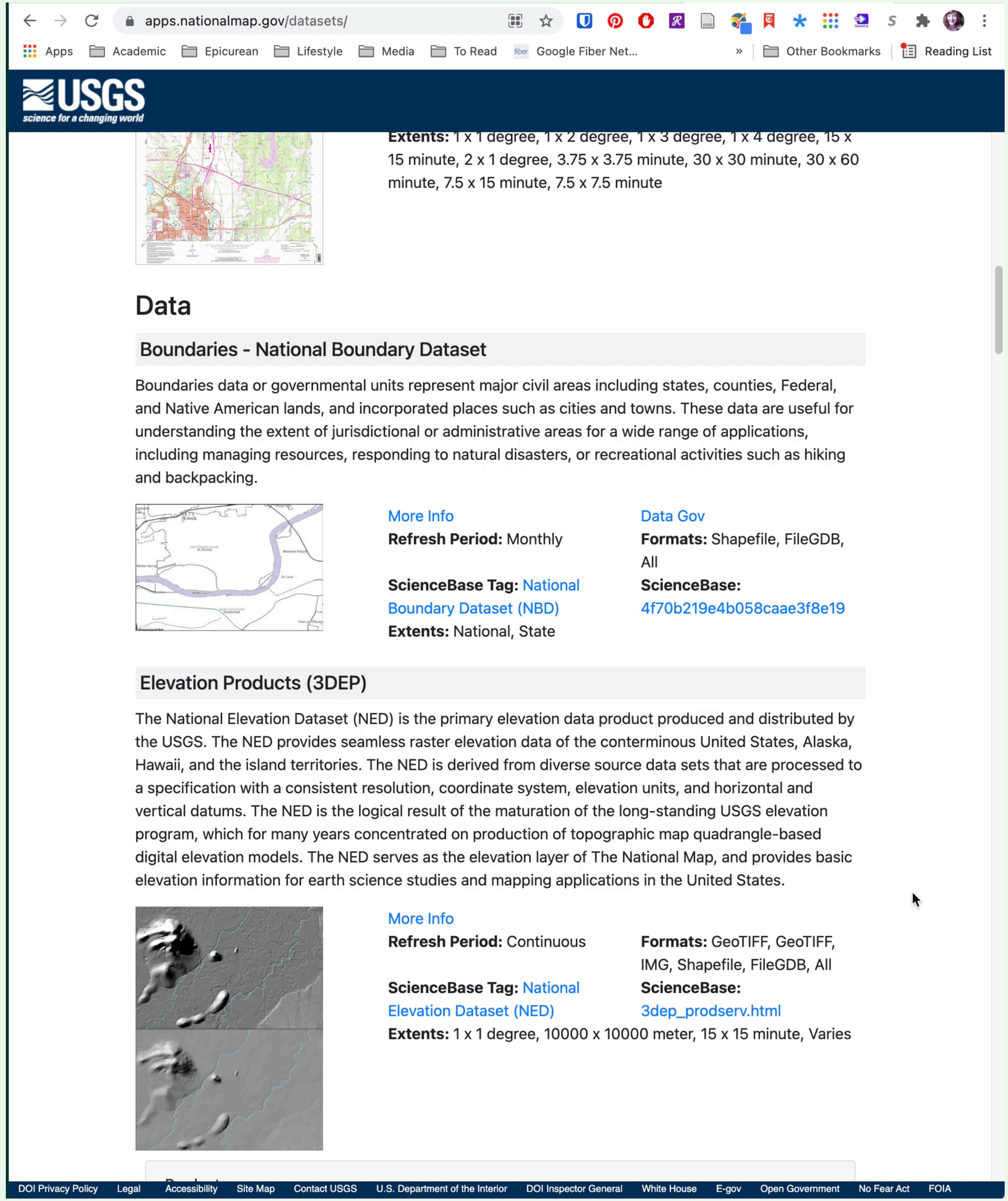
GEOG 358: Introduction to Geographic Information Systems

Digitizing & georeferencing

- Vector data
- Aerial and satellite imagery
- Paper maps
- Field survey
- There's a lot of data out there but it was all created by someone!

