FoCIS

Google Earth Engine Tutorial

Prepared by the NASA DEVELOP Summer 2019 New York Ecological Forecasting II Project for the Adirondack Park Invasive Plant Program

This tutorial guides users through specific steps related to adding new or updated datasets to the models, and generating the necessary rasters from raw data. In the handoff package, all data necessary to run the model is included, and the "readme" should be treated as a tutorial for the process of running the models "as is" and for understanding the steps of the model.

Outline

- Glossary of terms
- Table of inputs into the GEE code
- Creating point data compatible for GEE and a Random Forecast Model
- Reprojecting shapefiles into WGS 1984 (WKID 4326) for GEE compatibility
- Creating Distance to Stream data in ArcMap
- Creating Soil Acidity data in ArcMap
- Importing shapefiles and raster layers as assets into GEE
- Importing ancillary datasets into GEE
- Running and exporting model results
- Adjustments for future data in the GEE code
- Point of Contact

Glossary of Terms

- **Google Earth Engine Asset**
  - External dataset loaded into Google Earth Engine for analysis
- **Table**
  - Vector data in shapefile format
  - Example: Ground-trothed location data
- **Image**
  - Raster data composed of one or more bands
  - Example: Euclidean distance to stream
- **Image Collection**
  - A stack or time series of images
  - Example: Landsat 8 imagery

Table of Inputs

Variable	Asset
l8_SR	USGS Landsat 8 Surface Reflectance Tier 1 (LANDSAT/LC08/C01/T1_SR)
SRTM	SRTM Digital Elevation Data 30m (USGS/SRTMGL1_003)
SMAP	NASA-USDA SMAP Global Soil Moisture Data (NASA_USDA/HSL/SMAP_soil_moisture)
geometry	Polygon used for downloading data from region of interest
distStream	Euclidean distance to stream at 30 m resolution from New York State linear hydrography
apValids	Point data used to validate the Random Forest Model. These points indicate ground-truthed locations of hemlock in Adirondack Park.
apTrains	Point data used to train the Random Forest Model. These points indicate ground-truthed locations of hemlock in Adirondack Park.
apipp	Boundary of Adirondacks Park
NLCD	USGS National Land Cover Database (USGS/NLCD)
acidSoils	Soil pH levels as derived from NRCS SSURGO data
S2	Sentinel-2 MSI: MultiSpectral Instrument, Level-1C
NYS	Boundary of New York State

Creating Point Data for Import

- Ensure that shapefile's attribute table contains the following columns:
  - **Tsuga:** % cover of Tsuga canadensis if measured. If not, leave blank or "0"
  - **DominanceN:** Hemlock dominance 0-3
    - 0: No Hemlock
    - 1: Hemlock Sub Dominant
    - 2: Hemlock Dominant
    - 3: Pure Hemlock
  - **Plot:** Plot ID
  - **Doms:** Hemlock presence (1) or absence (0)

