Exercise: A simple analysis with QGIS

D G Rossiter
Cornell University, Section of Soil & Crop Sciences
ISRIC–World Soil Information
南京师范大学地理学学院

28-January-2020

Contents

1 Introduction **1**

2 Installing and starting QGIS **2**

3 Specifying a project coördinate reference system (CRS) **3**

4 Downloading and displaying sample data **4**
4.1 Agricultural districts 5
4.2 Other Tompkins County vector layers from Tompkins County GIS 9
4.3 Other Tompkins County vector layers from CUGIR 11
4.4 Soils 11

5 Manipulation **12**
5.1 Reprojecting 12
5.2 Joining an interpretive table: prime farmlands 16
5.3 Enabling the processing toolbox 21
5.4 Simplifying a classified map 22
5.5 Map reclassification: generalized land use/land cover 24
5.6 Processing Tools: Extract objects by attribute 26

6 Analysis **28**
6.1 Clipping a vector coverage 28
6.2 Creating a buffer 31

6.3 Finding the hot spots 31

7 Answers to in-text questions **35**

8 Assignment **37**

9 Challenge **38**

References **39**

A Appendix: more information on the problem setting **40**
A.1 Prime farmland 40

B Appendix: QGIS extensions **40**

C Appendix: Other QGIS operations **41**
C.1 Background images from XYZ tiles 41
C.2 Creating a new layer 44
C.3 A raster layer 45
C.4 Terrain analysis with the SAGA plugin 48
C.5 Graphical model builder 51
C.6 Python programming 54

1 Introduction

This exercise introduces the QGIS[1] open source Geographic Information System (GIS) and some of its more important spatial analysis capabilities. It is designed for a three-hour lab. period; some extra material, interesting but not needed for the analysis, is presented in an appendix.

> **Note:** QGIS is an active project with frequent version releases. Thus tutorial was last checked with QGIS 3.4.14 "Madeira" stable version.

Prerequisites You should have some familiarity with GIS concepts and operations.

Outcomes After completing this exercise you will be able to:

1. Install and start QGIS;
2. Download coverages (layers) from data repositories, add these to a QGIS project, and change how they are displayed;
3. Define a coördinate reference system (CRS);
4. Reproject vector layers to a different CRS;
5. View and navigate in the QGIS processing toolbox;
6. Perform a table join to add attributes to a vector layer from an attribute table;
7. Compute new attributes for a vector layer using the table calculator;
8. Create new layers by intersection of existing layers;
9. Compute polygon areas, and sum them by a categorical attribute;
10. Clip coverages to a restricted geographic area;
11. Carry out a simple spatial modelling exercise;
12. *Optional*: download and display a raster digital elevation model (DEM);
13. *Optional*: Use a QGIS plugin to display background maps from public servers;
14. *Optional*:

The *optional* outcomes are in the Appendix.

Objective We suppose that we have been requested by the City of Ithaca and Tompkins County, New York State (USA) to identify areas where agricultural activities may affect the water quality in the City of Ithaca reservoir, which is in the lower reaches of Six Mile Creek.

[1] http://qgis.org

We will first obtain relevant GIS coverages (§4):

1. Watershed boundaries;
2. Hydrology (streams, ponds, lakes);
3. Soil types, with attributes such as their hydrologic response group and prime farmland classification;
4. Land use/land cover;
5. Agricultural districts;
6. Roads (used only as geographic reference, not in analysis).

We will then do some GIS manipulations to prepare for analysis (§5).

Finally, we will then combine various layers in order to find land currently in agriculture, within a buffer distance of Six Mile Creek or its tributaries (§6).

What is QGIS? QGIS is a project of the Open Source Geospatial Foundation (OSGeo)[2] and is released under the GNU GPL license, meaning that it is free to use, distribute and modify, as long as it is not sold. It runs on Linux, Unix, Mac OS X, Windows and Android and can be extended by plugins written in Python or C++, many of which have been developed and which bring the functionality of other popular open-source GIS (e.g., SAGA, GRASS, Tau-DEM) to QGIS. It can work with vector and raster maps, as well as databases.

Resources Several documents are provided by the QGIS project[3]; these are also accessible from the `Help | Help Contents` menu item:

- QGIS User's Guide
- QGIS Training Manual
- A Gentle Introduction to GIS

The best single reference for QGIS 3.x is the QGIS User's Manual[4].

2 Installing and starting QGIS

The download page[5] explains how to download for various platforms.

Task 1 : If QGIS is not already installed on your computer, download and install it. •

Task 2 : Start QGIS. •

[2] http://www.osgeo.org
[3] http://qgis.org/en/docs/index.html
[4] https://docs.qgis.org/testing/en/docs/user_manual/index.html
[5] http://qgis.org/en/site/forusers/download.html

QGIS is started in the same way as any application program.

Working directory and project file The first order of business is to start a project.

TASK 3 : Set up a project directory for the data files and analysis results, and start a new QGIS project in that directory.

1. Create a new folder in your file system for the project.
2. Open QGIS and click the "create new project" button (or menu `Project | New`). A blank project will open.
3. Save the project into the project directory you just created, using the "save project" button (or menu `Project | Save`). As you save, give it a file name, without extension. This will create a file with extension `.qgz`.

•

3 Specifying a project coördinate reference system (CRS)

Every map in a GIS stores locations as ordered sets of numbers; for 2D maps these are pairs such as (Longitude, Latitude) or (Easting, Northing), referenced to some **coördinate reference system (CRS)** (坐标参考系统).

The CRS for **geographic** coördinates, i.e., (Longitude, Latitude), only requires an ellipsoid (mathematical figure of the Earth). Maps using metric coördinates i.e., **projected** CRS, must also specify the datum (origin of the horizontal coördinate system), the coördinate units, the projection, and parameters of the projection as required. QGIS is able to work directly with maps with any CRS, as long as it is specified for each map.

> **Note:** For details on CRS, see Iliffe and Lott [2] or Bugayevskiy and Snyder [1].

Each project has a default CRS, which is used in the map display. Maps with other CRS can be projected 'on-the-fly' to the project CRS, that is, for display but not changed internally.

> **Note:** "On-the-fly" (在飞行) refers to calculations that are made during map display, only when they are needed. The same CRS transformations are applied as are applied when a coverage is reprojected into a different CRS (see §5.1, below), but the coverage is not changed. This is sufficient to display coverages in different CRS as one map, but does not allow analysis which requires these to be in the same CRS.

In this project we choose to use the **Universal Transmercator (UTM) projection on the NAD83 datum** as the project CRS, because it is used by the majority of maps we will download, and because we want to use meters as the unit of measure in some of our analysis. This gives coördinates in meters from an origin $(500\,000\ E, 0\ N)$ at the equator (N) and the centre (E) of one of 60 6° wide zones, beginning with zone 1 centred on 177° W of Greenwich meridian. Central New York State is in UTM zone 18N, centred on 75° W

> **Note:** Ithaca is at approximately 76° 30' W, 42° 30' N[6]. The central point of Ithaca (State & Tioga streets) is at UTM 18N (376 890 E, 4 697 698 N). This is about 300 km south of 45° N (northern border of New York State), half-way from the equator (UTM 0 N) to the north pole (approximately 10 000 km, per the original definition of the metre). It is about 125 km west of the central meridian for this UTM zone (75° W), which passes just east of Oneonta.

Task 4 : Specify a project CRS.

1. Open the properties (menu `Project | Properties ...`) and give the project a name in the `Project file` text box.
2. Open the project properties; then click on the `CRS` tab on the left.
3. To find the right CRS, use the `filter` box, enter `NAD83 / UTM zone 18`
4. Several CRS are shown, differing slightly in their definitions of the NAD83 ellipsoid. We want the original one, with EPSG code 26918; select this and click the "OK" button. See Figure 1.
5. Save the project.

•

Most CRS are listed in the comprehensive EPSG database[7] maintained by the International Association of Oil & Gas Producers. This code is a useful shorthand for a long list of CRS parameters.

4 Downloading and displaying sample data

Many GIS coverages of Tompkins County are available via the Tompkins County GIS division[8], which links to the Tompkins County GIS Datasets at the Cornell University Geospatial Information Repository (CUGIR)[9]. The Tompkins County GIS data is also provided as an open data portal[10], see Figure 2.

Task 5 : Use a web browser to open the CUGIR page, and the Tompkins County GIS open data portal. •

[6] This point on the WGS84 datum is in Lansing, north of Burdick Hill Road.
[7] `http://www.epsg-registry.org`
[8] `http://tompkinscountyny.gov/gis`
[9] `https://cugir.library.cornell.edu/`
[10] `https://tcdata-tompkinscounty.opendata.arcgis.com/`

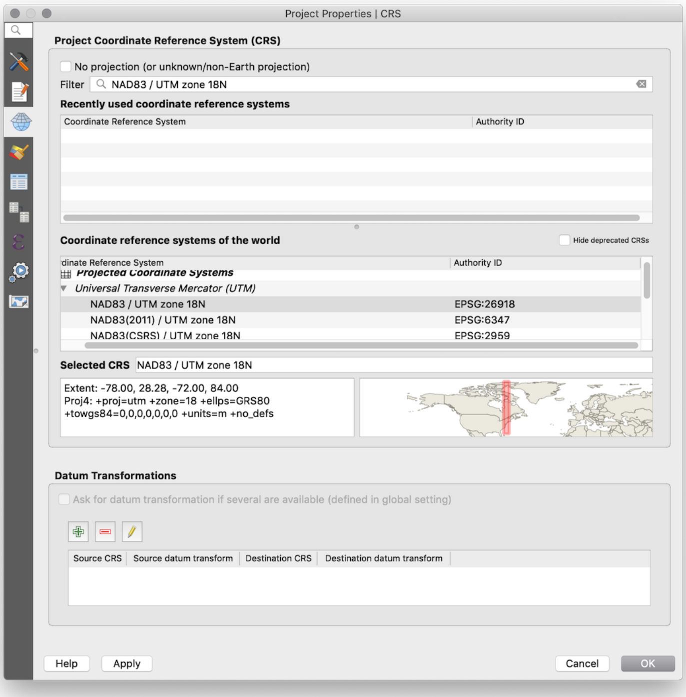


Figure 1: Specifying a project CRS

4.1 Agricultural districts

The first coverage we will download is that showing the agricultural districts. These are land areas that have certain legal protections and restrictions to promote farming[11]. They are based on the viability of current farming enterprises, not directly on the natural resources; landowners are free to join or not.

> **Note:** Not all land in an agricultural district is used for agriculture; some is woodland, some is residential, some is water.

TASK 6 : Find the entries at CUGIR for the agricultural districts coverage (see Figure 3). •

After selecting the entry for the 2018 agricultural districts, your screen should look like Figure 4. Of the formats available for ag. districts, the PDF only allows a view; the KML and KMZ are to display on Google Earth; the ESRI Shape File format (code SHP) is a format that can be used as a GIS layer. Note also that there is **metadata** for this dataset.

TASK 7 : Examine the metadata for the agricultural districts coverage, by

[11] https://agriculture.ny.gov/land-and-water/agricultural-districts

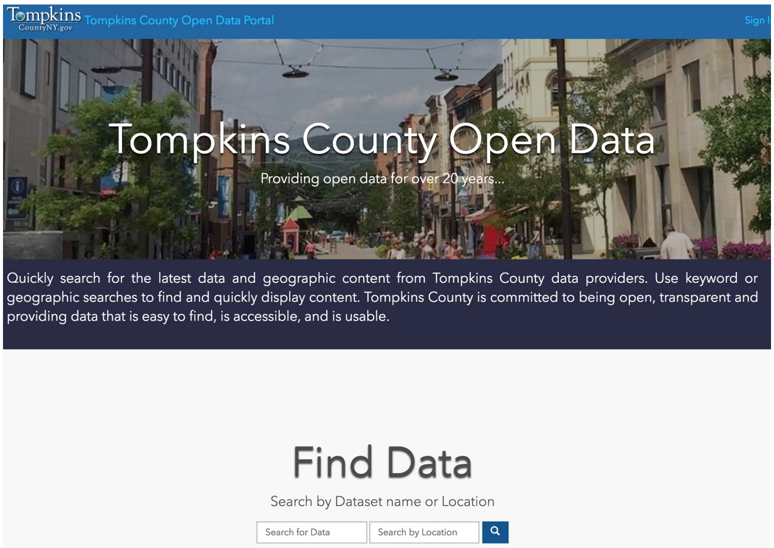


Figure 2: Tompkins County GIS open data portal

Figure 3: Search CUGIR for the Tompkins County agricultural districts

clicking on the ‘Metadata’ link at the bottom of the coverage information. •

See Figure 5.

This format follows the FGDC (US Federal Geographic Data Committee) metadata standard[12].

Q1 : *What is the coördinate reference system (CRS) for this layer? (Hint: look for* `Spatial_Reference_Information`*). Does it match the CRS specified for this QGIS project?* *Jump to A1* •

[12] https://www.fgdc.gov/metadata/geospatial-metadata-standards

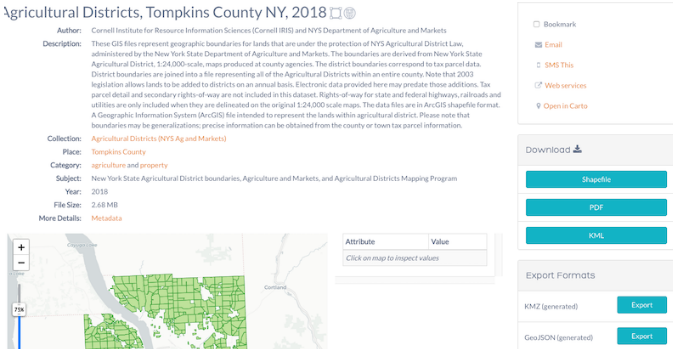


Figure 4: Datasets at CUGIR for the Tompkins County agricultural districts

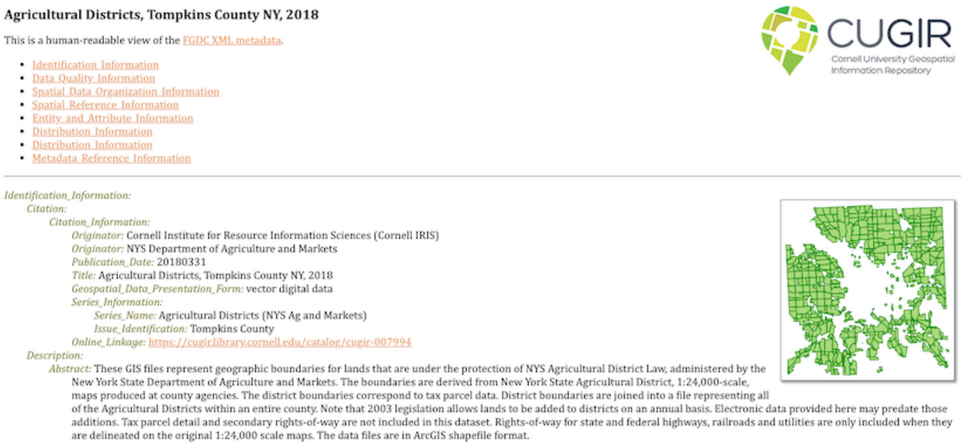


Figure 5: Tompkins County agricultural districts, metadata table of contents and identification information

---

Task 8 : Download the shapefile for this layer and move it to your project directory. •

CUGIR provides the shapefile as a compressed file.

