finished, click “View Details” and copy/paste the info under “Messages” into your text file.

**Step 4:** We are going to run the “Forest-based Classification and Regression” tool six more times. For each time, the inputs should be identical to those in Step 3 except for a single parameter which we will change each time:

- **Run 1 of 6:** Change Advanced Forest Options > Maximum Tree Depth to 5. Name the output “depth\_5\_output\_CC”
- **Run 2 of 6:** Change Advanced Forest Options > Maximum Tree Depth to 15. Name the output “depth\_15\_output\_CC”
- **Run 3 of 6:** Change Advanced Forest Options > Number of Randomly Sampled Variables to 1. Name the output “var\_1\_output\_CC”
- **Run 4 of 6:** Change Advanced Forest Options > Number of Randomly Sampled Variables to 2. Name the output “var\_2\_output\_CC”
- **Run 5 of 6:** Change Validation Options > Number of Runs for Validation to 2. Name the output “runs\_2\_output\_CC”
- **Run 6 of 6:** Change Validation Options > Number of Runs for Validation to 10. Name the output “runs\_10\_output\_CC”

**DO NOT FORGET TO CHANGE YOUR OTHER SETTINGS BACK TO THE DEFAULT BETWEEN EACH RUN.** If you want to be extra sure, open the “Forest-based Classification and Regression” tool from “History” where you ran your CROWN\_CLOS control (which has all the defaults) and do each run from there. You can check your inputs under “View Details” if you want to double-check.

When you are finished you should have 10 output rasters, 3 output confusion matrices, and 7 text files with your regression results.

**Step 5:** View all of your new rasters in your map and inspect the results. Symbolize them in a logical manner.

Your output rasters for FUEL\_TYPE may be numbers, even though the variable was a string! If your fuel type prediction rasters show numbers, do the following: on each of your sample feature classes, open the attribute table and summarize the unique values in the FUEL\_TYPE field using the same method as Task 2 Step 2. The OID in the summary table and the value in your corresponding output raster refer to the same fuel type, e.g., 0 is C-2, 1 is C-3, etc, as seen to the left. Change the labels for the legend in raster Symbology to the fuel type instead of the number when you make the map.

---

## Summary

There are many things to consider when undertaking machine learning with geospatial data. Although we only practiced with the random forest algorithm, many of these considerations are important for other methods as well. Machine learning algorithms work well for large datasets where you can randomly apportion testing, training and validation datasets. When applying these algorithms to spatial data, we often need to consider data density in the spatial and sometimes temporal domains. Interrogating spatial and temporal autocorrelation can help to identify the appropriate sampling strategy for your problem. As you learned in this lab, even the type of variable that you are predicting (continuous vs discrete) can require completely different methods for data handling, testing, and validation. Keep in mind that most of the data pre-processing was removed from the lab (see Data) so that you could focus on the analysis, though these efforts remain some of the most time-consuming to ensure a successful analysis.

Return to the **Deliverables** section to check off everything you need to submit for credit in the course management system

# Chapter 5

## Analyzing Green Equity Using Geographically Weighted Regression

Written by Paul Pickell

### Lab Overview

In this lab, you will be exploring a few different statistical approaches to modelling geographic data, including geographically weighted regression (GWR). GWR is the spatial extension of aspatial regression analysis and much more. Traditional regression analysis assumes that global statistics adequately describe local relationships that might exist in the data. For example, consider looking at the relationship between housing prices and the floor space, lot size, etc., of houses in the city of Vancouver. While we could develop a ‘global’ model that adequately describes the relationship between those variables, knowing what you do about housing prices in the city of Vancouver (e.g., that a house of similar dimensions, age, lot size, etc., in the east side of Vancouver will sell for hundreds of thousands of dollars less than an identical house in the west side of Vancouver), the utility of such a model when looking at neighborhood-level housing issues would be very doubtful. Nonetheless, for decades such models, such as hedonic models, have been normalized in real estate research.

Similarly, consider studying the relationship between rates of crime or diseases to environmental conditions, local conditions can be much more important than any global relationship that might be discovered via a traditional aspatial statistical approach. Using polygon or point data, GWR allows us to explore the local relationships amongst a set of variables and examine the results spatially

using ArcGIS Pro. It should be noted that in R you can find more sophisticated approaches to GWR than what is provided by ArcGIS Pro.

In this lab, you will explore the equity of green space for the city of Vancouver using Landsat imagery and demographic data from the 2021 Canadian census.

Why is access to green spaces so important? Human well-being, including physical and psychological well-being increase when residents are exposed to green space and urban forests. In addition, ecosystem services provided from green spaces include improved air quality, urban heat island mitigation, and opportunities for recreation. Yet, there is unequal access to green spaces across urban landscapes. The distribution of green space is often disproportionately present in affluent communities. So, you will test the hypothesis that there is less green space in marginalized communities. We cannot infer any causal relationships, but we can examine the relationship between the location of green spaces and demographic variables.

Vancouver is the most populous city in British Columbia, Canada with a population of 662,248 in 2021. Vancouver is an ideal study site because of the city's high level of heterogeneity among its demographic and green space structure.

---

## Learning Objectives

- Apply advanced SQL and PostGIS functions to a relational database to prepare high-dimensional census data for analysis
  - Calculate a vegetation index from Landsat imagery and report summary statistics over census dissemination areas
  - Evaluate different models and defend your model selection
  - Interpret charts and statistics of ordinary least squares and geographically weighted regression
  - Map geographically weighted regression results and interpret and defend your conclusions
- 

## Deliverables

Lab report with the following specification:

6 pages maximum PDF including figures, tables and references (3 points).  
Single-spaced, 12-point Times New Roman font (1 point).

Results should address the following questions and requirements (25 points):

- Describe the qualities of the census data and the Landsat image.
- How did you evaluate your models and select your final model? Report any relevant statistics that you used in your judgement.
- Why did you select the 10 characteristics that you choose in Task 3?
- A table with the ordinary least squares and geographically weighted regression results for the models that used the best subsets of your 10 characteristics.
- Describe all of the terms, coefficients, and explanatory variables of your selected GWR model.
- Justify your choice of a final set of independent variables for your GWR model.
- Compare and contrast the different NDVI statistics. Which statistic had the best model? What evidence do you have for that conclusion? Why do you think that relationship was the strongest?
- Interpret one of your independent variables in one of your three NDVI statistics using the Std.Error and Coefficient map. What spatial patterns do you see? What do you think could be influencing this relationship?
- Maps illustrating the standardized residuals and local  $R^2$  for each of your three different NDVI statistics.

Discussion should address the following questions and requirements (20 points):

- What other factors (spatial or aspatial) might be contributing or confounding your analysis? In other words, what other data sources might you add/calculate or what methods might you change to improve your results?
  - What can you conclude about green equity among dissemination areas in Vancouver?
  - What are your final recommendations to city council about green equity in Vancouver?
  - Reference to any peer reviewed sources as needed.
- 

## Data

All data for this lab are accessible via the UBC PostgreSQL server. Instructions for connecting to the server are given in the tasks below. We will be using data from the `greenequity` database.

Statistics Canada. 2022. Census Profile. 2021 Census. Statistics Canada Catalogue no. 98-316-X2021001. Ottawa. Released December 15, 2022. <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E>

We are using only a small subset of the national 2021 census data set for British Columbia: “Canada, provinces, territories, census divisions (CDs), census subdivisions (CSDs) and dissemination areas (DAs) - British Columbia only” (Statistics Canada Catalogue no. 98-401-X2021006).

The Statistics Canada 2021 spatial boundary files are maintained separately and available for download from here: <https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21>

The spatial data from Statistics Canada that we will be using:

Layer Name	Description
lcsd000b21a_e	Census subdivisions
lda_000b21a_e	Dissemination areas

If you are a student at UBC, these data have already been prepared and loaded into the UBC PostgreSQL server, so there is no need to download anything. The links above are only for reference.

Metadata for the 2021 spatial boundary files can be found here: <https://www150.statcan.gc.ca/n1/pub/92-160-g/92-160-g2021002-eng.htm>

The Dictionary for Census of Population 2021 can be found here: <https://www12.statcan.gc.ca/census-recensement/2021/ref/dict/index-eng.cfm>

---

## Task 1: Prepare census data

Statistics Canada census data are distributed in tables, which are great for working with in a relational database like PostgreSQL. These data can be particularly challenging to work with because they span multiple geographical hierarchies (e.g., national, provincial, municipal, etc.), multiple dates (the Canadian census occurs every 5 years), many demographic dimensions (e.g., population, age, education, language, etc.), and there are an enormous amount of enumerated areas.

The smallest geographic unit that census data are enumerated over are known as Dissemination Areas (DA). Statistics Canada gives the definition:

A dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons based on data from the

previous Census of Population Program. It is the smallest standard geographic area for which all census data are disseminated. DAs cover all the territory of Canada.

As of the 2021 Canadian census, there are 57,936 unique Dissemination Areas. Each Dissemination Area is described by 2,631 unique characteristics (total population, age, education, language, etc.). That is a whopping 152 million values for describing Canadians! Lucky for us, we will be working with DAs for Vancouver, British Columbia and only a handful of characteristics.

To make this a realistic exercise, we will be working with the raw Statistics Canada table for British Columbia and the national set of dissemination areas, which have been loaded into the UBC PostgreSQL server. These data are also publicly available if you want to replicate the lab on your own (see Data section)

**Step 1:** Open QGIS and connect to the `greenequity` database on the UBC PostgreSQL server using the credential that you have been provided. You might be tempted to add the layers `lda_000b21a_e` and `lcsd000b21a_e` to your map. You can do this, but it will probably slow down your computer as you are requesting all 57,936 dissemination areas and 5,161 census subdivisions from the server, respectively. Instead, we will use some PostGIS magic to filter on the server side before making our request for only the Vancouver DAs to our QGIS client.

Unfortunately, the only geographical identification attribute that Statistics Canada distributes with the dissemination areas layer (`lda_000b21a_e`) is “PRUID”, which limits us to querying on provinces and territories based on codes. Cities, like Vancouver, are available in the census subdivisions layer (`lcsd000b21a_e`), so we will need to do a simple overlay to extract only the dissemination areas in Vancouver.