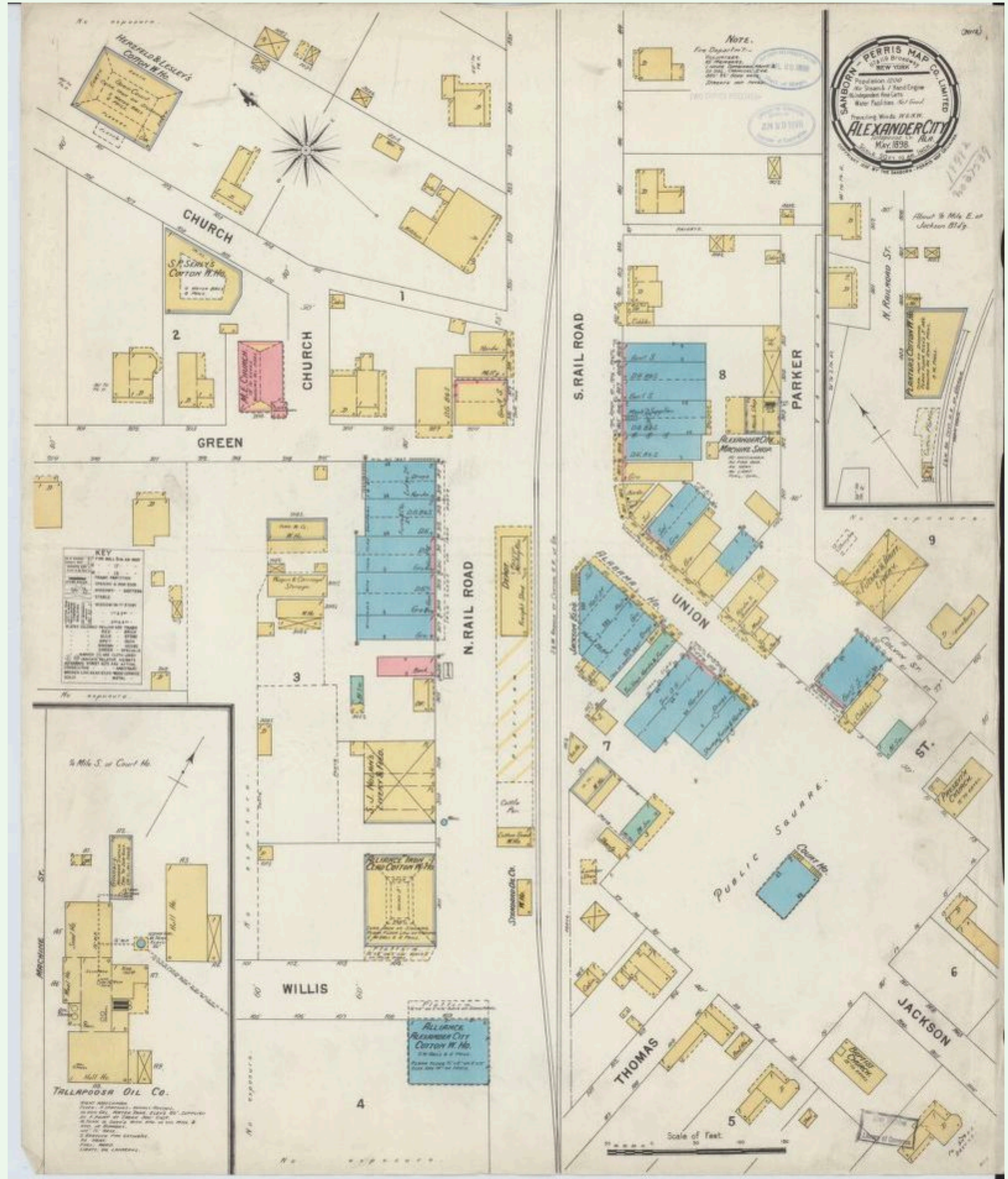
Existing data

- Vector and raster



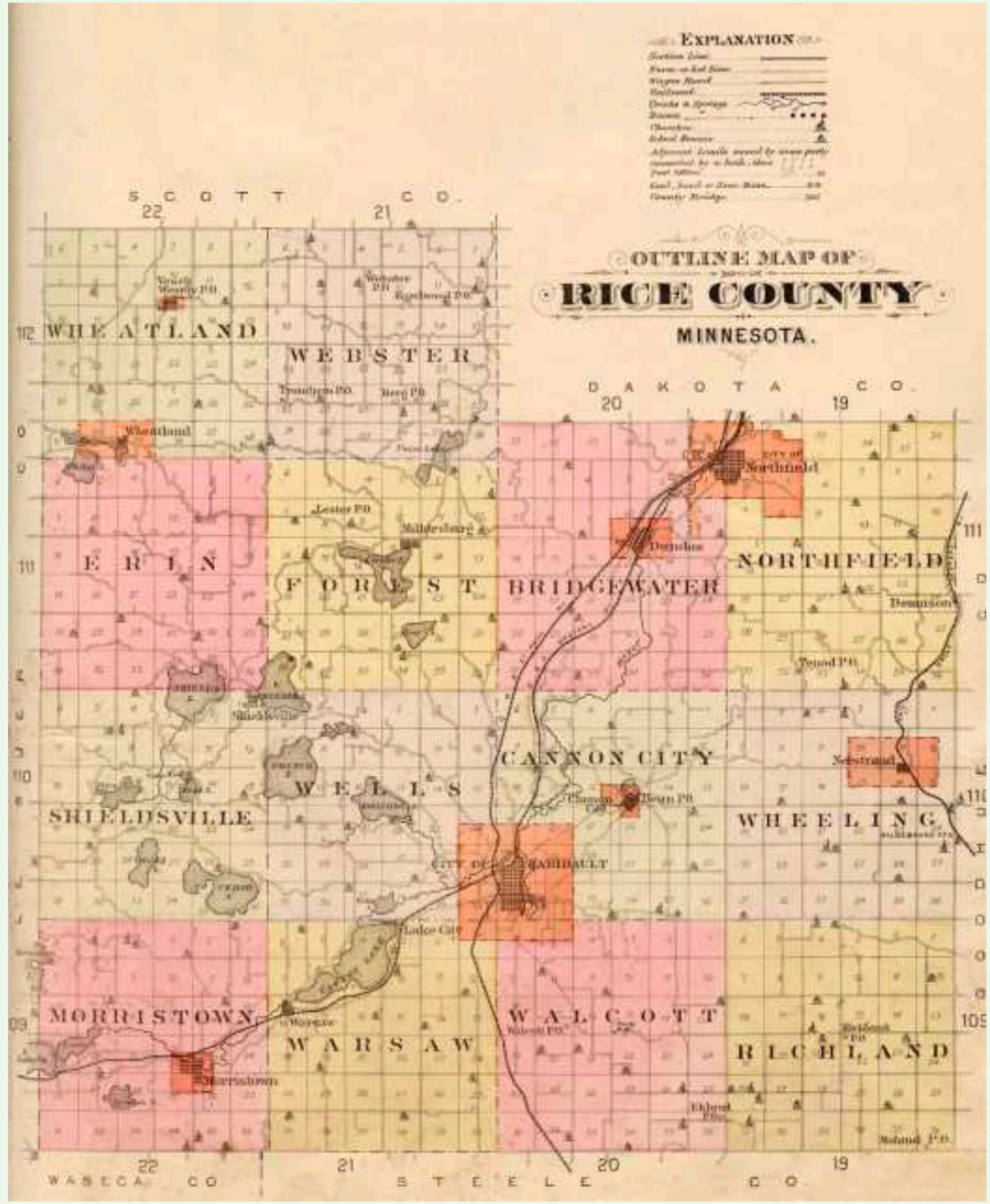
Paper maps

- Insurance maps
- Plat maps
- Historic documents



Paper maps

- Insurance maps
- Plat maps
- Historic documents



Historic documents

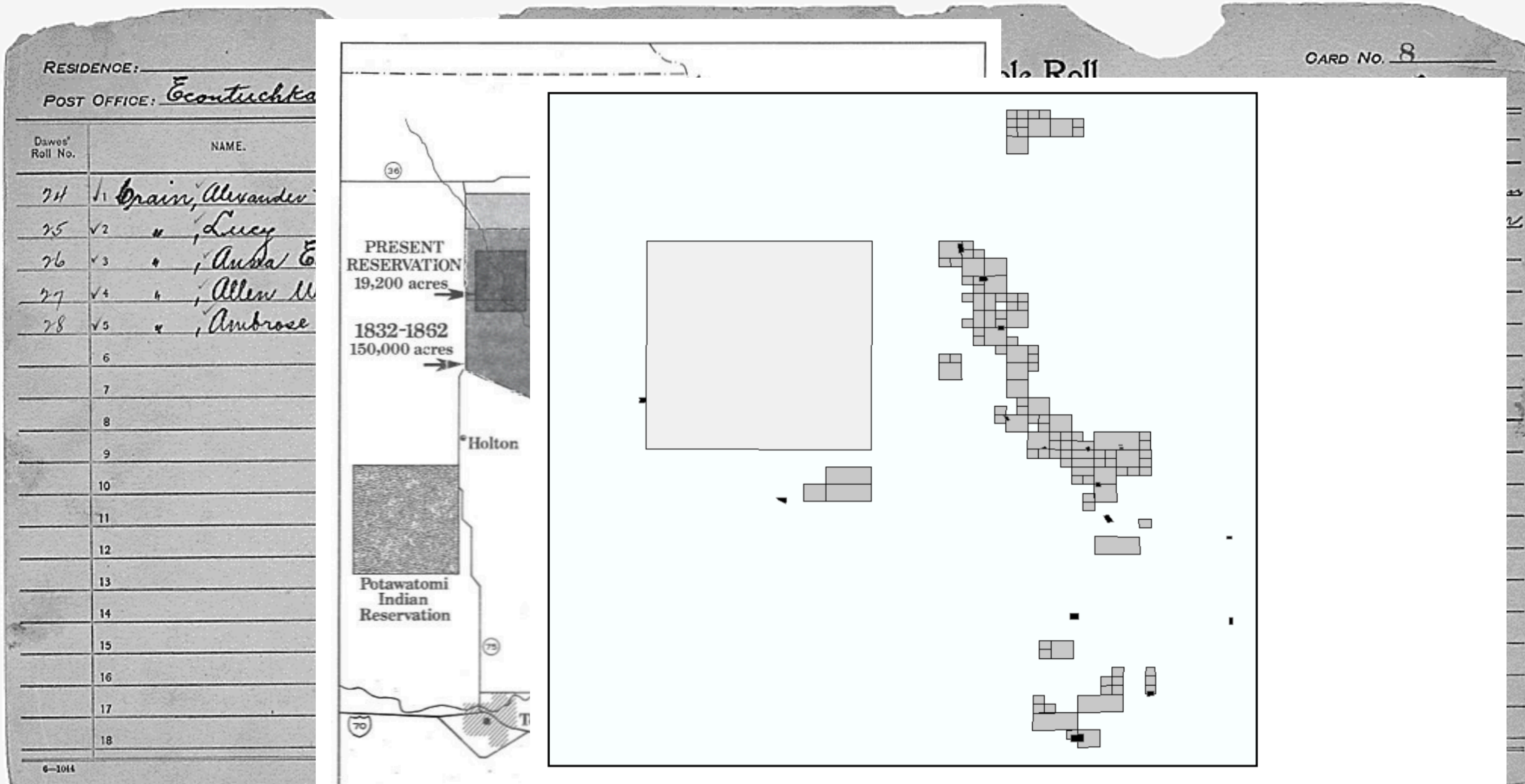
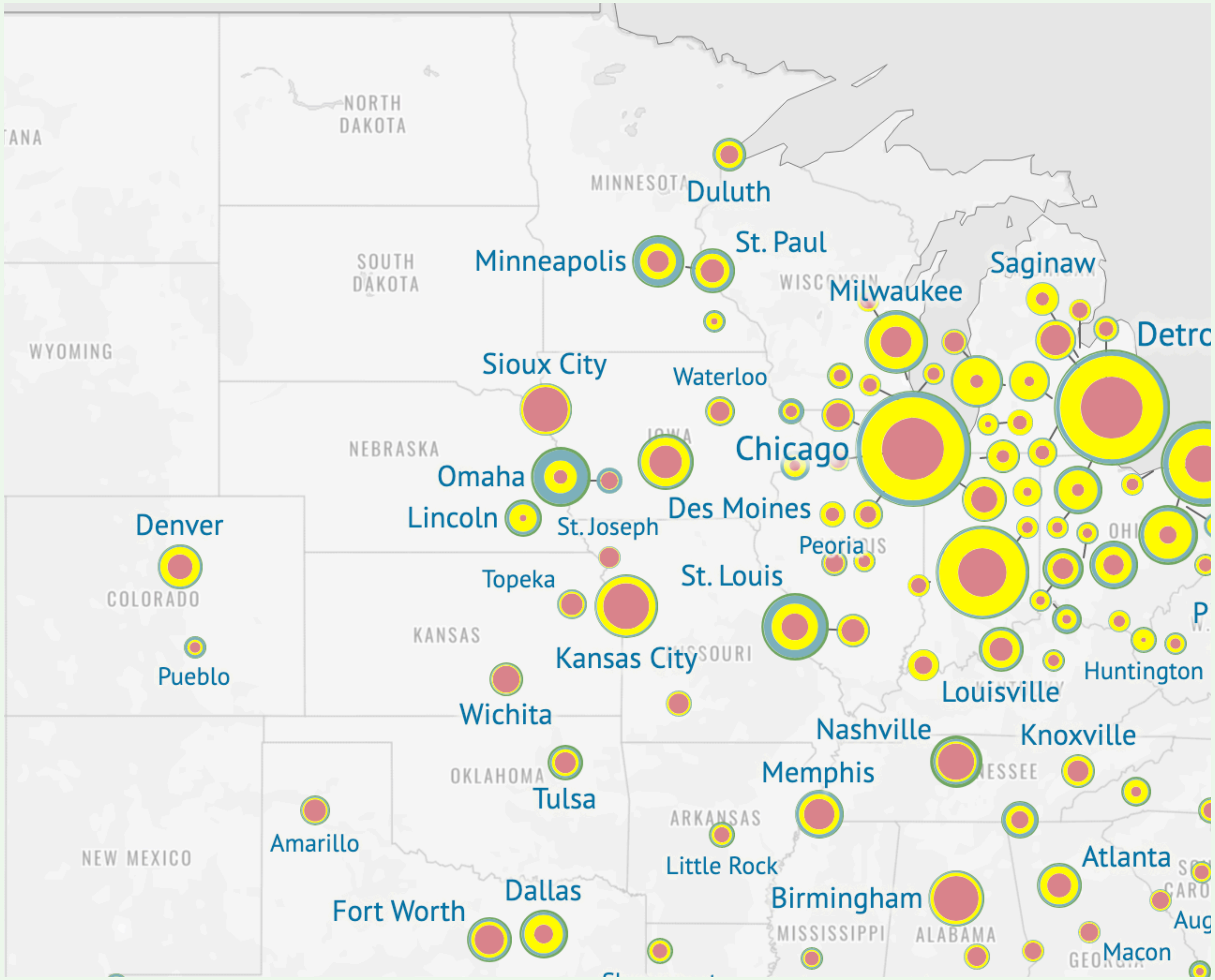


Figure 1. Map of diminishin
p. 72. Original cartography

Figure 3. Lands allotted as a result of the Treaty of 1862. The large square on the left (light gray) is the land allotted in common, while the smaller parcels on the right (mid-gray) are the allotments in severalty. The small black polygons are cultivated fields as mapped in the original land surveys. (Map by authors in ArcGIS.)

Egbert, S., & Smith, P. (2017). Great Frauds and Grievous Wrongs”: Mapping the Loss of Kickapoo Allotment Lands. In Native American Symposium, Representations and Realities; Southeastern Oklahoma State University: Durant, OK, USA.

Historic documents



Historic documents

Historic remote sensing

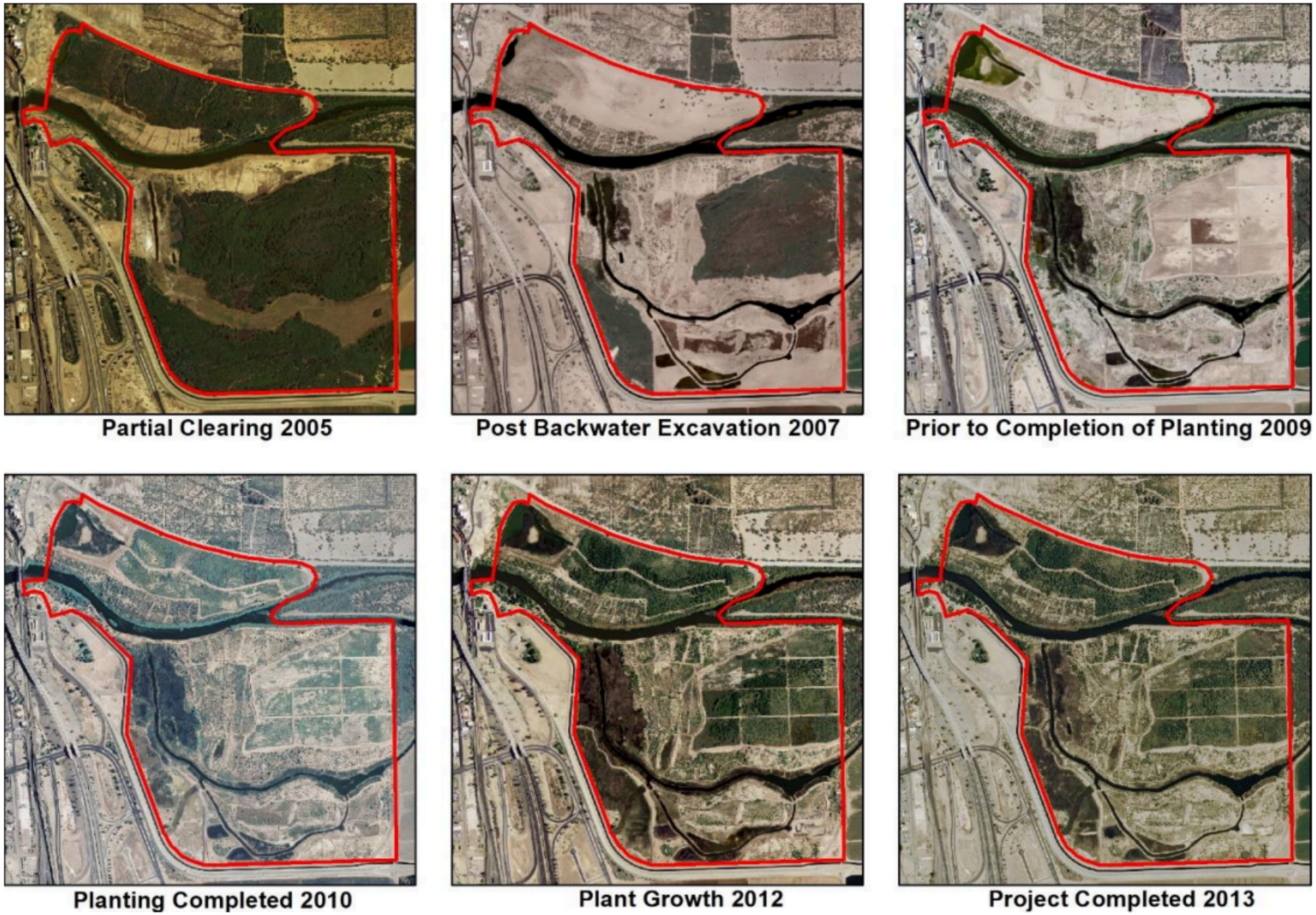
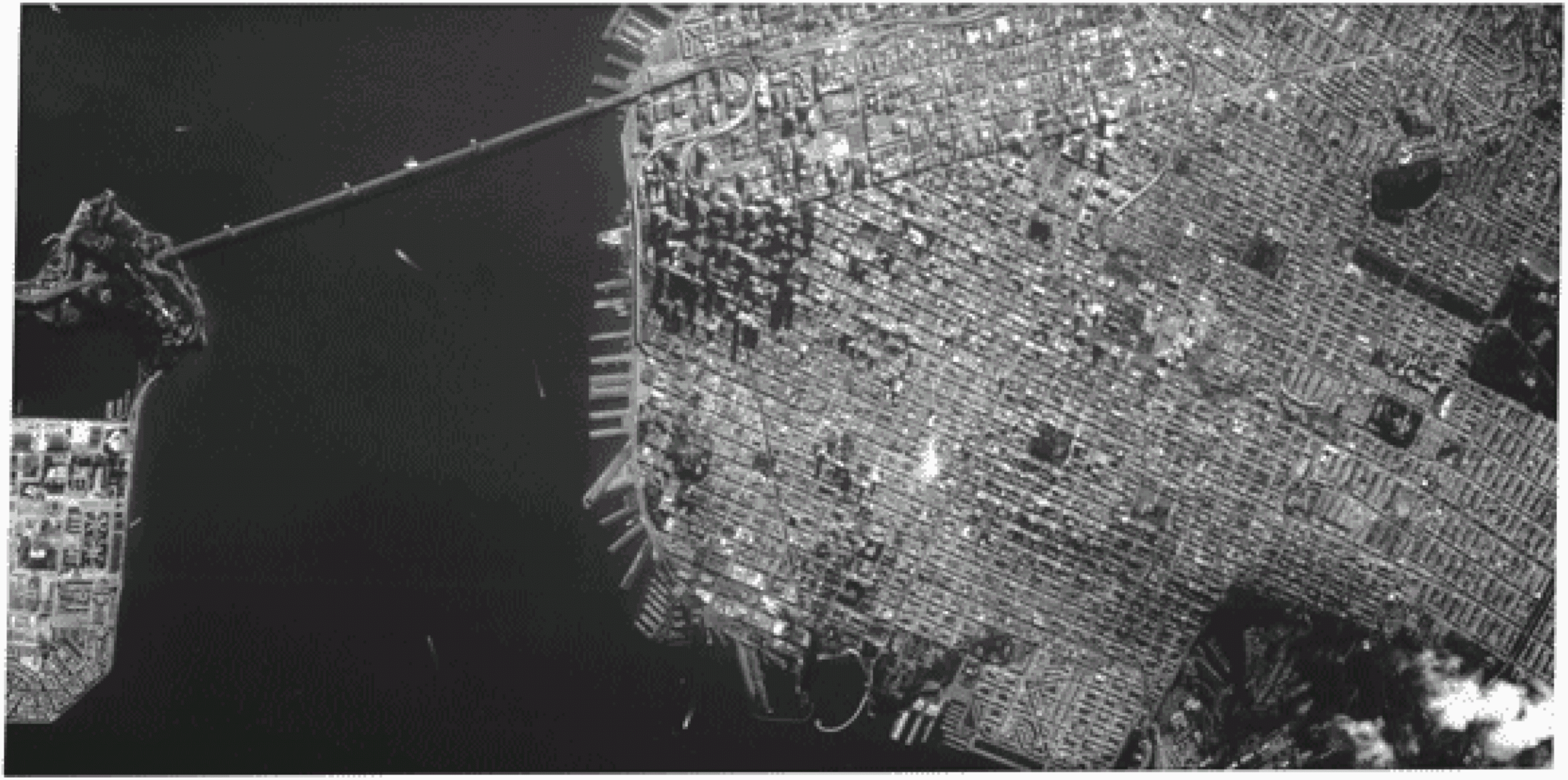