Task 9 : Uncompress the downloaded file; this will create a subdirectory of the same name, containing metadata (file `fgdc.html`) and the coverage as a shapefile (six files with the same file name but different file extensions). •

Task 10 : Add the agricultural districts layer to your QGIS project, using menu `Layer | Add Layer ...| Add Vector Layer ...` or the appropriate tool button; select the shapefile `agTOMP2018.shp`. •

You should see this layer displayed in the viewer and listed in the layer catalog at the lower left of the screen with the name `agTOMP2018`.

> **Note:** The shapefile bundle includes a metadata file with information on

the shapefile's CRS, therefore QGIS did not need to ask about it.

We change the display to show some information about the districts. Shapefiles consist of the **geometry** (here, the polygons defining each parcel) and **attributes**, i.e., a database which gives the values of the data items for each object.

TASK 11 : Right-click on the layer in the catalog and select the `Open attribute table` item. Examine the structure of the attribute table and then close it. •

Q2 : *What are the attributes recorded for each parcel?* *Jump to A2* •

We see that there are two districts in Tompkins County; each polygon is in one of these two.

We now see how to change the display style for a layer.

TASK 12 : Display this layer with different colours for the two districts.

1. Right-click on the layer in the catalog and select `Properties`, then the `Symbology` tab.
2. Change from "Single Symbol" to "Categorized", and select `DistCode` as the classification column.
3. Click on the "Classify" button; you should see the two classes listed, as well as a third category for "unknown".
4. You can accept the default colours or select ones of your choosing.
5. Change the "Legend" column to more meaningful names, e.g. "District 1" and "District 2".
6. Click the "Apply" button to see your changes.
7. Switch to the `Source` tab and change the layer name to "Agricultural Districts 2018".
8. When you are done, click the "OK" button to close the properties.

•

Your screen should look like Figure 6, maybe with different colours.

TASK 13 : Save the changed project; you should do this regularly each time you make a substantial change. •

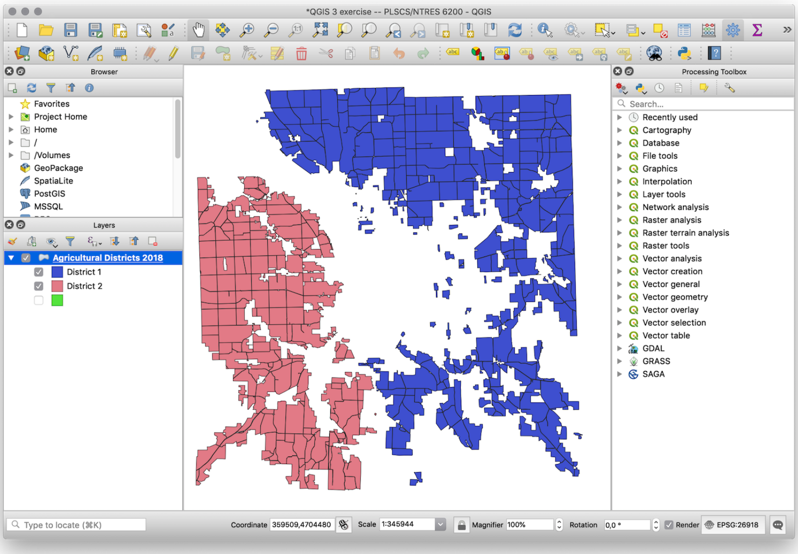


Figure 6: Map display after adding agricultural districts

4.2 Other Tompkins County vector layers from Tompkins County GIS

Task 14 : From the Tompkins County open data GIS page, search for the 'roads' coverages. •

The road centrelines are contained in dataset `TcRoads`; see Figure 7.

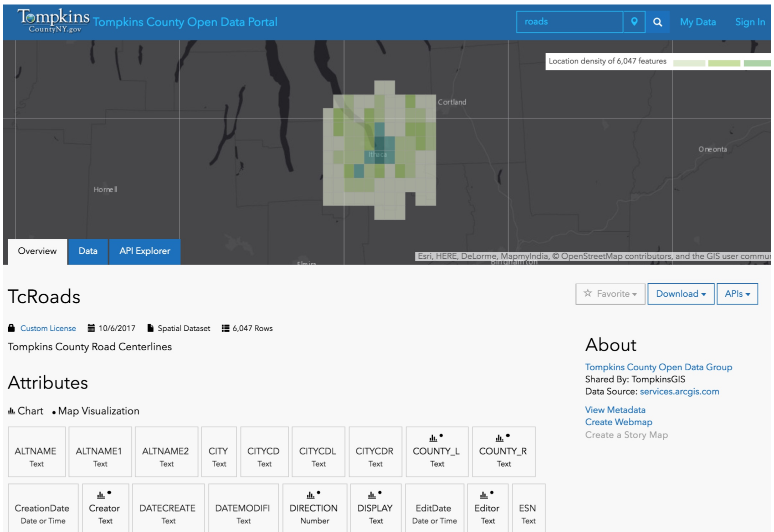


Figure 7: Catalog entry for Tompkins County roads

Task 15 : Download the shapefile for this layer. It will be packaged as a compressed file. Move the compressed file to the project directory and uncompress it. Open the `TcRoads.prj` file (this is the 'projection' file associated with a shapefile) in a **plain-text editor** (e.g., Notepad on Windows) or file viewer and examine its CRS information. •

The CRS should be listed as:

```
GEOGCS["GCS_WGS_1984",DATUM["D_WGS_1984",
SPHEROID["WGS_1984",6378137,298.257223563]],
PRIMEM["Greenwich",0],UNIT["Degree",0.017453292519943295]]
```

"`GEOGCS`" is ESRI's abbreviation for a "geographic coördinate system", that is, one using geographic coördinates (Long/Lat). The definition specifies the datum, spheroid, origin, and unit of measurement.[13] The GEOGCS name `GCS_WGS_1984` is equivalent to EPSG code `4326`; this is what ESRI calls the "well-known ID".

Task 16 : Add this layer to the project and display it. Change the symbology to show each road segment's maintainer (state, county, city ...) (field `Maintainanc`).

Your display should look like Figure 8, maybe with different colours.

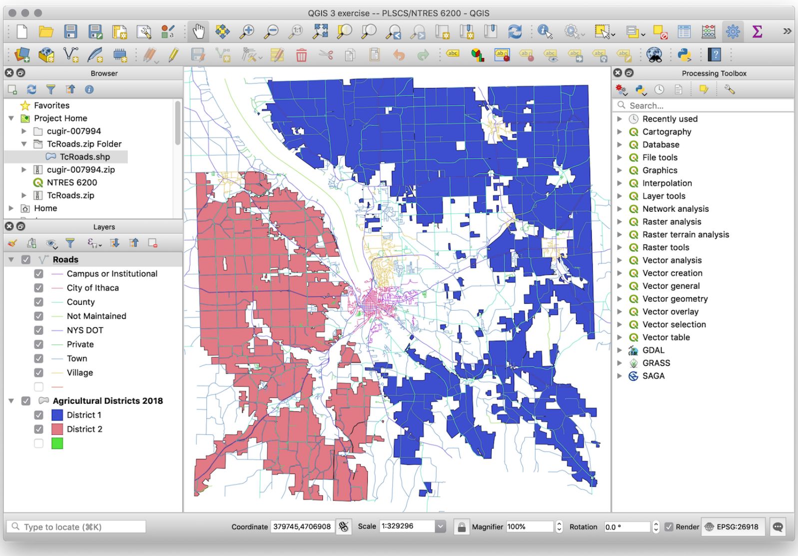


Figure 8: Map display after adding agricultural districts and road centrelines

•

Notice that this layer is shown correctly with the agricultural districts, even though the two layers have different CRS. This is because "on-the-fly" transformation has been enabled, and the CRS in which to display

[13] http://resources.arcgis.com/en/help/arcgis-rest-api/index.html#/Geographic_coordinate_systems/02r300000105000000/

(EPSG:26918) is selected; see the lower-right of the screen.

Q3 : *What is the geometric type of the roads layer? of the agricultural districts layer? Comment on their geometric relation: what should it be (logically), and do you see this in the map? (Hint: zoom in on some roads and parcel boundaries to see how well they match.)* *Jump to A3* •

4.3 Other Tompkins County vector layers from CUGIR

Task 17 : From the CUGIR home page, search for and select these coverages:

1. Tompkins County hydrography 2001;
2. Tompkins County watersheds 1991;
3. Tompkins County land use/land cover 2012.

Download the shapefiles for these layers. They will be packaged as compressed files. Move these compressed file to the project directory and uncompress them. •

Task 18 : Examine the metadata for each of the four coverages, either at the CUGIR page for the coverage or in the downloaded metadata file `fgdc.html`. •

Q4 : *What are the CRS of these coverages?*
These are listed in the Spatial_Reference_Information *section of the metadata for each coverage* *Jump to A4*
•

Task 19 : Add all these to the catalog and display them. Because of "on-the-fly" transformation, they should display correctly with the other coverages. **Do not** change the symbology yet, we will do that after reprojecting these coverages (see below, §5.1). •

4.4 Soils

Information on the soil geography of Tompkins County are provided by the Natural Resources Conservation Service (NRCS) of the USDA.

> **Note:** The national repository for NRCS soils data is the Web Soil Survey[14]; however CUGIR has copied the data for NY counties.

[14] https://websoilsurvey.sc.egov.usda.gov

Also, the California Soil Resource Lab has formatted the soil survey as SoilWeb[15], where you can see the soil map units in Google Earth or Google Maps, superimposed on terrrain and imagery.

Task 20 : Add the Tompkins County soil map polygons to the project.

1. Search for "soils (SSURGO) Tompkins" on the CUGIR page[16].
2. Download the shapefile; this will be a compressed `.zip` file;. Move this file to your project directory.
3. Unpack the `.zip` file; this will create a folder `NY109` (the code of the soil survey area). This folder has files for metadata and folder for the spatial and tabular (attribute) data.
4. Add the spatial shape file: `soilmu_a_ny109.shp` (Map Unit Polygons)[17] to the project. This file is in the `NY109/spatial` folder.

•

Note that the "helper" files associated with this shapefile are automatically read, including the CRS.

Q5 : *What is the CRS of the SSURGO polygons? You can find this in the coverage's general properties, or the metadata file supplied with the coverage. What are the units of measure? Is this the same as the project CRS?* *Jump to A5* •

5 Manipulation

In this section we see a few ways to manipulate the data within the GIS. This is not yet analysis; for that see the following §6.

5.1 Reprojecting

Although all the layers display properly, because we chose to allow on-the-fly CRS conversion for display, they are not all in metric coördinates, which would allow area computations in survey meters. Further, overlay operations require that two layers to be overlain must have the same CRS.

In this project, the agricultural districts layer uses CRS EPSG:26918 (UTM 18N on NAD83), which we chose as the project CRS. However, the SSURGO soils layer just added to the project uses CRS EPSG:4326 (geographic on WGS84).

[15] `https://casoilresource.lawr.ucdavis.edu/soilweb-apps`

[16] `https://cugir.library.cornell.edu`

[17] the `a` in the file name is an abbreviation of `area`; the other geometries are `l` for `line` features and `p` for `point` features, i.e., spot symbols

The 2012 land use/land cover layer, the hydrography, and the watersheds layer CRS use State Plane feet, on the 1983 State Plane coördinate system. The NAD83 datum and WGS84 datums are almost but not exactly identical; this is because the GRS80 ellipsoid used as the basis for the NAD83 datum is slightly different from the WGS84 ellipsoid.[18]

We choose to convert the SSURGO, the 2012 land use/land cover, the hydrography, and the watersheds to CRS EPSG:26918, i.e., the CRS we selected for the project; coördinates will then be in meters and all will have the same CRS.

In addition, we will save the reprojected coverages in an open geospatial file format, the GeoPackage, "an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information"[19], rather than as ESRI shapefiles (a proprietary format).

Task 21 : Convert the SSURGO soils layer to the CRS with EPSG code 26918 (i.e., UTM 18N on NAD83).

1. In the layer list, right-click on the layer to be converted, here SSURGO layer.
2. Select `Export | Save features as ...`; see Figure 9.
3. You should now see a screen like Figure 10.
4. Click on the three dots `...` at the right edge of the "File name" text box; this will open a file browser.
5. Navigate to a location where you want to save the file within your project directory, probably the same folder as the original soils files.
6. Name the file `SSURGO_26918`; close the "File name" text box and return to the previous dialogue box.
7. Choose to save as a GeoPackage.
8. Select the proper CRS (EPSG:26918); this will be in the drop-down list as the project CRS; see
9. Confirm with `OK`; the reprojected map will be added to the catalog and displayed.
10. Remove the original SSURGO layer from the layer list. Note that this does *not* remove the data itself, only the legend entry.
11. Change the symbology from a single symbol to categorized symbols, to show each map unit in a different colour, using field **MUSYM** (map unit symbol) to classify. You might want to use a **colour**

[18] There is an extensive discussion of datums for North America at http://gisgeography.com/geodetic-datums-nad27-nad83-wgs84/

[19] http://www.geopackage.org

ramp, e.g. “Spectral”, which gives soil map units with similar codes similar colours.

•

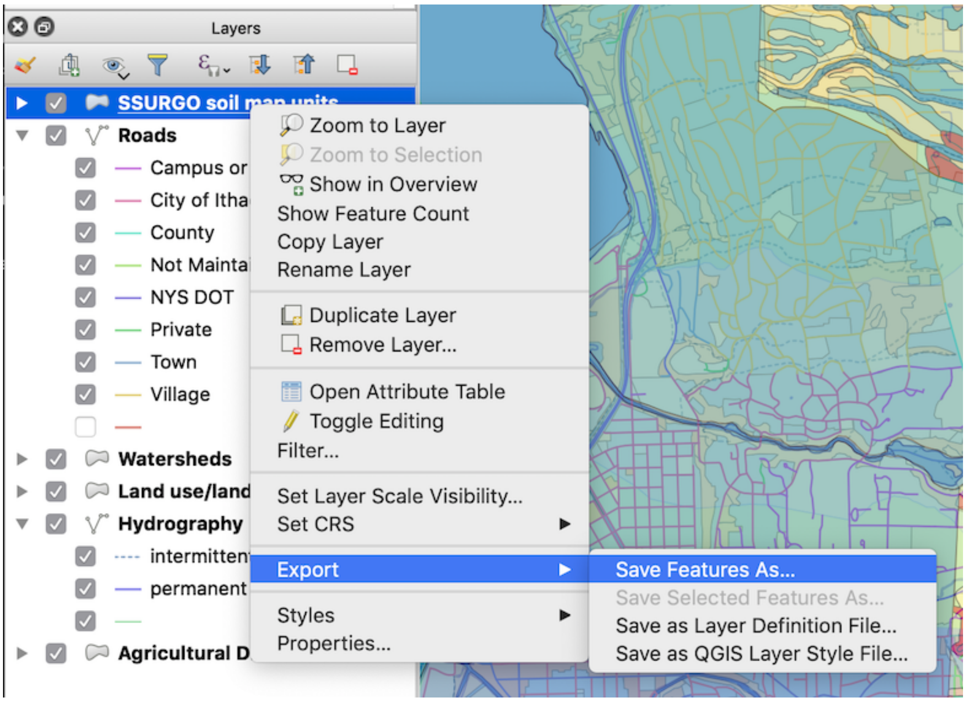


Figure 9: Saving a coverage in another format and/or CRS

For the following three tasks, remember to specify a folder in your project directory as the target.