**Step 2:** Open the Database Manager in QGIS (“Database” > “DB Manager”). Expand the “PostGIS” source provider on the left, expand the “`greenequity`” database, and finally expand the “public” schema. Click the button at the top to open a new SQL window. In the empty query space, paste the following SQL query:

```
WITH vancouver AS (
    SELECT wkb_geometry FROM lcsd000b21a_e
    WHERE CSDNAME = 'Vancouver'
)
SELECT lda_000b21a_e.wkb_geometry
FROM vancouver, lda_000b21a_e
WHERE ST_Within(lda_000b21a_e.wkb_geometry, vancouver.wkb_geometry);
```

This syntax should look familiar if you have completed the previous PostGIS labs. A common table expression (CTE) is used to get the census subdivision representing Vancouver, which is then used to get all the dissemination areas that represent Vancouver. We are introducing a new PostGIS function here

called `ST_Within`, which simply tests if the geometry of `l1da_000b21a_e` are completely within the geometry of `vancouver`. Click the “Execute” button and inspect the output on screen. You should see 1,016 rows returned. But in order to join the census data, which are stored separately, we will also need the DAUID and DGUID fields in this output layer.

**Step 3:** Modify the query above so that you return a table that holds the DAUID, DGUID, and the wkb\_geometry fields. Once successful, your table should look like the image below.

Execute		1016 rows, 1.102 seconds	Create a view	Clear
	david	dguid	wkb_geometry	
1	59150581	2021S051259150581	0106000020130...	
2	59150583	2021S051259150583	0106000020130...	
3	59150584	2021S051259150584	0106000020130...	
4	59150585	2021S051259150585	0106000020130...	
5	59150586	2021S051259150586	0106000020130...	
6	59150587	2021S051259150587	0106000020130...	
7	59150588	2021S051259150588	0106000020130...	
8	59150589	2021S051259150589	0106000020130...	

If you want, you can toggle on “Load as a new layer”, set “Column(s) with unique values” to “david”, set “Geometry column” to “wkb\_geometry”, and click the “Load” button to temporarily load the query into your map to inspect it. This is not going to be our final layer yet, as we still need to join the appropriate census data to the geographic features.

**Step 4:** Open the psql shell and connect to the `greenequity` database on the UBC PostgreSQL server using the credential that you have been provided. Use the following query to return the first 20 rows of the “census2021” table:

```
SELECT * FROM census2021 LIMIT 20;
```

Do you notice that the data are oddly formatted? Notably, the first eight columns have repeating information, including the primary key “GUID”! This data organization is known as **long-format** and is useful when storing highly-dimensional data, and the census data you are working with have 2,631 dimensions!

Think of long-format as a list. Each row represents a new entry that holds a value or several values for a given observation-dimension pair. The advantage here is that if we do not have a value for a particular dimension for a partic-

ular observation, then we do not need to record anything, the row does not exist. Another advantage is that most relational database management systems have a limit on the number of columns that can be stored in a single table. For PostgreSQL, that limit is 1,600 columns compared to 255 for Microsoft Access databases, 1,024 for Microsoft SQL Server databases, and 1,000 for Oracle databases, just to name a few. However, there is practically no upper limit to the number of rows that can be stored in any of these databases.

This is a bit odd, because for most data analysis we are accustomed to seeing **wide-format** data where each row represents a unique feature that is described by many dimensions across the columns.

**Step 5:** Use the following query to return a table of population in the first 10 DGUIDs:

```
SELECT alt_geo_code, MAX(c1_count_total) FILTER (WHERE characteristic_id = 1) AS population
FROM census2021
GROUP BY alt_geo_code
LIMIT 10;
```

In this query, we are aggregating and grouping the data values back into our customary wide-format. We use the `MAX` aggregation function, though any other aggregation function would work since for any given dimension-observation there is only one value, so nearly any aggregation function applied to a single value would return that value (e.g., `max`, `min`, `avg`, `sum`). The reason for using the aggregation function at all is because the `FILTER` keyword will only apply aggregation functions to the predicate that follows (`WHERE characteristic_id = 1`). Finally, `GROUP BY alt_geo_code` ensures that the output table lists the populations for each `alt_geo_code`.

### A word about `alt_geo_code`

The `alt_geo_code` is an identifier for every unique geographic division for census data in Canada. Some codes are shorter or longer than others, and this signifies the level in the geographic division hierarchy. For example, `alt_geo_code = 1` is the code for the entire country of Canada. So the first row you see in the output table from the query you ran above is the total population `characteristic_id = 1` of Canada in 2021. Since the `alt_geo_code` is numeric, it is more flexible to query with than the DGUID, which is alphanumerical. `alt_geo_code = 59` is the code for the province of British Columbia. Remember that the data we are working with is only a subset of the national census representing the province of British Columbia, so there are no other provincial codes represented here. The code structure is pretty easy to follow from here. Every other geographic subdivision in British Columbia will begin with 59 and depending on the level of the hierarchy, will contain a different number of digits:

- `alt_geo_code = 59` represents British Columbia in the **Provinces and**

### Territories Unique Identifier (PRUID)

- alt\_geo\_code = 5901 represents the first 01 **Census Division (CD)**
- alt\_geo\_code = 5901003 represents the third 003 **Census Subdivision (CSD)** in the first CD
- alt\_geo\_code = 59010100 represents the 100th 0100 **Dissemination Area (DA)** in the third CSD, in the first CD

Codes with 8 digits are therefore **Dissemination Area Unique Identifiers (DAUID)**, which you should recognize are the values we need to use to join the census data to the actual dissemination area polygons that we just extracted in QGIS.

**Step 6:** We can modify our earlier query slightly to ensure we are only dealing with DAs:

```
SELECT alt_geo_code, MAX(c1_count_total) FILTER (WHERE characteristic_id = 1) AS population
FROM census2021
WHERE geo_level = 'Dissemination area'
GROUP BY alt_geo_code
LIMIT 10;
```

**Step 7:** Now modify the query above so that it will produce an output table that contains all of the characteristics listed in the table below. Some IDs are given in the table below. For the remainder, you are expected to find the correct ID. You may need to refer to the readme file that is distributed with the 98-401-X2021006 table in order to identify the correct characteristic\_id, which is listed under “Definitions / Footnotes”, “Characteristic (2631)”, and “Member”. The number appearing to the far left of the readme file corresponds to the characteristic\_id in the database.

Characteristic_ID	Characteristic	Units	Name
	Population, 2021	Persons	population
	Population density per square kilometre	Persons per square kilometre	popdensity
	Total - Private households by household size - 100% data	Families	households
	Average household size	Persons	hhszie

Characteristic	ID	Characteristic	Units	Name
345		Prevalence of low income based on the Low-income measure, after tax (LIM-AT) (%)	%	lowincome
2008		Bachelor's degree or higher	Persons	education
		Average age of the population	Years	age
35		0 to 14 years	%	children
37		65 years and over	%	seniors
		Unemployment rate	%	unemployment
113		Median total income in 2020 among recipients (\$)	\$	medianincome
392		First official language spoken is neither English nor French	Persons	neitherenglishorfrench
398		Mother tongue is a non-official language	Persons	nonofficiallanguage
1536		Immigrants arriving in 2016-2021	Persons	immigrants

Your output table should look like the image below:

1016 rows, 7.465 seconds														Query History
	Execute	population	popdensity	households	hsize	lowincome	education	age	children	seniors	unemployment	medianincome	other	▼
1	59150581	390.0	3475.9	125.0	3.1	19.0	195.0	41.6	11.5	16.7	9.7	34800.0	50.0	
2	59150583	593.0	5410.6	265.0	2.3	11.2	285.0	45.4	11.8	24.4	4.0	52400.0	10.0	
3	59150584	604.0	6771.3	290.0	2.1	9.6	305.0	44.4	10.0	20.8	14.0	52400.0	10.0	
4	59150585	544.0	4437.2	195.0	2.8	8.4	230.0	40.7	15.6	16.5	3.7	50000.0	10.0	
5	59150586	554.0	3137.0	290.0	1.9	9.6	280.0	38.5	10.9	11.8	7.9	51200.0	5.0	
6	59150587	570.0	6521.7	255.0	2.3	8.0	275.0	40.8	10.5	16.7	16.3	49600.0	0.0	
7	59150588	581.0	5946.8	240.0	2.4	10.8	300.0	39.0	12.1	12.9	8.6	42000.0	15.0	
8	59150589	493.0	5082.5	190.0	2.7	12.8	260.0	41.3	14.3	18.4	9.5	40800.0	15.0	
9	59150590	552.0	5679.0	215.0	2.6	9.6	225.0	41.3	14.4	17.1	10.0	44400.0	15.0	▼

**Step 8:** Finally, we are going to put everything together into a single SQL statement that will query, intersect, aggregate, group, and join the spatial data with the tabular data. Save your SQL query in QGIS Database Manager or copy it and paste it somewhere, then clear the SQL query area and paste the following statement:

## 92CHAPTER 5. ANALYZING GREEN EQUITY USING GEOGRAPHICALLY WEIGHTED REGRESSION

```
WITH
```

```
-- Select the alt_geo_code and characteristics we want, aggregate and group them to create the final output table
characteristics AS (
    -- Insert your modified query from Step 7 here
),

-- Select the polygon geometry representing Vancouver from the Census Subdivisions (CSD)
vancouver AS (
    SELECT wkb_geometry
    FROM lcsd000b21a_e
    WHERE CSDNAME = 'Vancouver'
),

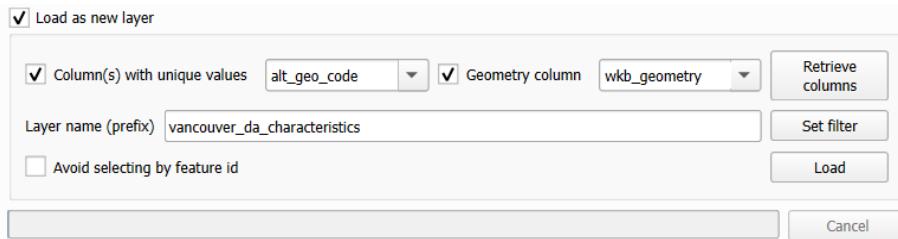
-- Select the Dissemination Areas (DA) that are within the Vancouver CSD
vancouver_da AS (
    SELECT lda_000b21a_e.DAUID, lda_000b21a_e.wkb_geometry
    FROM vancouver, lda_000b21a_e
    WHERE ST_Within(lda_000b21a_e.wkb_geometry, vancouver.wkb_geometry)
)

-- Perform the join of all the characteristics and the Vancouver DA geometries based on the DAUID
SELECT
    characteristics.alt_geo_code,
    characteristics.population,
    characteristics.popdensity,
    characteristics.households,
    characteristics.hhszie,
    characteristics.lowincome,
    characteristics.education,
    characteristics.age,
    characteristics.children,
    characteristics.seniors,
    characteristics.unemployment,
    characteristics.medianincome,
    characteristics.neitherenglishorfrench,
    characteristics.nonofficiallanguage,
    characteristics.immigrants,
    vancouver_da.wkb_geometry
FROM
    characteristics
```

**JOIN**

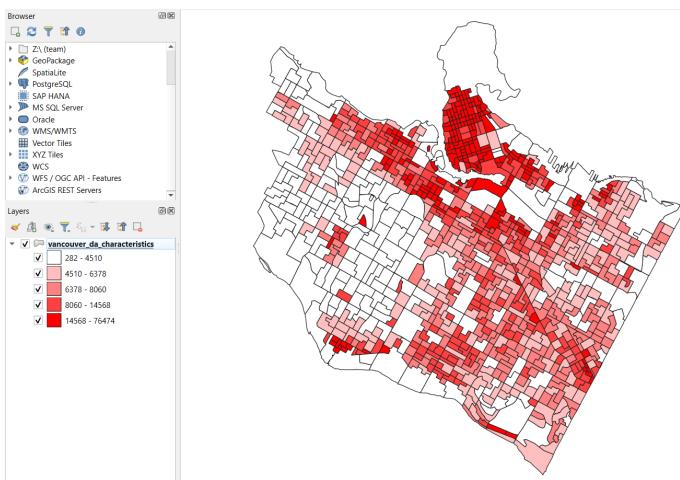
```
vancouver_da ON characteristics.alt_geo_code::integer = vancouver_da.DAUID::integer;
```

Be sure to add your modified statement from Step 7 where it is indicated. Then, the statement above should produce the final table we are looking for: a layer of dissemination areas for Vancouver that contain the values for the 14 characteristics we are interested in.



**Step 9:** When you are satisfied, export the layer to your QGIS map and save it in your project folder (right-click the layer, “Export” > “Save Features As...”). We suggest exporting as shapefile format to maintain compatibility with ArcGIS Pro.

You may also want to save your SQL statement somewhere, too. You can use it later to recover all of your work from this task at any time from the PostgreSQL server. In QGIS, you can now play around with symbolizing the different characteristics. The image below shows population density  $\frac{\text{persons}}{\text{km}^2}$  of Vancouver dissemination areas in 2021 (projected coordinate system is Lambert Conformal Conic).



**Step 10:** Save your QGIS project.

## Task 2: Practice geographically weighted regression

In this task, we are going to calculate the Normalized Difference Vegetation Index (NDVI) of each dissemination area in Vancouver and then select characteristics from the census data that explain the local variation in vegetation greenness, as expressed by NDVI. So the census characteristics are going to be our independent (explanatory) variables  $k$  and the calculated statistics of NDVI are going to be the dependent (response) variable  $y_i$  for our geographically weighted regression:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$$

$\beta_0(u_i, v_i)$  is the local model intercept at position  $(u_i, v_i)$

$\beta_k(u_i, v_i)$  is the local coefficient (slope) of the  $k$ -th independent variable (census characteristic) at position  $(u_i, v_i)$

$x_{ik}$  is the local  $i$ -th observation of the  $k$ -th independent variable (census characteristic)

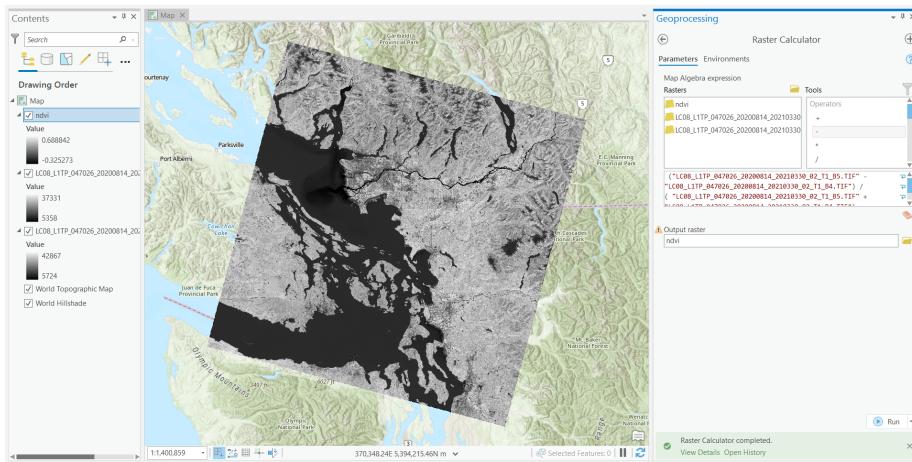
$\varepsilon_i$  is the local error term (residual) for the  $i$ -th prediction

All of the instructions that follow are written for ArcGIS Pro because we will be performing the geographically weighted regression in ArcGIS Pro. We suggest setting up a new project and copying the “vancouver\_da\_characteristics” shapefile that you produced in the previous task into that project folder.

**Step 1:** First we will save the Landsat tiles as tifs. Connect to the `greenequity` database on the UBC PostgreSQL server in QGIS. Add two Landsat raster images: `LC08_L1TP_047026_20200814_20210330_02_T1_B4` and `LC08_L1TP_047026_20200814_20210330_02_T1_B5`. You will need to set the CRS of the rasters to UTM Zone 10 (EPSG: 32610) and export to geotif. These images represent bands 4 (red) and 5 (near-infrared), respectively, and were acquired on August 14, 2020, which is approximately when the 2021 census data were collected.

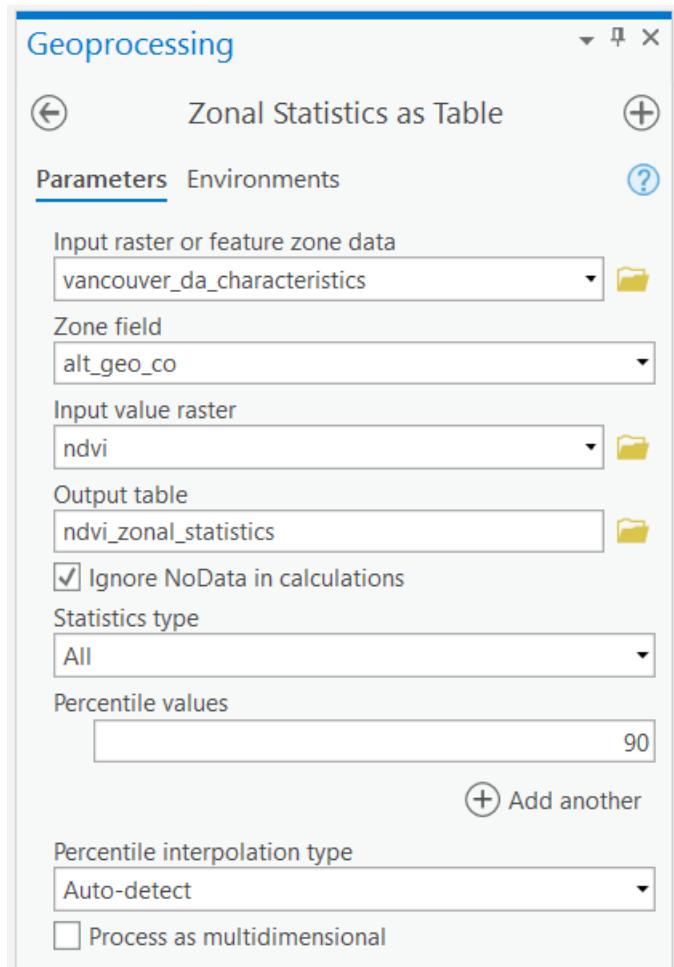
**Step 2:** Now in ArcGIS, open the geotiffs you just created. Open the “Raster Calculator” tool and calculate the Normalized Difference Vegetation Index (NDVI) and save the output in your project geodatabase simply as “ndvi”:

$$NDVI = \frac{Band5 - Band4}{Band5 + Band4}$$



**Step 3:** Now we need to summarize the NDVI values over the dissemination areas. Open the “Zonal Statistics as Table” tool and use “vancouver\_da\_characteristics” as the “Input raster or feature zone data”, select “alt\_geo\_co” as the “Zone field”, use “ndvi” as the “Input value raster”, and name the “Output table” as “ndvi\_zonal\_statistics”. Ensure that “Statistics type” is set to “All”, leave the other fields as default and run the tool.

