NRM435 Spring 2018 Accuracy Assessment of GIS Data

GIS data always contains errors…hopefully the errors are so small that will do not significantly affect the results of the GIS analysis. In this lab, you will assess positional and attribute errors of features and rasters.

You will estimate the following positional errors:

1) Assess the precision and accuracy of simulated GPS point estimates.
2) Assess the relative accuracy of a highway line theme.
3) Check a polygon theme for sliver errors.
4) Assess the effect of positional error in rasters on change estimtes

And you will work with polygons and rasters to:

1) Check for single and relational attribute errors.
2) Check for spatial relational attribute errors
3) Develop class error estimates by comparing ground truth points with a land cover raster.

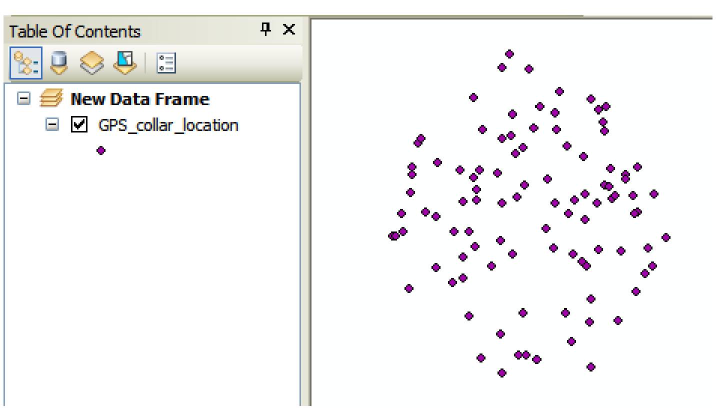
Download and unzip the file ***accuracy__lab.zip*** from:

http://dverbyla.net/nrm435/data/

All GIS data have error..the question is how large or small is the error? There are two basic types of errors in GIS: positional (X,Y,Z,M) errors and attribute errors. We will assess these errors today using fictional data.

Point Positional Errors

Sometimes you do not have known exact locations as benchmarks to compare with and you at least want to assess the variability in your GIS locations. For example, you are going to use radio-collars on some animal or radio-tag some fish or use GPS in the field…what is the precision of these locations? You purchase GPS collars for locating animals and the collars are advertised as having "2-5 meter accuracy". What will be the accuracy and precision of your GPS position estimates? You place the collar on a surveyed benchmark and record the estimated location every 5 seconds for 500 seconds, resulting in 100 estimates.

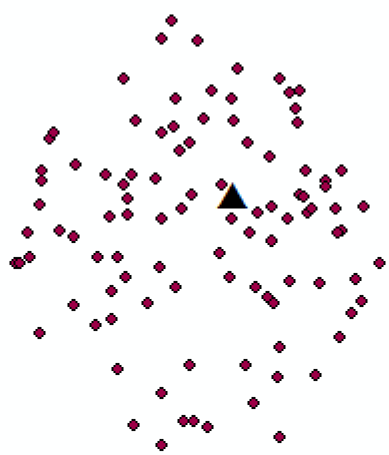


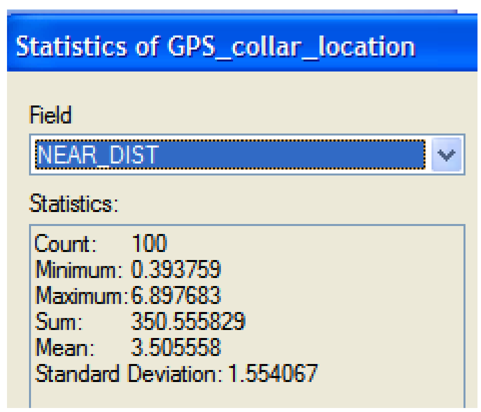
What is the precision (or repeatability) of these estimates?   Precision means how variable are these estimates…one metric of variability is standard deviation, so use the **summary statistics** tool to determine the standard deviation in the X and Y direction. First, use the **AddXY** tool to add the point x,y properties to the point attribute table.

FREQUENCY	STD_POINT_X	STD_POINT_Y
100	2.57	2.60

So the standard deviation is about 2.6 meters in the X and Y directions.

What is the accuracy (or error in relation to the true location) of these estmates? To answer that question, you need to know the true location where the collar was located. Let's say we placed the collar on a benchmark that was surveyed to have the coordinates of
X= 500,001.05 Y= 7,181,001.36.
For each GPS location estimate, compute the distance from the estimated location to the actual location. You can do this by using the ***Make XY Event Layer*** and ***Generate Near Table*** tools.



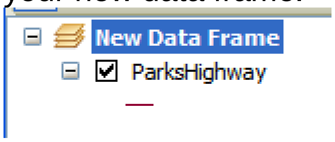


So the best estimate was .4 meters away from the true location, and the worst estimate was 6.9 meters away from the true location.

Line Relative Errors

Sometimes you do not know the true location, so you can not estimate actual positional error. But if you know what the actual distance was, you can compute estimate relative positional error. For example, let's say that we have a highway with a fixed distance of 6 meters from the centerline to the shoulder.

Create a new data frame and add the line theme representing the Parks Highway to your new data frame.



ParksHighway

Shape *	Id	Type
Polyline	1	Shoulder
Polyline	6	CenterLine

The distance from the road centerline to the shoulder ***should be 6 meters***. Use the ***Make Feature Layer*** and ***Feature Vertices To Points*** tool to create points along the shoulder line. Determine the distance from each point to the nearest center line…then create a table summarizing the errors as 6 meters – GIS distance.