Task 22 : Reproject the 2012 land use/land cover layer to the project CRS, i.e., EPSG:26918, and save as a GeoPackage. Remove the original 2012 land use/land cover layer from the project. •

Task 23 : Reproject the hydrography layer to the project CRS, and save as a GeoPackage. This allows computations in meters, for example, to compute a buffer around water features. Remove the original hydrography layer from the project. •

Task 24 : Reproject the watershed layer to the project CRS, and save as a GeoPackage. Remove the original watershed layer from the project. •

Task 25 : Right-click on the newly-added layers to display a context-sensitive menu, select the `Rename` menu item, and rename them appropriately. •

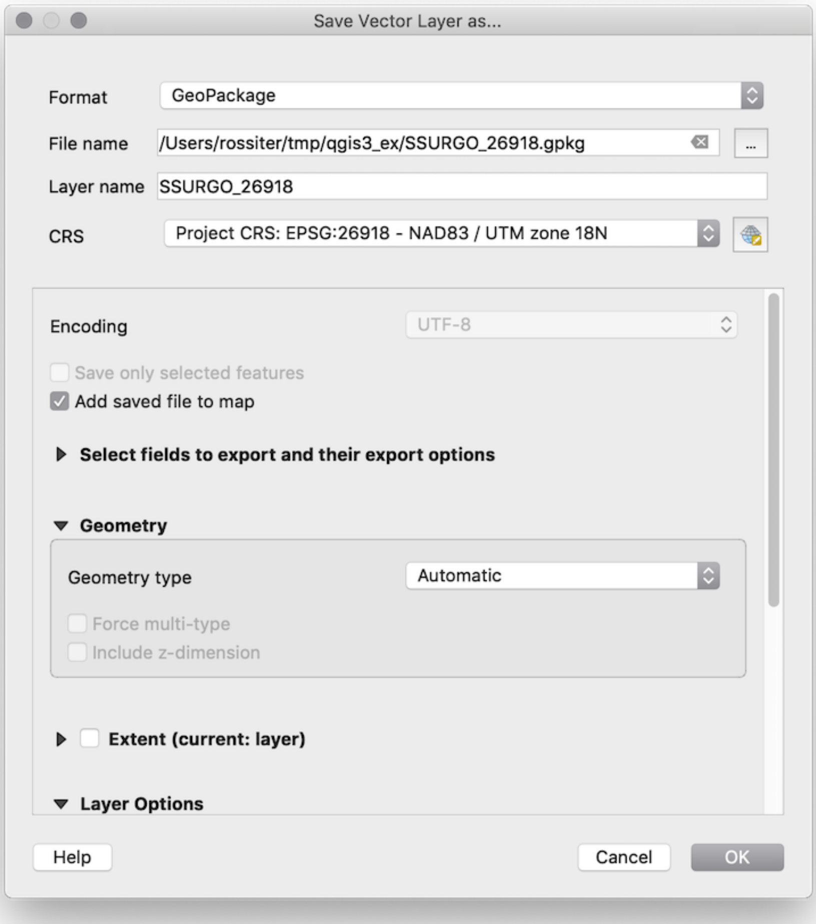


Figure 10: Specifying a GeoPackage and CRS

Task 26 : Change the symbology for Land Use/Land Cover to show the different 2012 land use classes as a "categorized" legend, based on field `LU`.

> Hint: a colour ramp like "RdYlBu' works well here, because labels in alphabetic order correspond to similar land uses.

Decrease the opacity (i.e., increase the transparency) of the LU/LC coverage to allow layers below this to be shown. This is under the "Layer Rendering" section of "Symbology". See Figure 11. Do the same for watersheds. •

Task 27 : Change the symbology for the watersheds layer from a single symbol to categorized symbols, to show each watershed in a different colour (field `WATERSHED`). Since there is no natural ordering, a random colour palette is appropriate for this layer. Decrease the opacity. •

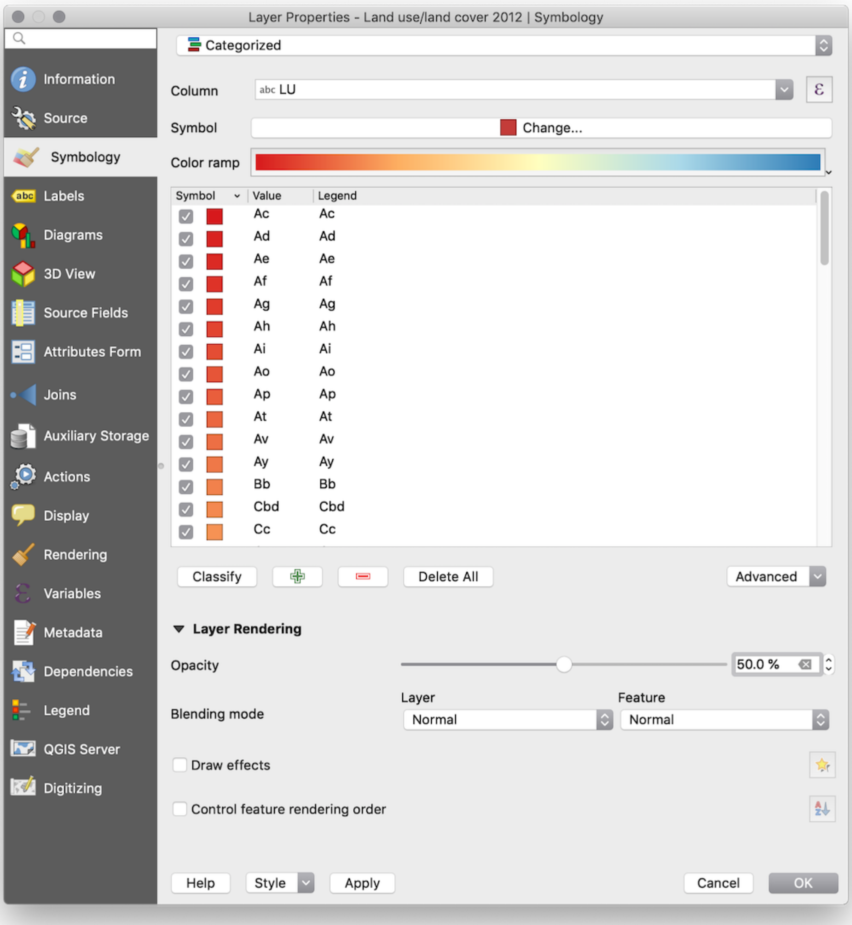


Figure 11: Changing the colour ramp and opacity

The Tompkins County roads layer does not need to be re-projected, because we will not use it in the analysis, only for visualization.

Task 28 : Adjust the transparency of each layer, and the layer ordering, to your taste. •

Your display should look something like Figure 12.

You can spend a lot of time improving the map display! But we want to move on to other data sources and spatial analysis.

Remember to **Save** the project after each major change.

5.2 Joining an interpretive table: prime farmlands

The soil map units by themselves are not so useful; we'd like a map of interpretations, e.g., the suitability of soils for various uses. SSURGO is a

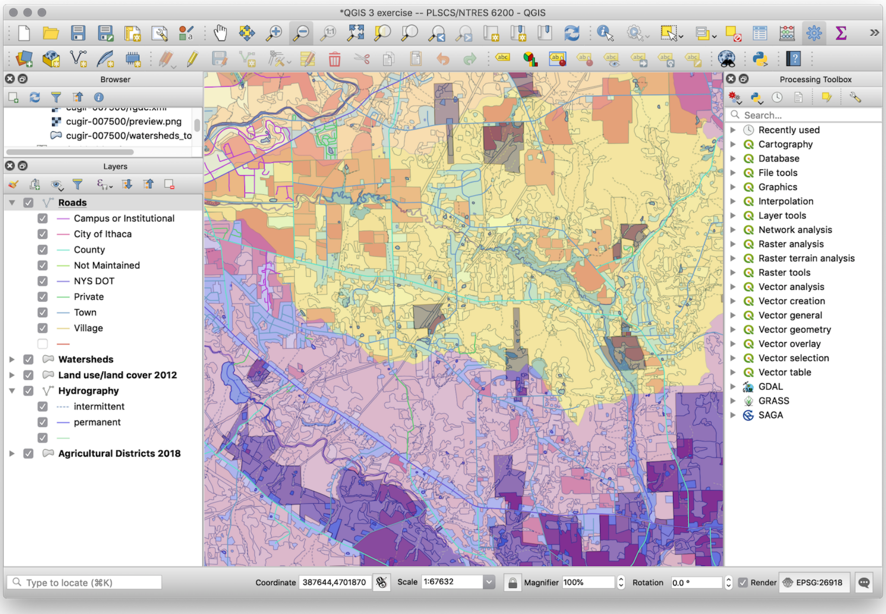


Figure 12: Initial display of five layers

relational database, meaning that there are many tables related by keys, where each table contains one kind of information.

Unfortunately, the metadata provided in the unpacked folder `NY109` does not list the tables; it refers to the National Soil Information System (NASIS) data base structure. Some digging finds the web page[20] which links to the data structure diagrams. Here we won't look at these in detail; we just use one of the tables to show how to add tables to QGIS and do table joins.

Task 29 : Locate the text file `mapunit.txt` in your file system; this is in the `NY109/tabular` folder. Open it with a plain-text editor (e.g., Notepad) and examine its structure. •

Q6 : *What character is used to delimit the fields?* *Jump to A6* •

Q7 : *Does the file include names for the fields? I.e., does the file include a header line with field names?* *Jump to A7* •

Q8 : *What is the meaning of the first field? How can that be related to the SSURGO polygons just added to the project?* *Jump to A8* •

[20] http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/tools/?cid=nrcs142p2_053631

Task 30 : Add this text file to the project:

1. Use the `Layer | Add Layer | Add Delimited Text Layer` menu or the corresponding toolbar button (looks like a big comma).
2. Find the file in your file system as "File name".
3. This table does not describe geographic entities; rather it describes feature-space entities, i.e., map units, as opposed to delineations of these map units. So, expand the "Geometry Definition" tab and specify "no geometry (attribute only table)".
4. This table does not include the names of the data fields, so expand the "Records and Field Options" tab and uncheck the `First record has field names` box.
5. Check the option to discard empty fields.
6. Expand the "File Format" tab and specify the custom delimiter "Others" as `|` (a vertical bar).
7. Name the layer `SSURGO map units`; see Figure 13.
8. Click the `Add` button to add the table to the project.

•

Task 31 : Open the attribute table of the `SSURGO map units` table, i.e., the one you just added, by right-clicking on its entry in the layers panel and selecting "Open Attribute Table". Review its structure, then close it.

•

Q9 : *What is the theme of `field_5`? (Infer from its data values).* *Jump to A9* •

We now have some thematic information about each SSURGO polygon; these are related by the map unit symbol. Each polygon belongs to some map unit (a **many-to-one** relation), and each map unit has various attributes. So by **joining** the thematic attributes of the map units to the polygon table, we can associate those attributes with each polygon. This adds two new data fields (attributes of each polygon) to the SSURGO **polygons** table, taken from the map units **tabular** file.

We chose to add information on **prime farmland**.

In the USA, the USDA has defined the concept of "prime farmland". It is a reclassification of soil map units according to their **potential** for agriculture – not their actual land use; for that we refer to the land use/land cover layer (§4.3). This layer shows which soils are preferred for farming – others can be used, and these need not be used, for farming. See Appendix A.1.

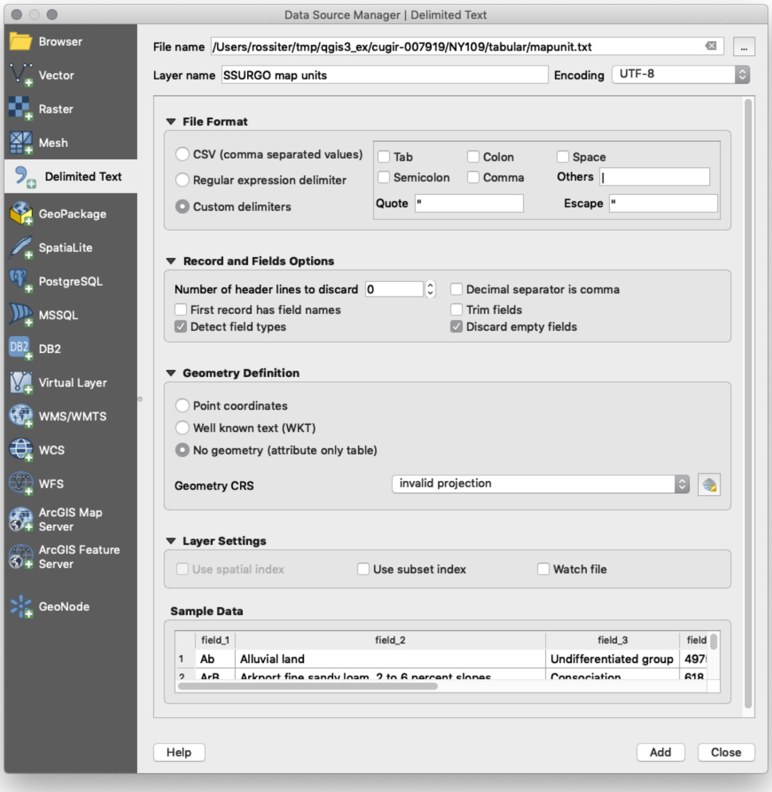


Figure 13: Adding the SSURGO map unit description table to the project

---

Task 32 : Join the information on prime farmland (in the map units table) to the SSURGO attribute table of the spatial entities (polygons):

1. Open the properties of the `SSURGO_26918` **polygon coverage** map units entry in the catalog;
2. Select the `Joins` tab;
3. Click the + "add joins" button;
4. The "join layer" is the SSURGO map units **table**;
5. The "join field" is `field_1`; this is the map unit symbol in the SSURGO map units table;
6. The "target field" is field `MUSYM` in the table associated with the `SSURGO_26918` polygons;
7. Check the "Joined Fields" box; a list of the fields is shown. Select only fields 2 (full name) and 5 (prime farmland classification).
8. Check the "Cache join layer in virtual memory" box; this makes the display much faster;
9. Click the "OK" button; See Figure 14.

10. Select the `Source Fields` tab of the properties; you will see the two added columns (fields), i.e., the two added from the map units table; See Figure 15

•

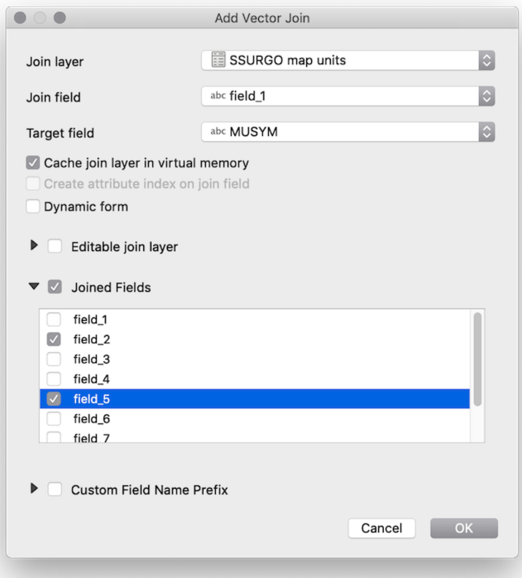


Figure 14: Specifying a table join

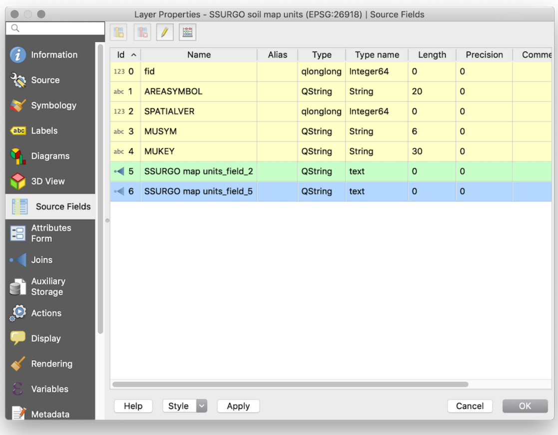


Figure 15: Results of a table join

Task 33 : Make a new layer showing prime farmland:

1. Duplicate the `SSURGO_26918` layer: right-click layer in catalog, select "Duplicate Layer";
2. Rename the duplicated layer `prime farmland`.