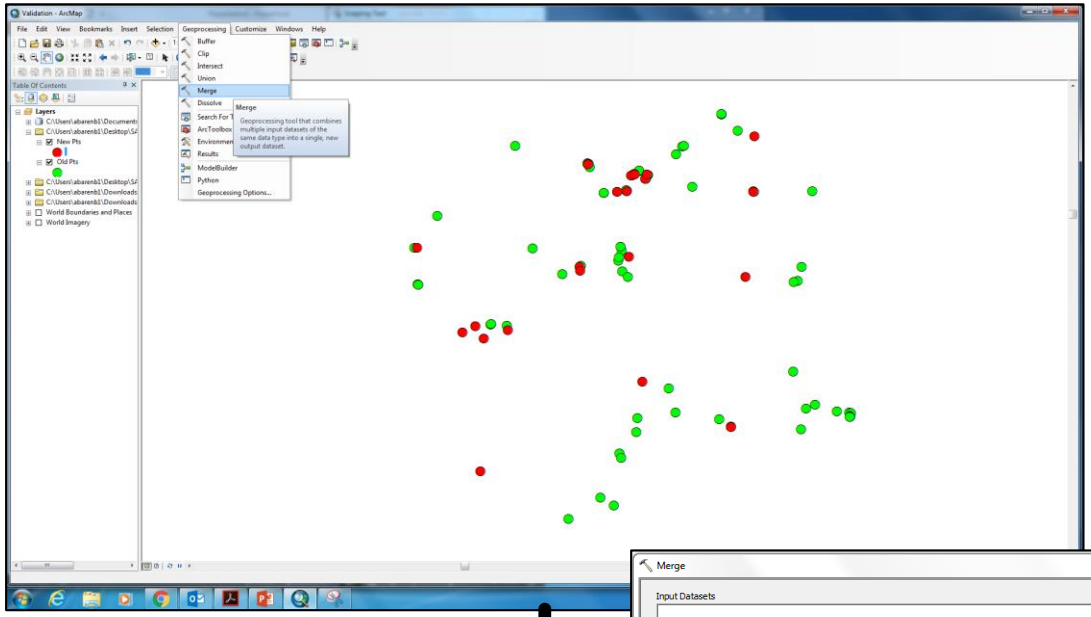
Table

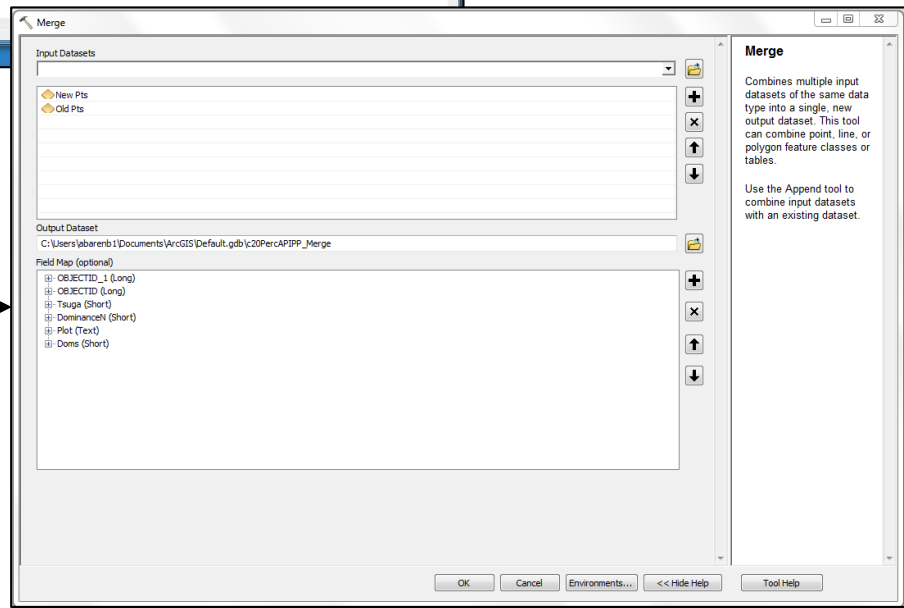
ALLPtsTrue_Clip

Shape *	OBJECTID	Tsuga	DominanceN	Plot	Doms
Point	37	0	2	Hemlock Dominant 013	1
Point	38	0	2	Hemlock Dominant 017	1
Point	39	0	2	Hemlock Dominant 018	1
Point	40	0	2	Hemlock Dominant 020	1
Point	41	0	2	Hemlock dominant 021	1
Point	42	0	2	Hemlock Dominant 024	1
Point	43	0	2	Hemlock dominant 028	1
Point	44	0	2	Hemlock Dominant 201	1
Point	45	0	2	Hemlock Dominant 205	1
Point	46	0	2	Hemlock Dominant 206	1
Point	47	0	1	Hemlock Sub Dominant 007	1
Point	48	0	1	Hemlock Sub Dominant 008	1
Point	49	0	1	HEmlock Sub Dominant 010	1
Point	50	0	1	Hemlock Sub Dominant 011	1
Point	51	0	1	Hemlock Sub Dominant 019	1