FREQUENCY	MIN_Error	MAX_Error	MEAN_Error	STD_Error
81	-0.45	5.61	2.43	1.74

The mean error was 2.4 meters, but notice that our errors were biased to overestimate the distance…only 9 of the 81 points had a distance to the nearest centerline of less than 6 meters:

Shoulder_Points

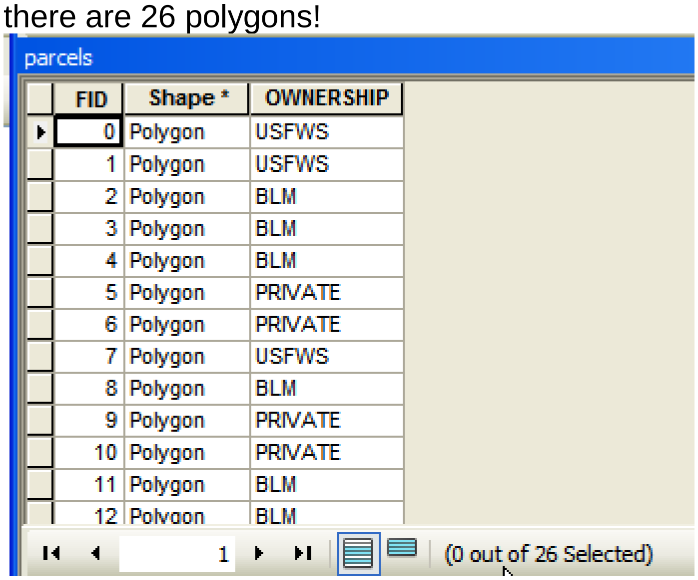
FID	Shape *	Id	Type	Error
11	Point	1	Shoulder	-0.448645
8	Point	1	Shoulder	-0.330059
10	Point	1	Shoulder	-0.288599
9	Point	1	Shoulder	-0.258016
14	Point	1	Shoulder	-0.241581
7	Point	1	Shoulder	-0.107604
5	Point	1	Shoulder	-0.052872
12	Point	1	Shoulder	-0.020379
6	Point	1	Shoulder	-0.000658
4	Point	1	Shoulder	0.024732
15	Point	1	Shoulder	0.046941
3	Point	1	Shoulder	0.088332
2	Point	1	Shoulder	0.105048

0 (9 out of 81 Selected)

Polygon Sliver Errors

Polygon slivers are created when adjacent polygons do not share exactly the same border X,Y coordinates. As an example, create a new data frame and add the polygon theme Parcels to your new data frame.

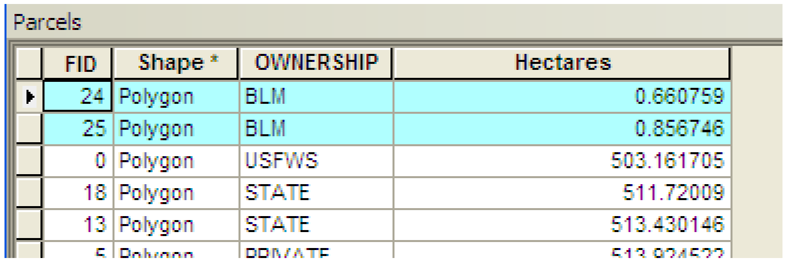
There should be 6 X 4 = 24 parcel polygons…yet according to the polygon attribute table, there are 26 polygons!


parcels

FID	Shape *	OWNERSHIP
0	Polygon	USFWS
1	Polygon	USFWS
2	Polygon	BLM
3	Polygon	BLM
4	Polygon	BLM
5	Polygon	PRIVATE
6	Polygon	PRIVATE
7	Polygon	USFWS
8	Polygon	BLM
9	Polygon	PRIVATE
10	Polygon	PRIVATE
11	Polygon	BLM
12	Polygon	BLM

1 (0 out of 26 Selected)

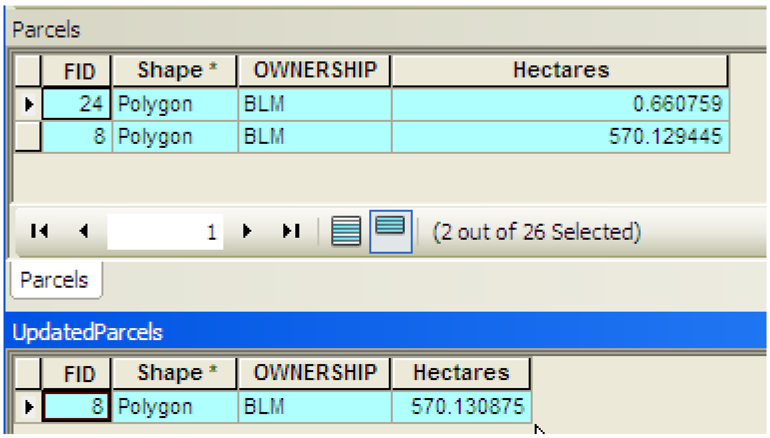
There are 2 “sliver” polygons with tiny areas…



Parcels

FID	Shape *	OWNERSHIP	Hectares
24	Polygon	BLM	0.660759
25	Polygon	BLM	0.856746
0	Polygon	USFWS	503.161705
18	Polygon	STATE	511.72009
13	Polygon	STATE	513.430146
5	Polygon	PRIVATE	513.924522

Use the ***Eliminate*** tool to eliminate these two selected polygons…the first sliver polygon is completely eliminated, since it is completely inside the BLM parcel polygon.