3. Change the symbology:

   (a) remove the categories that were copied with the layer (i.e., soil map units, field `MUSYM`);

   (b) categorize by the join field, i.e., prime farmland (`SSURGO map units field 5`), using any colour scheme.

•

My display at this point is shown in Figure 16.

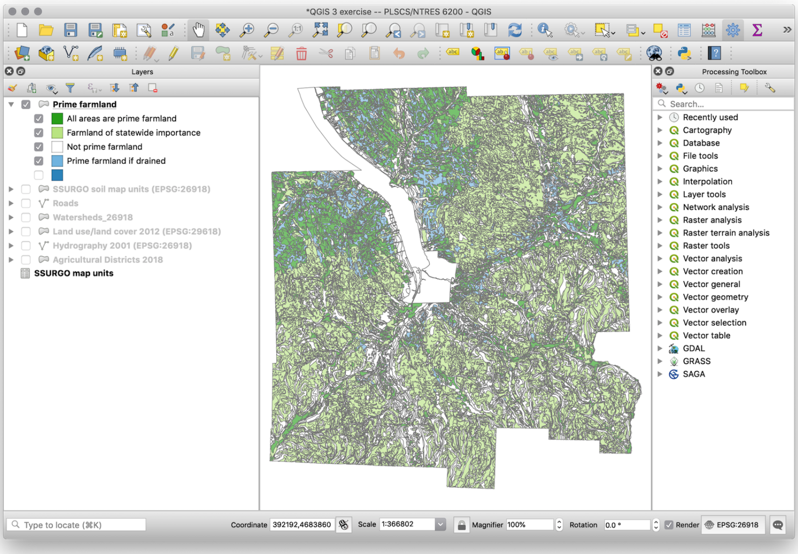


Figure 16: QGIS display showing prime farmland

5.3 Enabling the processing toolbox

QGIS **processing tools** are used to process and analyze one or more layers, and can produce both tabular and map output. The tools are collected in a "Toolbox" panel.

Task 34 : If the "Processing Toolbox" panel is not displayed, display it by selecting the `View | Panels` menu item, and checking the `Processing Toolbox` item. •

You should see various collections of processing algorithms; See Figure 17.

The various QGIS geoalgorithms are always installed; others such as "GDAL", "GRASS", and "SAGA" are usually installed by default; see Appendix B for more information on QGIS Plugins.

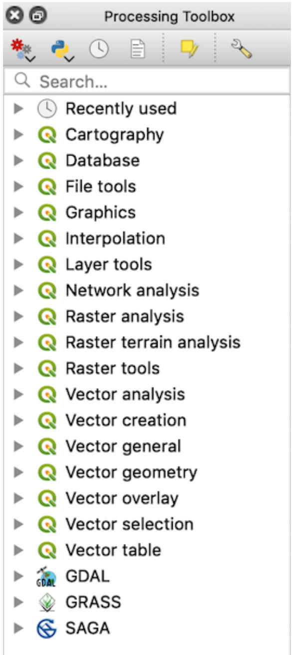


Figure 17: QGIS Processing Toolbox

5.4 Simplifying a classified map

When we categorized the soil map units by their prime farmland type, most of the boundary lines separating different soil map units became irrelevant, because map units on both sides of the boundary are in the the same prime farmland category. So the boundary between such polygons is not meaningful for prime farmland. In this situation it is usual to **dissolve** the coverage, i.e., merge polygons with the same class across boundaries. This is an example of a typical vector GIS geometric operation.

TASK 35 : Create a prime farmlands layer by dissolving the reclassified soil map units according to the prime farmland class.

1. Select “Vector geometry”, “Dissolve”.
2. Input layer is the prime farmlands soils.
3. Click on the `...` at the right of “Dissolve field(s) [optional]”’ this will expand into a list of fields to be used for the dissolve.
4. Select `SSURGO map units_field_5` as the dissolve field; this is the field which will be used to decide if adjacent polygons with the same value in the field should be merged.
5. Accept the default to `[Create temporary layer]` and to open the output file after running the algorithm; see Figure 18.

•

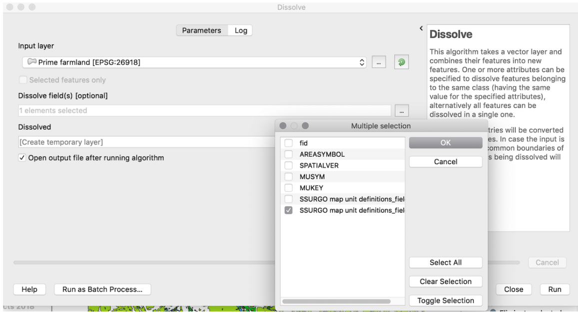


Figure 18: Specifying a dissolved Prime Farmland layer

This processing will take several minutes.

Caution: You may get an error message "Feature (5221) has invalid geometry and has been skipped. Please fix the geometry or change the Processing setting to the 'Ignore invalid input features' option."

If so, open the `Settings | Options ...` menu item, select the `Processing` tab, and then expand the `General` tab, and set the `Invalid features filtering` option to `Ignore features with invalid geometries`; see Figure 19.

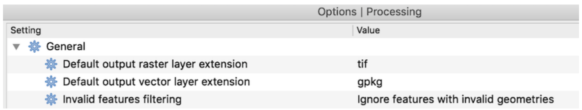


Figure 19: Processing option for invalid geometry

Note: The geometry checker `Vector geometry | Check validity` can be used to find invalid geometries, which can then be fixed with editing tools. However, checking this input layer does not find any errors.

TASK 36 : Change this layer's symbology to show categorized prime farmland.
The classification field is the joined field `SSURGO map units_field_5`.

Select appropriate display colours – dark green for "All areas . . . ", light green for "Statewide . . . ", light blue for ". . . if drained", and no fill for "Not prime farmland".

For this latter, select "Fill Style" as "No brush". •

Figure 20 shows the large polygons that result from dissolving the redundant boundaries of this layer.

TASK 37 : Save this temporary layer by right-clicking on its catalog entry and selecting "Make permanent". Save as a GeoPackage with name `PrimeFarmlandSoils`. •

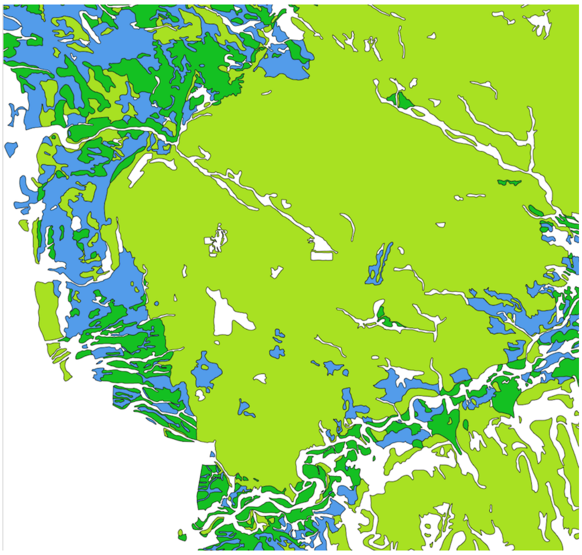


Figure 20: Portion of the dissolved Prime Farmland layer

5.5 Map reclassification: generalized land use/land cover

The 2012 land use/land cover map is categorically detailed; for some types analysis we do not need such detail, just a simpler classification of agricultural vs. non-agricultural land.

Q10 : *How many different land use/land cover types are shown in this layer?* *Jump to A10* •

To answer this, you can count the legend entries in the catalog, or you can use the `Vector analysis | List unique values ...` tool in the processing toolbox; the `LU` field of the attribute table contains the land use codes. See Figure 21.

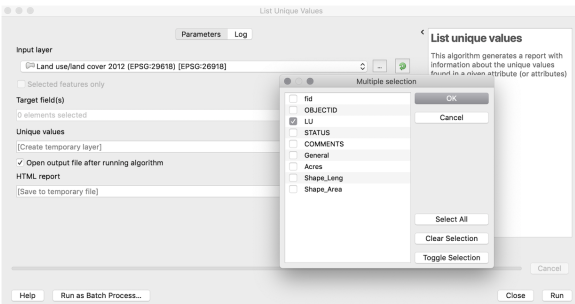


Figure 21: Processing Toolbox: List unique values of Land Use/Land Cover

To determine the meaning of each code, we have to refer to the metadata. Recall that the metadata was downloaded with the dataset.

Task 38 : Open the metadata file `fgdc.html` in the directory where you downloaded the LU/LC layer and examine the *Entity and Attribute Information*. •

You should see metadata as shown in Figure 22.

Entity_and_Attribute_Information:
  Detailed_Description:
    Entity_Type:
      Entity_Type_Label: tomp_lulc_2012
      Entity_Type_Definition: Land use or land cover class, 2012
      Entity_Type_Definition_Source: Classification system is based on LUNR, but has been adapted to meet the needs of the Tompkins County Planning Department
    Attribute:
      Attribute_Label: OBJECTID
      Attribute_Definition: Internal ID
      Attribute_Definition_Source: Tompkins County Planning Department and Tompkins County ITS GIS Division
      Attribute_Domain_Values:
        Unrepresentable_Domain: Positive integers that are automatically generated.
    Attribute:
      Attribute_Label: LU
      Attribute_Definition: Land Use and Land Cover classes devised by the staff in the Tompkins County Planning Department and the Tompkins County ITS GIS Division, based in part on the LUNR classification system from 1968.
      Attribute_Definition_Source: Tompkins County Planning Department and Tompkins County ITS GIS Division, 1999
      Attribute_Domain_Values:
        Enumerated_Domain:
          Enumerated_Domain_Value: Ac
          Enumerated_Domain_Value_Definition: Cropland: Tillable land used for growing cultivated field crops, forage crops, grain, beans, etc. Hedgerows separating defined Ac areas were delineated as separate classes (typically Fd, Fm, Fb or Fc) if they are greater than 20 meters wide
          Enumerated_Domain_Value_Definition_Source: Tompkins County Planning Department and Tompkins County ITS GIS Division, 1999.

Figure 22: Entity and attribute metadata for 2012 Land use/land cover

---

Q11 : *What is the meaning of code* Ac*?* *Jump to A11* •

Q12 : *What is the format of all the codes for agricultural land uses?* *Jump to A12* •

We now use the **field calculator** to add a computed field to an attribute table; this field will have the simplified land use type.

Task 39 : Create a **virtual field** with just the major land use / land cover type. A "virtual" field is one that can be computed on-the-fly as needed, and is not stored in the table.

1. Open the attribute table for the 2012 land use / land cover layer.
2. Open the field calculator by clicking the appropriate toolbar button or pressing <CTRL+I>.
3. Enter the expression:

   ```
   substr("LU",1,1)
   ```

   This extracts the first character from the land use field. The `1,1` means to extract from character at position 1 (the first), to the same, i.e., just one character.

   > **Note:** You can also build this expression by finding the `substr` operator in the function list (under the "String" functions) and the field name from which to extract a substring, from the "Fields and Values" functions.

4. Make the output field type "Text (string)". The output field length is automatic, because it is a virtual field.
5. Name the output field `LU_general`.
6. Make sure the "Create virtual field" box is checked; See Figure 23.
7. Click `OK` to create the new field.

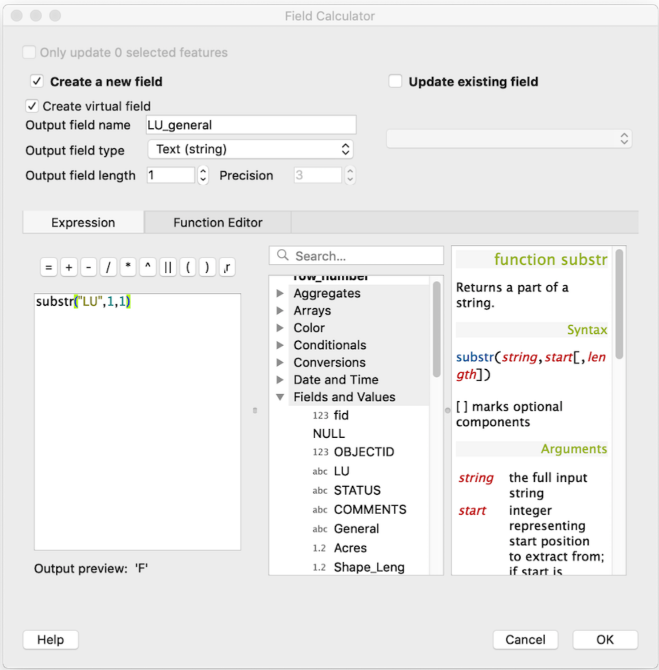


Figure 23: Field calculator to generalize land use/land cover

You will now see it displayed as the last field in the attribute table. The new field has single-letter codes: `A` = agriculture, `F` = forest, `W` = water, etc.

5.6 Processing Tools: Extract objects by attribute

Now we have an attribute of the land use/land cover that specifies whether the use is agriculture or not. We can display this, but the main use is to **clip** other coverages, i.e., reduce their area to only agricultural areas. For this purpose we need to make a new coverage with only the agricultural polygons.

TASK 40 : Create a layer with only the polygons with agricultural land use:

1. In the Processing Toolbox select "Vector selection", then "Extract by attribute".

2. Input vector layer: projected land use/land cover.
3. Selection attribute: field LU_general.
4. Operator =.
5. Value A (no quotation marks). Recall, this is the code we assigned to all agricultural land.
6. Extracted (attribute): accept the default of a temporary file [Create temporary layer].
7. Click "Run"; this will create the new coverage and display it.
8. In the catalog, rename the coverage to a meaningful name, e.g., "Agricultural areas".
9. Adjust the style to show the detailed land use class (type of agriculture); a good colour ramp is "Greens"; this can be inverted so that the darkest colour corresponds to class Ac (cropland).
10. Save as permanent file by right-clicking the catalog entry of the temporary file and selecting "Make permanent..."; save as a Geopackage with base name LULC_IsAg.

•

See Figure 24.

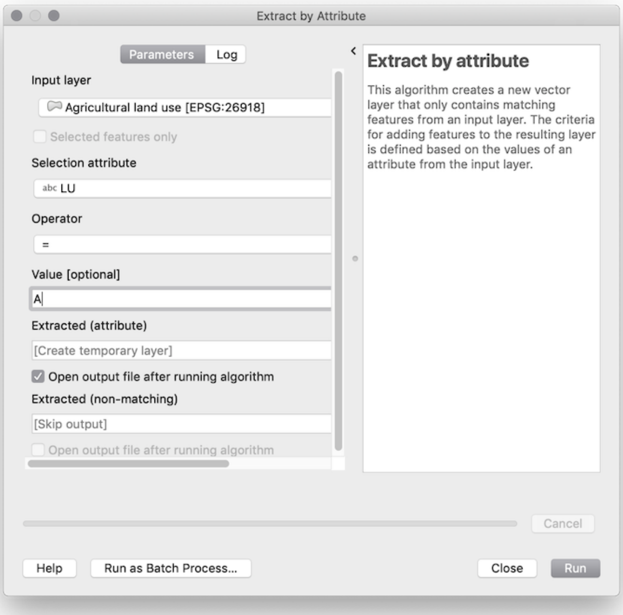


Figure 24: Creating a coverage with agricultural vs. non-agricultural lands

My screen at this point is shown in Figure 25.