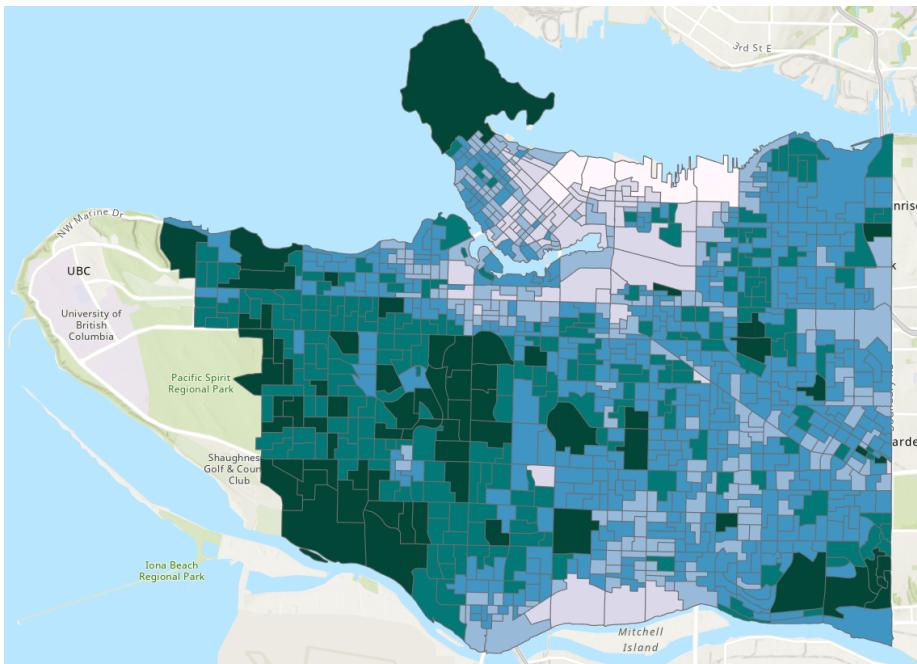
## 96CHAPTER 5. ANALYZING GREEN EQUITY USING GEOGRAPHICALLY WEIGHTED REGRESSION



This will produce a table that should look like the image below. The table contains summary statistics of NDVI calculated for each dissemination area. Now we need to join this table to the polygon feature class.

	OBJECTID *	alt_geo_co	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM	MEDIAN	PCT90
1	59150307	344	309600	0.040504	0.490431	0.449928	0.255403	0.078936	87.858482	0.241773	0.362556	
2	59150308	130	117000	0.12102	0.301768	0.180747	0.199177	0.041597	25.892972	0.190761	0.260801	
3	59150309	120	108000	0.027569	0.37528	0.347711	0.191453	0.075939	22.974345	0.200279	0.276725	
4	59150310	121	108900	0.023724	0.395474	0.37175	0.174382	0.09216	21.100902	0.176804	0.297772	
5	59150311	88	79200	0.137352	0.413987	0.276634	0.248319	0.069086	21.852113	0.235089	0.352274	
6	59150312	92	82800	0.085763	0.421946	0.336183	0.234682	0.089918	21.590789	0.207706	0.382812	
7	59150313	189	170100	0.066428	0.593144	0.526716	0.30131	0.13854	56.947558	0.261707	0.554626	
8	59150328	262	235800	0.012696	0.455591	0.442895	0.153418	0.103814	40.195575	0.132325	0.302454	
9	59150329	327	294300	0.006419	0.581047	0.574628	0.278462	0.148368	91.056925	0.236215	0.480618	
10	59150330	94	84600	0.091012	0.319956	0.228943	0.190786	0.051581	17.934152	0.185233	0.262784	
11	59150331	159	143100	0.030479	0.344333	0.313854	0.188262	0.05554	29.933607	0.188427	0.254612	

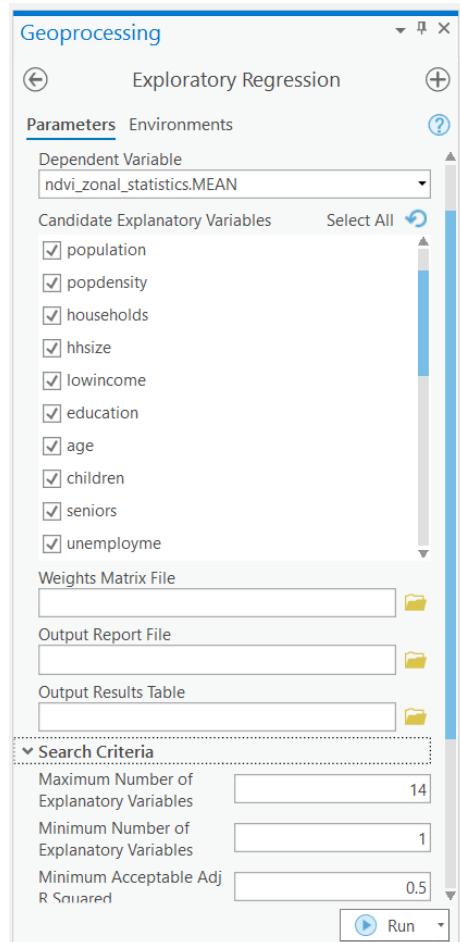
**Step 4:** Right-click on the “vancouver\_da\_characteristics” layer in your Contents Pane and select “Joins and Relates”, then “Add Join”. The “Input Table” is “vancouver\_da\_characteristics” and the “Join Table” is “ndvi\_zonal\_statistics”. Select the correct keys to join the tables. This is a one-to-one join. Map the output, below is an example of average NDVI.



It is important to initially analyze our census characteristics  $k$  to determine which independent variables and combination of these variables have the strongest relationship with our dependent variable, NDVI  $y$ . To conduct the initial analysis, we will use a tool called “Exploratory Regression”, which is part of the Spatial Statistics Toolbox.

**Step 5:** Open the “Exploratory Regression” tool. Select “vancouver\_da\_characteristics” as your “Input Features” and “ndvi\_zonal\_statistics.MEAN” as your dependent variable. Select all of the census characteristics as your “Candidate Explanatory Variables”, expand “Search Criteria” and change the “Maximum Number of Explanatory Variables” to 14 then run the tool with other fields as default.

## 98CHAPTER 5. ANALYZING GREEN EQUITY USING GEOGRAPHICALLY WEIGHTED REGRESSION



*NOTE: Some fields were truncated when we saved the “vancouver\_da\_characteristics” shapefile. It should still be apparent which fields to select, but some of the original names will not perfectly match.*

**Step 6:** When the tool has finished running, click “View Details” at the bottom and then click “Messages”. Under the heading “Highest adjusted R-squared results”, you can explore the modeled relationship between one or more independent variables and the dependent variable. You should see the adjusted R-squared plateau at 0.41 when using a model with seven independent variables. We can use the statistics like Akaike’s Information Criterion (AICc), Jarque-Bera p-value (JB), and Max Variance Inflation Factor (VIF) to choose between similar models with different sets of independent variables. Be sure to copy-paste this output message to a notepad so that you can reference it later in your report.

Exploratory Regression (Spatial Statistics Tools)

Started: Today at  
Completed: Today at  
Elapsed Time: 24 Seconds

Parameters Environments Messages (49)

Choose 1 of 14 Summary

Highest Adjusted R-Squared Results

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.22	-3191.13	0.00	0.89	1.00	0.00	+VANCOUVER_DA_CHARACTERISTICS.CHILDREN***
0.17	-3125.26	0.00	0.04	1.00	0.00	-VANCOUVER_DA_CHARACTERISTICS.POPDENSITY***
0.15	-3102.69	0.00	1.00	1.00	0.00	+VANCOUVER_DA_CHARACTERISTICS.HHSIZE***

Passing Models

AdjR2	AICc	JB	K(BP)	VIF	SA	Model
-------	------	----	-------	-----	----	-------

Choose 2 of 14 Summary

Highest Adjusted R-Squared Results

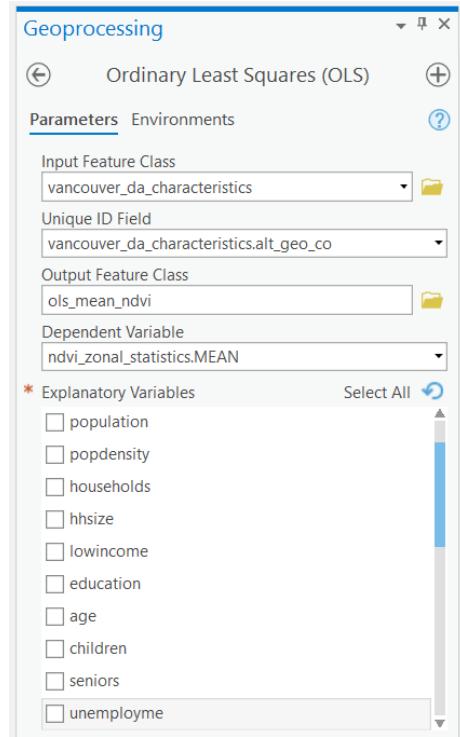
AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.31	-3303.94	0.00	0.09	1.01	0.00	+VANCOUVER_DA_CHARACTERISTICS.CHILDREN*** +VANCOUVER_DA_CHARACTERISTICS.SENIORS***
0.29	-3279.00	0.00	0.06	1.10	0.00	+VANCOUVER_DA_CHARACTERISTICS.AGE** +VANCOUVER_DA_CHARACTERISTICS.CHILDREN***
0.28	-3264.40	0.00	0.00	1.22	0.00	-VANCOUVER_DA_CHARACTERISTICS.POPDENSITY*** +VANCOUVER_DA_CHARACTERISTICS.CHILDREN**

Passing Models

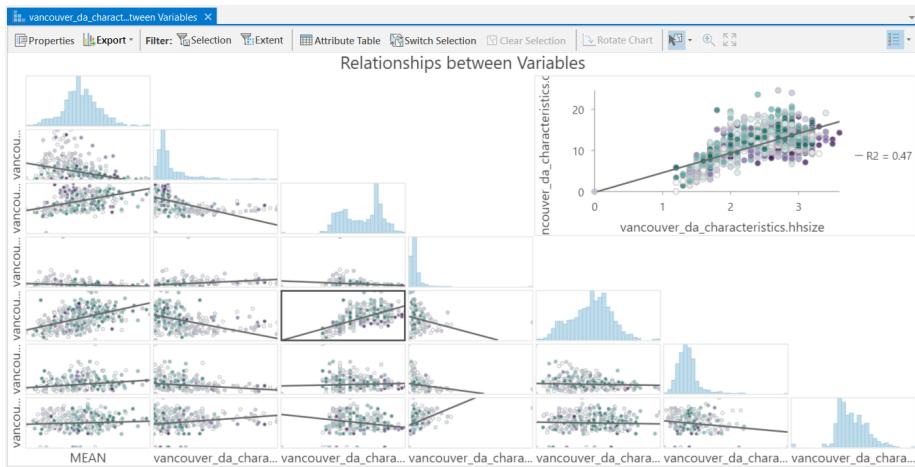
AdjR2	AICc	JB	K(BP)	VIF	SA	Model
-------	------	----	-------	-----	----	-------

**Step 7:** Once you have selected a model, write down the independent variables that are used in the model. Open the “Ordinary Least Squares (OLS)” tool. The “Input Feature Class” is “vancouver\_da\_characteristics”, “Unique ID Field” is “vancouver\_da\_characteristics.alt\_geo\_co”, name the “Output Feature Class” as “ols\_mean\_ndvi”, set the “Dependent Variable” to “ndvi\_zonal\_statistics.MEAN”, and then select all of the independent variables that you wrote down from the last step. Run the tool, then click “View Details”, select “Messages”, and copy-paste the output to a notepad to reference it later in your report.