Point	52	0	1	Hemlock sub dominant 025	1
Point	53	0	1	Hemlock Sub Dominant 027	1
Point	54	0	1	Hemlock Sub Dominant 029	1
Point	55	0	1	Hemlock Sub Dominant 200	1
Point	56	0	1	Hemlock Sub Dominant 204	1
Point	57	0	1	Hemlock Sub Dominant 208	1
Point	58	0	1	Hemlock Sub Dominant 209	1
Point	59	0	1	Hemlock Sub Dominant 210	1
Point	60	0	1	Hemlock Sub Dominant 211	1
Point	61	0	1	Hemlock Sub Dominant 212	1
Point	62	0	1	Hemlock Sub Dominant 215	1

Merge New Point Data to Old Point Data

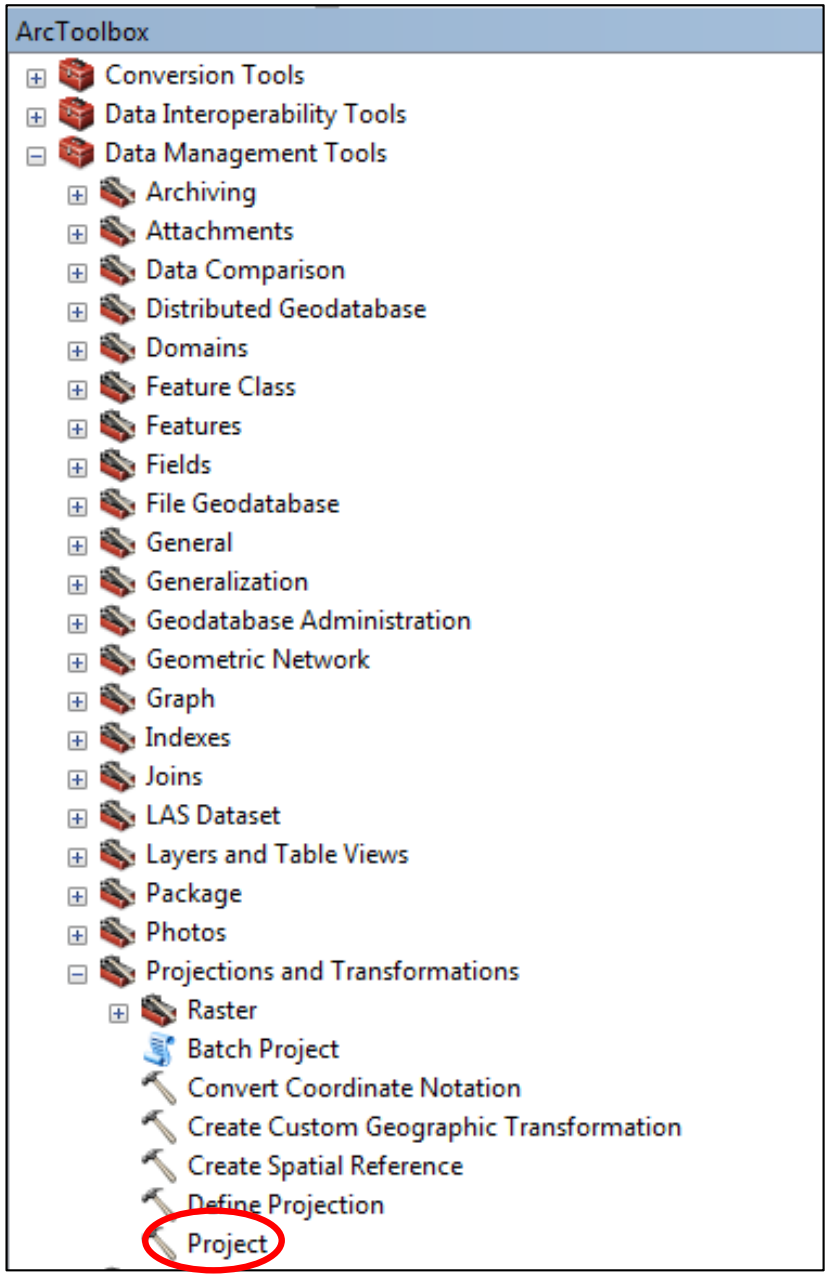
1. In ArcMap, go to Geoprocessing and select Merge
2. For “Input Datasets” select new and old point shapefiles
3. Click “okay”
4. Allow ArcMap to add new shapefile to map