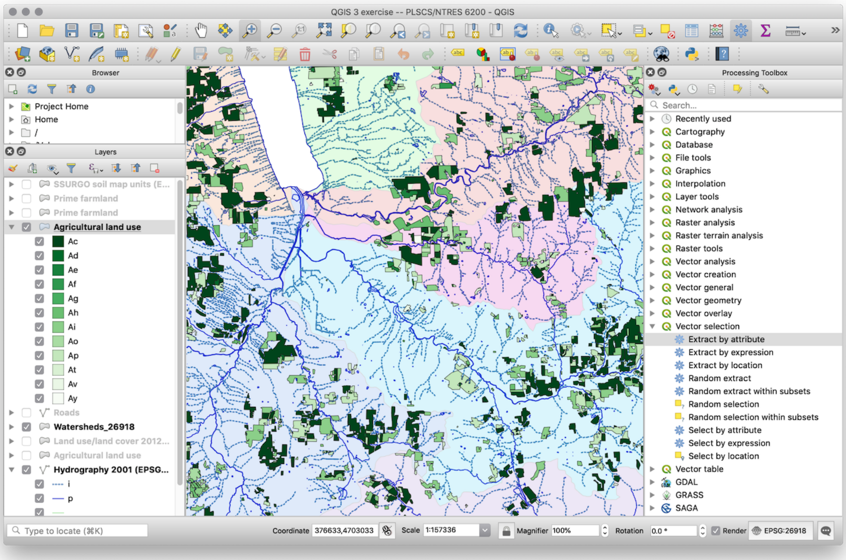


Figure 25: QGIS project after creating a layer of agricultural areas

6 Analysis

Here we show some of the analytic capabilities of QGIS, with a simple analysis of water pollution hazard, which will be expanded in the assignment (§8). Agricultural activities can lead to water pollution by accelerated erosion (sediment) and rapid runoff (pesticides, fertilisers) into streams, and these pollutants are then inputs water treatment plants.

Recall the objective from §1: We have been requested by the City of Ithaca and Tompkins County, New York State (USA) to identify areas where agricultural activities may affect the water quality in the City of Ithaca reservoir (intake to Ithaca water treatment plant at east edge of Third Dam, [379 845E, 469 7153N] in the project CRS), which is in the lower reaches of Six Mile Creek.

In this analysis we find potential pollution hot spots, thereby identifying priority areas for soil conservation and best management practices. We define these hot spots as:

1. land being used for agriculture in 2012; and
2. within 100 m of a stream or pond.

Note: As part of the assignment (§8) you will be asked whether this model is appropriate, over-simplified, missing factors, etc.

6.1 Clipping a vector coverage

We first create a vector layer with just the Six Mile Creek watershed polygon, to use as the clipping layer, i.e., a layer to **clip** other coverages to a limited area:

Task 41 : Create a layer with only the polygon of the Six Mile Creek watershed.

1. In the Processing Toolbox select “Vector selection, then “Extract by attribute”.
2. Input vector layer: projected watersheds.
3. Selection attribute: field `WATERSHED`.
4. Operator =.
5. Value `Six Mile Creek` (no quotation marks). (To see the field and watershed names, you can examine the attribute table.)
6. Extracted (attribute): accept the default of a temporary file `[Create temporary layer]`.
7. Click “Run”; this will create the new coverage and display it.
8. In the catalog, rename the coverage to a meaningful name, e.g., “Six Mile Creek watershed”.
9. Make this layer permanent as a Geopackage, named something like ‘`SixMileCreekWatershed`

•

My results are shown in Figure 26.

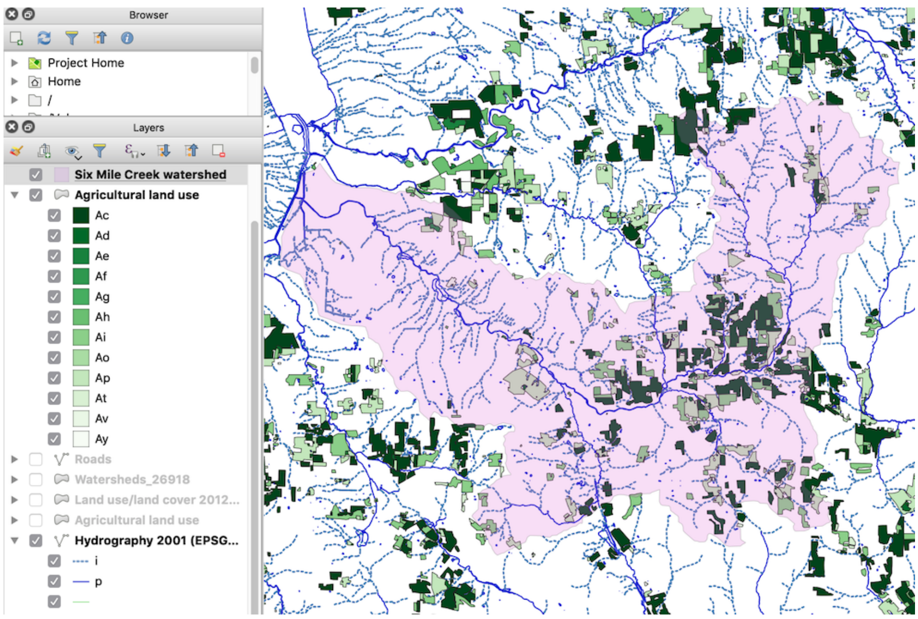


Figure 26: QGIS display showing Six Mile Creek watershed

Now we can use this to **clip** other layers. This reduces the geographic area covered by the layer to that specified by the clipping layer.

Task 42 : Clip the agricultural lands layer to the Six Mile Creek watershed:

1. In the Processing Toolbox select “Vector overlay”, then “Clip”.
2. Input vector layer: agricultural areas.
3. Clip layer: Six Mile Creek watershed.
4. Extracted (attribute): accept the default of a temporary file `[Create temporary layer]`.
5. Click “Run”; this will create the new coverage and display it.
6. In the catalog, rename the coverage to a meaningful name, e.g., “Six Mile Creek agriculture.
7. Make this layer permanent as a Geopackage, named something like `LULC_IsAg_SixMile`.
8. Display and symbolize as for the source layer, i.e., agricultural land use for Tompkins County.

•

Task 43 : Using the same procedure, clip the water features layer to the Six Mile Creek watershed and save the resulting coverage. You may want to change the display to show permanent and intermittent streams with different symbology; see field `CLASS`. •

My results are shown in Figure 27.

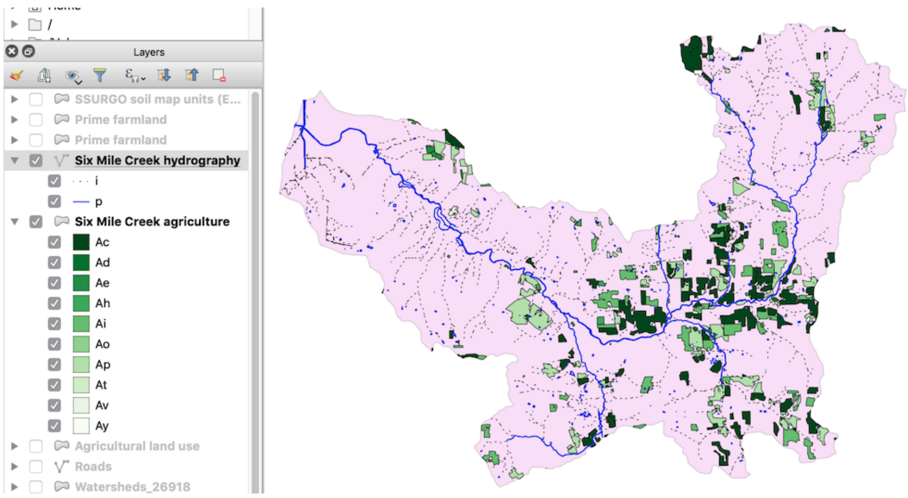


Figure 27: QGIS display showing agricultural areas hydrography, Six Mile Creek watershed

6.2 Creating a buffer

Task 44 : Create a vector layer of the land within 100 m of a water feature, as represented in the (reprojected) Six Mile Creek hydrography layer.

1. In the Processing Toolbox select "Vector geometry", and then "Buffer".
2. Input layer is the reprojected hydrography layer for Six Mile Creek.
3. Distance is 100; this is meters, because the CRS of the reprojected layer uses UTM meters.
4. Accept the default 5 segments to approximate a quarter circle at the upper end of buffers.
5. Check the "dissolve result" box, to merge adjacent polygons (e.g., where streams join).
6. Save to a permanent Geopackage shapefile with an appropriate name.
7. Display the results over the 2012 land use/land cover layer; choose an interesting symbology.

•

Task 45 : Display the agricultural areas along with the hydrography, roads, and water feature buffer for Six Mile Creek. •

See Figure 28.

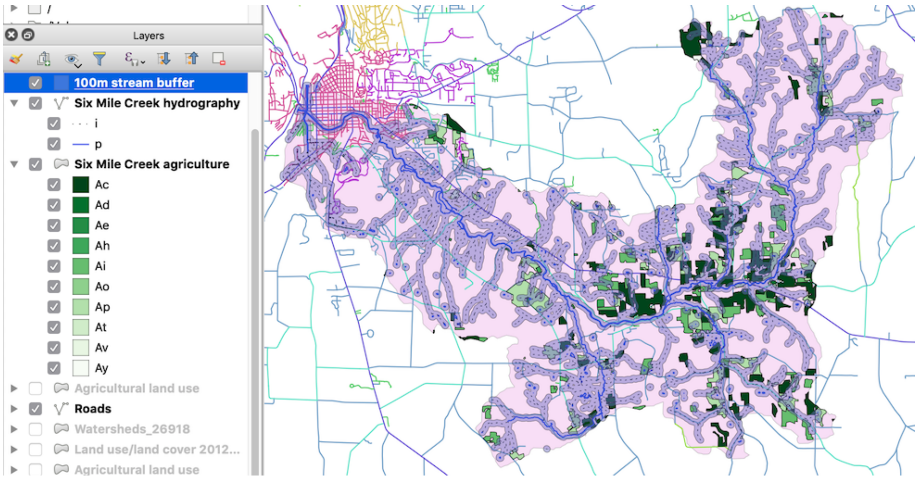


Figure 28: QGIS display showing agricultural areas and a 100 m buffer around water features

6.3 Finding the hot spots

The "hot spots" are defined as those agricultural lands in the buffer zone.

Task 46 : Clip the 2012 agricultural lands in the Six Mile Creek watershed to the 100 m buffer zone; save as a Geopackage; display. •

My results are shown in Figure 29.

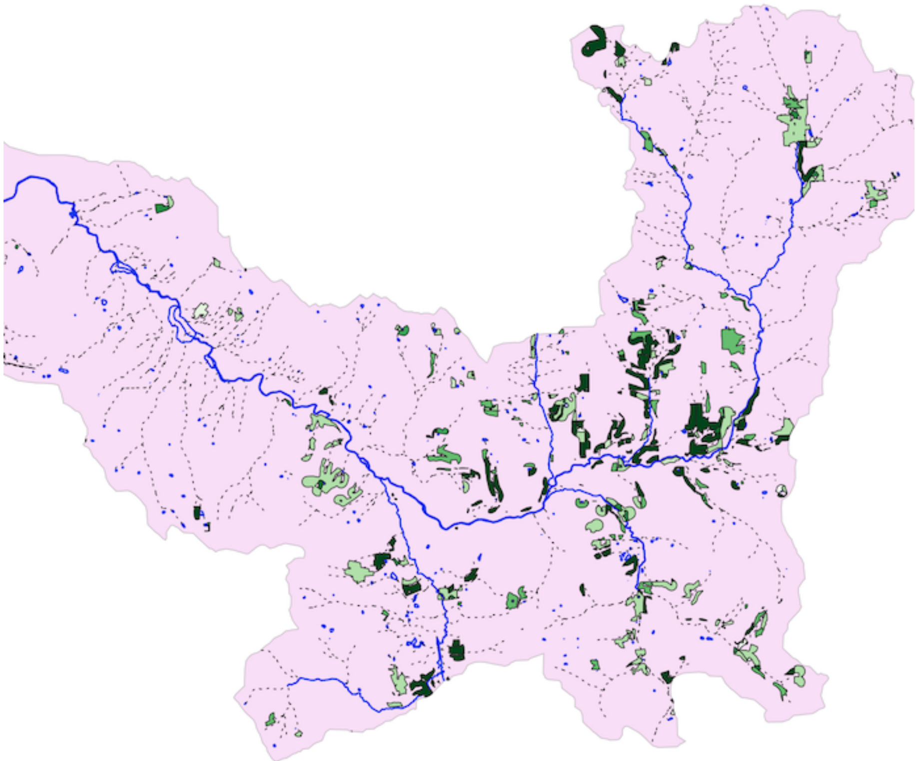


Figure 29: QGIS display showing agricultural pollution hot spots, Six Mile Creek watershed

Now we want to know the area of each hot spot and each land use class.

Task 47 : Open the attribute table for the doubly-clipped layer (i.e., agricultural areas within 100 m of water) and examine the fields. •

The attribute table for the doubly-clipped layer has fields `Acres`[21] and `Shape_Area` in ft^2, but these refer to the polygons *before* clipping, not the current geometry, and in addition they use obsolete units of measure[22].

So, we need to re-compute the area, for two reasons: (1) changed polygon sizes; (2) reprojected, so changed units of measure.

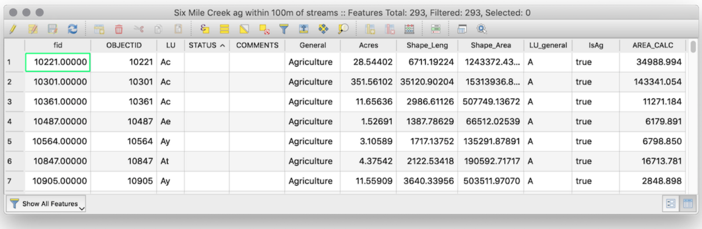
Task 48 : Add a field showing the area of each polygon to the layer table. With this table displayed (see previous task):

[21] defined as an area enclosed by one furlong (220 yards) and one *chain* (22 yards), i.e. 4840yards^2 = 4046.856m^2

[22] The USA signed the Metre Convention in 1878

1. Open the field calculator by clicking the appropriate toolbar button or pressing <CTRL+I>.
2. Check the ‘Create a new field’ and ‘Create virtual field’ check boxes.
3. Expand the ‘geometry’ function list.
4. Double-click on `$area`; this will be displayed in the ‘expression’ box and you will see a preview of the output. This expression returns the area of each polygon.
5. Name the output field `AREA_CALC`.
6. Specify “Decimal number (double)” for the output field type.
7. Click ‘OK’, this will create the field; it will be displayed in the table. See Figure 30.
8. Close the attribute table.