Parcels

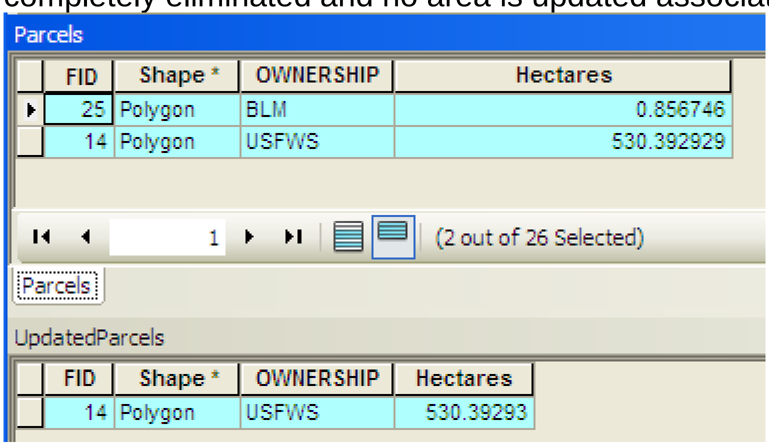
FID	Shape *	OWNERSHIP	Hectares
24	Polygon	BLM	0.660759
8	Polygon	BLM	570.129445

(2 out of 26 Selected)

UpdatedParcels

FID	Shape *	OWNERSHIP	Hectares
8	Polygon	BLM	570.130875

the other sliver polygon is completely inside an existing USFWS polygon, so it is also completely eliminated and no area is updated associated with this polygon.



Parcels

FID	Shape *	OWNERSHIP	Hectares
25	Polygon	BLM	0.856746
14	Polygon	USFWS	530.392929

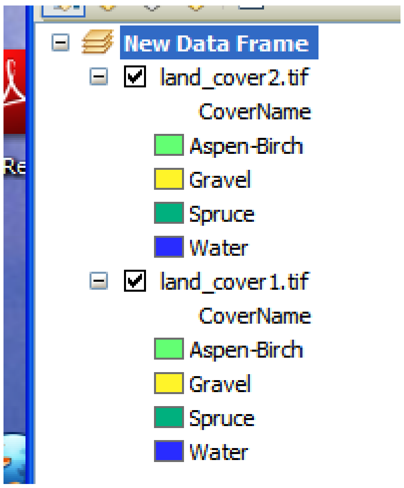
(2 out of 26 Selected)

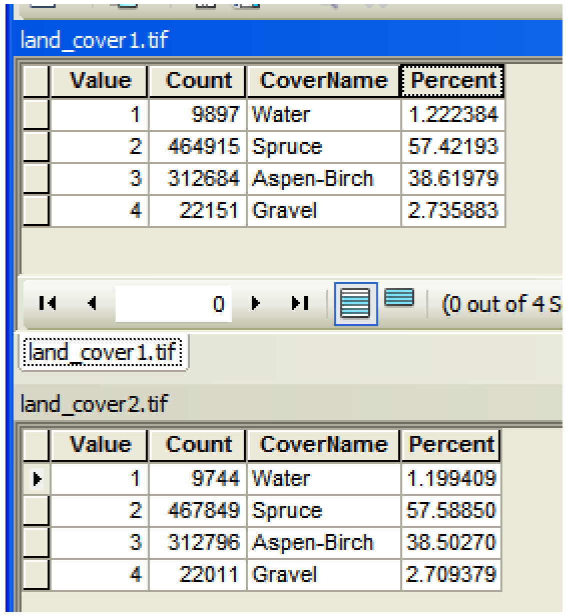
UpdatedParcels

FID	Shape *	OWNERSHIP	Hectares
14	Polygon	USFWS	530.39293

Positional Error in Rasters

Like points, lines, and polygons, there is always some positional error in the location of raster pixels. Change analysis using land cover rasters is especially sensitive to positional errors. Make a new data frame and add the 2 land cover rasters to your new data frame.





land_cover1.tif

Value	Count	CoverName	Percent
1	9897	Water	1.222384
2	464915	Spruce	57.42193
3	312684	Aspen-Birch	38.61979
4	22151	Gravel	2.735883

(0 out of 4 S

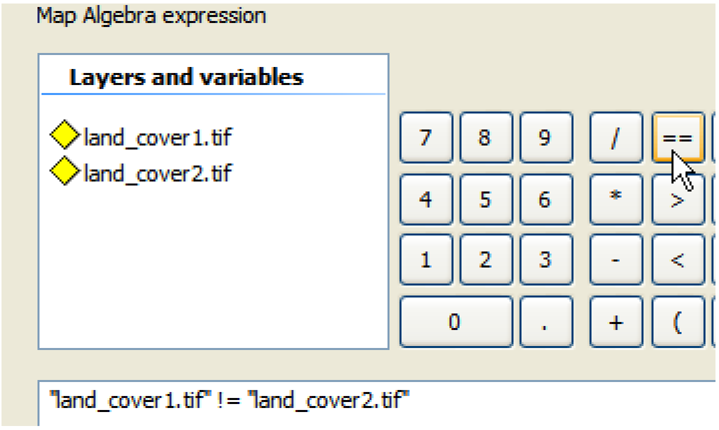
land_cover2.tif

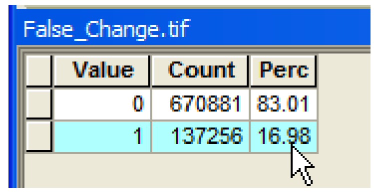
Value	Count	CoverName	Percent
1	9744	Water	1.199409
2	467849	Spruce	57.58850
3	312796	Aspen-Birch	38.50270
4	22011	Gravel	2.709379

Notice that the 2 rasters have very similar percents in terms of land cover…they were both produced using the same source data, but the positional alignment between the 2 rasters is on the average off by 1 pixel.