Figure 3. Project Timeline Using Available NAIP Imagery

Historic remote sensing

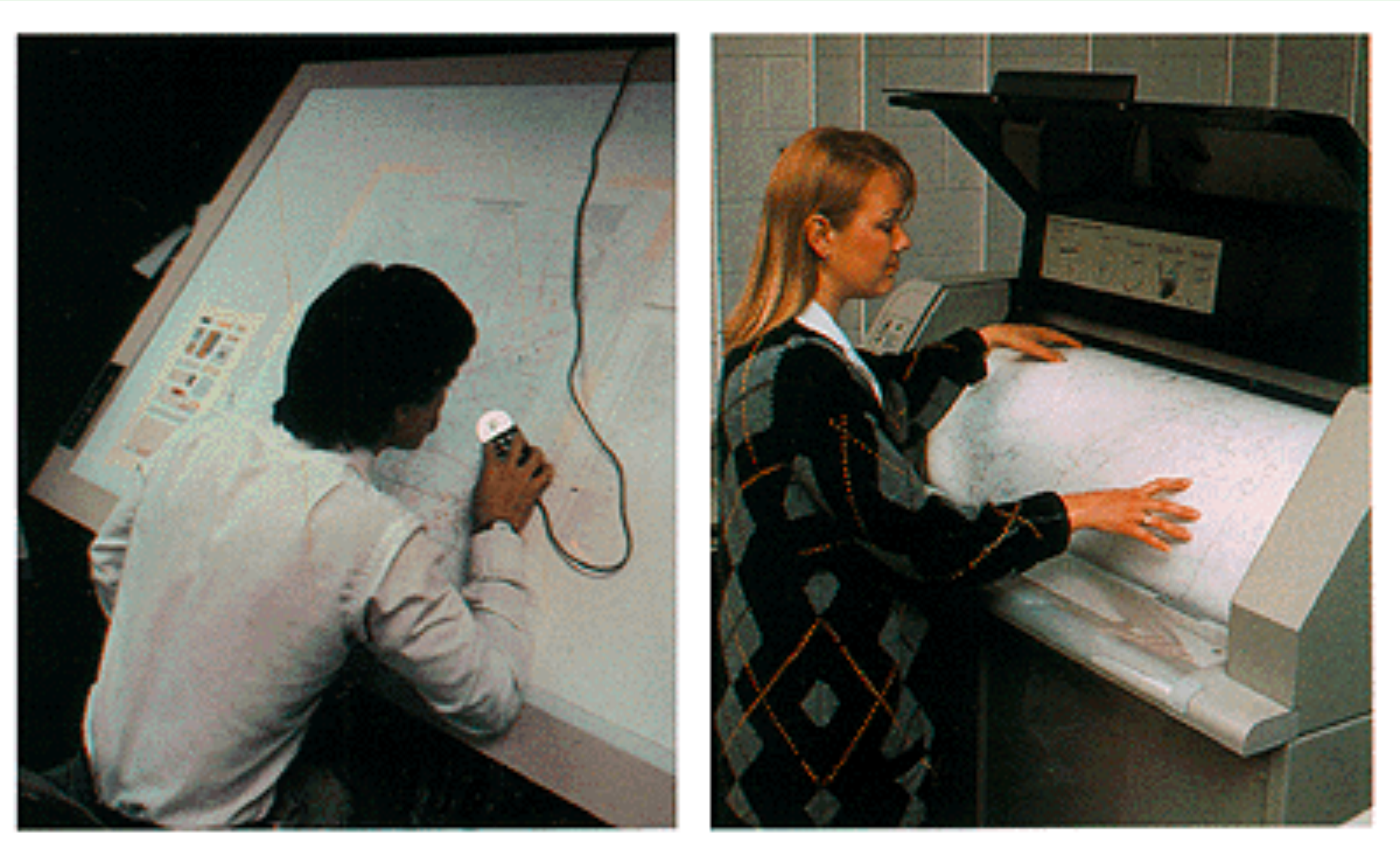


Field Survey

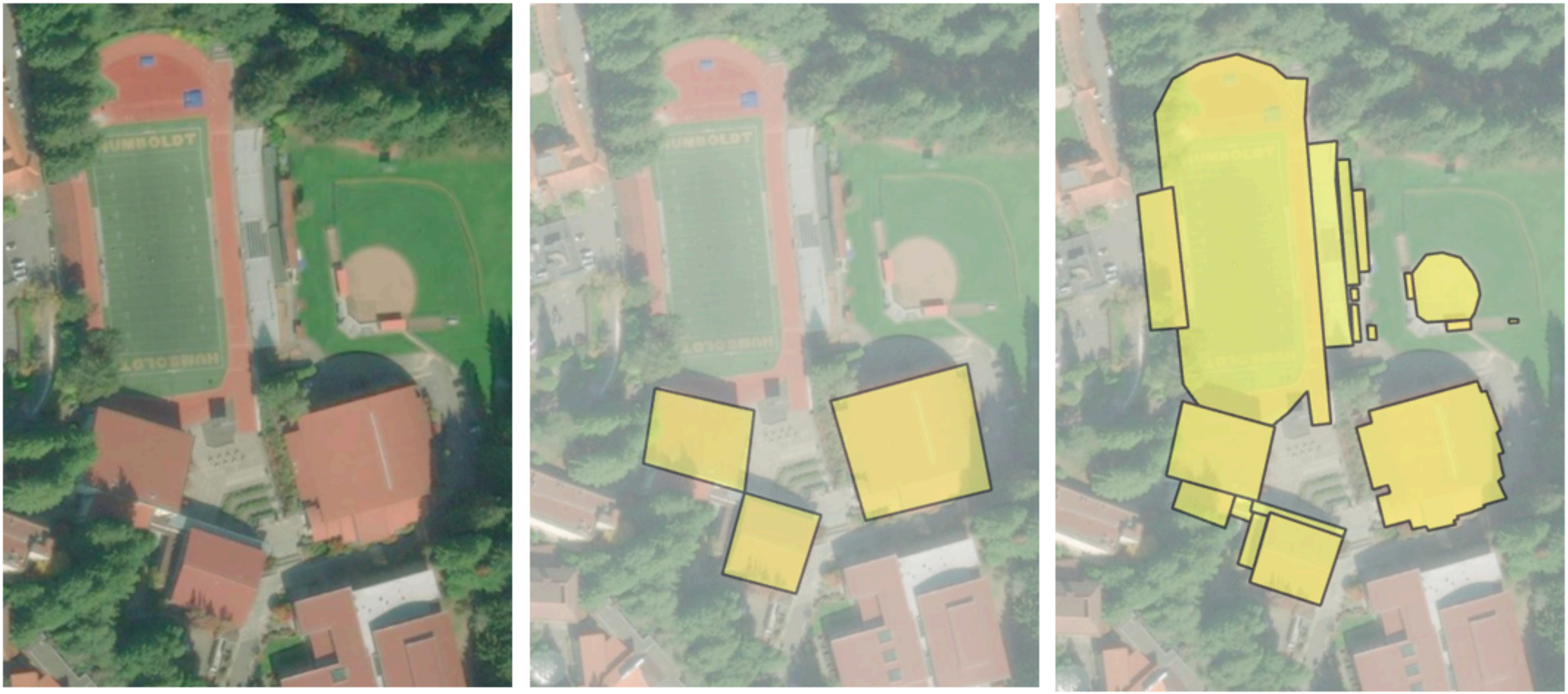
Digitizing

- Creating digital geographical features
  - Convert analogue features to digital features
    - Paper maps
  - Identify and record features
    - From images or recorded field data
- Two approaches
  - Digitizing tablets
  - On-screen digitization