•



Six Mile Creek ag within 100m of streams :: Features Total: 293, Filtered: 293, Selected: 0

	fid	OBJECTID	LU	STATUS	COMMENTS	General	Acres	Shape_Leng	Shape_Area	LU_general	IsAg	AREA_CALC
1	10221.00000	10221	Ac			Agriculture	28.54402	6711.19224	1243372.43...	A	true	34988.994
2	10301.00000	10301	Ac			Agriculture	351.56102	35120.90204	15313936.8...	A	true	143341.054
3	10361.00000	10361	Ac			Agriculture	11.65636	2986.61126	507749.13672	A	true	11271.184
4	10487.00000	10487	Ae			Agriculture	1.52691	1387.78629	66512.02539	A	true	6179.891
5	10564.00000	10564	Ay			Agriculture	3.10589	1717.13752	135291.87891	A	true	6798.850
6	10847.00000	10847	At			Agriculture	4.37542	2122.53418	190592.71717	A	true	16713.781
7	10905.00000	10905	Ay			Agriculture	11.55909	3640.33956	503511.97070	A	true	2848.898

Show All Features

Figure 30: Area in m^2 of hot spots

Task 49 : Compute the area of each hot spot, and then summarize by agricultural land use class:

1. In the Processing Toolbox select “Vector analysis”, then “Statistics by categories”.
2. Input vector layer: same as the previous step.
3. Field to calculate statistics on: `AREA_CALC`
4. Field with categories: `LU`, i.e., the type of agricultural land use.
5. In the ‘Statistics by category’ input box, click the three dots ‘. . . ’, select `Save to file...`, select format `.csv` (comma-separated values); name a file for the results, e.g. `HotSpotStatistics.csv`.
6. Click ‘Run’. A table named `Statistics by category` will be added to the layer list.
7. Change its name to something that indicates its source, e.g., ‘Hot Spot Statistics’.

8. Open this attribute table to see the statistics (minimum, maximum, mean, standard deviation, sum, polygon count) for each category. You can sort by any column by clicking the column heading.

•

My results are shown in Figure 31, sorted by land use class code.

Hot spot statistics :: Features Total: 10, Filtered: 10, Selected: 0

	LU	count	unique	min	max	range	sum	mean	median	stddev	minority	majority	q1	q3	iqr
1	Ac	85	85	19.876	143341.054	143321.178	2443637....	28748.671	17578.7...	30369.54...	19.876	19.876	7365.518	39883.634	32518.116
2	Ad	9	9	15.278	11612.63	11597.352	41893.694	4654.85...	2938.892	3710.645...	15.278	15.278	2281.641	6060.709	3779.068
3	Ae	27	27	32.443	30686.995	30654.552	141438.0...	5238.44...	3099.2...	5907.568...	32.443	32.443	2309.7...	6650.047	4340.2985
4	Ah	12	12	271.301	21566.175	21294.874	64052.271	5337.68...	2439.8...	6668.835...	271.301	271.301	1654.4...	4892.632	3238.1675
5	Ai	48	48	0.002	166821.94	166821.938	1081351....	22528.15...	16760.6...	28731.91...	0.002	0.002	4971.31...	25693.9485	20722.636
6	Ao	2	2	2175.126	3440.182	1265.056	5615.308	2807.654	2807.654	632.528	2175.1...	2175.126	2175.126	3440.182	1265.056
7	Ap	76	76	1.432	352152.966	352151.534	2709341....	35649.2...	15562....	60388.97...	1.432	1.432	5444.0...	32095.128	26651.0575
8	At	15	15	22.989	26071.793	26048.804	115328.6...	7688.5778	6226.121	7482.490...	22.989	22.989	1378.731	12457.9865	11079.2555
9	Av	2	2	4471.924	51898.687	47426.763	56370.611	28185.3...	28185....	23713.38...	4471.9...	4471.9...	4471.924	51898.687	47426.763
10	Ay	17	17	97.93	14996.7	14898.77	87227.757	5131.044...	3564.821	4243.152...	97.93	97.93	2240.8...	6698.552	4457.654

Show All Features

Figure 31: Agricultural pollution hot spots, Six Mile Creek watershed, statistics by land use type

---

Q13 : *What statistics are calculated? What is the unit of measure for the areas?* *Jump to A13* •

Q14 : *What parts of the watershed have most of the hot spots? What is the dominant agriculture class in the hot spots? (See the metadata for the Land Use/Land Cover layer, `Entity_and_Attribute_Information` for the definitions of the land use classes.)* *Jump to A14* •

7 Answers to in-text questions

A1 : *Informally, the CRS is UTM Zone 18N projected meters on the NAD83 datum. Formally, the metadata is as shown in Figure 32.*

It matches the project CRS. Return to Q1 •

```
Spatial_Reference_Information:
    Horizontal_Coordinate_System_Definition:
        Planar:
            Grid_Coordinate_System:
                Grid_Coordinate_System_Name: Universal Transverse Mercator
                Universal_Transverse_Mercator:
                    UTM_Zone_Number: 18
                    Transverse_Mercator:
                        Scale_Factor_at_Central_Meridian: 0.999600
                        Longitude_of_Central_Meridian: -75.000000
                        Latitude_of_Projection_Origin: 0.000000
                        False_Easting: 500000.000000
                        False_Northing: 0.000000
            Planar_Coordinate_Information:
                Planar_Coordinate_Encoding_Method: coordinate pair
                Coordinate_Representation:
                    Abscissa_Resolution: 0.000001
                    Ordinate_Resolution: 0.000001
                Planar_Distance_Units: meters
        Geodetic_Model:
            Horizontal_Datum_Name: North American Datum of 1983
            Ellipsoid_Name: Geodetic Reference System 80
            Semi-major_Axis: 6378137.000000
            Denominator_of_Flattening_Ratio: 298.257222
```

Figure 32: Tompkins County agricultural districts, spatial reference information

---

A2 : *The county (here, always Tompkins), the district number, the creation and certification dates, the towns included, and a code.* Return to Q2 •

A3 : *All three use the NAD83 datum specified for this project, but their unit of measure is State Plane feet and their projection is Transverse Mercator, but* ***not*** *UTM! Notice the different central meridian (-76.583333, as opposed to -75 for UTM 18N), false E and N, projection origin and scale factor.*

> ***Note:*** *This projection's central meridian goes through Lucifer Falls in Enfield Glen (Robert Treman State Park), while the origin latitude is near York PA. The false Easting places the $(0, 0)$ about 245 km southwest of Ithaca, near Connellsville in southwestern PA.*

Note*: the State Plane system is on the original NAD83 datum,* not *the High Accuracy Reference Network (HARN) version.* Return to Q3 •

A4 : *The roads are polylines, the ag. districts polygons. Ag. district polygons should end at road boundaries, since roads are public lands. Thus the road polylines should always be between two district polygons, and exactly in the middle of the space between them. This is not always the case; see for example Bailor Rd. uphill from Central Chapel Rd. (UTM 38933E, 4691965N).* Return to Q4 •

A5 : *The SSURGO map units CRS is EPSG code 4326, i.e., geographic coörd-inates (Long/Lat) on the WGS84 ellipsoid. This is not the project CRS, EPSG code 26918, which is UTM zone 18N meters on the NAD83 ellipsoid.* *Return to Q5* •

A6 : *The | (vertical bar) character is used to delimit the fields.* *Return to Q6* •

A7 : *The file does not include names for the fields.* *Return to Q7* •

A8 : *The first field is the map unit symbol. This is also given for each SSURGO polygons.* *Return to Q8* •

A9 : `field_5` *classifies each map unit according to its "prime farmland" status. Some of the data values are* `Not prime farmland`, `All areas are prime farmland`, `Farmland of statewide importance`. *Return to Q9* •

A10 : *There are 70 unique values (classes) including the* **NULL** *class.* *Return to Q10* •

A11 : *"Cropland: Tillable land used for growing cultivated field crops, forage crops, grain, beans, etc."* *Return to Q11* •

A12 : *They all have a first upper-case letter* `A` *"agriculture", followed by a lower-case letter indicating the type.* *Return to Q12* •

A13 : *Statistics are the number of polygons of the category, the minimum, maximum, range, sum, mean, median, standard deviation, minority, majority, the first and third quartiles, and the inter-quartile range.*

Since the linear units of this layer are UTM meters, the area units are m^2. *Return to Q13* •

A14 : *Most of the hot spots are in the upper watershed along tributaries to the main stream. Class* `Ap` *(pasture) is the most common agricultural land use in the hot spots (76 polygons, total 270.9 ha); class* `Ac` *(cropland) is also substantial (85 polygons, 244.4 ha).* *Return to Q14* •

A15 : *The "layer cake" effect on steep hillsides is an artefact of the DEM and does not represent reality.*

This is also apparent in the histogram: notice the "sawtooth" pattern of the frequencies.

A clue to the cause is given by the metadata (see link in the CUGIR catalog): the map was created by some interpolation from contour lines, which were manually drawn using a stereo-plotter. *Return to Q15* •

8 Assignment

You have been asked to prepare a map showing where current agriculture might lead to water pollution by accelerated erosion (sediment) and rapid runoff (pesticides, fertilisers). As stated in §1, we suppose that we have been requested by the City of Ithaca and Tompkins County, New York State (USA) to identify areas where agricultural activities may affect the water quality in the City of Ithaca reservoir, which is in the lower reaches of the Six Mile Creek. This water is input to the Ithaca City water treatment plant. The intake is at the Third Dam, [379 845E, 469 7153N] in the project CRS.

As in §6, we define these hot spots as land being used for agriculture (i.e., from the actual land use in 2012, not from the agricultural district definition nor from the prime farmland definition), within 100 m of a stream or pond.

1. Prepare a **map display** showing the Six Mile Creek watershed, with the "hot spots" of potential water pollution from agriculture as practiced in 2012. The various kinds of agriculture (as of 2012) should be shown with different colours or symbols.

   Also show the agricultural districts (hint: use transparency in the display style), so we can see which hot spots are in the district and which not. Note these are not included in the analysis, only the display.

   Display the roads and hydrography to help orient the viewer.

   Although there is a sophisticated map composer in QGIS, we haven't learned to use it; so just export the map display using the `Project | Import/Export | Export map to image ...` menu item; save in PDF, PNG or JPG format with 300 dpi resolution.

2. Prepare a **table** showing the area, in hectares, of the hot spots, categorized by type of agricultural land use, and as a total.

3. Optional: from this, compute the proportion of the total from each land use and show it in the same table.

4. Submit the saved map image and table as part of the document you write in answer to the following question.

5. Suppose this map is used for planning interventions, e.g., restricting agricultural uses in the mapped "hot spots" in order to ensure water quality in the City of Ithaca Six Mile Creek reservoir. Answer these questions:

   (a) Which of the input layers to **this** model of hot spots is most likely to not completely correspond to reality?

   (b) Which, if any, of the modelling **assumptions**, for **this** model using these layers, is most likely to be incorrect or over-simplified? How so?

(c) Which of the layers used in **this** model contain insufficient information to improve the model according to the previous answer? How would the extra information improve this model?

(d) Which, if any, **other layers** (available from CUGIR or not) should be included in the model and reflected on the map? How would these improve the model?

9 Challenge

We have three coverages which are related to agriculture:

1. Agricultural districts with certain protections for farming;
2. Prime farmland suitable for agriculture (three suitability classes);
3. Actual agricultural land use.

An obvious question is how these are related. For example:

1. Question: How much of the agricultural districts are actually farmed?
2. Question: How much actual agriculture is *not* protected by agricultural districts?
3. Question: What areas of actual agriculture are *not* on prime farmland?
4. Question: Which prime farmland areas are *not* used for agriculture? For what are they used?
5. Question: Which areas are in agricultural districts, prime farmland, and used for agriculture?

Prepare a set of maps and tables to answer these and similar questions. What do you conclude about agriculture in Tompkins County? Are there implications for land use zoning or planning?

References

[1] Lev M Bugayevskiy and John P. Snyder. *Map projections : a reference manual*. Taylor & Francis, 1995.

[2] Jonathan Iliffe and Roger Lott. *Datums and map projections for remote sensing, GIS, and surveying*. Whittles Pub.; CRC Press, 2008. ISBN 9781420070415.

[3] Soil Survey Division Staff. *Soil Survey Manual, 4th edition*. US Department of Agriculture, 2017. URL http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/planners/?cid=nrcs142p2_054262.

[4] Soil survey staff. *National Soil Survey Handbook*. United States Department of Agriculture, Natural Resources Conservation Service, Feb 2015. URL https://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/soils/ref/?cid=nrcs142p2_054242.

A Appendix: more information on the problem setting

A.1 Prime farmland

The Soil Survey Manual [3] defines prime farmland as a technical soil grouping in support of national (USA) legislation related to the environment and agricultural production. The following classes are defined in §657.5 of the National Soil Survey Handbook [4], which contains the formal definitions, based on Title 7, Volume 6, Part 657 of the Code of Federal Regulations.

Simplified definitions are:

prime farmland : "...land that has the best combination of physical and chemical characteristics for producing food, feed, forage, fiber, and oilseed crops. It must also be available for these uses. It has the soil quality, growing season, and moisture supply needed to produce economically sustained high yields of crops when treated and managed according to acceptable farming methods."

farmland of statewide importance : "...land, in addition to prime ...farmlands, that is of statewide importance for the production of food, feed, fiber, forage, and oil seed crops. Criteria for defining and delineating this land are to be determined by the appropriate State agency ...Generally, additional farmlands of statewide importance include those that are nearly prime farmland and that economically produce high yields of crops when treated and managed according to acceptable farming methods."

prime farmland if drained : The only restriction on being classified as prime farmland is wetness. If this restriction is removed the land would be classified as prime farmland.

not prime farmland : Either has physical characteristics that are not favourable to farming, or can not be used for farming (e.g., urban areas).

State soil scientists examine the definitions of each map unit and classify it accordingly. Thus "prime" farmland is defined nationally; each state may have other lands that it considers "prime" within its context but are not so considered in the national (USA) context.

B Appendix: QGIS extensions

QGIS is easily expanded by **plugins**. These are extensions written by contributors (not only the QGIS project) to add features and functions to QGIS. You can see the installed plugins in the Processing Toolbox. Some are installed by default with a fresh install of QGIS, including GDAL, GRASS and SAGA.

You can add others with the `Plugins | Manage and Install Plugins ...` menu item; see Figure 33.

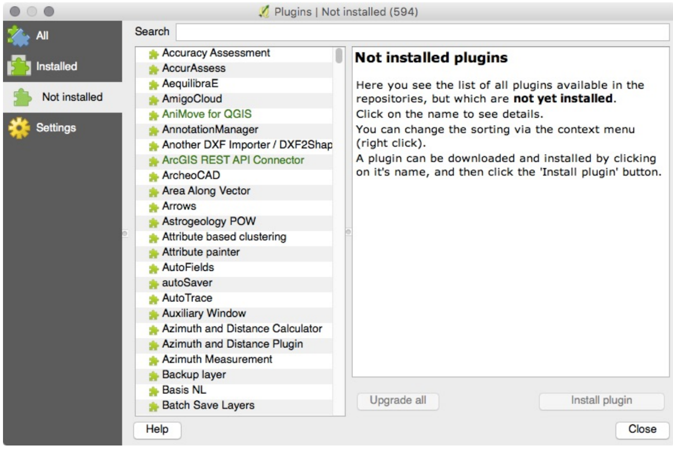


Figure 33: Managing plugins

The plugins listed depend on **providers**. You can see these with the `Settings | Options ...` menu item, `Processing` tab, and then by expanding the `Providers` tab; see Figure 34.

