GE2215 Lecture 7
Spatial Data Quality and Spatial Data Editing

Dr. Yan Yingwei

Department of Geography

National University of Singapore

Recap: Why geometric transformation?

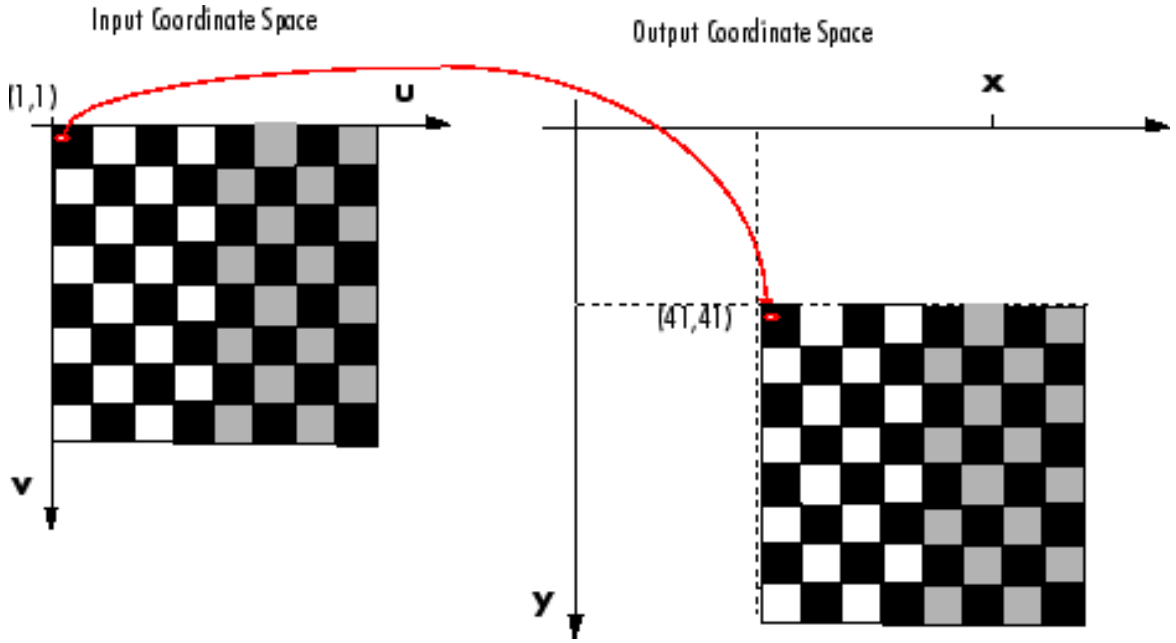
- Geometric transformation can be used to:
  - Assign **coordinate systems** to **images** and digitized **maps**
    - **Scanned paper maps**
    - **Remote sensing images**
  - Correct **random** **(non-systematic)** distortion (deformation)

Recap: Types of geometric transformation

- Map-to-map transformation
  - Apply to a **digitized map**
  - Assign a projected coordinate system to the digitized map
  - Convert the map coordinates to projected coordinates
- Image-to-map transformation
  - Apply to **remotely sensed (RS) images**
  - Assign a projected coordinate system to the RS map
  - The original RS image may contain some **distortions**
  - Transformation may change the image size depending on the resampling size
- **Georeferencing**

Recap: Essence of geometric transformation

- Essence of geometric transformation
  - Building the mapping relationships between a map coordinate (u,v) before transformation and a map coordinate (x, y) after transformation