Digitizing with a tablet

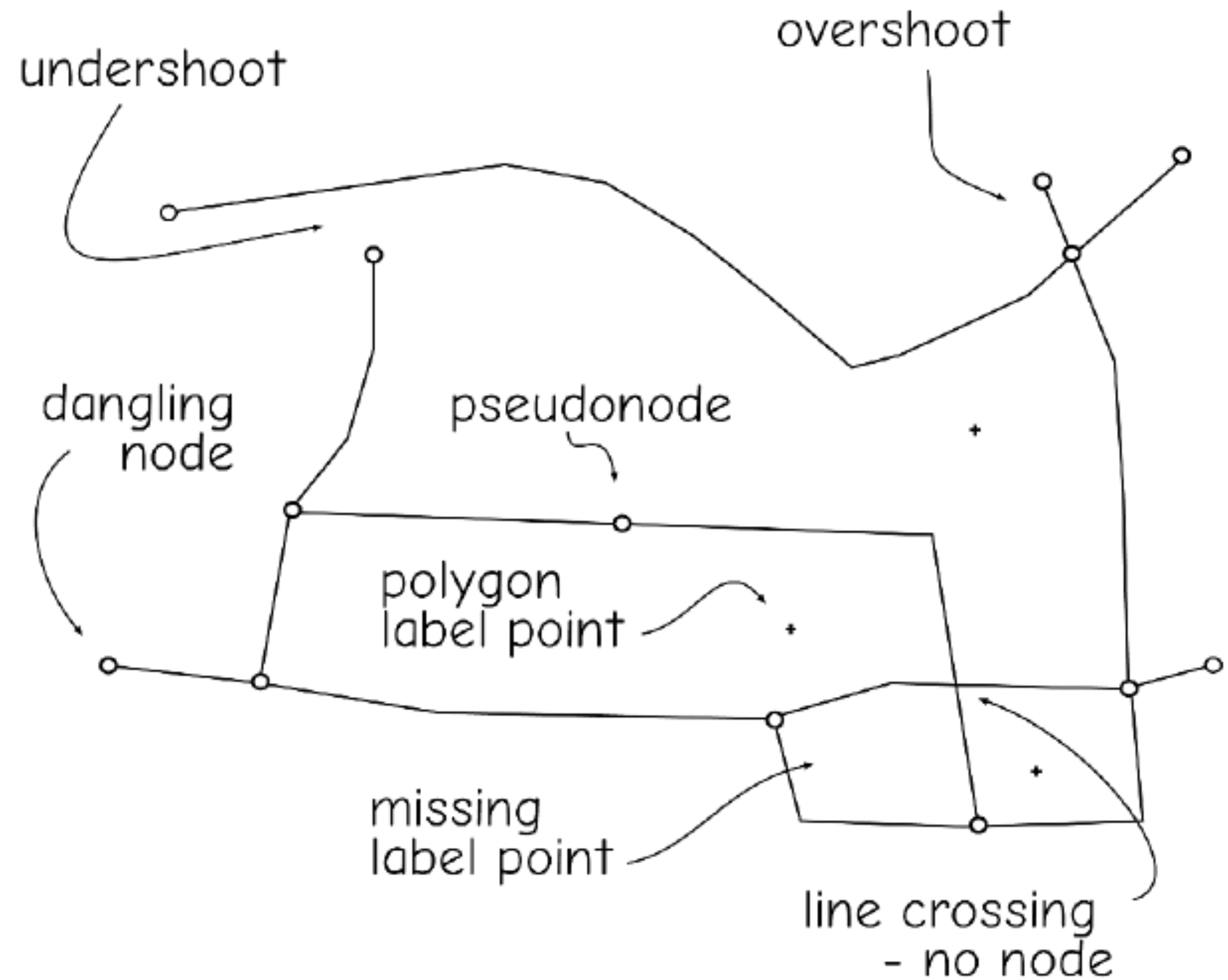


Digitizing onscreen



Common Errors in Manual Digitizing

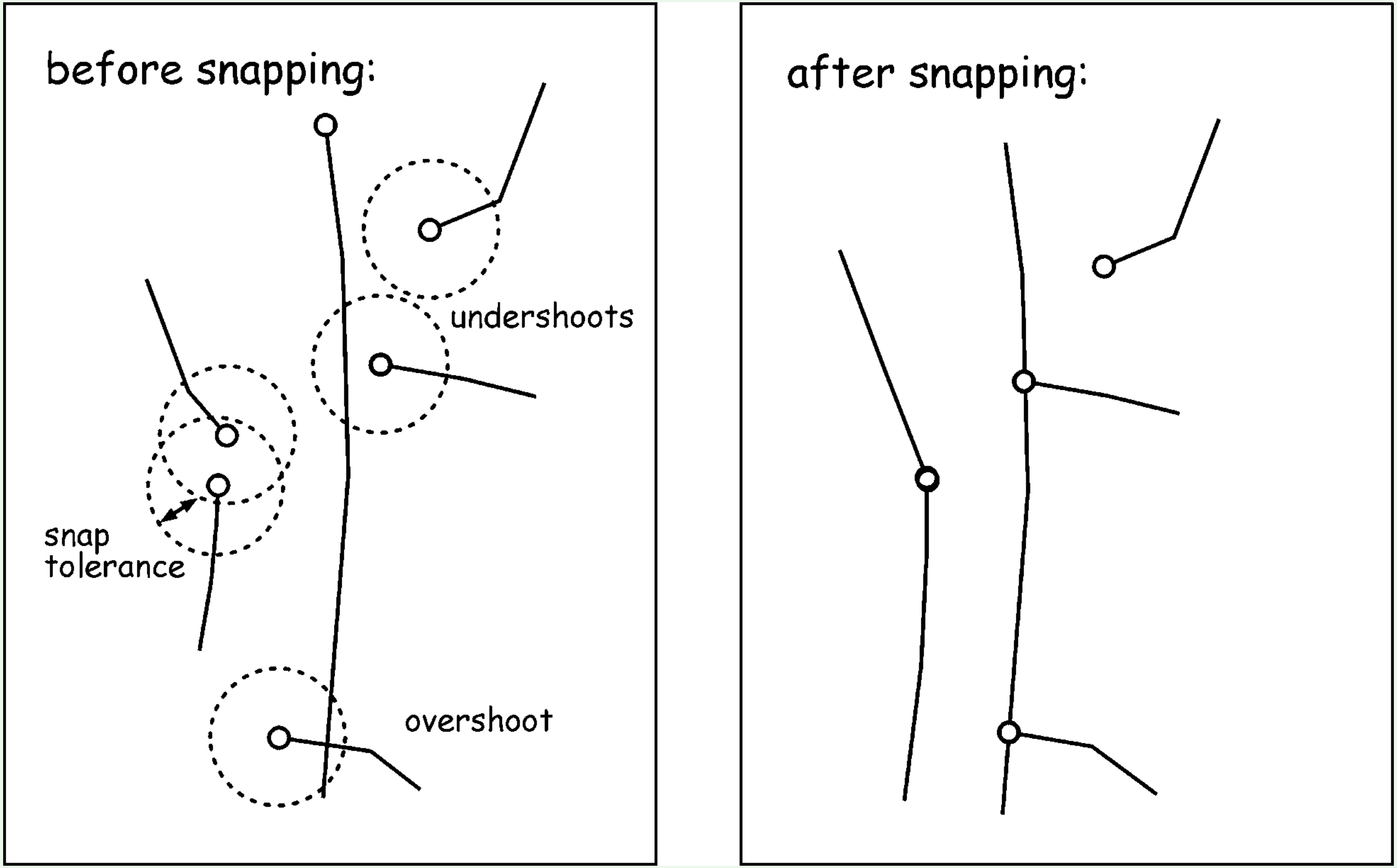
Figure 4-15: Common digitizing errors.



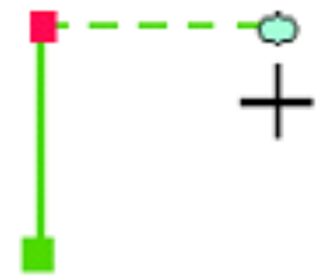
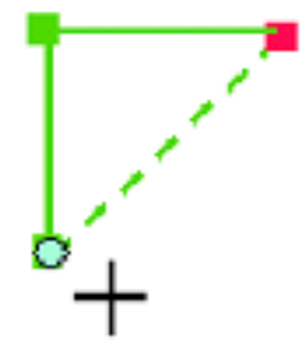
Editing—Removing Errors

- Some errors could be automatically detected - topography
  - Missing intersections
  - Pseudo nodes
  - Dangle lines (lines shorter than a specified dangle distance)
- Interactive editing
  - Line and point locations are adjusted on computer screen
- Snapping function helps reduce some errors
  - Points which fall within a specified distance (snap distance) of each other are snapped together
  - During digitizing or in editing

Snapping


before snapping:
undershoots
snap
tolerance
overshoot
after snapping:

Layer and Sketch Snapping Properties

Layer snapping properties		Sketch snapping properties	
Vertex	Snaps to each vertex of the features in that layer. 	Perpendicular to sketch	Lets you create a segment that will be perpendicular to the previous. 
Edge	Snaps to the entire outline (both segments and vertices) of each feature in that layer. 	Edit sketch vertices	Snaps to the vertices of the sketch. 
Endpoint	Snaps to the first vertex and the last vertex in a line feature. 	Edit sketch edges	Snaps to the entire outline (both segments and vertices) of the sketch.

Line Densify and Thin

- Densify
  - Too few vertices
  - A spline is a set of polynomial functions that join smoothly
- Thin or generalize
  - Too many vertices

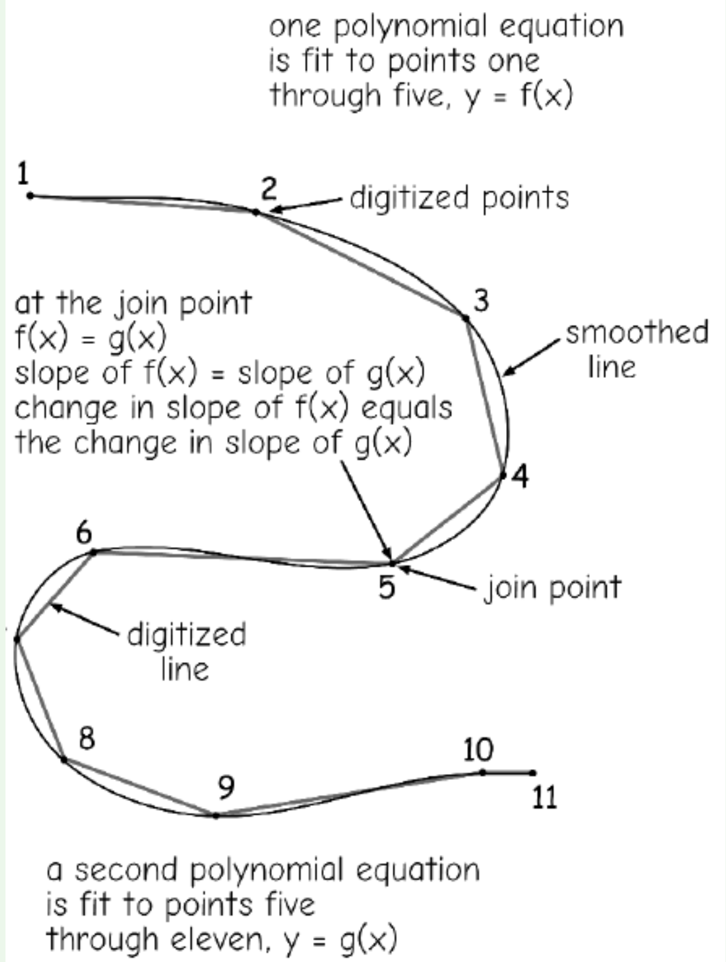


Figure 4-17: Spline interpolation to smooth digitized lines.