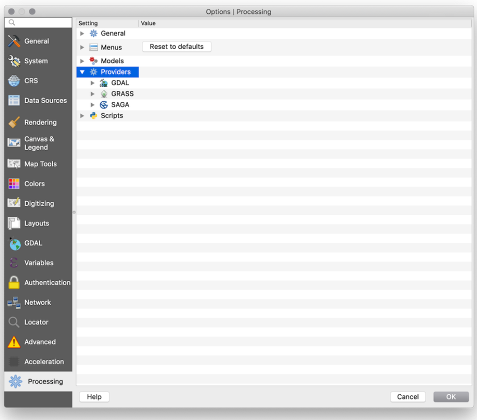


Figure 34: Managing plugin providers

You can also download plugins from the QGIS plugins web page[23].

C Appendix: Other QGIS operations

C.1 Background images from XYZ tiles

There are many map servers on the internet that will, upon request, supply images which can be used as background. Various protocols have been specified for this. One is called WMS (Web Map Service)[24], and is a standard from the Open Geospatial Consortium (OGC)[25]. These maps are

[23] https://plugins.qgis.org
[24] http://www.opengeospatial.org/standards/wms
[25] http://www.opengeospatial.org

not stored locally; rather, the WMS serves them as needed according to the current view. They can not be used in analysis, only for visualization.

> **Note:** The related WFS (Web Feature Service) and WCS (Web Coverage Service) do serve vector and raster layers, respectively, that can be used in analysis.

Another method is the "Tiled web map" service[26], which displays a map by joining individually-requested image files; these are dynamically served as the user zooms or pans the map display. These image files are prepared by by a "Tile Map Service"[27]; a long list of these is at provided by NextGIS[28]; this site also lists WMS, WFS, and GeoJSON services.

QGIS can access these in several ways; the simplest is using the QuickMapServices plugin.

Install this plugin with the `Plugins | Manage and Install Plugins ...` menu item. Search for this plugin by name and then install it. It may then be accessed by the `Web | QuickMapServices` menu item.

After selecting QuickMapServices, you will see some pre-selected services, for example "OSM" for Open Street Maps.

To get a full range of services, select the `Settings | More services` tab, and then click on "get contributed pack"; these are services contributed by users[29]. These include general services such as Google (maps, terrain, satellite ...) and Bing, and specialized services such as NASA fires.

Task 50 : Use the QuickMapServices plugin to add a background image of your choice. Move it to the bottom of the layer list so that themes are displayed over it. Make display themes transparent enough to see the background. •

Figure 35 shows Bing Satellite images as a background at the border of Tompkins and Seneca Counties, with a semi-transparent land use / land cover map overlain.

You can turn the background on or off, as you can for any layer, with the checkbox next to it in the layer list.

[26] https://en.wikipedia.org/wiki/Tiled_web_map
[27] https://en.wikipedia.org/wiki/Tile_Map_Service
[28] https://qms.nextgis.com
[29] Listed at https://github.com/nextgis/quickmapservices_contrib

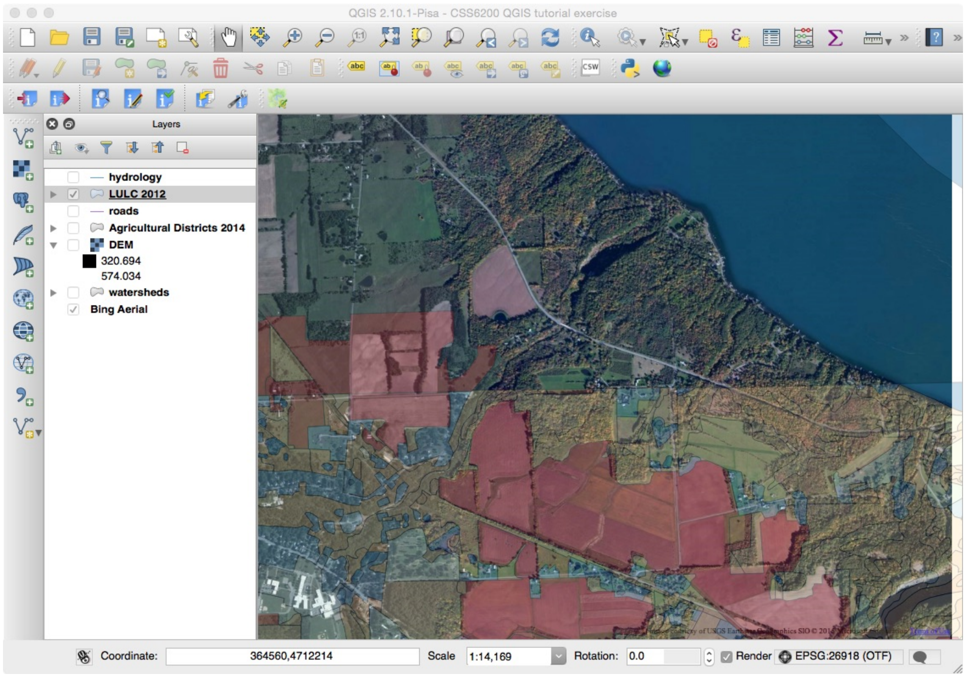


Figure 35: Bing Aerial imagery background, under the 2012 Land Use/Land Cover

C.2 Creating a new layer

You can create and edit QGIS layers with a full set of editing tools. In this section we show how to add a point-geometry shapefile with one point: the location of the Ithaca City water treatment intake at the Third Dam, [379 845E, 469 7153N] in the project CRS.

Task 51 : Display the hydrography layer and zoom in to this location. You will see the outlines of the reservoir, a centre line of the stream, and the dam as a transverse line across the end of the reservoir. See Figure 36. •

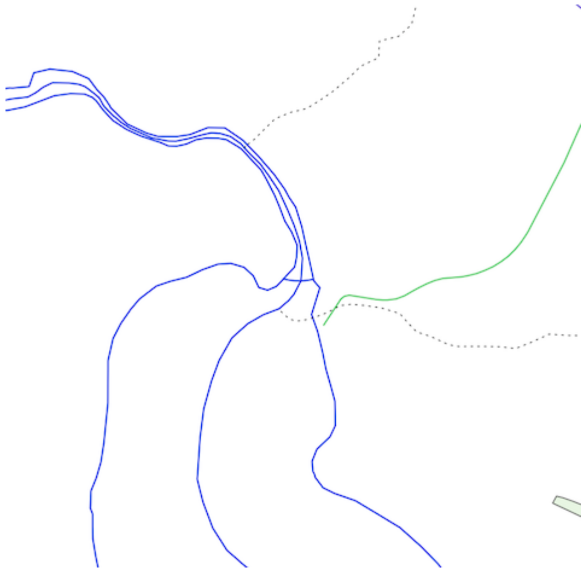


Figure 36: Detail of City of Ithaca reservoir hydrography (access road in green)

Task 52 : Create a point-geometry shapefile for the water intake.

1. Select menu item `Layer | Create Layer | New Shapefile Layer ...`; see Figure 37.
2. Name the file `WaterIntakeShape`, with point geometry, in the project CRS, and with no fields other than the point ID; see Figure 38. Save the file.
3. Right-click on the catalog entry for this layer and select `Toggle editing`. Several buttons in the editing toolbar (under the standard toolbar at the top of the screen) are now active.
4. Select menu item `Edit | Add Point Feature`[30]; see Figure 39.
5. In the map display, move the cursor to the point at the right edge of the dam and click to add the point.

[30] there is also an edit toolbar button as well as a keyboard shortcut for this

6. Enter any ID number, this will never be used.
7. Right-click on the catalog entry and select `Save layer edits`, and then `Toggle editing` to stop editing.
8. Change the display to make the single point more prominent in the map display.

•

My display at this point is shown in Figure 40.

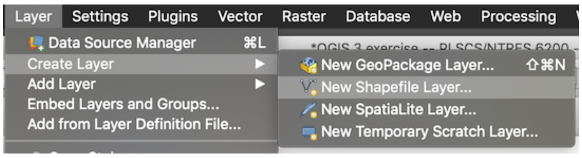


Figure 37: Menu selection to add a new shapefile

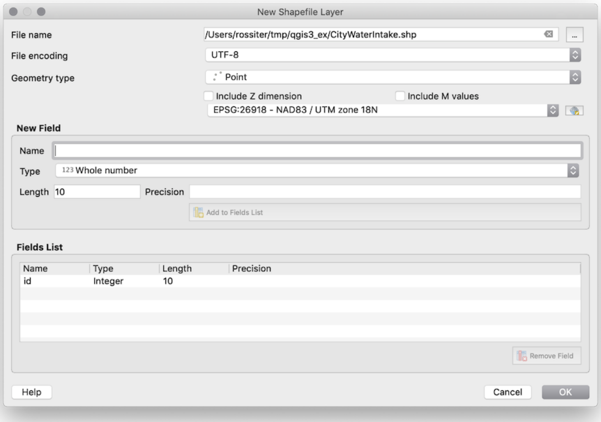


Figure 38: Specifying parameters for the new shapefile

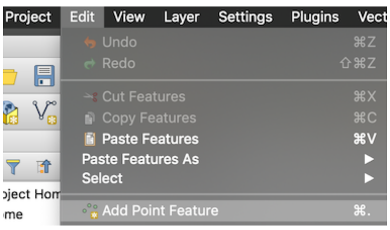


Figure 39: Menu item to add a point feature

This procedure can be used for point, line and polygon features.

C.3 A raster layer

QGIS can also work with raster layer. These are much larger than vector layers. This § shows how to load and display a raster layer.

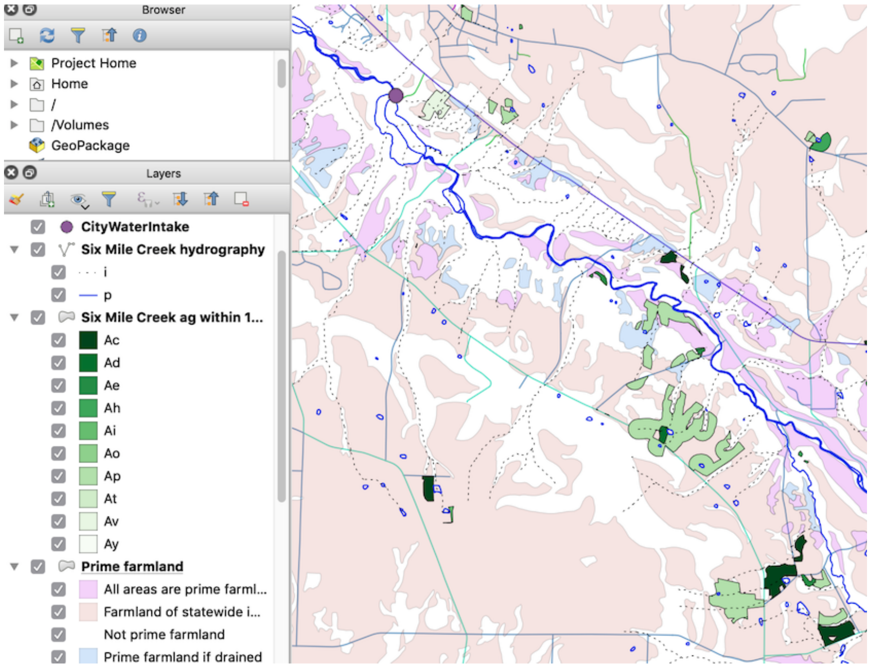


Figure 40: Point feature at the water treatment intake (purple dot in upper-left corner

---

Task 53 : At the CUGIR web page[31], search for "DEM". See Figure 41. Open the index map, zoom in to the Dryden area, and select that tile. See Figure 42. Download it and move the compressed file to your working directory. •

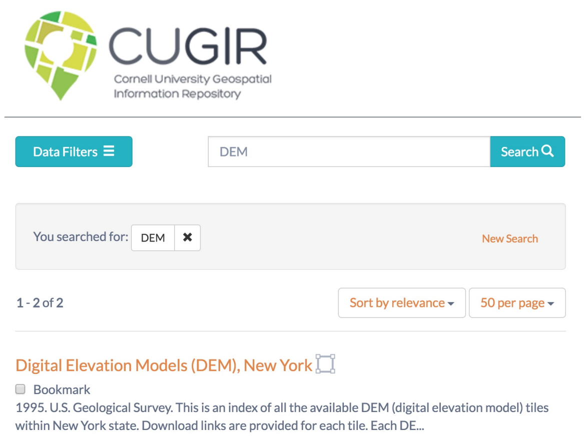


Figure 41: Search CUGIR for raster DEM

Task 54 : Add the DEM to the project:

1. Unpack the coverage from the `.zip` file provided by CUGIR. The file

[31] https://cugir.library.cornell.edu/

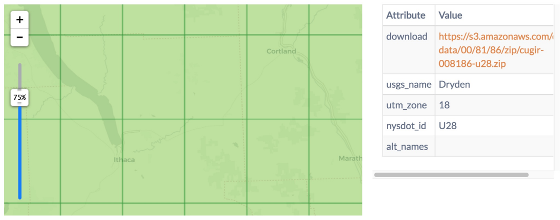


Figure 42: Selecting the Dryden DEM

name should be `u28elu.dem`.

2. Add the unpacked file to the project with the `Layer | Add Layer ...| Add Raster Layer ...` menu item or the appropriate button. Rename it in the layers catalog.

3. Zoom to the layer.

•

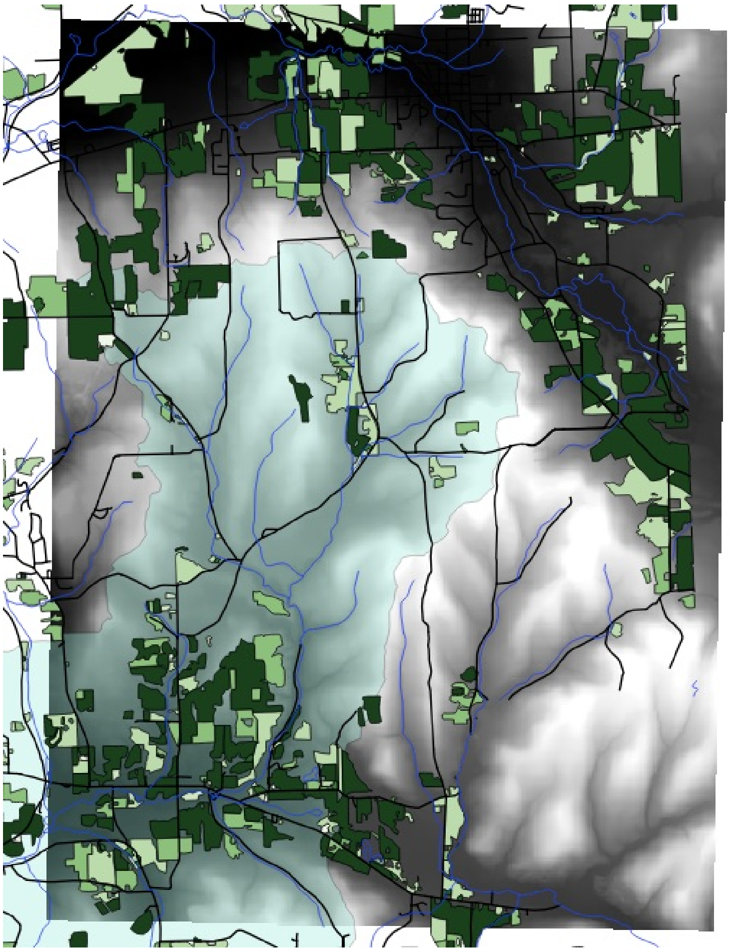


Figure 43: QGIS display after adding a raster layer

Note: The DEM specifies a CRS, which is UTM but on the older NAD27 datum. This displays correctly because of on-the-fly transformation.