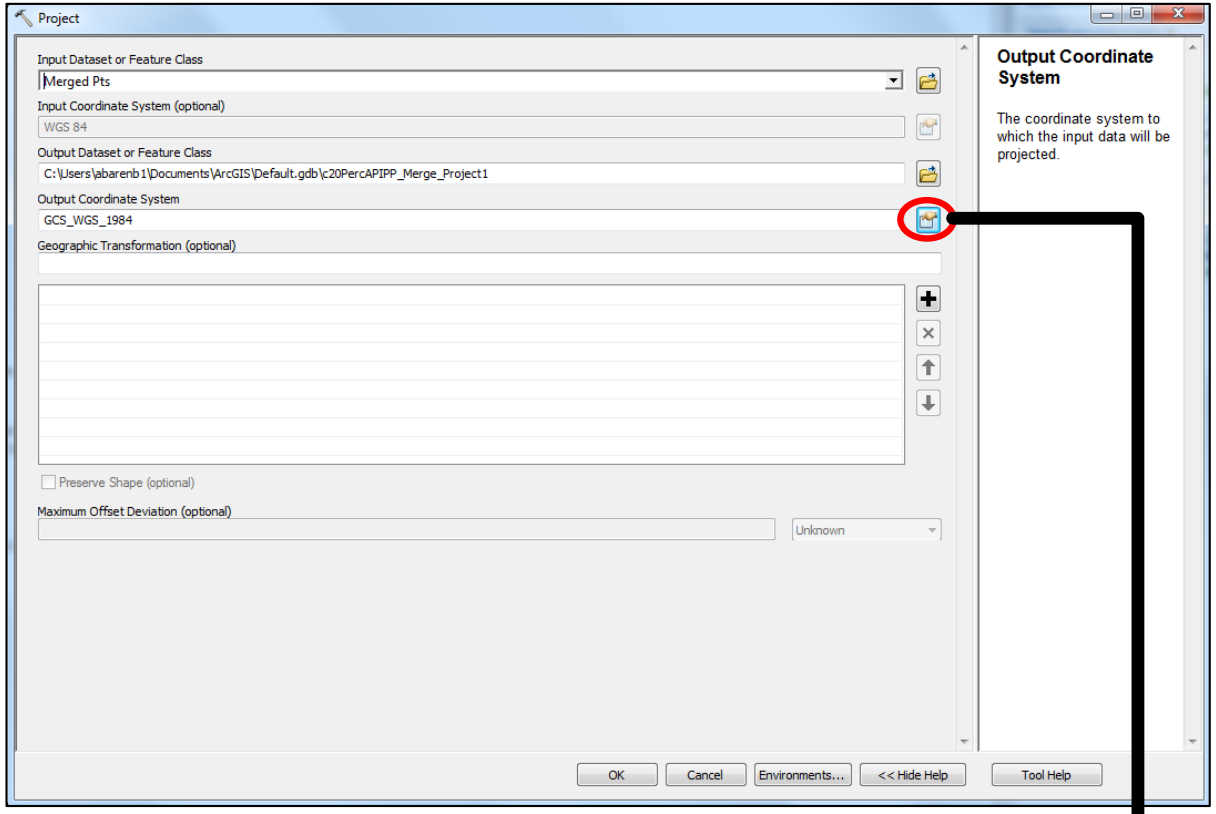
Reprojecting Shapefiles

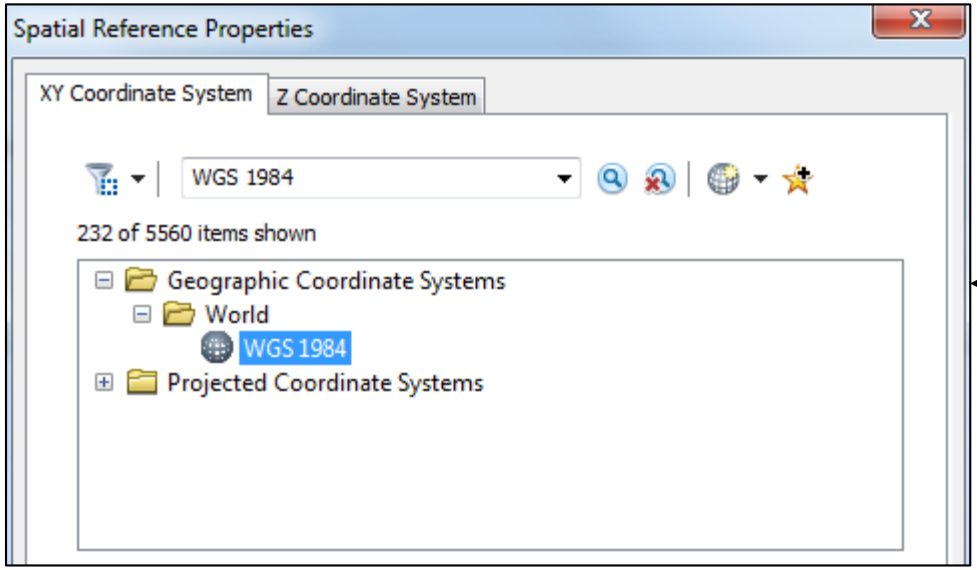
1. Under ArcToolbox navigate to “Project”
   - Projections and Transformations
     - Project



Reprojecting Shapefiles

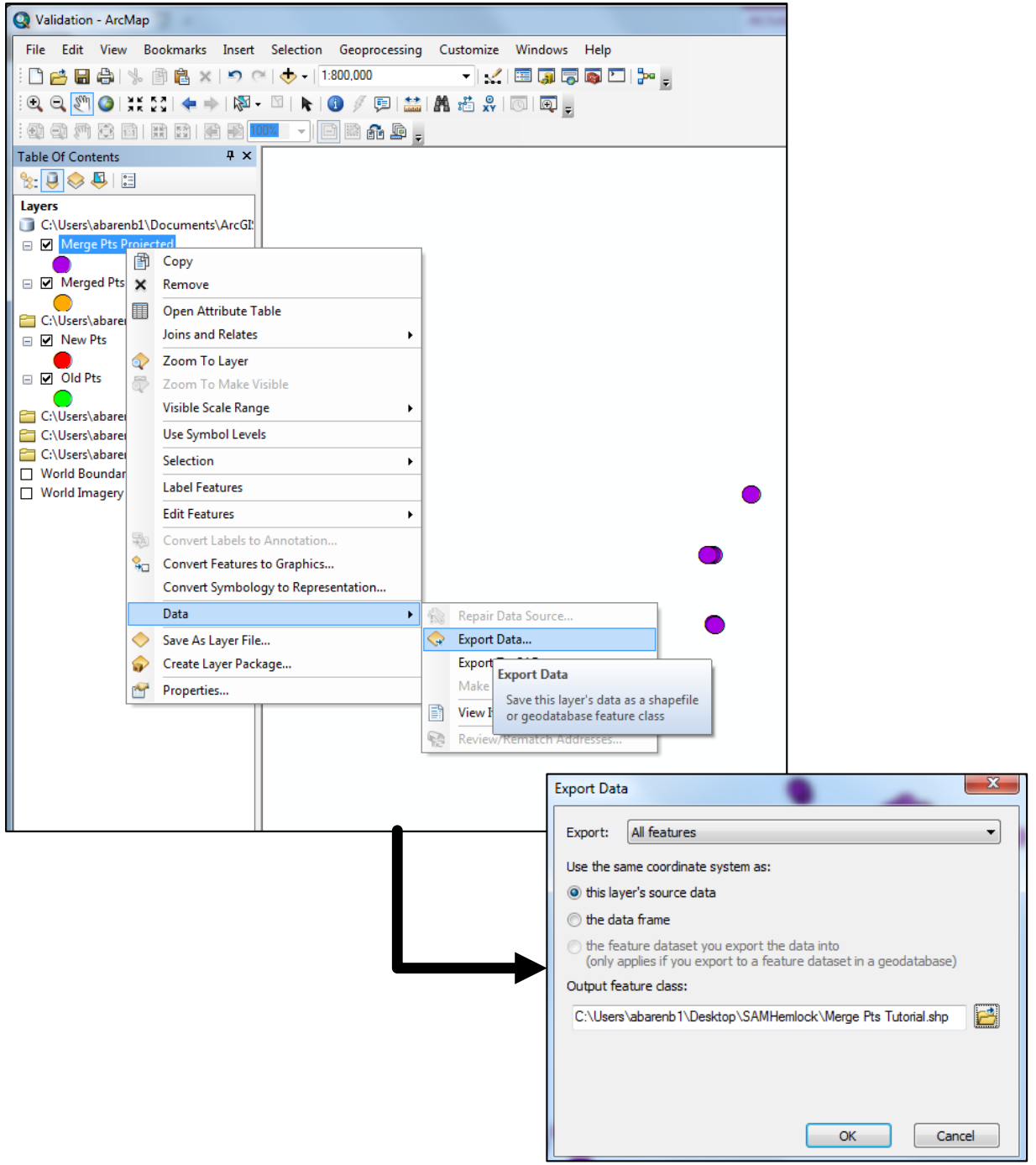
2. For “Input Datasets” select the shapefile of interest
3. Rename the file under “Output Dataset or Feature Class”
4. Select the button next to “Output Coordinate System”
5. In the search bar, search “WGS 1984”
6. Expand the “Geographic Coordinate Systems” folder, then expand the “World” folder and select “WGS 1984”





Reprojecting Shapefiles

7. Export the new point layer as a shapefile
8. Click on the new layer, navigate to “Data” and select “Export Data”
9. Rename the shapefile and ensure it is in your folder of interest