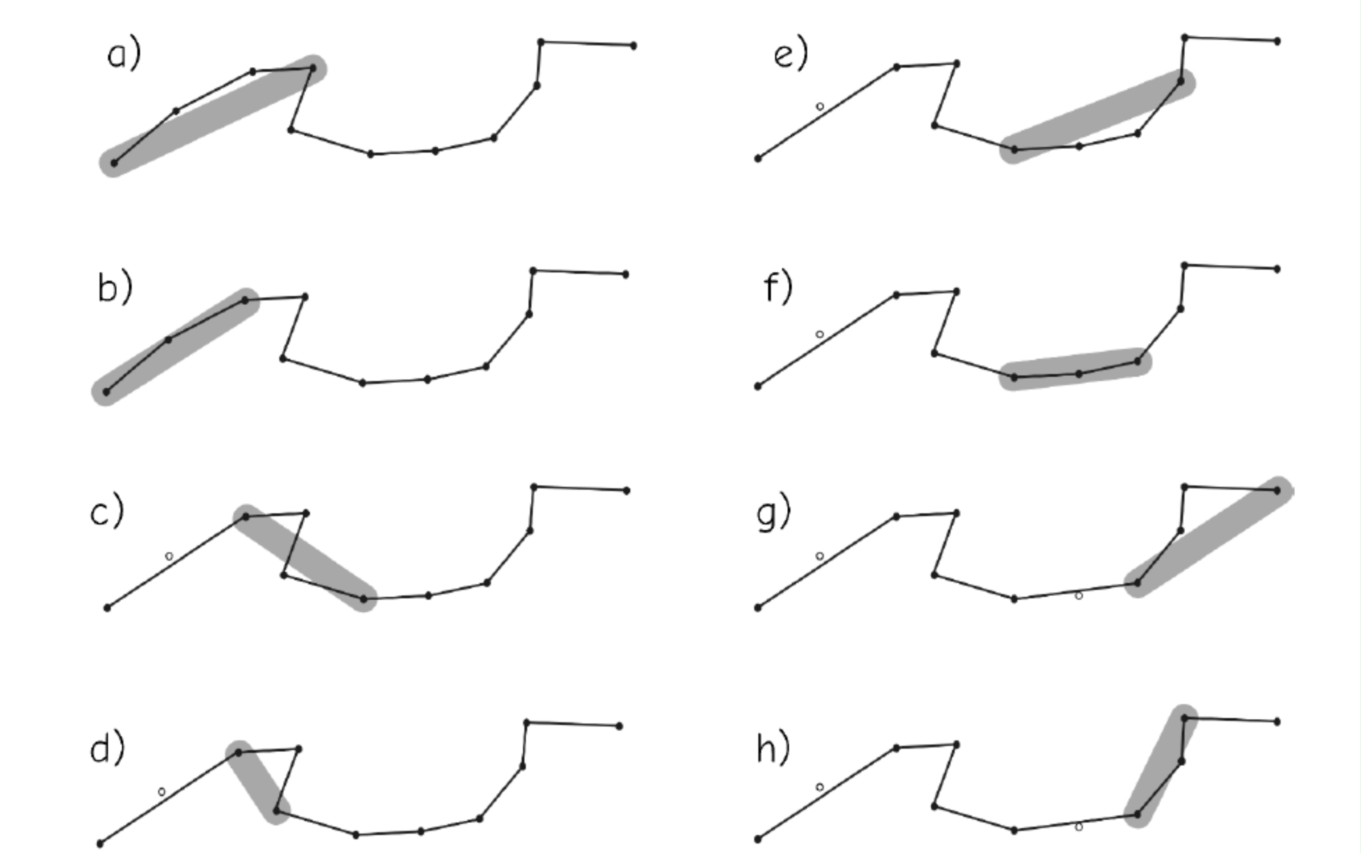
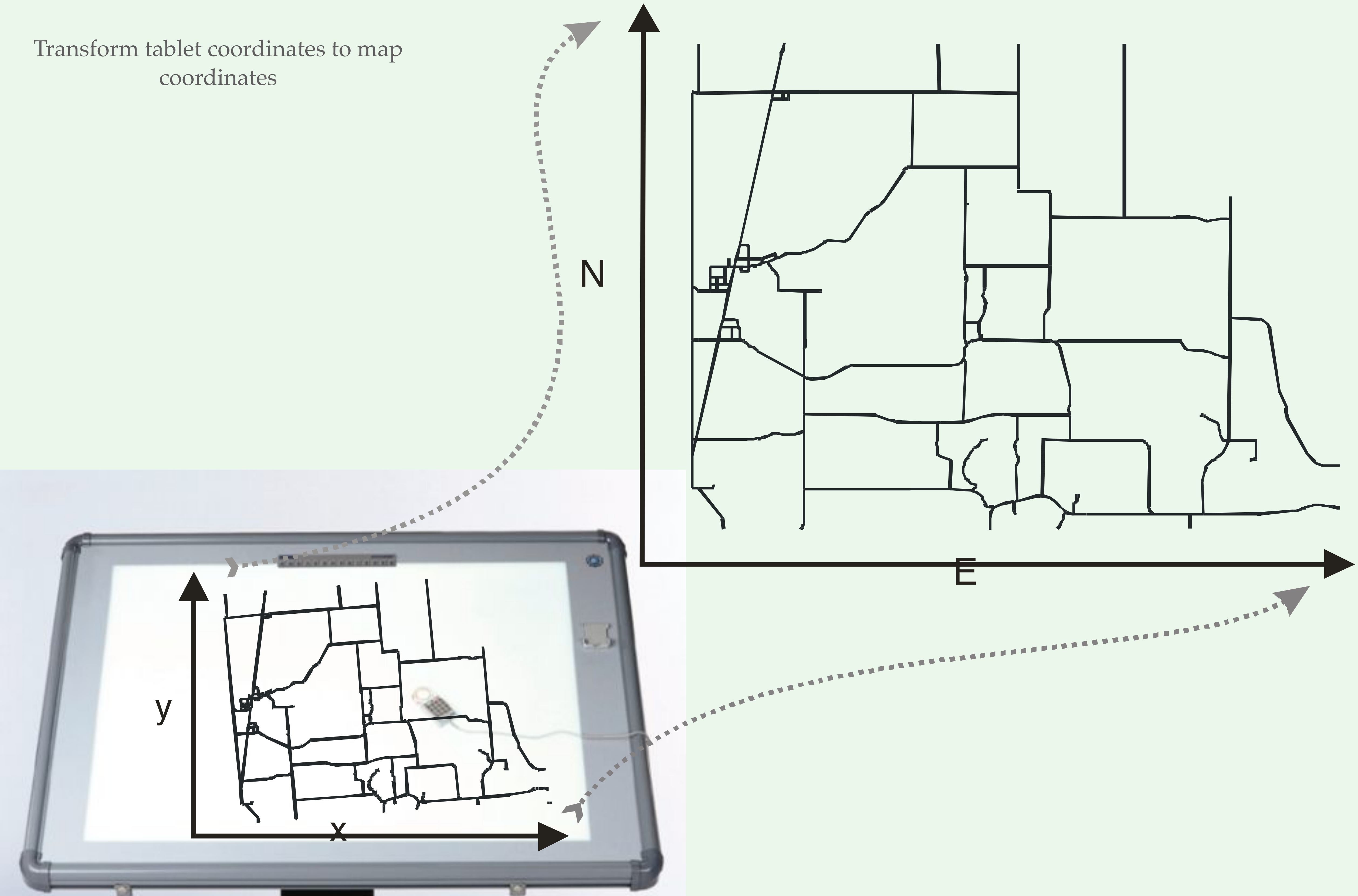


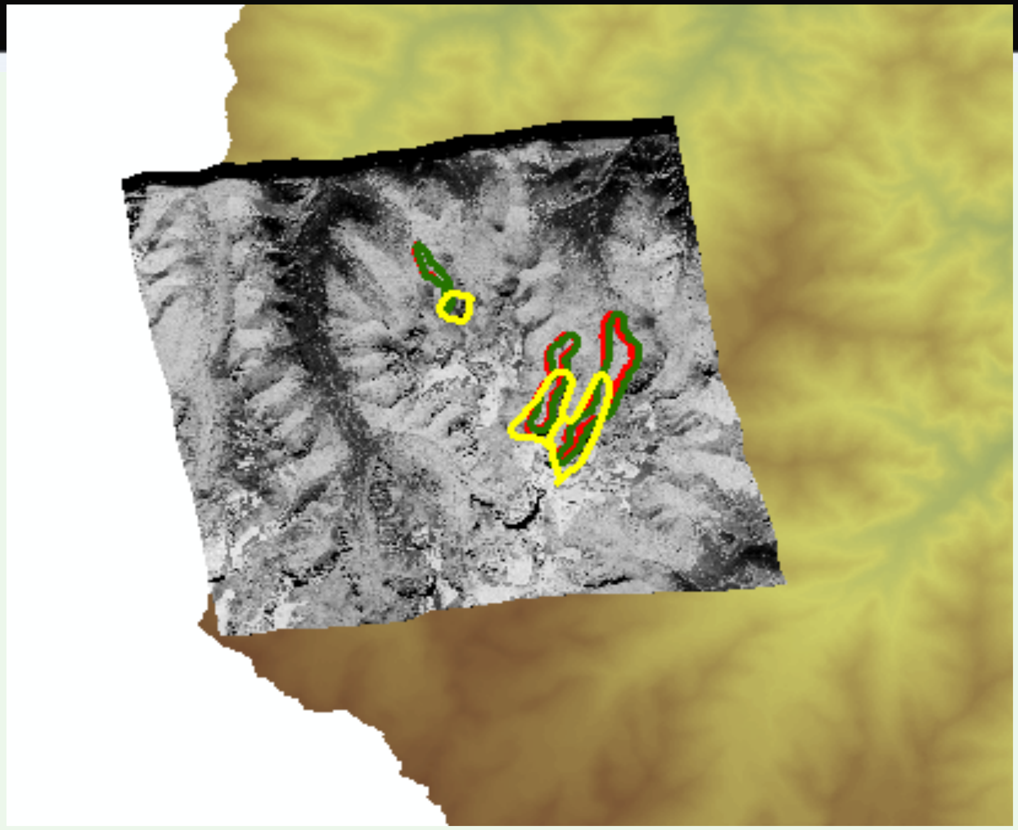
Figure 4-18:The Lang algorithm is a common line-thinning method. In the Lang method, vertices are removed, or thinned, when they are within a weed distance to a spanning line (adapted from Weibel, 1997).

Coordinate Transformation

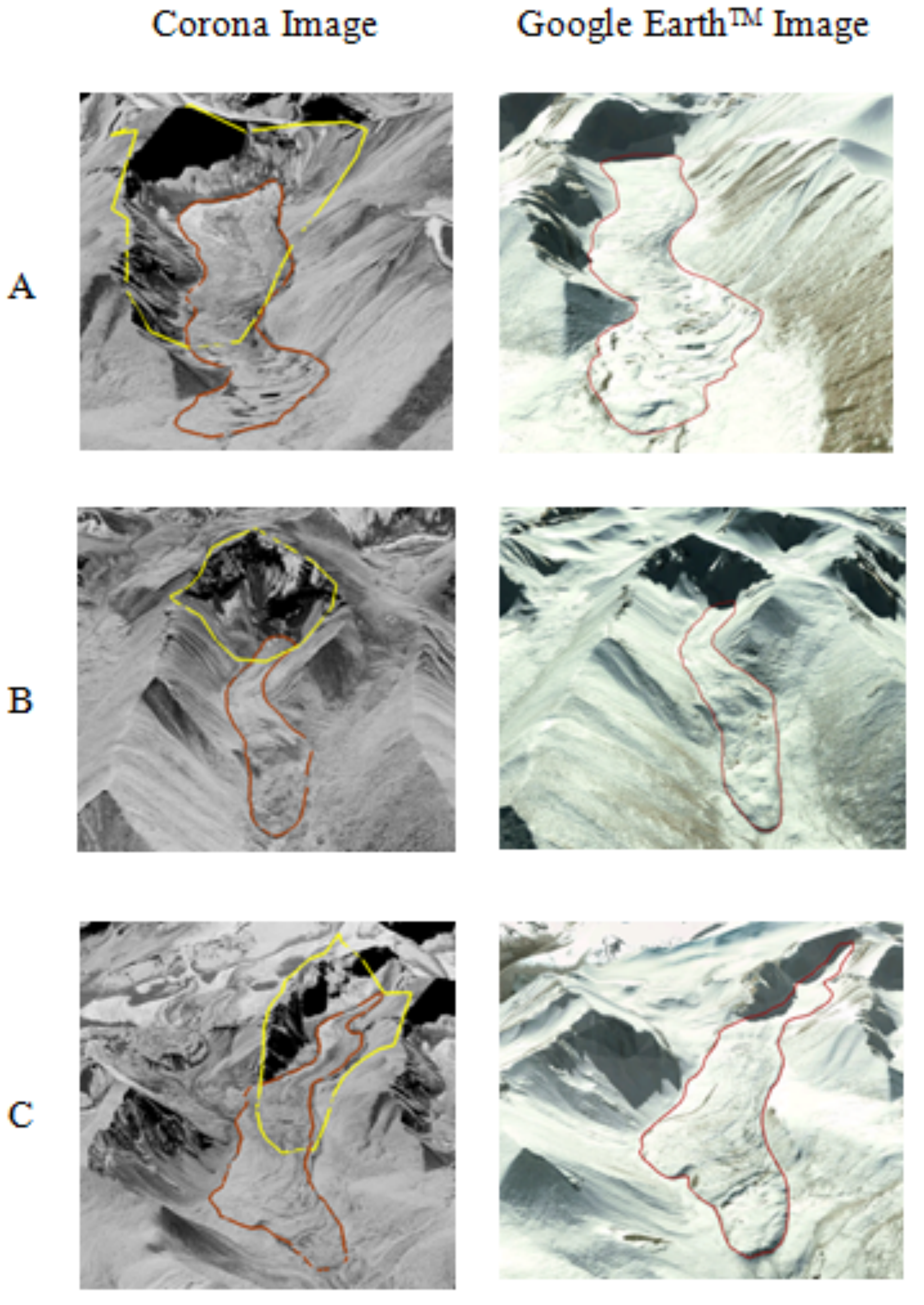
Transform tablet coordinates to map coordinates