How does a mean positional error of 1 pixel affect land cover change estimates? In this example, there is no real change since the source data was the same satellite image, however there will be false artificial change because of positional error in the rasters.

Use the ***Raster Calculator*** or **Combine** tool to find pixels that do not have the same cover class value, and compute the percentage of false change.





So in this case we have 17 percent false change. Change estimates are always biased as overestimates due to postional error between rasters.

Attribute Errors

One method of assessing attribute errors in natural resource information is to compare mapped class values with ground truthed values. This is typically portrayed in an error matrix with ground truth as columns and mapped classes as rows. For example, imagine you have a permafrost map and 100 ground truth points. The associated error matrix might look something like:

	Ground Truth Points Permafrost:		
Mapped as:	Present	Absent	Percent
Permafrost	**30**	20	60
No Permafrost	10	**40**	80
Percent:	75	66.7	**70**

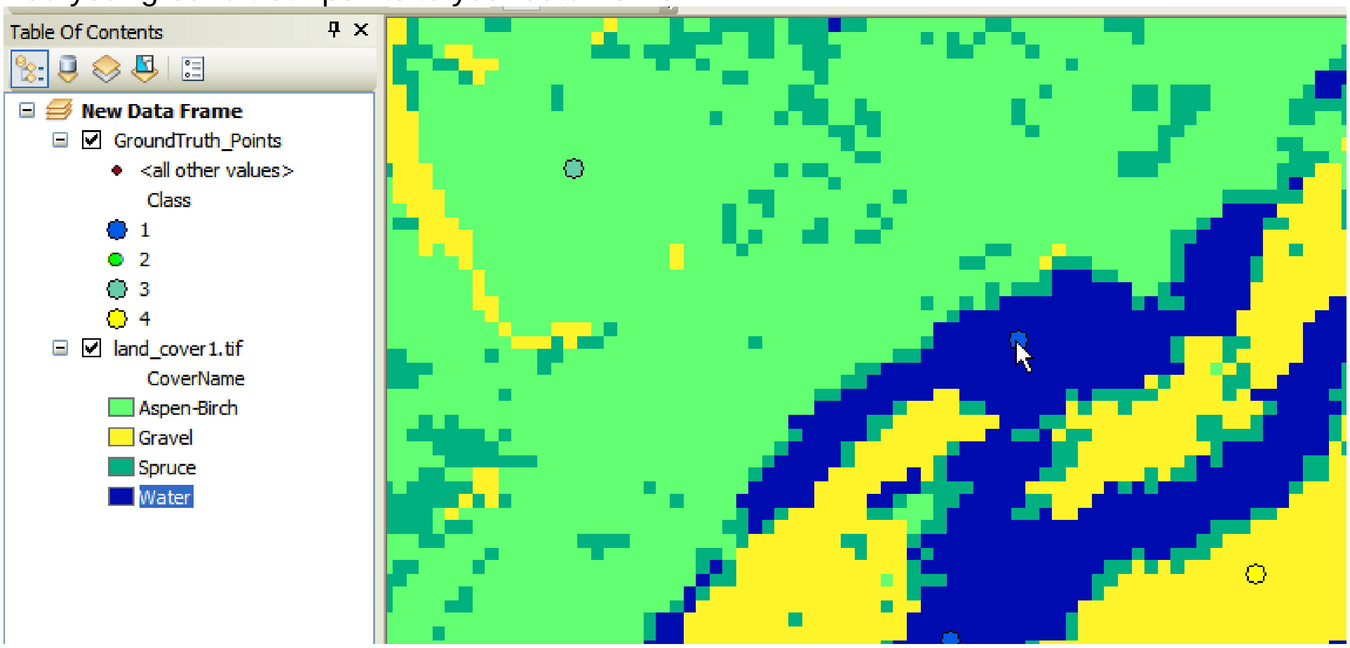
The overall classification accuracy is 70 correct/100 total ground truth points = 70 percent.

The class accuracy depends on your perspective, for example for presence of permafrost 75 percent of the permafrost ground truth locations were correctly mapped (30 of 40 permafrost ground truth locations), but 60 percent of all the points that were mapped as permafrost were really permafrost (30 of 50 pixels mapped as permafrost pixels).

Likewise, 80 percent of the points mapped as absent of permafrost were really absent of permafrost(40/50) but of the 60 ground truth points that were absent of permafrost 66 percent were correctly mapped (40/60).

There are 2 common ways to produce an error matrix of classification errors: 1) Extract pixel values to your ground truth points, or 2) convert your ground truth points to pixels, and use the Combine tool. We will do both to assess the accuracy of land_cover1.tif.