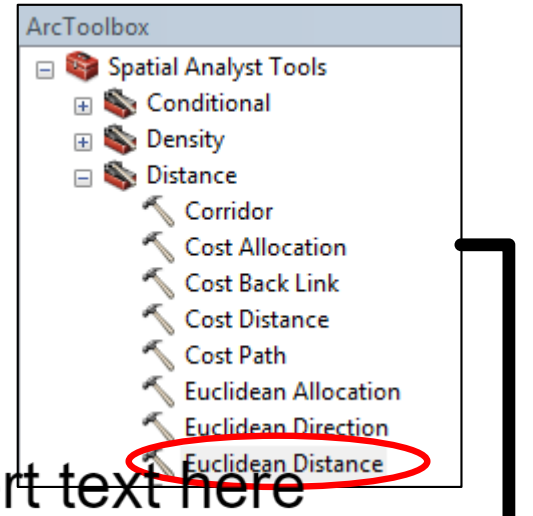


Creating Distance to Stream Raster

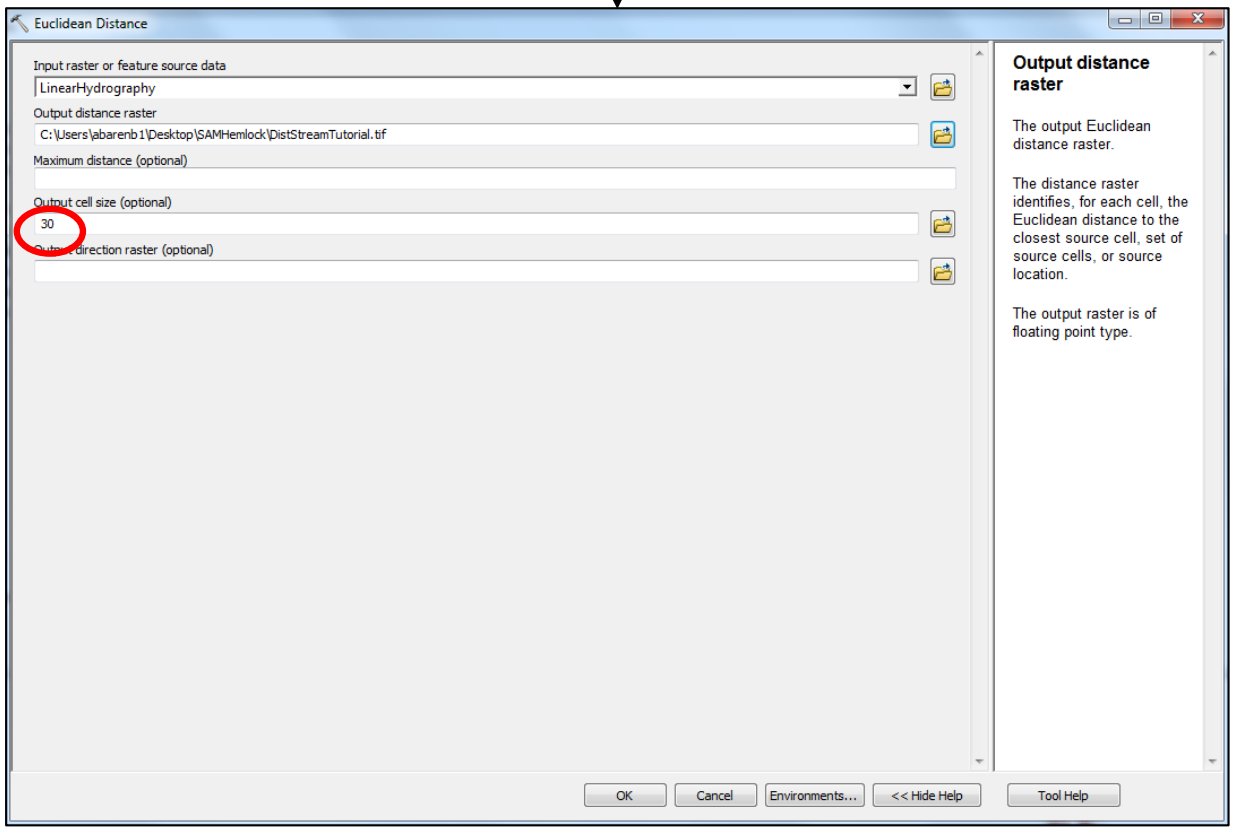
- We included naturally flowing water in our analysis (FCC: H10, H11, H12, H13)
- http://gis.ny.gov/gisdata/supportfiles/org_522_cscic_data_dictionary.zip

1. Under ArcToolbox navigate to "Euclidean Distance"
   - Spatial Analyst Tools
     - Distance
       - Euclidean Distance
2. For "Input raster or feature source data" select a linear hydrography dataset for New York
3. Change "Output cell size" to 30 meters to match Landsat 8 resolution
4. Click "OK" and allow tool to run
5. Export new raster layer as a ".tif" file following steps 8-9 from "Reprojecting Shapefiles"

A distance to streams raster is included in the handoff folder. If you would like to build a new raster based off of different data, follow these instructions.