## 100CHAPTER 5. ANALYZING GREEN EQUITY USING GEOGRAPHICALLY WEIGHTED REGRESSION



**Step 8:** Open the “Geographically Weighted Regression (GWR)” tool and parameterize it the same as you did in the last step for OLS, but change “Neighborhood Type” to “Number of neighbors”, change “Neighborhood Selection Method” to “Golden search”, and set “Minimum Number of Neighbors” to 50 and “Maximum Number of Neighbors” to 250. Name the “Output Features” as “gwr\_mean\_ndvi”. Again, select all of the independent variables that you wrote down from earlier then run the tool. The output will automatically be added to your map along with several charts. Doubling-clicking on a chart will open it.



**Step 9:** Save your ArcGIS Pro project.

---

### Task 3: Analyze your own census characteristics

Your final task for this lab is to repeat Steps 8 and 9 of Task 1, but this time, select 10 of your own census characteristics and then perform your own analysis using what you have learned in Task 2. We strongly recommend referring to the readme file and other metadata that are linked in the **Data section** of the lab.

**Step 1:** Select 10 different census characteristics other than the 14 we worked with and repeat Steps 8 and 9 of Task 1. You will be expected to rationalize why you selected these 10 characteristics in your report.

**Step 2:** Run all 10 characteristics through the “Exploratory Regression” tool. Repeat this process with three different NDVI statistics (e.g., mean, max, min, standard deviation, etc.). Record the output message and refer to any statistics here to justify your choice of a final set of independent variables for your regressions.

**Step 3:** Using your selected subset of independent variables, run the “Ordinary Least Squares (OLS)” tool for each of the three dependent variables, NDVI statistics that you chose in Step 2. Record the output message and refer to these statistics when you compare your OLS results to your GWR results.

**Step 4:** Using your selected subset of independent variables, run the “Geographically Weighted Regression (GWR)” tool for each of the three dependent variables, NDVI statistics that you chose in Step 2. Record the output message and refer to these statistics when you compare your GWR results to your OLS results.

**Step 5:** Explore your GWR results and make maps of the following for each of the three different NDVI statistics that you will reference in your report: - Standardized residuals - Local R<sup>2</sup>

**Step 6:** Explore your GWR results and make a map of one “Coefficient” and one “Std.Error” for one of your independent variables. Reference these maps when you are explaining your results in your report.

**Step 7:** Answer the following questions in your report and refer to all the maps, tables, and figures you made in the previous steps:

- Compare and contrast the different NDVI statistics. Which statistic had the best model? What evidence do you have for that conclusion? Why do you think that relationship was the strongest?
- Interpret one of your independent variables in one of your three NDVI statistics using the Std.Error and Coefficient map. What spatial patterns do you see? What do you think could be influencing this relationship?
- What other factors (spatial or aspatial) might be contributing or confounding your analysis? In other words, what other data sources might you add/calculate or what methods might you change to improve your results?
- What can you conclude about green equity among dissemination areas in Vancouver?

You will be expected to rationalize why you selected these 10 characteristics in your report.

---

## Summary

Geographically weighted regression can be a powerful tool for exploring spatial relationships. It takes some care and practice learning to interpret the many statistics along the journey, but it is one of the statistical methods that is rewarding to map and visualize. You should think of geographically weighted regression as a first approach at looking at a problem. It is great for exploring relationships, but not necessarily testing them. As you have seen, geographically weighted regression is a wonderful way to generate spatial hypotheses about data and explore the underlying tendencies of different relationships. Along the way, you have also learned how to wield high-dimensional census data in a database. Census data pair well with a wide variety of spatial analyses once you have decoded and unlocked their spatial mysteries.

Return to the **Deliverables** section to check off everything you need to submit for credit in the course management system.

# **Chapter 6**

# **Web Mapping**

Written by Paul Pickell

## **Lab Overview**

In this lab you will learn how to make web maps. Web maps are maps that are formatted and stylized for consumption through a web browser. Usually, they include interactive features that simulate the feeling of using other GIS software (e.g., panning, zooming, toggling layers on/off, etc), but with limited analytical capacity. The aim of a good web map is no different than other map: we want to communicate an idea or story. Web mapping gives more control to the end user to explore and visualize the data. So as web cartographers, our role is primarily reduced to decisions about what data to present (e.g., features, attributes, and values) rather than how (i.e., style). We will explore web mapping through ArcGIS StoryMaps. You will learn to publish web maps through ArcGIS StoryMaps, which is integrated with ArcGIS Online. You will create a StoryMap for your FCOR 599 project. The final product will be used to archive your project results on the MGEM Project Library.

---

## **Learning Objectives**

- Prepare and manage geographic data for display on a web map
  - Design and style layers and other elements in an ArcGIS StoryMap
-

## Deliverables

Link to your published ArcGIS StoryMap with the following requirements (50 points):

- A cover slide with a title, subtitle, author name, etc. Picture optional depending on which cover format you choose. (5 points)
  - All major text sections of your final report: Introduction, Study Area, Methods, Results, Discussion, and Conclusions/Recommendations. (10 points)
  - The sidecar explaining three key concepts/steps from either your Methods, Results, or Discussion sections. (10 points)
  - An Express Map of your choosing with a title – can be placed anywhere (sidecar, intro, etc.) (10 points)
  - At least two Custom Maps with titles showing your key results. (10 points)
  - Reference to any peer reviewed sources as needed, images are properly attributed and captioned. (5 points)
- 

## Data

Data for this lab will come from your FCOR 599 final project.

---

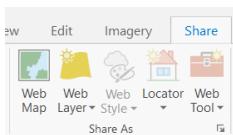
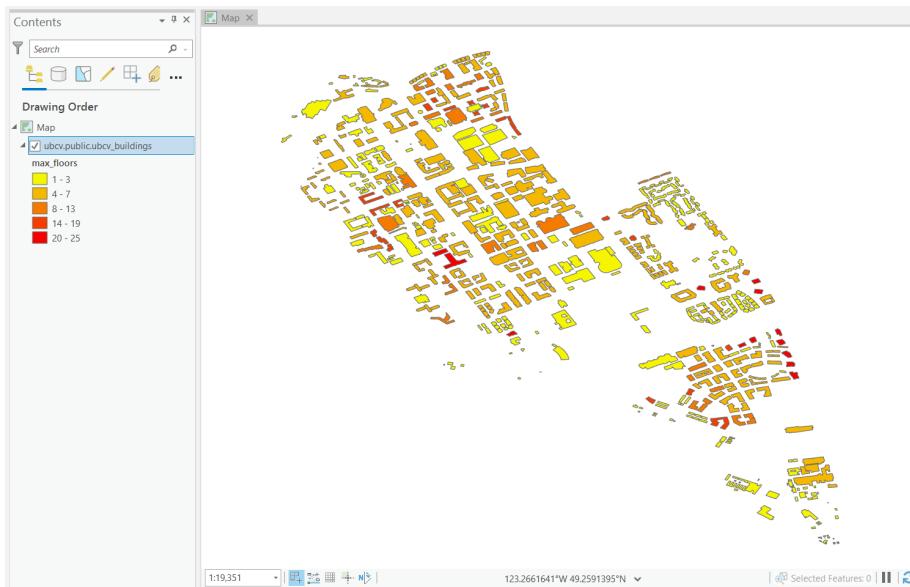
### Task 1: Adding content to your ArcGIS Online profile

The mapping functionality available in ArcGIS Online has developed rapidly in recent years, moving from mostly visualization-oriented tools to an expanded suite of processing, analysis, and visual editing tools. ArcGIS Online provides portability, shareability, and more rapid workflow for computers with limited processing power. In this task, we will explore the capability of ArcGIS Online for hosting web layers, which will ultimately allow us to interact with those web layers in web maps.

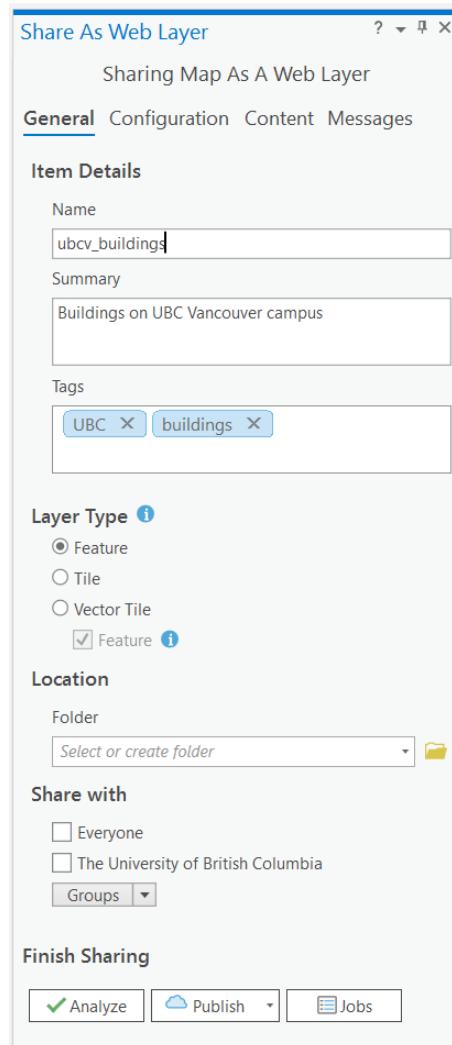
If you would like to explore the interface and functionality of ArcGIS Online, try adding in some raw or partially processed data from your project, test out some tools, and then make a map from scratch. There are several in-depth tutorials and useful references on the ESRI Website. The resources below are a great starting point:

Web Maps Help Page: <https://doc.arcgis.com/en/arcgis-online/reference/what-is-web-map.htm> Try ArcGIS Online: <https://learn.arcgis.com/en/parts/try-arcgis-online/> Try the New Map Viewer: <https://learn.arcgis.com/en/parts/try-the-new-map-viewer/>

**Step 1:** Open ArcGIS Pro and load in one vector layer from your project. Ensure that the layer has a Geographic Coordinate System of WGS 1984 and is not projected. Symbolize it however you like and remove all other layers from the map.