Add your ground truth points to your data frame:



Use the ***Extract Values To Points*** tool to extract the class of each pixel that each point is sitting in.

pixel_values

FID	Shape *	Class	RASTERVALU	CoverName
0	Point	2	2	Spruce
1	Point	2	2	Spruce
2	Point	2	2	Spruce
3	Point	2	2	Spruce
4	Point	2	2	Spruce
5	Point	2	2	Spruce

Next add more descriptive fields of Predict and Truth and calculate these fields from RasterValu and Class.

pixel_values

FID	Shape *	CoverName	Truth	Predict
38	Point	Spruce	2	1
39	Point	Aspen-Birch	3	1
41	Point	Spruce	2	1
42	Point	Aspen-Birch	3	1
43	Point	Spruce	2	1

1 (21 out of 120 Selected)

There are 21 of the 120 points that were incorrectly classified when compared to the ground truth values. So the overall classification accuracy is (120-21)/120 * 100 = 82.5%

The next step is to create a classification error matrix to determine the accuracy of each class. For each unique Truth/Predict class determine the number of points (use the ***Frequency*** tool)

GroundTruth_Frequency

Truth	Predict	FREQUENCY
1	1	16
1	2	11
1	3	3
2	2	30
3	3	30
4	1	1
4	2	4
4	3	2
4	4	23

Next, use the **Pivot Table** tool to reformat this information with the ground truth classes as columns and the predict classes as rows. Here is an example from the help:

Input Table

EntID	LinkType	TableID	LinkValue
1	A	X1	20
1	A	X2	21
2	A	X1	23
2	A	X2	29
2	B	X1	80
2	B	X2	77

Input Fields (EntID, LinkType) — Pivot Field (TableID) — Value Field (LinkValue)

Output Table

EntID	LinkType	X1	X2
1	A	20	21
2	A	23	29
2	B	80	77

GroundTruth_Frequency_PivotT

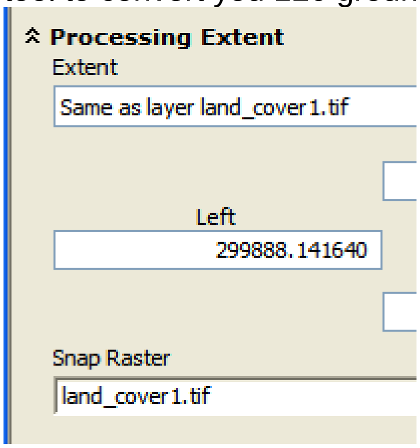
CoverName	Predict	Truth1	Truth2	Truth3	Truth4
Water	1	16	0	0	1
Spruce	2	11	30	0	4
Aspen-Birch	3	3	0	30	2
Gravel	4	0	0	0	23

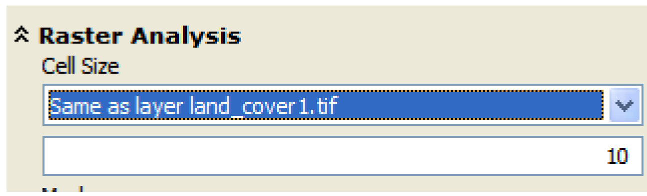
Finally, since your error matrix is a dbf table, use Excel to modify this table by computing the percent correct for each class:

	A	B	C	D	E	F
1		Water	Spruce	Aspen-Birc	Gravel	
2	Predict	Truth1	Truth2	Truth3	Truth4	***Predict Percent***
3	Water	**16**	0	0	1	***94.1***
4	Spruce	11	**30**	0	4	***66.7***
5	Aspen-Birch	3	0	**30**	2	***85.7***
6	Gravel	0	0	0	**23**	***100.0***
7	***Truth Percent***	***53.3***	***100.0***	***100.0***	***76.7***	
8						

Notice that for water, 16 of the 30 water ground truth points were correctly mapped (53.3%), yet of the 17 water pixels, 16 were mapped as water pixels (94.1% correct).

You could also create an error matrix with a ground truth raster. First, set your geoprocessing environments so they perfectly match the land cover raster. Then use the ***Point To Raster*** tool to convert you 120 ground truth points to 120 pixels in a ground truth raster.





GrounthTruth.tif

Value	Count
1	30
2	30
3	30
4	30

Next, use the ***Combine*** tool to compare your ground truth raster with your land cover raster.

Truth_Predict.tif

Value	Count	GrounthTru	land_cover
1	30	3	3
2	4	4	2
3	30	2	2
4	11	1	2
5	16	1	1
6	3	1	3
7	23	4	4
8	2	4	3
9	1	4	1

And then use the ***Copy Rows*** and ***Pivot Table*** tools to create your error matrix dbf, which matches our original error matrix created from the ground truth points:

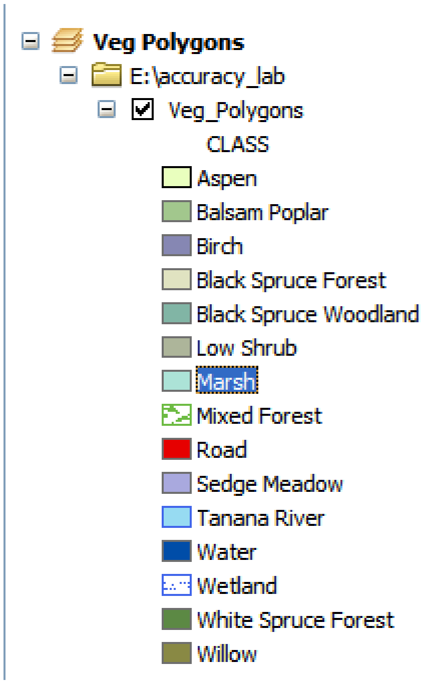
GroundTruth_Frequency_PivotT

CoverName	Predict	Truth1	Truth2	Truth3	Truth4
Water	1	16	0	0	1
Spruce	2	11	30	0	4
Aspen-Birch	3	3	0	30	2
Gravel	4	0	0	0	23