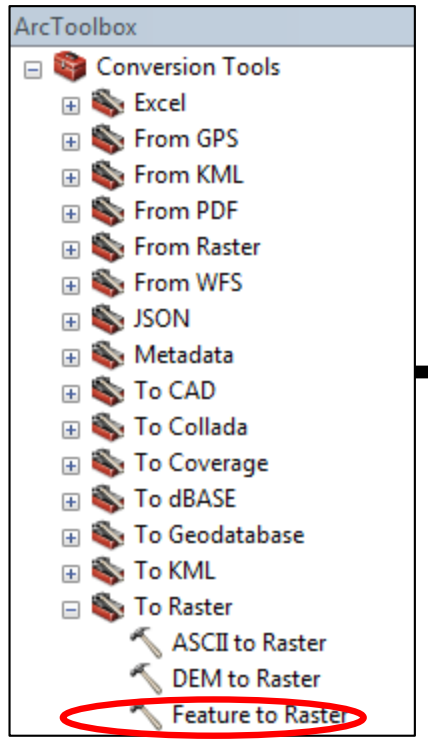
Insert text here

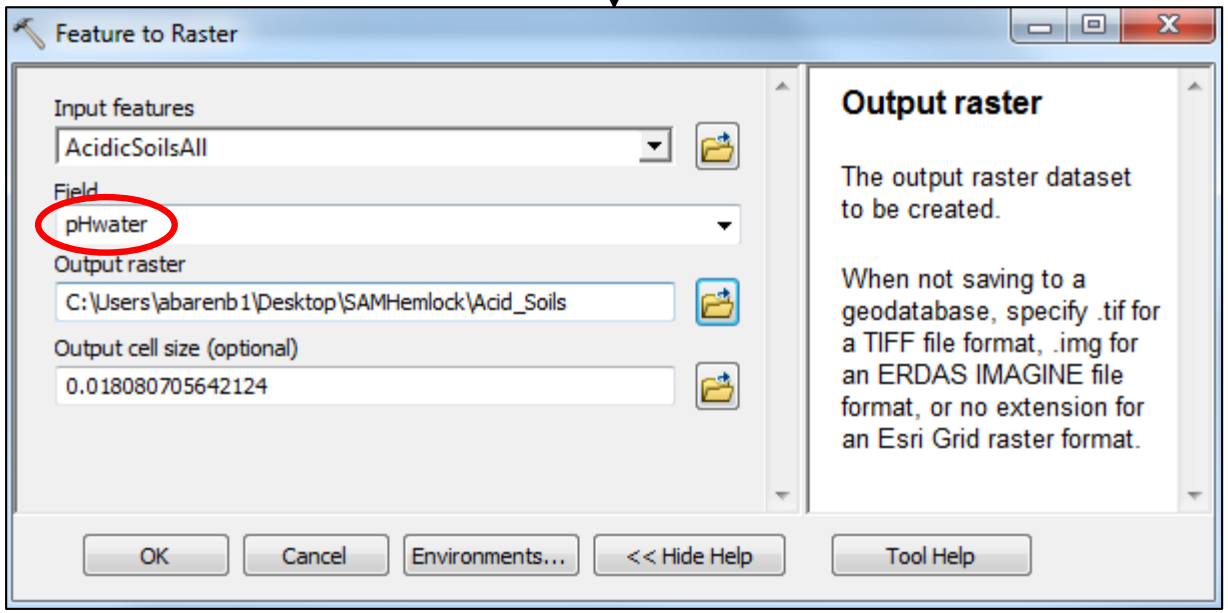


Creating Soil Acidity Raster from Vector Data

1. Under ArcToolbox navigate to “Feature to Raster”
   - Conversion Tools
     - To Raster
       - Feature to Raster
2. For “Input features” select soil acidity shapefile
3. For “Field”, use the dropdown menu to select “pHwater”
4. Keep “Output cell size” as the default
5. Click “OK” and allow tool to run
6. Export new raster layer as a “.tif” file following steps 8-9 from “Reprojecting Shapefiles”

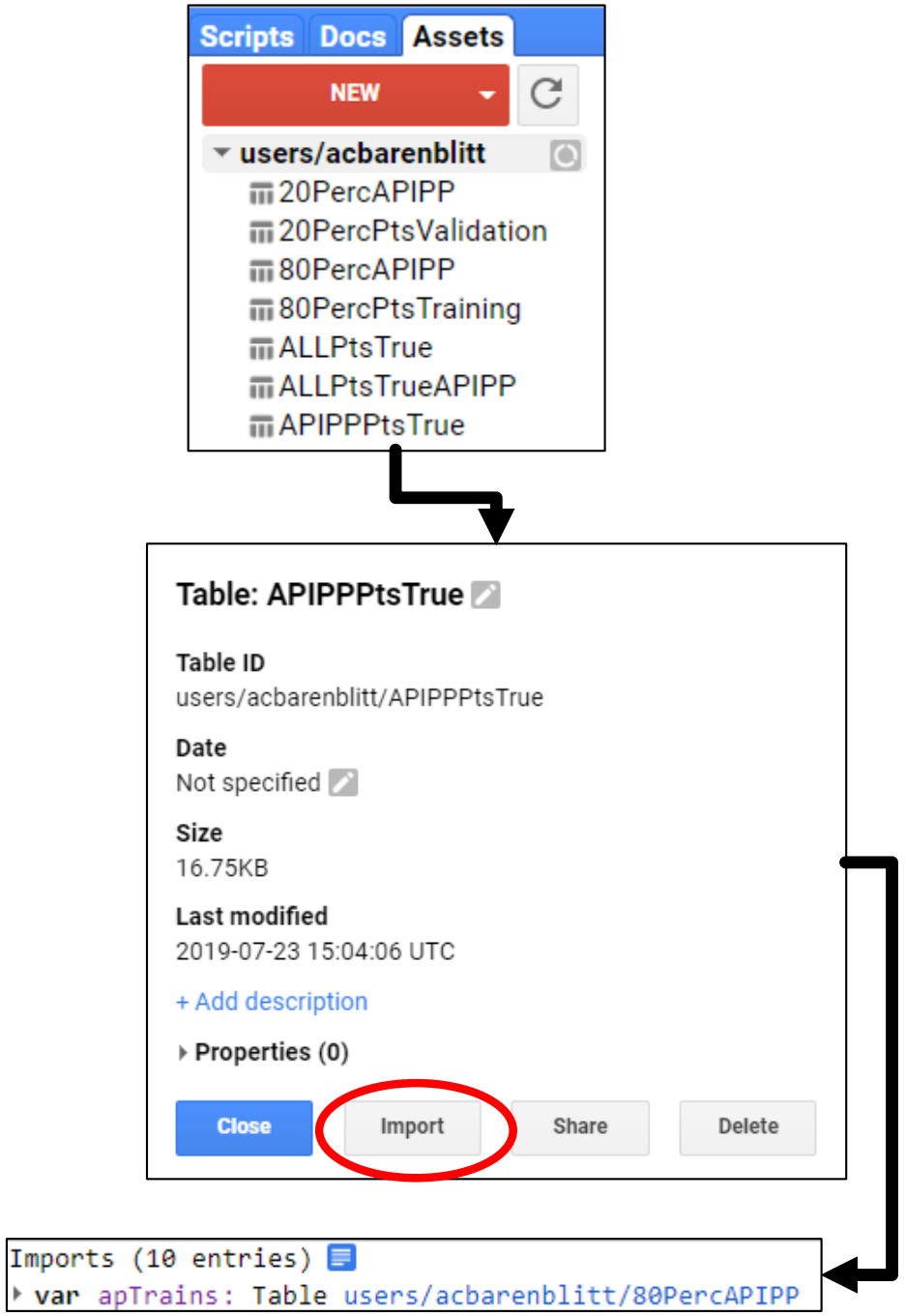
A soil acidity raster is included in the handoff folder. If you would like to build a new soil acidity based on updated data, follow these instructions