My display at this point is shown in Figure 43; this has the 6-Mile Creek watershed and agricultural land uses overlaid.

C.4 Terrain analysis with the SAGA plugin

A plugin that provides many algorithms for terrain analysis of raster DEM is from the SAGA GIS.[32]; SAGA is the abbreviation for "System for Automated Geoscientific Analyses" from a research group first at Göttingen (D) and now at Hamburg (D). So, the full functionality of SAGA geoprocessing is available within QGIS.

Note: The GDAL, GRASS and Tau-DEM plugins also provide terrain analysis algorithms.

Task 55 : Add a raster layer showing the slope gradient within the Dryden DEM tile. •

1. In the Processing Toolbox select "SAGA", then "Terrain analysis – Morphometry", then "Slope, aspect, curvature".
2. Input layer is the Dryden DEM tile.
3. Select a method; unless you have a special reason, accept the default (Zevenbergen & Thorne 1987, a commonly-used algorithm), but notice the many choices.
4. Choose to compute slope in percent (as commonly used in soil survey).
5. Check 'Open output file after running algorithm' only for slope. Uncheck all of the others; notice all the other parameters you could compute.
6. Run; examine the output on the map display.
7. Open the properties of the slope layer and compute its histogram. This is shown in Figure 44.

Note: Slope gradient can also be calculated with the GDAL/OGR plugin, "Raster analysis" group, "Slope" tool, and with the GRASS plugin, "Raster" group, `r.slope.aspect` tool.

Q15 : *Comment on the appearance of the raster layer. Does the histogram give any insight into this? What could be the cause?* *Jump to A15* •

We can reduce the noise in the slope layer with a simple filter.

Task 56 : Smooth the slope raster. •

[32] http://www.saga-gis.org

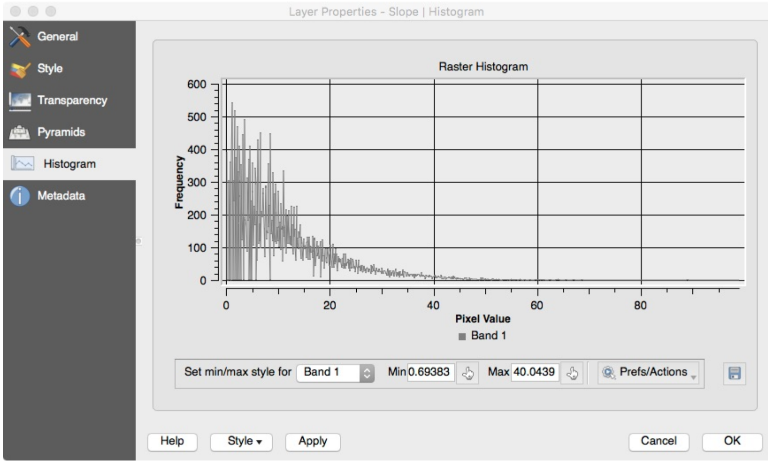


Figure 44: Histogram of the slopes computed for the Dryden 7.5” quad

1. In the Processing Toolbox select “SAGA”, then “Raster filter”, then “Simple filter”.
2. Input layer is the slope raster just created.
3. Select a search mode, either ‘square’ or ‘circle’.
4. Select the ‘smooth’ filter; this will average the values in the neighbourhood.
5. Select a radius, perhaps 4 pixels.
6. Check ‘Open output file after running algorithm’.
7. Run; examine the output on the map display.
8. Open the properties of the slope layer and compute its histogram.
9. Make the file permanent by right-clicking and then `Export | Save As ....` The format must be a GeoTIFF. A reasonable name would be `u28elu_slope_filtered.tiff`. Add to the catalog and remove the temporary slope maps (original and filtered).

The slope map is now much smoother. See Figure 45.

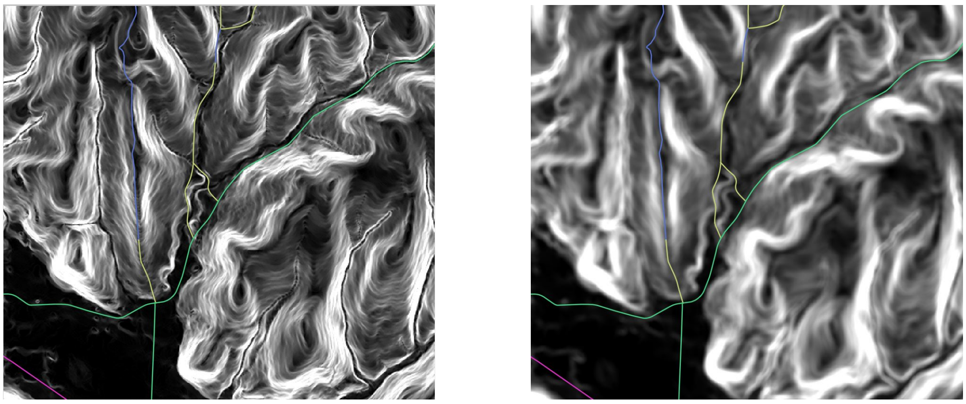


Figure 45: Slopes computed for the Dryden 7.5” quad: original (left), filtered (right)

C.5 Graphical model builder

QGIS includes a **graphical modeler**, with which you can define a **workflow**, i.e., a set of linked procedures, and then run it to produce one or more outputs. This is accessed from the `Processing | Graphical Modeler ...` menu item. A graphical model can also be run as a batch to process several of inputs of the same type.

A simple example is given at the "QGIS Tutorials and Tips" website[33].

Here we give a small example from the current project, assuming you've completed the processing through §5.1 "Reprojecting".

This model identifies highway flash flooding risk locations within a watershed, defined as:

1. intermittent streams (seasonal, can rapidly rise in heavy rainfall);
2. crossing roads;
3. within a single watershed.

This model identifies those crossings with and without culverts, and also keeps track of which entity maintains the road at each crossing. This would be useful planning information for the respective highway departments.

We build the model and then run with specific inputs: Tompkins County roads and hydrology, and the Six Mile Creek watershed.

Task 57 : Select menu item `Processing | Graphical modeller ...` to create a new graphical model. Name it `Flash flood risk`. See Figure 46. You will see a blank canvas and an "input" box, with two types of inputs (tabs below the box): "Parameters" (e.g., data layers) and "Algorithms" (i.e., how to manipulate or combine layers). See Figure 47. •

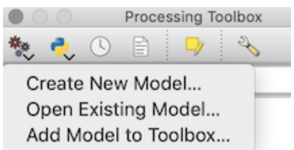


Figure 46: Creating a new processing model

Task 58 : Save the model in the default location, which is where QGIS finds models to display in the processing toolbox. Note this is *not* in your project, it is a model that can be used from any project. After each addition re-save. •

[33] `http://www.qgistutorials.com/en/docs/3/processing_graphical_modeler.html`

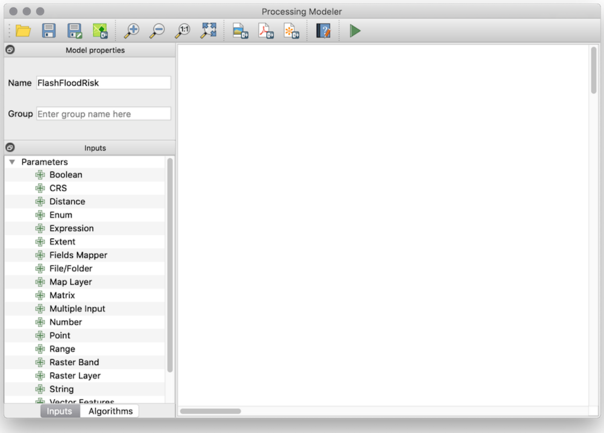


Figure 47: A blank new processing model

Task 59 : Add three "Vector Layer" input parameters: road network, stream network, and watershed. These must have geometry type `line`, `line` and `polygon`, respectively.

To add the parameter, click the plus-sign for "Vector Layer" in the input parameter list; you will then be asked for the name and type of vector layer. After you specify these, the new layer will appear in the modeller box. •

Note that as we *build* the model, we are *not* specifying an existing layer, just the type of layer to be input to the model. When we *run* the model, we will specify an existing layer which matches this specification. This way the model can be run on many combinations of inputs.

Task 60 : Add the algorithm `Vector overlay | Clip` to clip the two networks by the watershed. Give meaningful descriptions, e.g., "watershed roads" and "watershed streams", but do not add a name for the result, unless you want to see this intermediate result. •

Task 61 : Add the algorithm `Vector overlay | Line intersections` to intersect these two clipped line coverages to find where roads cross streams. Give a meaningful name, e.g. "road/stream intersections". Only keep the hydrography field `CLASS` (i.e., permanent or intermittent) and `BRIDGE` (i.e., does this stream segment go under a bridge or culvert?) and the roads field `Maintenanc` (i.e., which entity maintains the road).

To specify several fields to save, separate the names with a semicolon, in this case `CLASS;BRIDGE`. •

Task 62 : Add the algorithm `Vector selection | extract by attribute` to select the road crossings of intermittent streams (field `CLASS`, value `i` "intermittent"). •

Task 63 : Add the algorithm `Vector selection | extract by attribute` to select the road crossings of intermittent streams (i.e., *positive* result of previous selection) with and without culverts (field `BRIDGE`, value `cu` "culvert" . Save the results. •

Figure 48 shows the final graphical model.

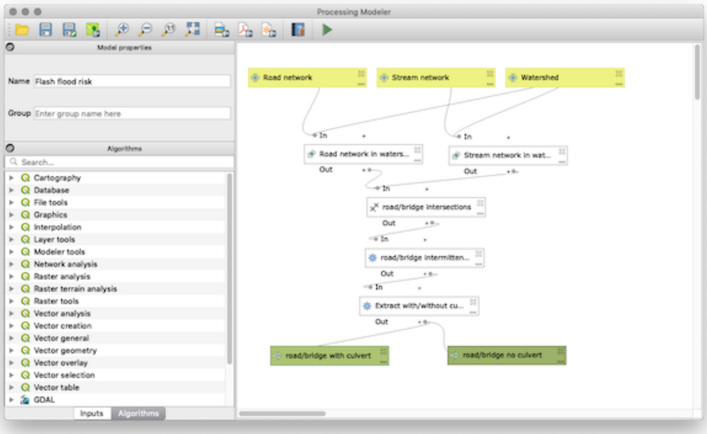


Figure 48: Final model

Task 64 : Run the model, with as inputs (1) Tompkins County roads; (2) the Tompkins County hydrography; (3) the 6-Mile Creek watershed boundary. Display the output in the catalog. See Figure 49. •

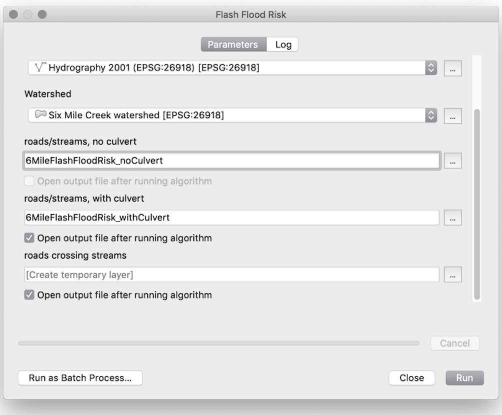


Figure 49: Running the model

Task 65 : Display the result.

Change the symbology to show the with/without culvert crossings in appropriate colours and with a large enough size, and also show the maintenance responsibility for each. •

My results are shown in Figures 50 (all road/intermittent stream crossings in the watershed) and 51 (just the locations without culverts – maybe the highway department should see the risk at these locations?).

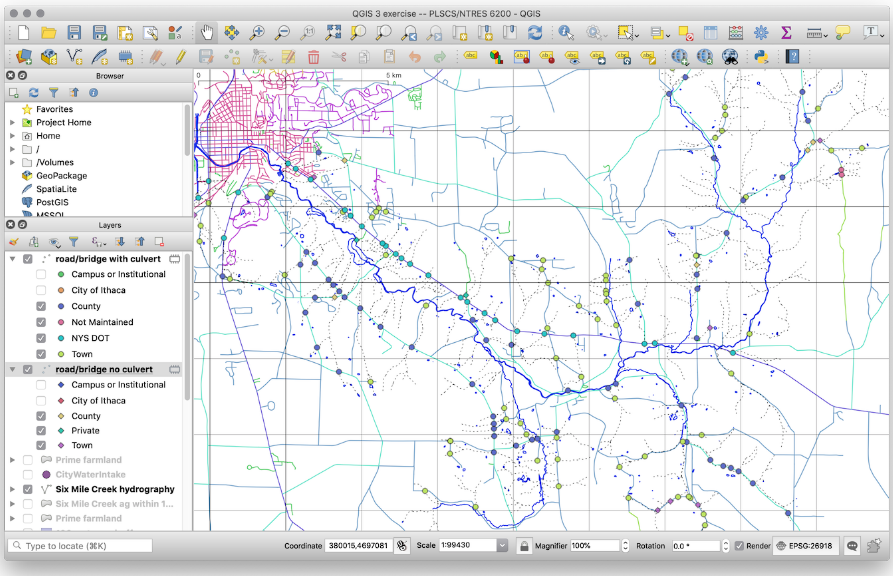


Figure 50: All road/intermittent stream crossings in the watershed

C.6 Python programming

QGIS has a built-in application programming interface (API) to the Python scripting language[34]. This can be used interactively in a console to perform GIS operations, or used to write scripts to automate processing, and even to create standalone applications.

The "QGIS Tutorials and Tips" website[35] has several good tutorials on this, see the "Python Scripting (PyQGIS)" section of that page. This includes a fairly comprehensive "Customizing QGIS with Python"[36] one-day course.

[34] https://www.python.org
[35] https://www.qgistutorials.com/en/index.html
[36] https://courses.spatialthoughts.com/pyqgis-in-a-day.html

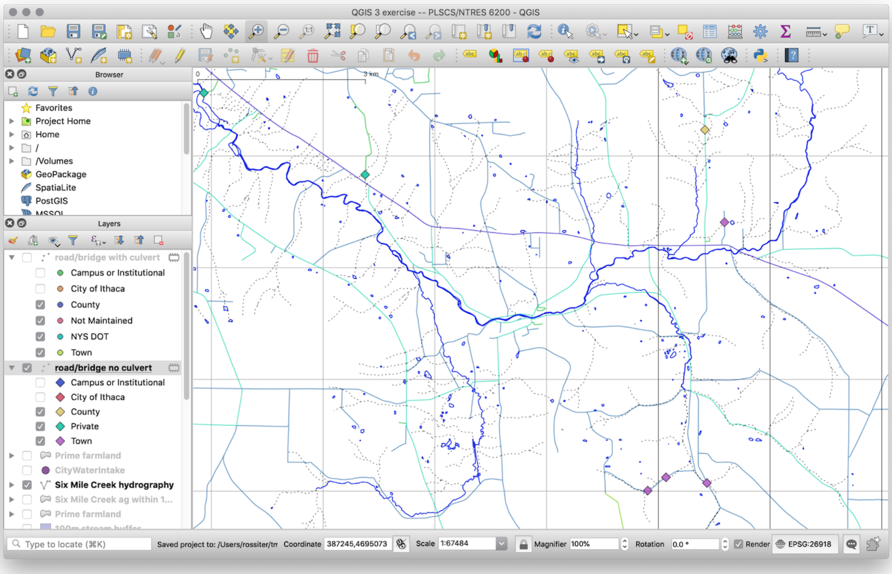


Figure 51: Locations without culverts