**Step 2:** Under the Share tab on the top ribbon, you will see two options: “Web Map” and “Web Layer”. Select “Web Layer” and “Publish Web Layer” under the Share tab. The default name will be the name of your map, but give the layer a meaningful name, add a few words of summary, and add some relevant tags. Select “Feature” as your “Layer Type”, then click “Analyze”. If you do not get any warnings, then click “Publish”. You must be connected to the internet in order to publish the content to your ArcGIS Online profile.



If you get warnings saying “24078 Layer’s data source is not supported” then ensure that you have removed the default basemap and hillshade from your map. Also ensure that your feature layer is in the WGS 1984 coordinate system and not projected.

**Step 3:** Repeat Step 2 with a raster layer from your project, but change the “Layer Type” to “Tile”. Again, ensure that the layer is in the WGS 1984 Geographic Coordinate System and is not projected. “Analyze” and then “Publish” the layer if there are no warnings.

**Step 4:** After completing Steps 2 and 3, a green message should appear with a URL saying “Manage the web layer”. You can click the URL to take you to the hosted feature/tile layer on ArcGIS Online or you can click here to log in

to your ArcGIS Online account. Navigate to the Content tab. You should see both your vector and raster layers listed there as “Feature Layer (hosted)” and “Tile Layer (hosted)”.

**Step 5:** Click each of your hosted layers and explore their page. Under the Data tab for the Feature Layer, you can look at the attribute table for your layer. You will notice several options on the right side for exporting, updating, and sharing your data, amongst others. Click the “Open in Map Viewer” option. This will take you to ArcGIS Online’s web mapping tool. Take a moment to explore the ways you can process and visualize your data.

**Step 6:** Navigate back to ArcGIS Pro. Instead of selecting Web Layer, this time select “Web Map”. Give your map a name, summary, and tags, and select your configuration to be “Editable”. Click “Analyze”. You will likely get an error about your projection. This is because all Web Maps on ArcGIS online are required to be in the same projection: WGS 1984 Web Mercator (auxiliary sphere). Make this change for your map and “Analyze” again. Note that your individual layers should not need to have their geographic coordinate system changed. If this error has been fixed, click “Publish”.

**Step 7:** Go back to ArcGIS Online and go to your Content tab again. You will see several new additions – a Web Map as well as hosted layers and service definitions for everything in your map. Click on your map and view it in the Map Viewer.

Now that you have some layers and maps available, you are ready to make put them in your StoryMap in the next task!

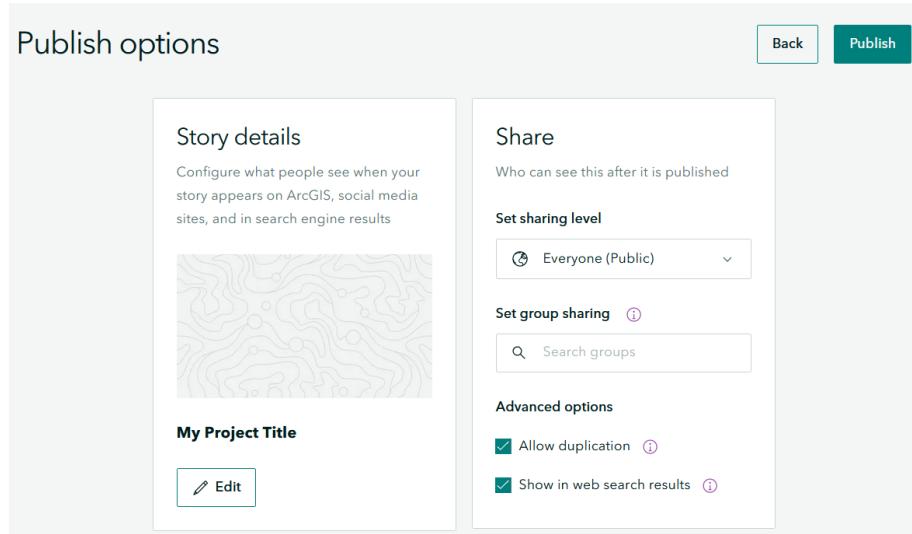
## Task 2: Creating your ArcGIS StoryMap

Your Story Map should be written and designed for a lay audience. All Story Maps will be published publicly on the MGEM Project Library website, so this is what people will see when they look at your project.

While you are making your map, there are a few things you should keep in mind. First, this is a standalone presentation (i.e., you will not be orally presenting it to anyone), so all the information you want to convey should be on the screen. Second, remember your audience is science literate but not expert-level. Make your message clear, interesting, and minimize the use of jargon. Thirdly, your product should be polished and professional, with good grammar and clean formatting.

**Step 1:** Log into to your ArcGIS Online account. Click the Content tab. All online content you have created or imported (maps, scenes, layers, etc.) will appear here.

**Step 2:** Click the “Create App” button in the top left corner and select StoryMaps or click here to begin creating a new StoryMap. Click the “Publish” button at the top and select “Everyone (public)”. This will allow you to link your story when its finished. Be sure to tick on “Allow duplication”, which will allow your story map to be copied to another user account for archival purposes.



**Step 3:** Click the “Design” tab at the top. Toggle the “Cover” selection to see how they change and choose one that appeals to you. Next, choose a theme. There are several presets, or you can create your own using the Browse Themes tab. Leave the “Credits” option turned on as this is where you will place your bibliography. Close the “Design” tab.

×

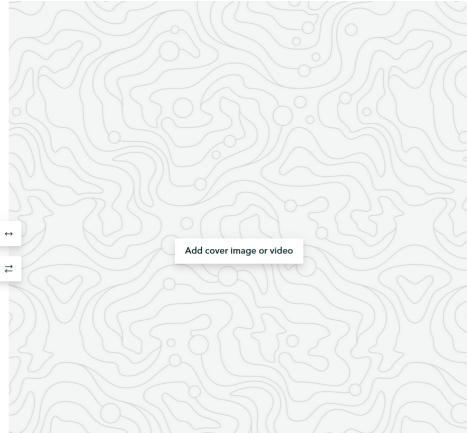
## Design

Change the look of your story

**Cover**

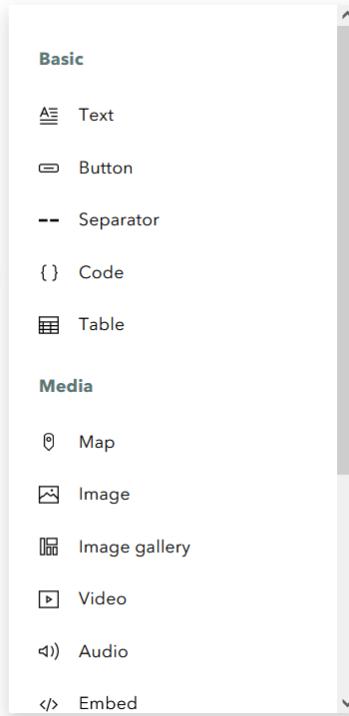


**Step 4:** Give your map a title and subtitle. The subtitle should be a punchy one sentence summary of your project. The word “Draft” underneath will disappear once your StoryMap is published. You can also edit your name if you wish.



**Tell your story...**

**Step 5:** Click the plus sign and select “Text”. Add some text that introduces your project together with a problem statement. You can use some of the text from your final report, but it should be modified to suit a public audience. Do not just copy and paste your final report text into your StoryMap! Repeat this step to add and arrange text for your methods, results and discussion sections.



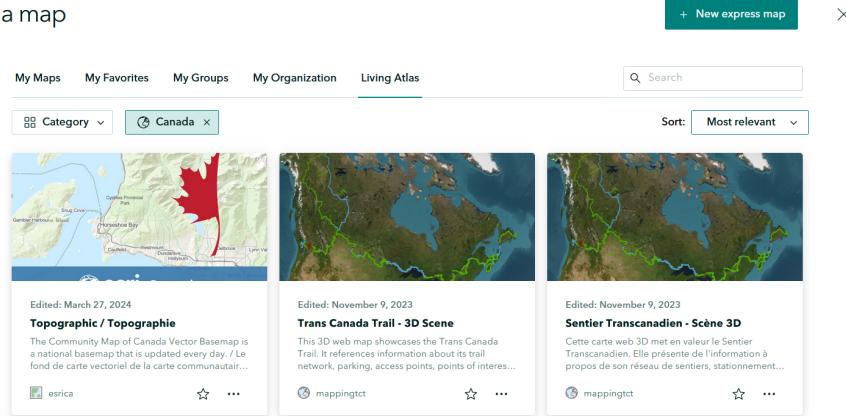
Double-click or highlight the phrase to bring up the text editing bar. This allows you to bold or italicize your text, change the size, add bullets, and a few more. Play around with these options and see what they do.

**Step 6:** You can add images by clicking the plus sign again and selecting “Image”. Click your image to bring up the display options and toggle between small, medium, and large. “Float” allows you to move your image around. The six dots should appear again when “Float” is selected. For any image that you add, you need to clearly indicate the source and properly attribute the author. Be sure to caption all images and properly attribute any material that is not your own.