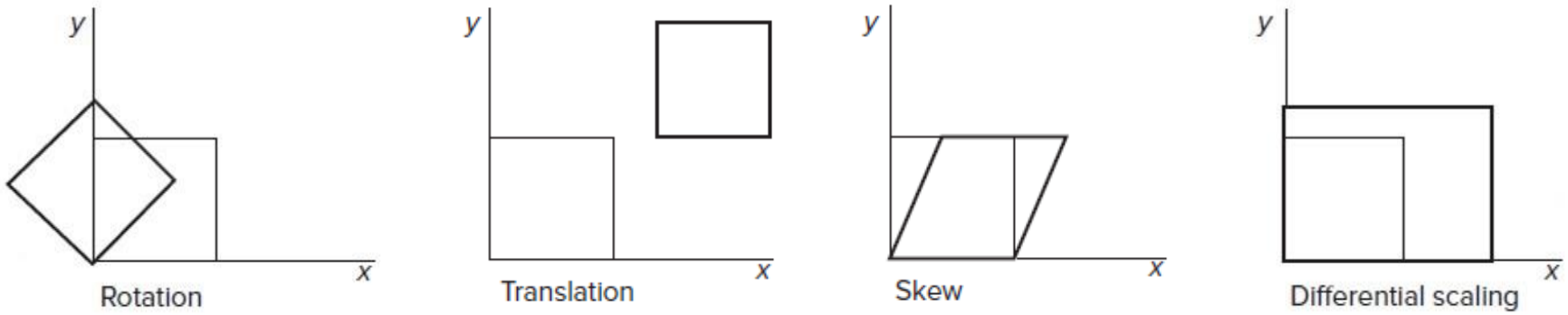


$$x = f_1(u,v) \quad y = f_2(u,v)$$

- f_1 and f_2 are the geometric transformation functions with a number of parameters

Recap: Affine transformation

- The affine transformation allows rotation, translation, skew and differential scaling, while preserving line parallelism. It is also called rubber sheeting



- The affine transformation assumes uniformly distorted input

Recap: Control points

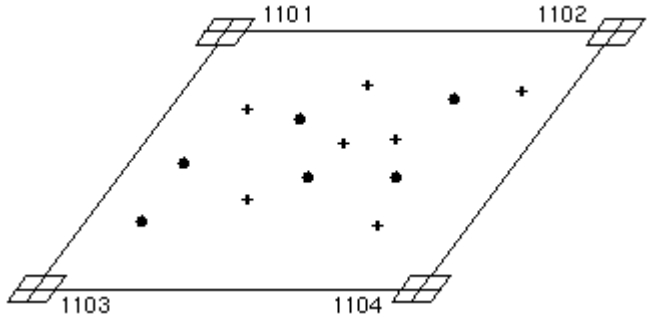
- **Control points** play a key role in determining the accuracy of geometric transformation

Ground control points (GCPs) ↓ Corresponding points ↓

Transformation	Point with input coordinates	Point with output coordinates
Map-to-map	Selected on the source map (usually map intersections)	• Points with known real-world coordinates • Points selected on the reference map
Image-to-map	Selected on the image (features that show up clearly as single distinct pixels, e.g., road intersections, small ponds)	• Points with known real-world coordinates • Points on the reference map • GPS points

Recap: Principles for selecting control points

- Easily identified on both images and maps
  - Maps: Tic points
  - Images: Road intersections, bends of rivers, small prominent features
- Evenly distributed on the images or maps
- Closer to the map features of interest (e.g., GCPs near to NUS as the Are of Interest)