Coordinate Transformation



Coordinate Transformation

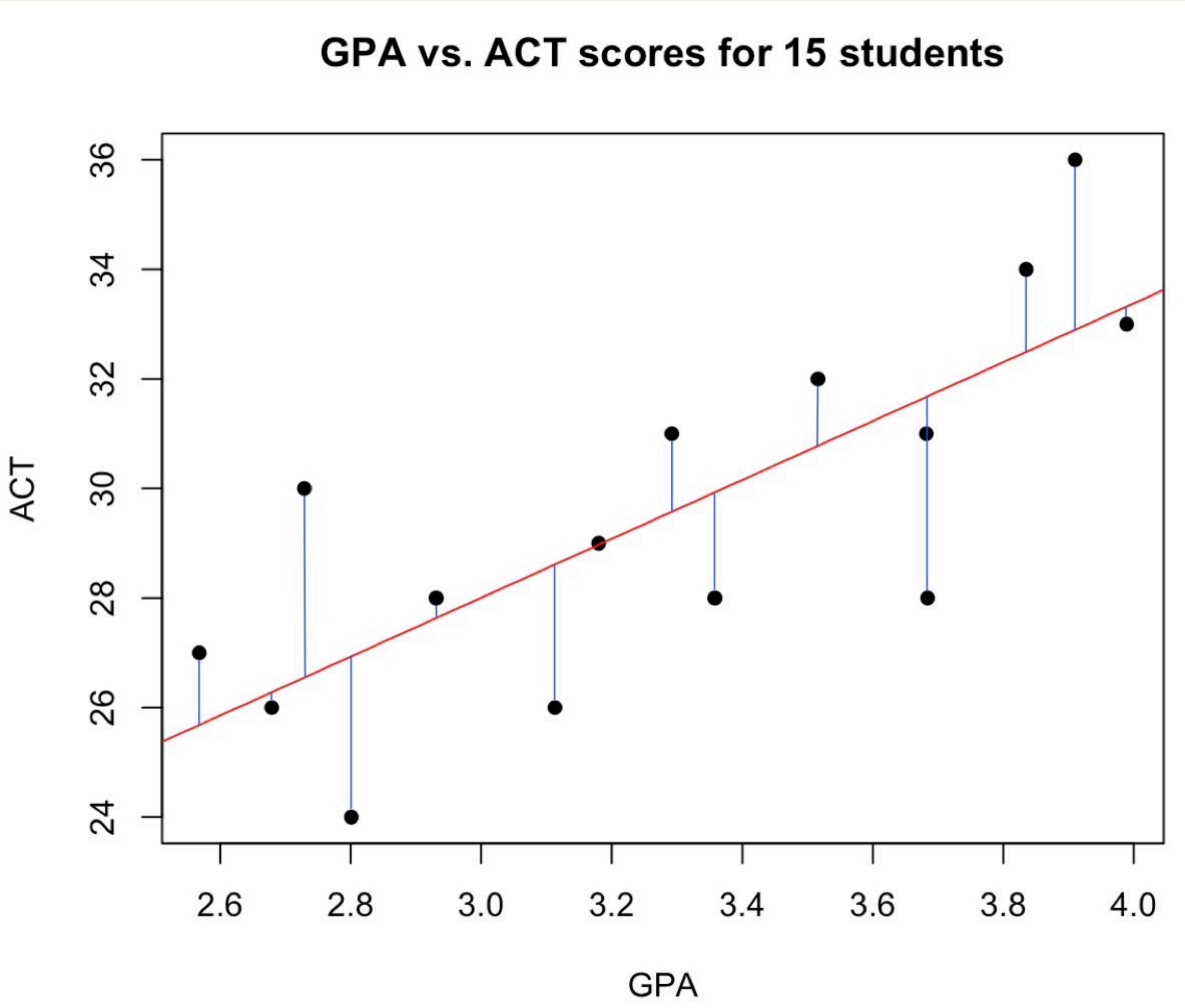


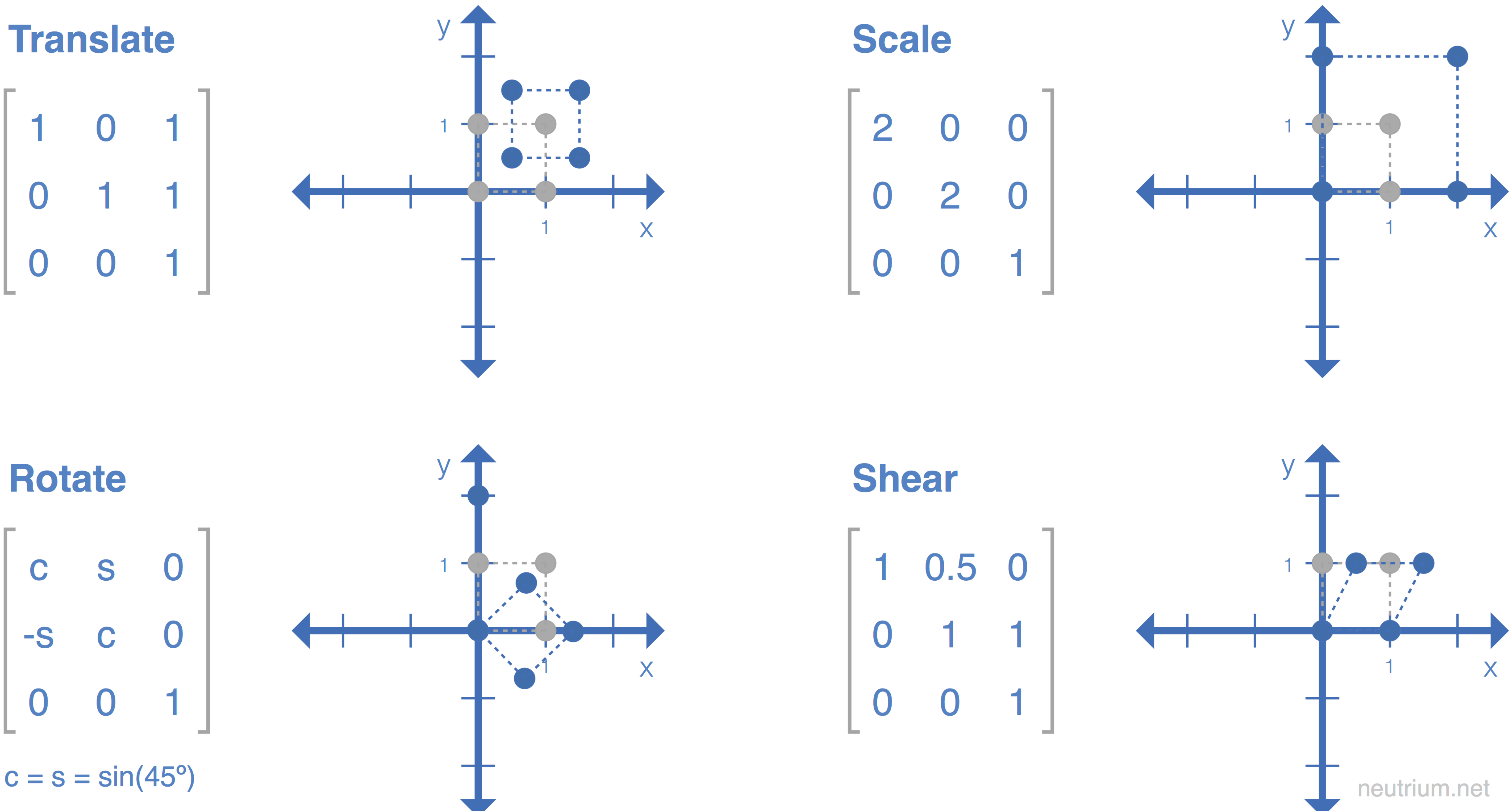
Coordinate Transformation

- Convert tablet or image coordinates to map coordinate systems
- When do you need it?
  - Digitizing maps on a tablet (inches)
  - Digitizing scanned images on screen (pixels)
- Transformation is typically done before digitizing starts
  - Transform the source (usually images) then digitize
- Also called image registration, rectification, or georeferencing when transform images

Coordinate Transformation

- The relationship between the coordinate systems is established using data fitting techniques
  - Similar to the least square method for linear regression on two variables
- Model parameters are obtained through a mathematical model and a set of points (control points) with known coordinates in both systems
- Should not be used to transform map projections or datums

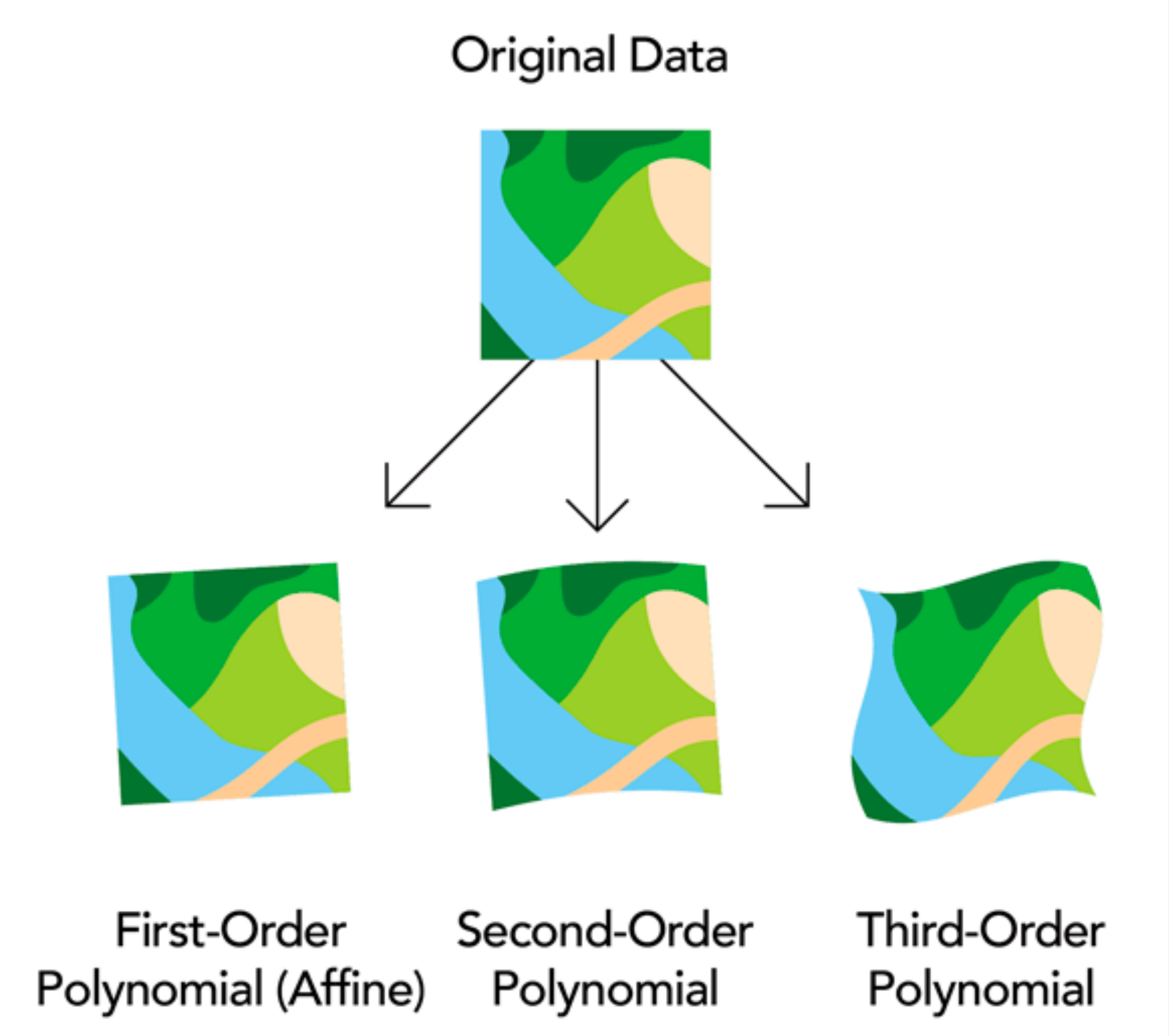




- Affine method is commonly used to transform digitized maps (paper maps or images) to a map coordinate system.
  - Rotation, shift (translation) in x and y, skew, scale in x and y
  - $E = a_0 + a_1x + a_2y$
  - $N = b_0 + b_1x + b_2y$

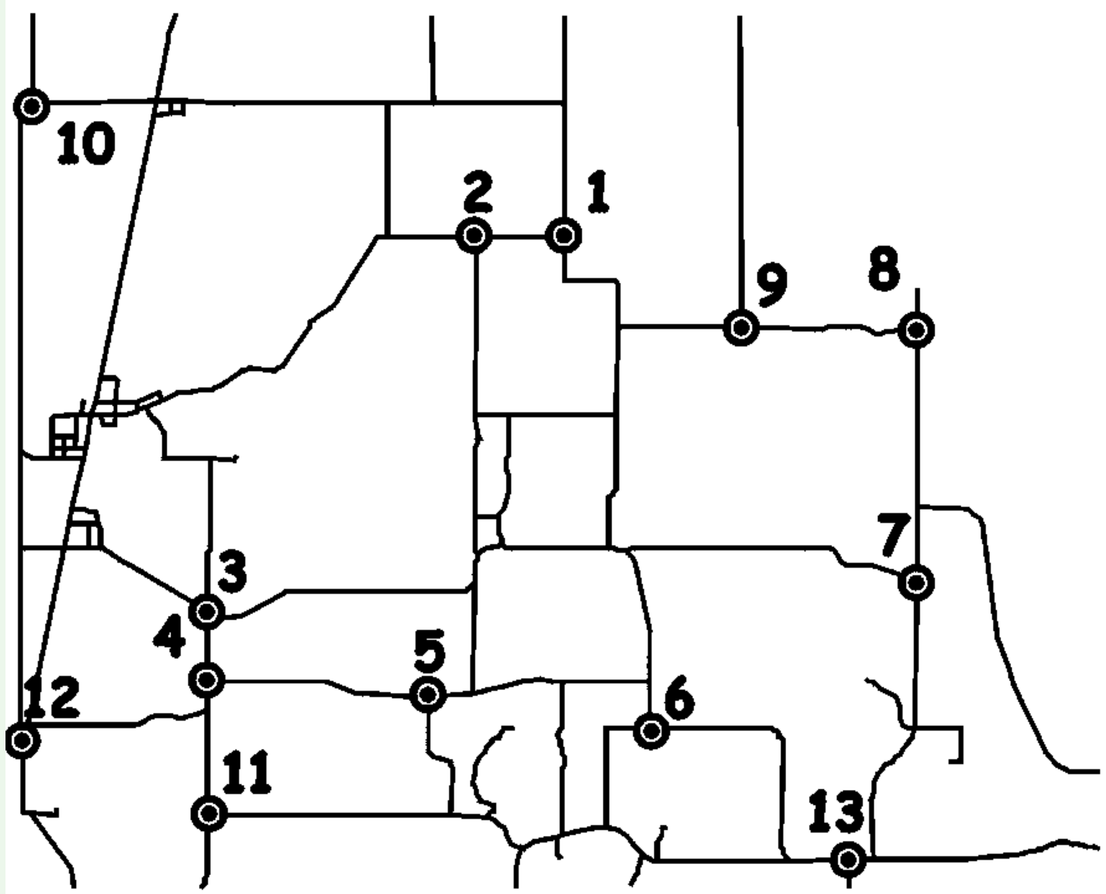
- Coefficients are obtained by control points through the least-square process.
- A minimum number of 3 control points are needed to solve for the 6 unknown parameters