Feature Attribute Errors

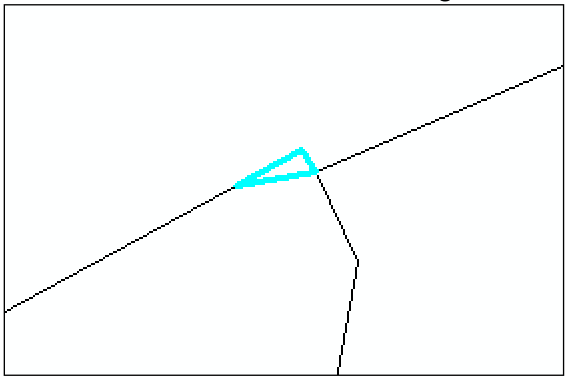
Typical checks for feature (point, line or polygon) attributes are check for: 1) values out of range, 2) errors in relationships between attributes, and 3) attribute errors in spatial relationships.

Create a new data frame and add the vegetation polygons theme to your new data frame.



Out of range errors. Examples of out of range errors for shape error might be tiny sliver polygons, or if there is an extent polygon of the entire area as a “vegetation” polygon.

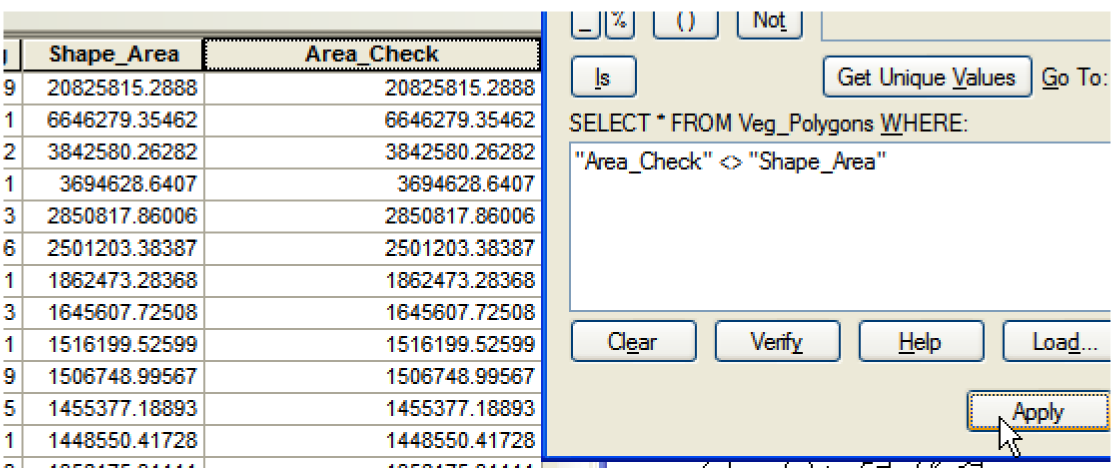
A simple and quick check to sort ascending, select the “smallest” polygon and the zoom to selected. Then sort descending, select and zoom to the largest polygon.



likely a sliver polygon error!

large black spruce woodland…likely not an attribute error

Shapefile length and areas are not updated following operations such as Clip, Intersect, Update, or editing, so it is always wise to check the area and length values whenever you receive shapefiles!



	Shape_Area	Area_Check
9	20825815.2888	20825815.2888
1	6646279.35462	6646279.35462
2	3842580.26282	3842580.26282
1	3694628.6407	3694628.6407
3	2850817.86006	2850817.86006
6	2501203.38387	2501203.38387
1	1862473.28368	1862473.28368
3	1645607.72508	1645607.72508
1	1516199.52599	1516199.52599
9	1506748.99567	1506748.99567
5	1455377.18893	1455377.18893
1	1448550.41728	1448550.41728

Relationship errors. Relationship errors are class values that do not make sense. For example, a sawtimber willow polygon, or a wetland aspen polygon.

Use the Class and SizeClass fields to check that

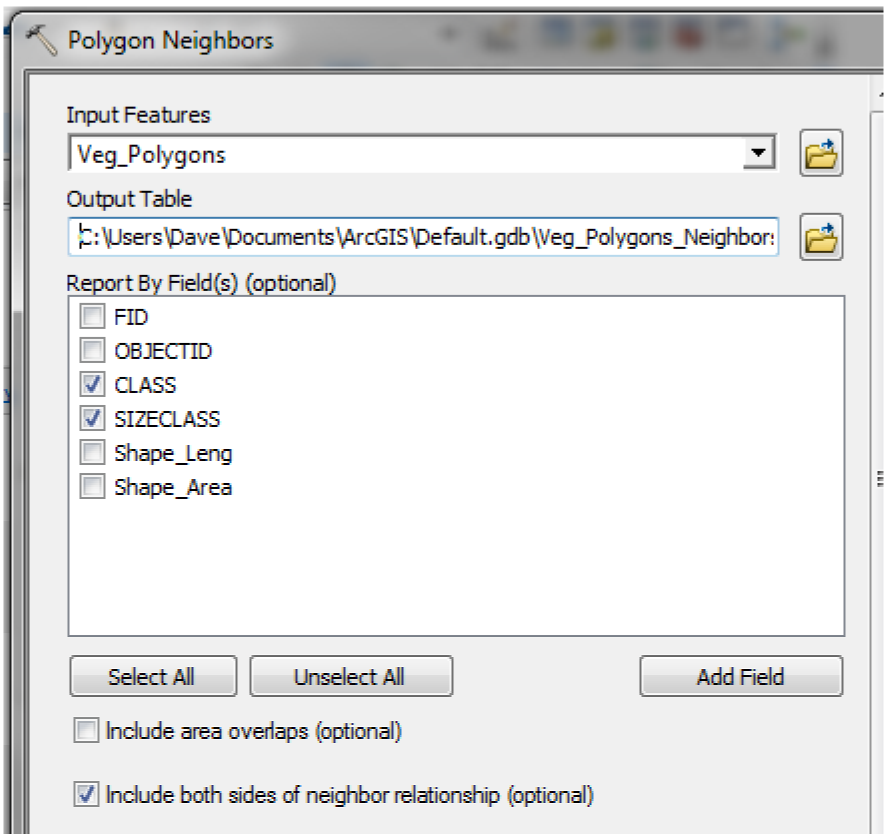
- ✓ all Wetland Polygons are in the wetland cover classes:

"SIZECLASS" = 'W' AND "CLASS" NOT In ('Wetland', 'Water', 'Tanana River', 'Sedge Meadow', 'Marsh')

- ✓ all shrub polygons are in size class of reproduction ("R")

("CLASS" in ('Low Shrub', 'Willow')) and ("SIZECLASS" <> 'R')

Attribute errors in spatial relationships. Some attribute errors are not evident until you look at spatial relationships. For example, are there any polygons that have the same Class and SizeClass that are adjacent to each other? If so, there is likely an attribute error since these should belong to the same polygon.



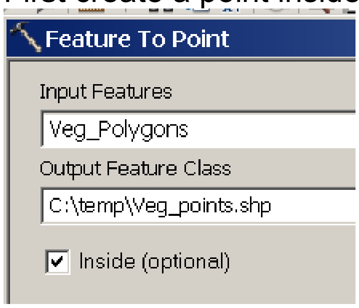
Veg_Polygons_Neighbors

src_CLASS	nbr_CLASS	src_SIZECLASS	nbr_SIZECLASS	LENGTH	NODE_COUNT
Birch	Birch	P	P	0	2
Black Spruce Forest	Black Spruce Forest	R	R	2608.945259	0
Black Spruce Woodland	Black Spruce Woodland	R	R	35016.875099	0
Black Spruce Woodland	Black Spruce Woodland	S	S	0	2
Low Shrub	Low Shrub	R	R	1267.211667	0
Willow	Willow	R	R	0	2

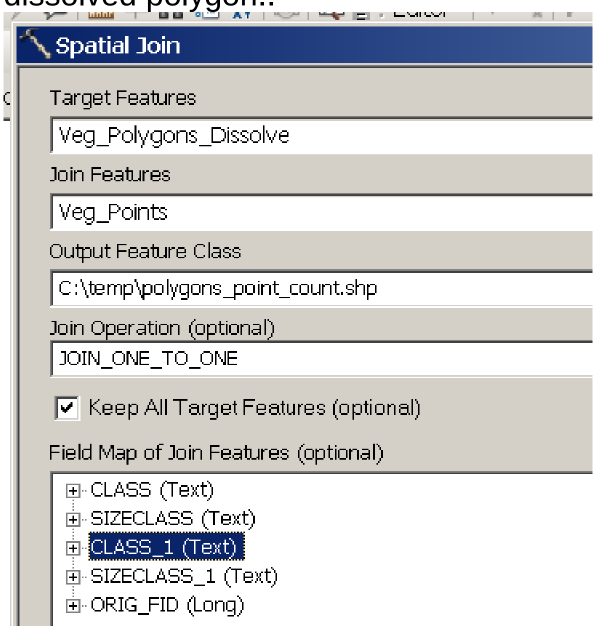
Use the ***Dissolve*** tool to dissolve adjacent polygons if they have the same Class, SizeClass values (do ***not*** create multipart polygons)…

Where are these adjacent polygons that had the same Class, SizeClass values? Use the ***Spatial Join*** tool to find these polygons…

First create a point inside each original vegetation polygon…



The use Spatial Join to determine how many original polygon points are inside each dissolved polygon..