Uploading Shapefiles for Model Training into GEE

1. Go to https://code.earthengine.google.com/
2. In the top left select “Assets”
3. Under “Assets”, click NEW and select “Table upload” from the dropdown
4. Click “select” and navigate to the file of interest. Select all files associated with your shapefile EXCEPT the “.sbx” and “.xml” files
5. Click “OK”. Your file will begin to ingest. You can track this under the “Tasks” tab. This will take 5-10 minutes
6. When the table is finished uploading, click the new table under “Assets” and click “Import”. The asset will appear as a new table under “Imports” in the central panel. Rename the “table” to “apTrains”



Uploading Shapefiles for Model to GEE

- Upload the APIPP boundary as an asset using steps 1-5.
- Rename the variable from “table” to “apipp”
- Repeat these steps for validation points (“apValids”)

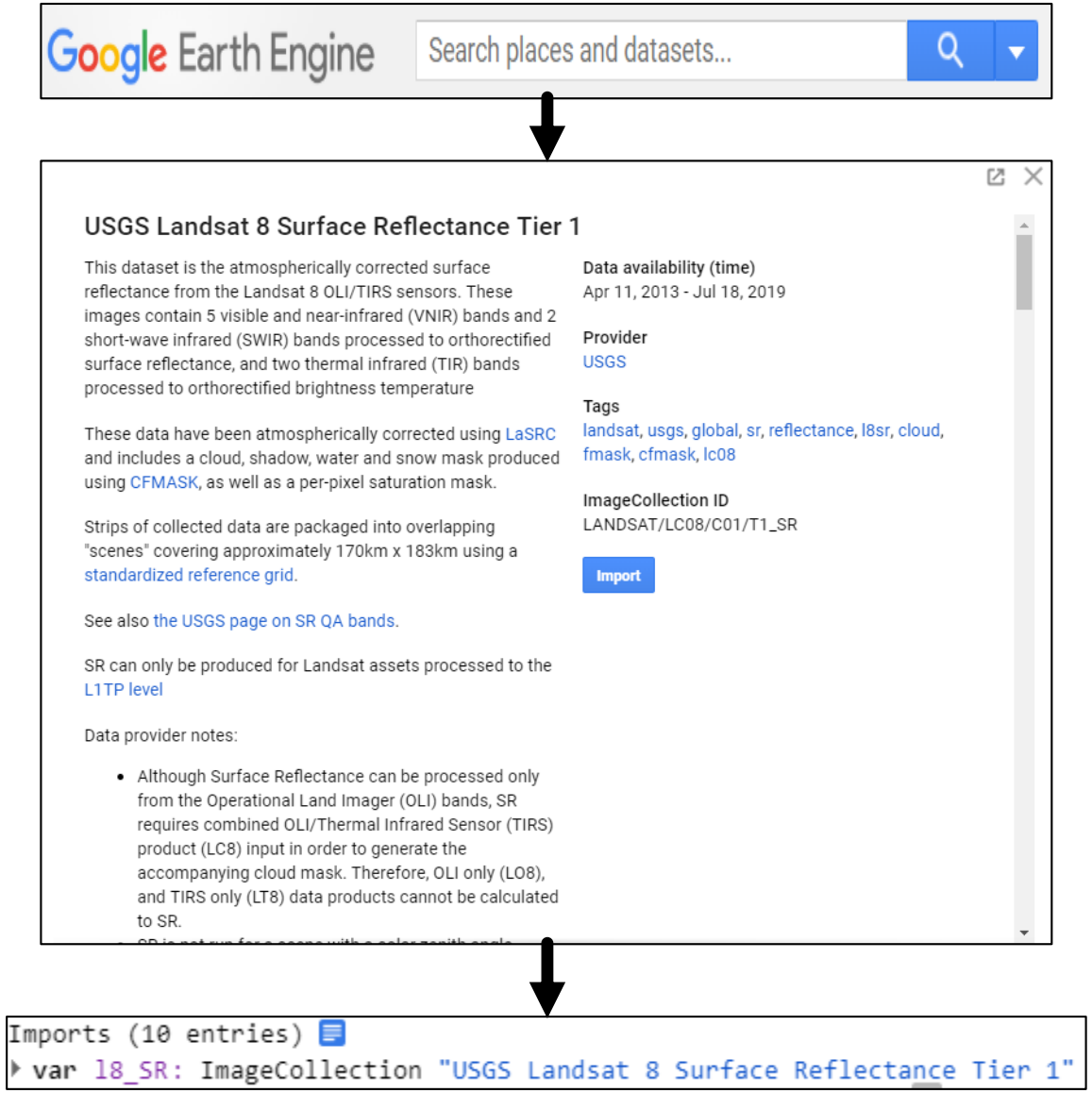
Uploading Raster Data to GEE

- Repeat steps 1-5 for uploading a shapefile, but click “Image Upload” and select a “.tif” file for Distance to Stream (“distStream”) and Acidic Soils (“acidSoils”)
- NOTE: Raster data tends to take longer to upload. Don’t be worried if your file takes 40 min-1 hour to upload!

Importing GEE Datasets into GEE

- We included Landsat 8 OLI imagery, SRTM Digital Elevation Data, SMAP, and NLCD from the GEE Dataset catalog

1. Go to https://code.earthengine.google.com/
2. At the top of the page, use the search bar to search for “Landsat 8 Surface Reflectance Tier 1”
3. Select the appropriate dataset and click “Import”
4. The asset will appear as a new table under “Imports” in the central panel. Rename the Image Collection to “l8_SR”

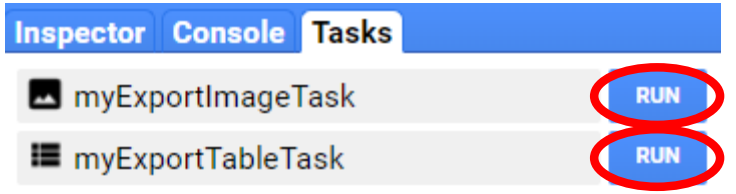


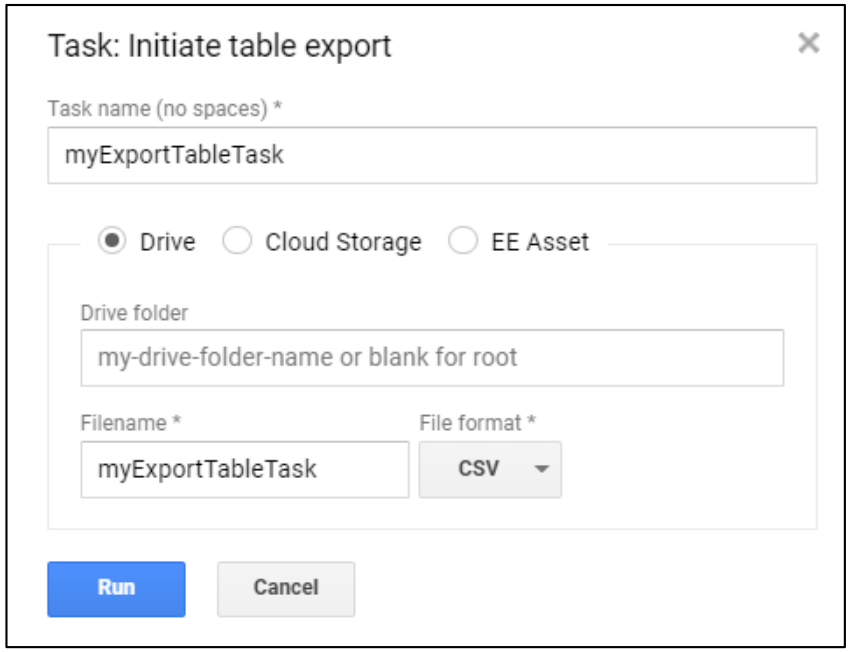
Importing GEE Datasets for Model to GEE

- Add SRTM to Imports following steps 1-3.
- Rename the variable from “image” to “SRTM”
- Repeat these steps for SMAP (“SMAP”) and NLCD (“NLCD”)

Running and Exporting Model Results

1. Once all datasets are uploaded to GEE click the "Run" button at the top of the middle panel
2. Navigate to the "Tasks" tab at the top of the right-hand panel
3. You will see two gray tasks in this panel:
    1. myExportImageTask: Raster output of habitat distribution
    2. myExportTableTask: Table output of hemlock dominance values associate with validation data based on model results
4. Click "Run". A new window will pop up. Rename the file under "Filename"
5. The exported file will be added to your Google Drive account





Adjustments for Future Data

1. When new ground-truthed data is ready for analysis, prepare and upload by following slides 3-10
2. To update years included in the “leaf-off” and “leaf-on” data, refer to lines 26-37 and 152-162 in the GEE code. Simply change the years included in these lines to reflect the dates of the study period during the “leaf-on” and “leaf-off” periods.

Point of Contact

Please direct any questions you may have toward Rya Inman at ryainman16@gmail.com and Ryan Hammock at ryan.hammock@ssaihq.cpm