Coordinate transformations



Control Points

Control Points



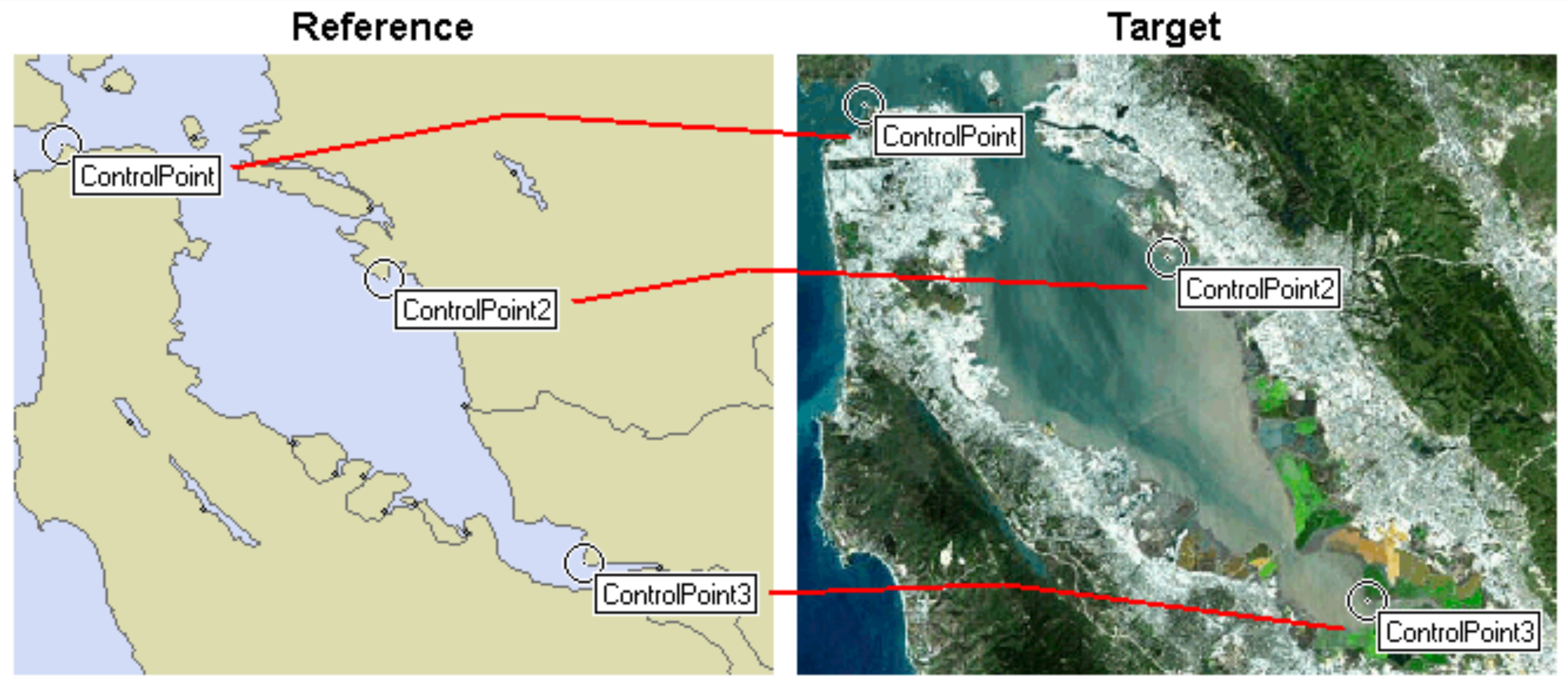
Digitizer Coordinates

ID	x	y
1.0	103.0	-100.1
2.0	0.8	-69.1
3.0	-20.0	-69.0
4.0	-60.0	-47.0
5.0	-102.0	-47.2
6.0	-101.7	10.8
7.0	-86.0	75.8
8.0	-40.0	45.7
9.0	11.0	36.8
10.0	63.0	34.0
11.0	63.0	17.7
12.0	63.0	64.3
13.0	106.0	47.7

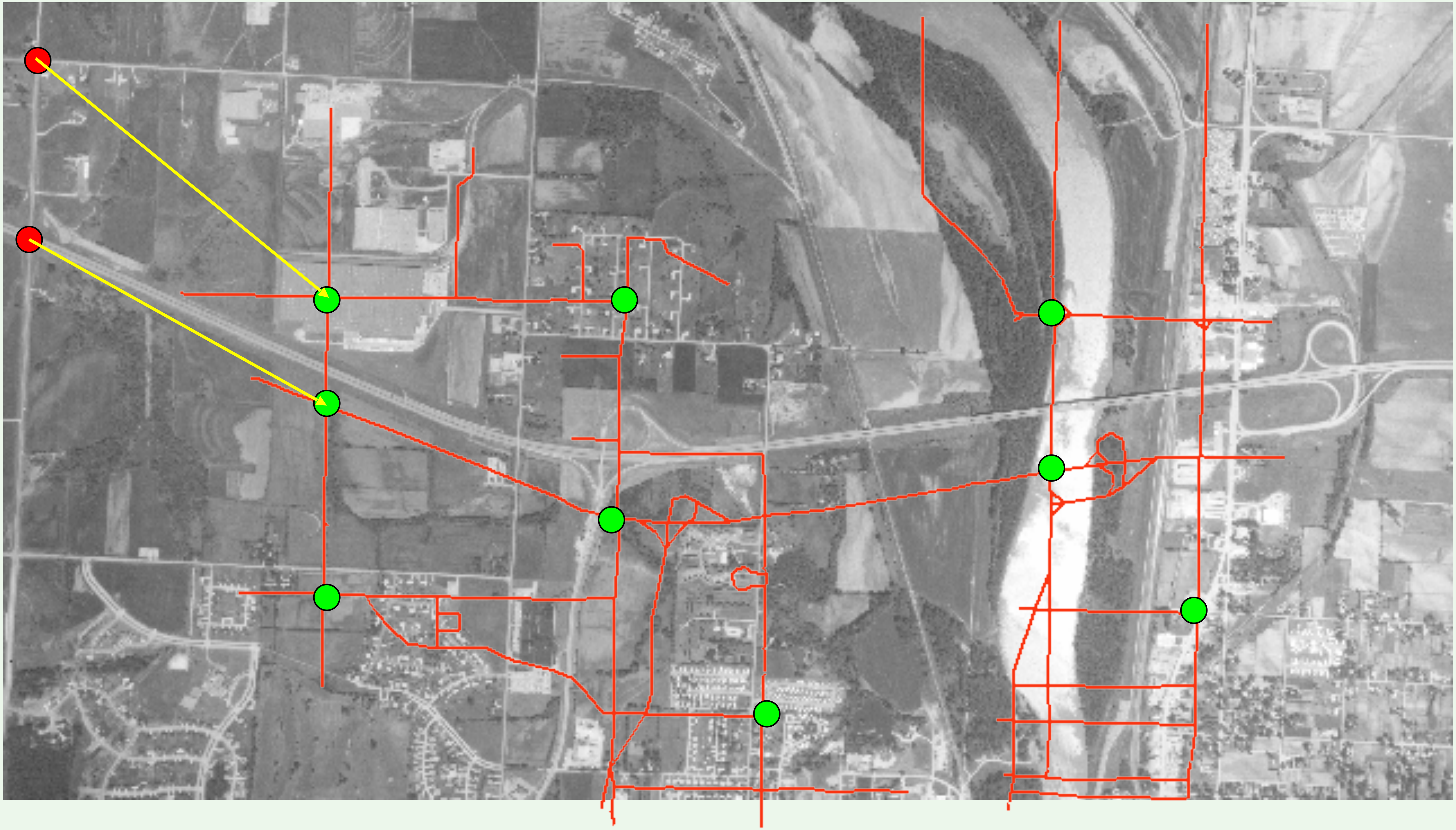
Projection Coordinates (UTM)

E	N
500,083.4	5,003,683.5
504,092.3	5,002,499.5
504,907.5	5,002,499.5
506,493.3	5,001,673.5
508,101.3	5,001,651.0
508,090.1	4,999,384.0
507,475.9	4,996,849.0
505,689.2	4,998,022.0
503,679.2	4,998,368.0
501,657.9	4,998,479.5
501,669.1	4,999,116.0
501,680.3	4,997,296.0
500,005.3	4,997,943.5

Control Points



Collect Control Points from Existing Digital Data



Transformation Errors

- There is always some lack-of-fit in the process
- Reasons
  - Pointing error
  - Blunders
  - Math model
- How to evaluate (quantify) the error?

Root Mean Square Error (RMSE)

- Each control point has an actual and estimated x and y in map coordinates system
- Error (residual) for a control point
  - X_a—actual; X_e--estimated

$$Error = \sqrt{(Xa - Xe)^2 + (Ya - Ye)^2} = e_i$$

- RMSE--average error of the transformation

$$\sqrt{\frac{\sum e_i^2}{N}}$$

- Possible reasons having a high RMSE
  - inaccurate control points
  - inappropriate math model

Georeferencing--an Iteration Process

Model Fit 1:

1	518,687.6	5,015,347.0	513,734.1	5,007,087.4	3.07
2	516,907.3	5,013,549.1	511,355.8	5,004,707.2	8.13
3	516,952.2	5,017,965.3	511,438.3	5,010,573.9	4.38

E = 1.:
+ 0.
- 2(

N = -
+
-

RMSI

Exam
& 1
blur
moc

Model Fit 2:

1	518,687.6	5,015,347.0	513,734.1	5,007,087.4	2.56
2	516,907.3	5,013,549.1	511,355.8	5,004,707.2	7.22
3	516,952.2	5,017,965.3	511,438.3	5,010,573.9	3.21

E =

N =

RM

Exc
&
bl
m

Model Fit 3:

E = 1.33118637 * x
+ 0.0056629 * y
- 205490.3

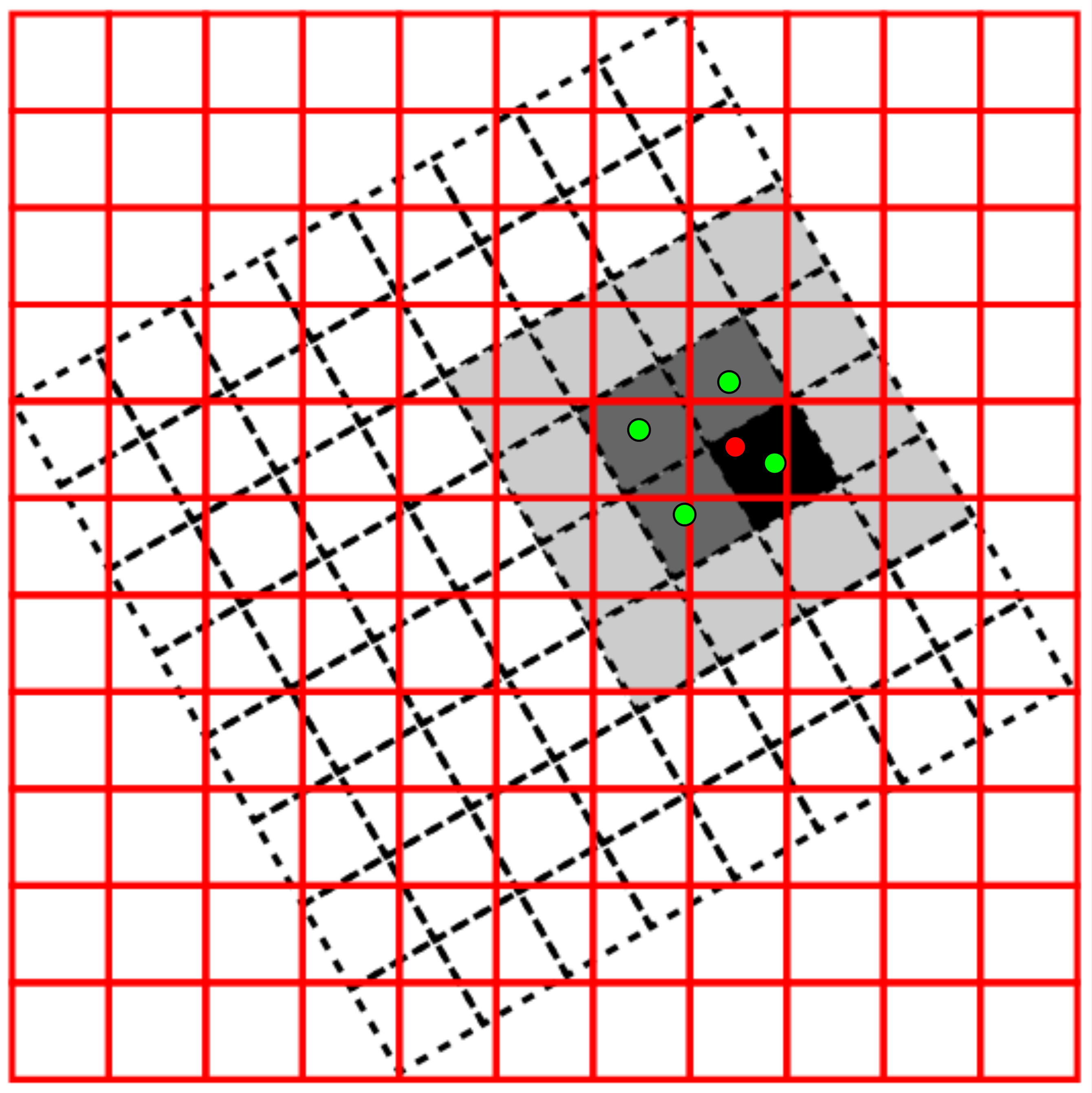
N = - 0.003516 * x
+ 1.3297296 * y
- 1160143

RMSE = 6.78

Examine points, no more blunders found.

1	518,687.6	5,015,347.0	513,734.1	5,007,087.4	2.48
2	516,907.3	5,013,549.1	511,355.8	5,004,707.2	5.63
3	516,952.2	5,017,965.3	511,438.3	5,010,573.9	3.84
4	518,699.6	5,014,396.8	513,739.3	5,005,831.0	6.62
5	518,099.6	5,013,576.2	512,938.9	5,004,733.6	2.71
6	518,992.6	5,017,306.0	514,144.0	5,009,699.3	8.40
7	519,150.0	5,013,556.6	514,331.9	5,004,709.3	6.55
8	519,259.8	5,013,600.0	514,482.8	5,004,764.0	1.52
9	516,916.8	5,016,528.9	511,378.9	5,008,669.6	3.30
10	516,659.6	5,018,093.8	511,043.8	5,010,744.1	5.27
11	519,474.3	5,018,046.9	514,807.0	5,010,675.2	11.78
12	519,549.2	5,014,375.9	514,873.0	5,005,798.0	3.54
13	518,089.4	5,014,478.2	512,938.6	5,005,931.0	9.34
14	518,087.4	5,014,755.2	512,936.0	5,006,299.0	8.30
15	518,079.1	5,016,483.3	512,921.1	5,008,596.5	5.67
16	516,947.5	5,017,736.1	511,424.5	5,010,277.6	6.73
17	517,015.8	5,014,443.1	511,495.1	5,005,894.9	3.30
18	517,785.1	5,017,492.6	512,542.4	5,009,954.0	8.86
19	519,435.7	5,017,340.7	514,736.0	5,009,735.7	4.13
20	518,710.3	5,016,544.2	513,778.7	5,008,679.7	9.55
21	518,984.0	5,016,548.6	514,127.8	5,008,678.2	9.53
22	516,717.8	5,014,546.4	511,106.4	5,006,028.8	9.01

Resampling (Creating a new image)

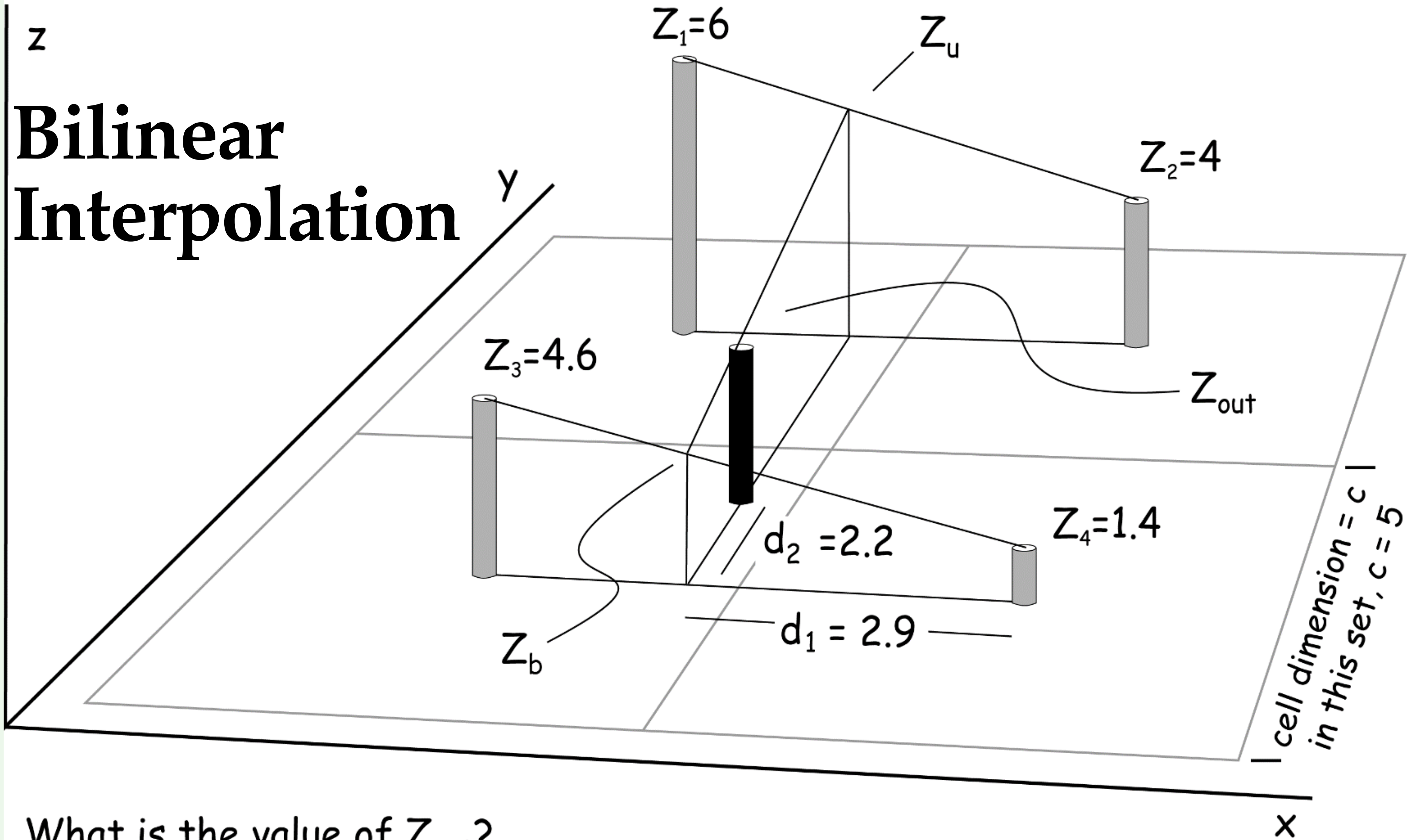


nearest neighbor assigns an output cell value from the nearest corresponding input cell

bilinear interpolation uses the nearest cell and next three closest cells in a weighted average

cubic-convolution includes the cells used for bilinear interpolation and the next 12 closest cells to compute a weighted average for each output cell

Bilinear Interpolation



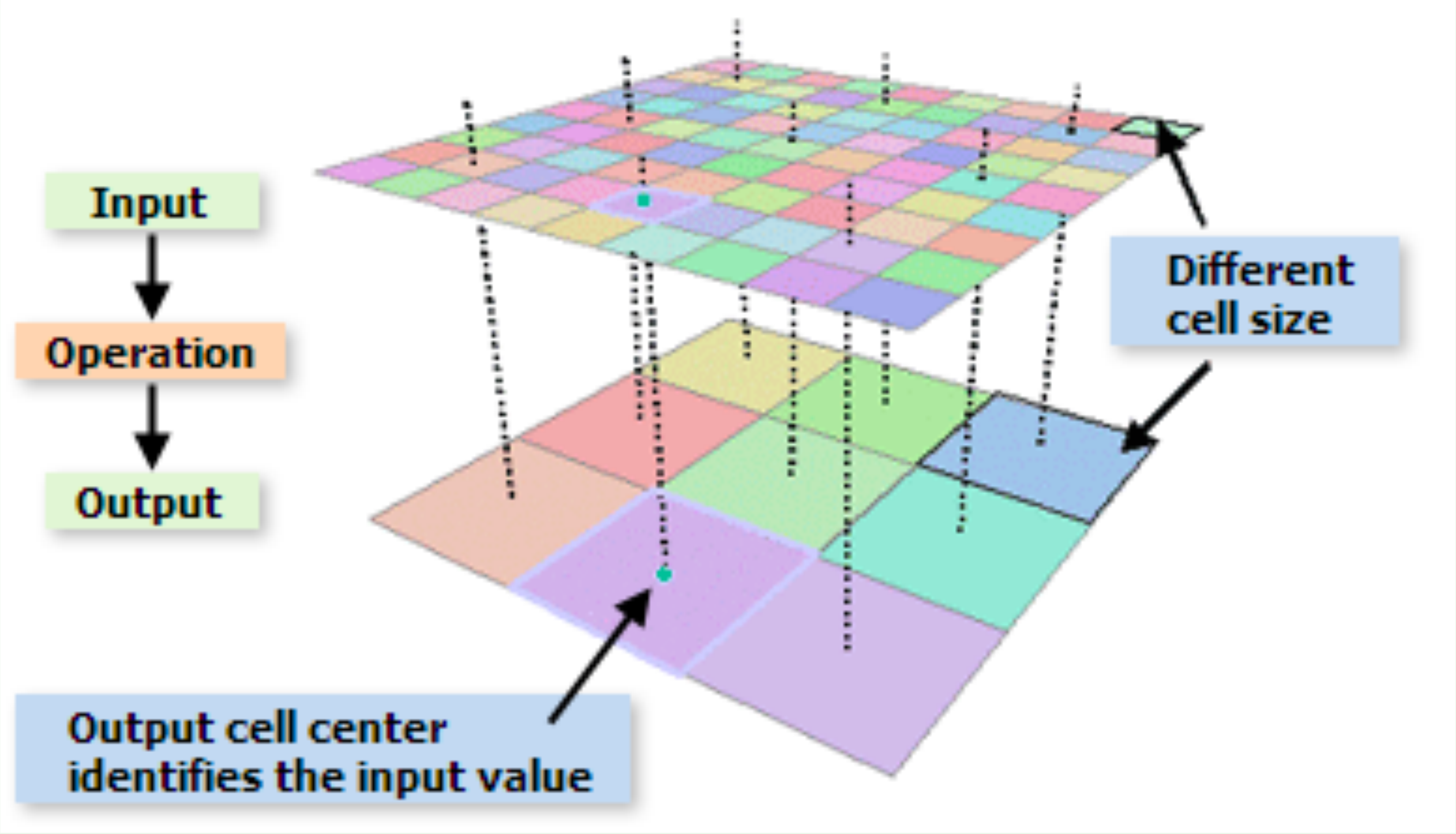
What is the value of Z_{out}?

$$Z_b = Z_4 + \frac{(Z_3 - Z_4)}{c} * d_1 \qquad Z_b = 1.4 + \frac{(4.6 - 1.4)}{5} * 2.9 = 3.26$$

$$Z_u = Z_2 + \frac{(Z_1 - Z_2)}{c} * d_1 \qquad Z_u = 4 + \frac{(6 - 4)}{5} * 2.9 = 5.16$$

$$Z_{out} = Z_b + \frac{(Z_u - Z_b)}{c} * d_2 \qquad Z_{out} = 3.26 + \frac{(5.16 - 3.26)}{5} * 2.2 = 4.1$$

Upscaling


Input
Operation
Output
Different
cell size
Output cell center
identifies the input value