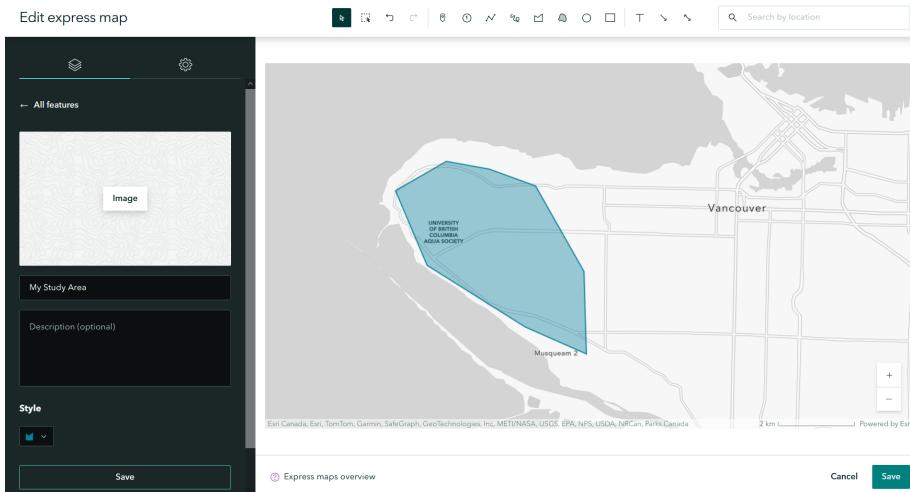
**Step 7:** Next, we will try adding some maps. There are a few different ways to do this, so we have listed some common scenarios below based on what kind of map you plan to use.

**Living Atlas:** the ESRI Living Atlas is a collection of maps with pre-made layers on terrain, climate, population areas, imagery and more. Similar to Express Maps, you can use the Living Atlas to add context to your overall study area. You can use layers from the Living Atlas as a standalone or as part of a custom map or express map.

## Add a map



**Express Map:** these are quick and simple reference maps used to draw boundaries or call attention to general locations. You can use an express map to show your study area and call out simple features by drawing polygons, lines, and points and adding text. You can select different base maps and add a legend if needed, but we recommend keeping express maps very simple. Click the “New Express Map” button at the top right to create an Express Map.



**My Maps:** these are maps published to your ArcGIS Online account from ArcGIS Pro or created in ArcGIS Online. Once they are published on your ArcGIS Online account, they can be used in your StoryMap interactively! These maps are very useful for sharing information that requires a lot of processing and analysis, unlike an Express Map. For your StoryMap, we expect that you will create

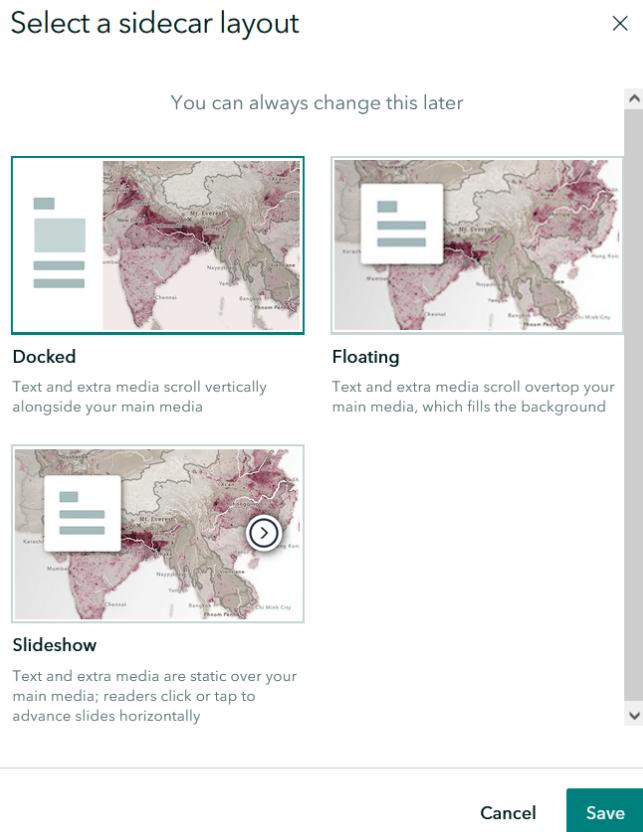
**Step 8:** In the top right corner, search for a city or town near your study area.

On the left side where the editing box is, click the gear symbol for the settings and change the base map to your liking.

**Step 9:** The top bar gives you options to add points, draw lines or polygons, and add text or arrows. Draw a polygon (freehand) around the area and use the “Style” button to adjust the colors. You will notice there are very few options to customize. Add a text box with the name of your study area. Click “Done” at the bottom of the page to add this map to your StoryMap. Explore the options available for displaying the map on your page.

**Step 10:** Create a map in ArcGIS Pro, publish it to your ArcGIS Online account, and then add it to your StoryMap. You can also make a web map from ArcGIS Online, but you will still need to upload your feature or raster data to your account first and ensure it is published publicly. Click the plus sign and select “Map”, but this time, select one of your own.

**Step 11:** Click the plus sign and select “Sidecar” near the bottom. Sidecars are miniature presentations within your story. If you have one topic with multiple pieces of media or are demonstrating a process within your step-by-step, sidecars are a great way to lay this out. You can add slides to your sidecar by clicking the plus button on the lower right, and delete/duplicate/hide individual slides by clicking on the three-dot menu on each slide and selecting your desired option. Click and drag the slides to reorder.



**Step 12:** Select three main concepts/steps you would like to showcase from your analysis. Using the Sidecar function, provide a quick demo or explanation of each concept/step in a sidecar slide. Each slide should have some form of media (image, diagram, map, screenshot, etc) and text explaining what you are showing.

**Step 13:** Continue to add any additional elements or other web maps that you want to use to help feature your project. Be sure to refer to the Deliverables section above to ensure your StoryMap meets the minimum criteria we expect for credit. When you are done, ensure that your “Publish” setting is set to “Everyone (public)”. Submit the URL to your StoryMap to the assignment page on the course management system for credit.

If you are told there are permissions issues with your map layers, navigate to your custom maps under the Content tab and click “Share” then “Everyone”, then try to publish your StoryMap again. If this issue persists, also share your layers with “Share” then “Everyone” in the Contents tab.

## Summary

In this lab, you were introduced to some of the functionality of ArcGIS Online and ArcGIS StoryMaps. In this example, you were shown how to host your own geospatial data on ArcGIS Online and then access those hosted layers via web maps that can be embedded into an ArcGIS StoryMap. We have really only scratched the surface with web mapping in this short lab. If you are interested in learning more, you should consider reading about leaflet and open layers, two popular free and open source software packages that can be used to make web and publish maps. As well, ArcGIS Pro supports the ability to publish Web Scenes, which is a 3D extension of the 2D web mapping capability that we covered in this lab.

Return to the **Deliverables** section to check off everything you need to submit for credit in the course management system.

## Chapter 7

# Writing Metadata and Documentation

Written by Paul Pickell and Francois du Toit

## Lab Overview

The aim of this lab is to practice good geospatial data management habits and produce robust metadata for your final project. Metadata describe your data so it can be used, shared, and understood widely. Creating metadata can be time consuming, however bad metadata negatively affects the integrity, discoverability, preservability, and usability of your data. If you do not create good metadata, it will have significant negative effects on your data and your credibility.

As you work through this lab, you should reference the UBC Library Research Commons Research Data Management website and specifically the pages related to depositing geospatial data to Dataverse.

---

## Learning Objectives

- Create metadata for geospatial data assets
  - Archive project outcomes with the UBC Scholars Portal Dataverse
  - Share and feature your achievements with the MGEM Project Library
-

## Deliverables

The main deliverables for this lab will be uploading your final project data products, as well as an associated README, and metadata to the Scholars Portal Dataverse and also to the course management system. These metadata will be used to create the public page that features your project achievements on the MGEM Project Library.

Answers to the metadata questions on the course management system (50 points):

What is the DOI URL for your project on the MGEM Dataverse? (5 points)

Upload an image that will be used to feature your project on the MGEM Project Library (5 points)

Upload your README file (5 points)

Select 6 keywords separated by commas to describe your project that you also used for the MGEM Dataverse (6 points)

Write the title for your project (5 points)

Write the abstract for your project (5 point)

Who is your faculty mentor or community partner? What is their affiliation? (5 points)

What is the URL to your project StoryMap? (5 points)

Upload a polygon GeoJSON file of your study area (5 points)

What is the maximum latitude (top) of your study area in decimal degrees? (1 point)

What is the minimum latitude (bottom) of your study area in decimal degrees? (1 point)

What is the minimum longitude (left) of your study area in decimal degrees? (1 point)

What is the maximum longitude (right) of your study area in decimal degrees? (1 point)

---

## Data

There are no data for this lab, but the information needed will come from your FCOR 599 final project.

---

## Task 1: Depositing in the Dataverse

In order to deposit data, you will first need to log in to the Dataverse.

**Step 1:** Navigate to the MGEM Dataverse, and log in via the top right corner. Be sure to select UBC, and you will be directed to use your Campus Wide Login (CWL).

Log In

Log in or sign up with your institutional account — [more information about account creation](#). Leaving your institution? Please contact [Borealis Support](#) for assistance.

Your Institution

University of British Columbia

Please select... Continue

Allow me to type the name of my institution

Institution not listed?

Username/Email Sign Up

**Step 2:** Once you are logged in, you can click “+Add Data” then select “New Dataset” from the drop-down menu. Important, make sure that you are on the MGEM Dataverse landing page and not the general UBC Dataverse! See screenshot below.

borealis

Master of Geomatics for Environmental Management MGEM Program Website

(University of British Columbia)

Borealis > UBC Dataverse Collection > UBC Faculty of Forestry >

Contact Share Edit

Welcome to the **MGEM Dataverse** All MGEM student project research data including final reports, maps, and spatial data are stored here for public access. For questions about specific projects, please contact the student directly who submitted the dataset.