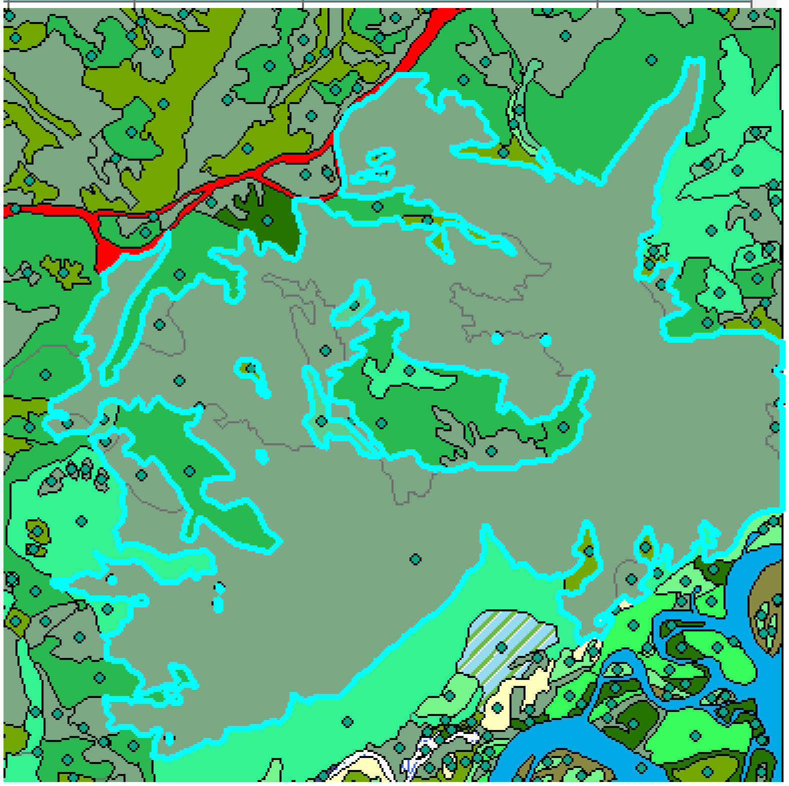


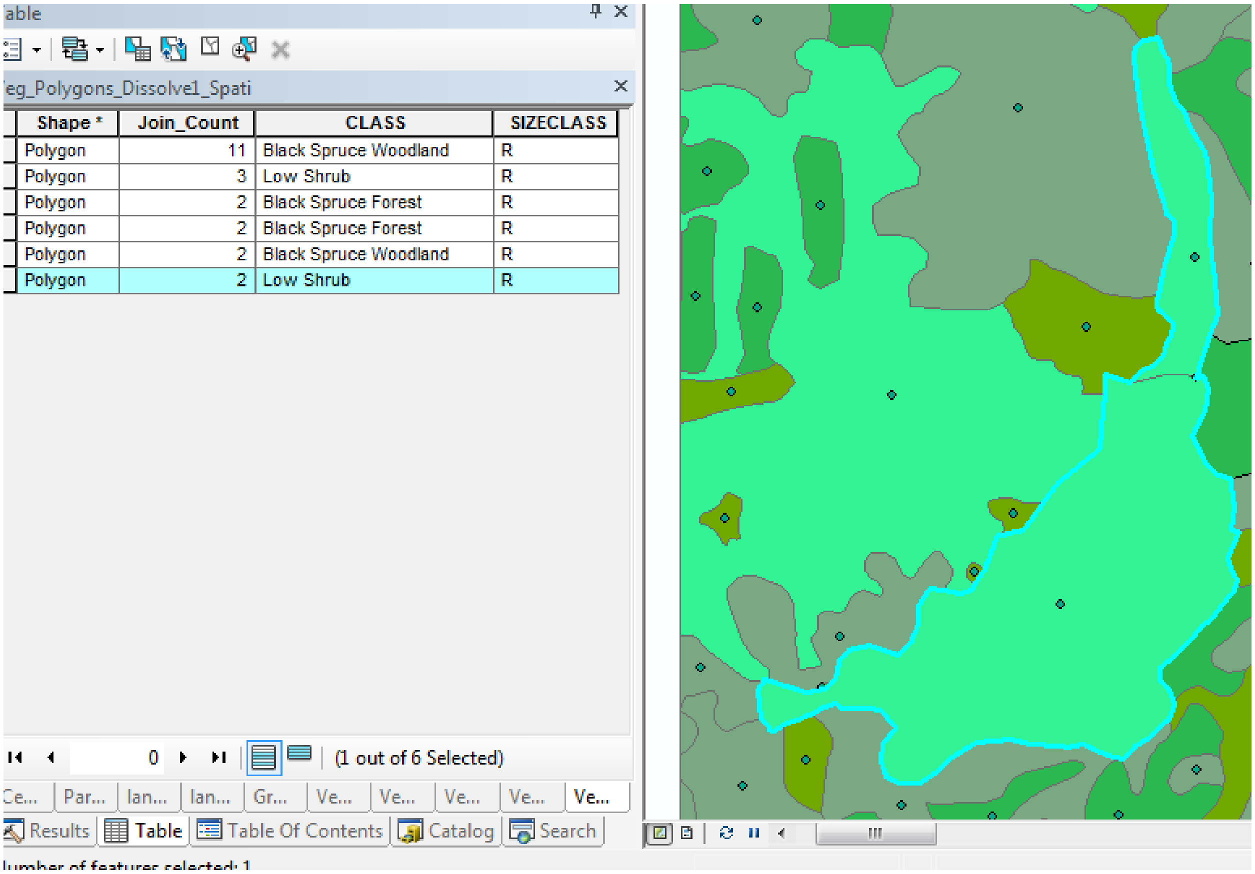
Shape *	Join_Count	CLASS	SIZECLASS	CLASS_1	SIZECLASS_
Polygon	11	Black Spruce Woodland	R	Black Spruce Woodland	R
Polygon	3	Low Shrub	R	Low Shrub	R
Polygon	2	Black Spruce Forest	R	Black Spruce Forest	R
Polygon	2	Black Spruce Forest	R	Black Spruce Forest	R
Polygon	2	Black Spruce Woodland	R	Black Spruce Woodland	R
Polygon	2	Low Shrub	R	Low Shrub	R
Polygon	1	Aspen	P	Aspen	P

Join_Count is the count of original polygons within each dissolved polygon. For example, there were 2 polygons that were adjacent to each other that had the same Class, SizeClass values…either one of the polygons had an attribute error, or these 2 polygons should have been represented by one polygon.

Here are 11 polygons adjacent to each other all having the same Class and SizeClass values…

Shape *	Join_Count	CLASS	SIZECLASS
Polygon	11	Black Spruce Woodland	R





Shape *	Join_Count	CLASS	SIZECLASS
Polygon	11	Black Spruce Woodland	R
Polygon	3	Low Shrub	R
Polygon	2	Black Spruce Forest	R
Polygon	2	Black Spruce Forest	R
Polygon	2	Black Spruce Woodland	R
Polygon	2	Low Shrub	R

And 2 adjacent polygons that were Class Low Shrub, Size Class R