+ Add Data

Ensure that the “Host Dataverse” field is pre-populated with “Master of Geomatics for Environmental Management”. If you see anything else, go back to the URL above to access the MGEM Dataverse and try again.

**Step 3:** You might also notice that you can choose a “Dataset Template”. “CC Attribution 4.0 International (CC BY 4.0)” is the default license that most MGEM datasets are licensed under, but you have the option to select any license you prefer from the drop-down menu. Fill in all of the required “Citation Metadata” fields below. In the “Description” field, paste your abstract. Add your “Keywords”, individually in the respective fields. Other non-required fields can be left as-is for now.

Borealis > UBC Dataverse Collection > UBC Faculty of Forestry > Master of Geomatics for Environmental Management >

<b>Host Dataverse</b>	Changing the host dataverse will clear any fields you may have entered data into.
<input type="text" value="Master of Geomatics for Environmental Management"/>	
<b>Dataset Template</b>	Changing the template will clear any fields you may have entered data into.
<input type="text" value="CC Attribution 4.0 International (CC BY 4.0)"/>	

\*Asterisks indicate required fields

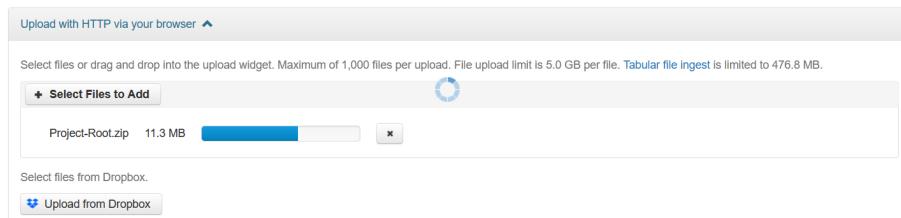
It is important that you discuss with your Topical Mentor the appropriate license for your project data as some research data may have a license that takes precedence over any derivatives that you have created for your project. You can learn more about the various Creative Commons (CC) licenses here: <https://creativecommons.org/share-your-work/cclicenses/>. You may also want to use this license chooser tool: <https://creativecommons.org/>.

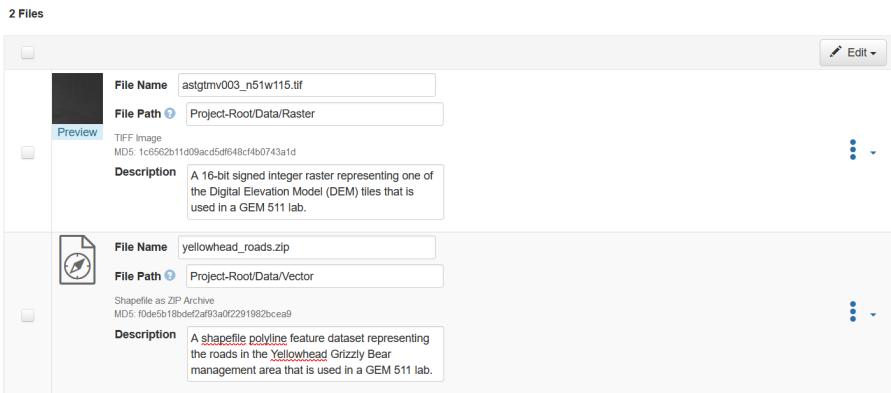
**Step 4:** Once you have filled in the metadata for your project, it is time to prepare your files for upload. Ensure that you are applying best practices:

- Followed file naming conventions
- Converted your spatial data to the appropriate archival formats
- Organized everything in the proper directory structure.

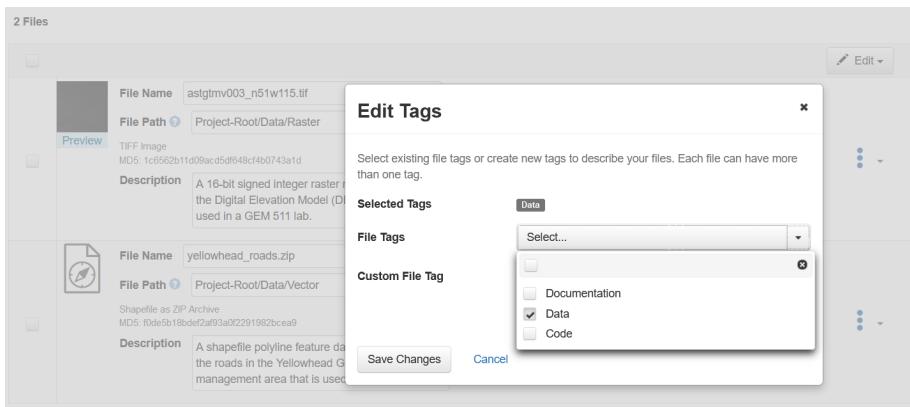
When uploading to Borealis, the first zipped level will be unpacked into whatever files and directories you have in the zip file. Any zip files within this structure (i.e., double-zipped) will be preserved as a zip file, which has some advantages for some file types and sub-directories.

Some geospatial data should be zipped or even double-zipped. Refer to the UBC Library Research Commons Geospatial Data Deposit guide for information specific to your type of research data. You can preserve the directory structure of your project by zipping the directory into a single zip file and uploading this to Borealis. However, you may want to zip or double-zip some sub-directories, like one that contains the contents of a shapefile. Double-zipping files also makes it simpler to describe a file with metadata on Borealis.





**Step 5:** At a minimum, upload your final report, as well as all of your final data outputs (derivatives only, no “raw” data), and your README file. You may also optionally add any other data that might be relevant to your project such as (well-annotated) code/scripts, videos, presentation slides, posters, and other multimedia. Add metadata descriptions for each file in your project and adjust file names as necessary. Click the ellipse to the right to add pre-set tags to each file or create your own custom tags.



**Success!** – This dataset has been created.

**Info** – This draft version needs to be published. When ready for sharing, please [publish](#) it so that others can see these changes.

## This template uses the CC Attribution 4.0 International (CC BY 4.0) license

[Draft](#) [Unpublished](#)

Pickell, Paul, 2024, "This template uses the CC Attribution 4.0 International (CC BY 4.0) license", <https://doi.org/10.5683/SP3/MA6HQ>, Borealis, DRAFT VERSION [?](#)

Cite Dataset [Learn about Data Citation Standards.](#)

Access Dataset ▾  
Publish Dataset ▾  
Edit Dataset ▾  
Contact Owner Share  
Dataset Metrics [?](#)  
0 Downloads [?](#)

Description <a href="#">?</a>	My abstract goes here.
Subject <a href="#">?</a>	Earth and Environmental Sciences
Keyword <a href="#">?</a>	Keyword1, Keyword2, Keyword3, Keyword4, Keyword5, Keyword6
License/Data Use Agreement	CC-BY 4.0

Files Metadata Terms Versions  
Change View Table Tree  
Search this dataset...

**Step 6:** When you are satisfied, click “Save Dataset”. You will be presented with a landing page that is a DRAFT of your final project archive. Scroll down a bit and click on the “Metadata” tab that appears just above your data files. Then click on the “Add + Edit Metadata” button in the top right. This will expose all of the available metadata fields, many of which were not available when you created the draft of your submission.

Add + Edit Metadata

### Citation Metadata ▾

Persistent Identifier <a href="#">?</a>	doi:10.5683/SP3/MA6HQ
Title <a href="#">?</a>	This template uses the CC Attribution 4.0 International (CC BY 4.0) license
Author <a href="#">?</a>	Pickell, Paul (University of British Columbia)
Point of Contact <a href="#">?</a>	Use email button above to contact. Pickell, Paul (University of British Columbia)
Description <a href="#">?</a>	My abstract goes here.
Subject <a href="#">?</a>	Earth and Environmental Sciences
Keyword <a href="#">?</a>	Keyword1 Keyword2 Keyword3 Keyword4 Keyword5 Keyword6
Depositor <a href="#">?</a>	Pickell, Paul
Deposit Date <a href="#">?</a>	2024-03-28

**Step 7:** Scroll to the bottom and expand the “Geospatial Metadata” and then fill in all of these fields. Take note that these values will also be submitted for this lab assignment on the course management system, so write them down or add them to the assignment page at the same time that you record them in

Borealis. Populate as many of the other fields as you can. For example, you should be able to add information about the “Language”, “Data Type”, and “Software” fields. Not all of the fields will apply to your project, which is why they are not all required, but you should take care to fill in as many as possible.

Remember, you are the only expert of your research data and the only person in the world who can properly describe your research data, so make it count!

The screenshot shows a 'Geospatial Metadata' form. It has three main sections: 'Geographic Coverage', 'Geographic Unit', and 'Geographic Bounding Box'. The 'Geographic Coverage' section contains fields for 'Country / Nation' (with a dropdown menu labeled 'Select...'), 'State / Province' (with a text input field), 'City' (with a text input field), and 'Other' (with a text input field). There is also a plus sign button. The 'Geographic Unit' section has a single text input field and a plus sign button. The 'Geographic Bounding Box' section contains four text input fields: 'West Longitude', 'East Longitude', 'North Latitude', and 'South Latitude', each with a plus sign button.

**Step 8:** Once you are satisfied with your metadata fields, click “Save Changes” at the bottom. From the project data landing page, in the top-right, select “Submit for Review”. This action will send your draft dataset to the admins of the MGEM Dataverse, which include the Research Data Librarian and the instructor. We only will review your submission for missing required information or make suggestions to improve your file naming convention, data organization or file types. You should proof-read your submission to ensure there are no typos or other major omissions.

## Summary

Writing metadata and documentation is a labor-intensive process that completes the life cycle of any project. Geospatial data have some unique characteristics that enable those data to be discovered via spatial filtering using bounding boxes or geographic and place name keywords. The important thing to remember is that you are the expert of your research or project data and no one will be more qualified to produce those metadata than yourself. Consider thinking back to all the data sources that you perused for your project. If those data did not have sufficient metadata to discover them, then your project might never have been possible. That is the power of good metadata.

Return to the **Deliverables** section to check off everything you need to submit for credit in the course management system.