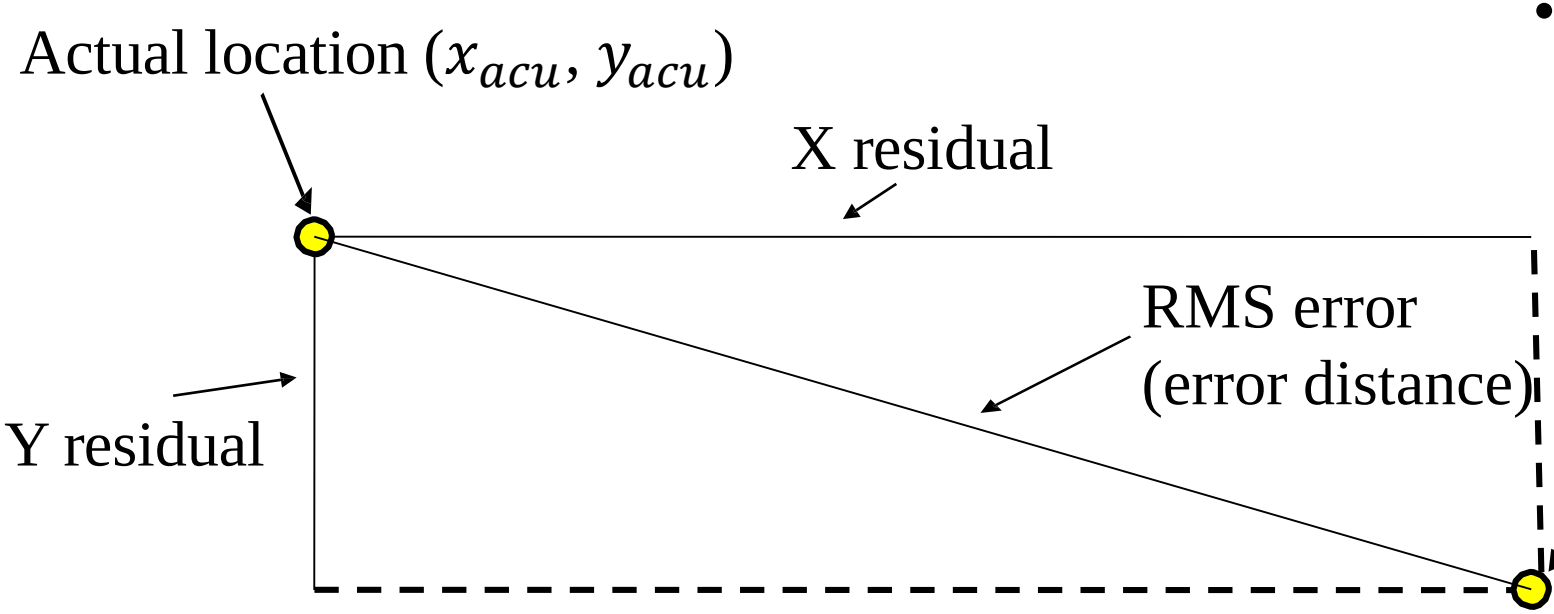
Recap: Root Mean Square error

- What are the actual (true) and estimated locations?

$$x = Au + Bv + C$$
$$y = Du + Ev + F$$

Output coordinates

Input coordinates



- The estimated location can deviate from its actual location with the transformation equation

Recap: Root Mean Square error

- The total RMS error:

Output $$\sqrt{\left[\sum_{I=1}^{n}(x_{acu,i}-x_{est,i})^2+\sum_{I=1}^{n}(y_{acu,i}-y_{est,i})^2\right]/n}$$

X residual

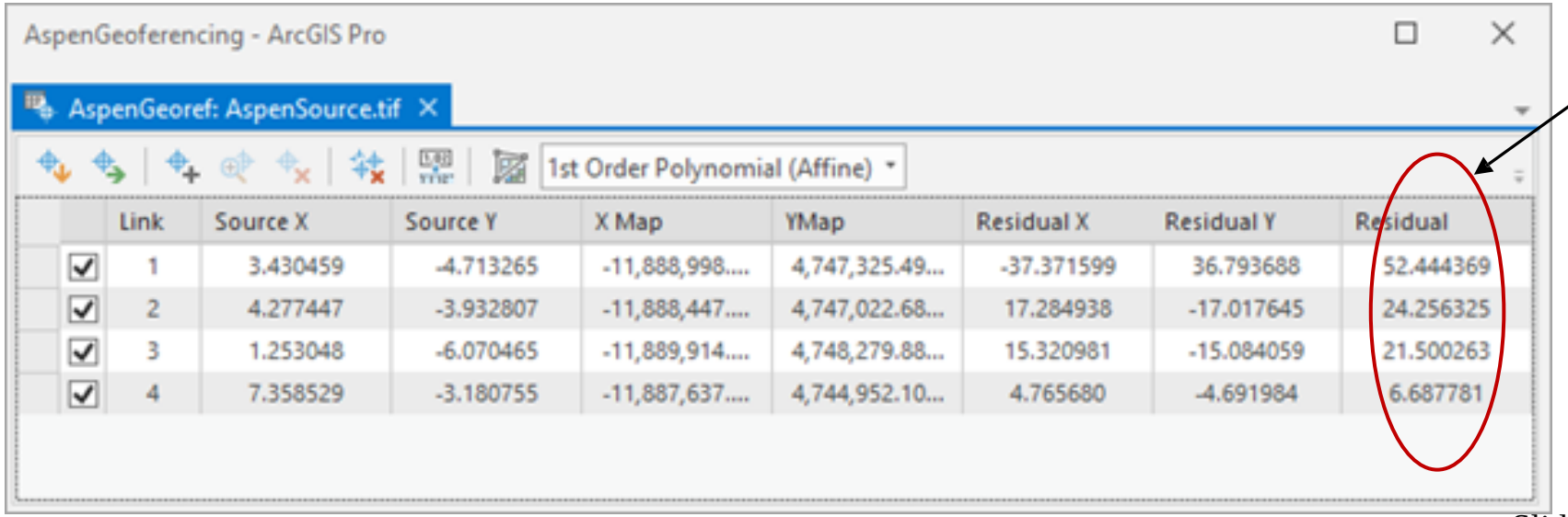
Y residual

- RMS error can only be computed when the number of GCPs are more than the minimum number required

Recap: RMS error tolerance

- How to reduce RMS errors?
  1. Choose better control points
  2. Drop the ones with large RMS errors
  3. Choose higher level model
  4. Add more control points

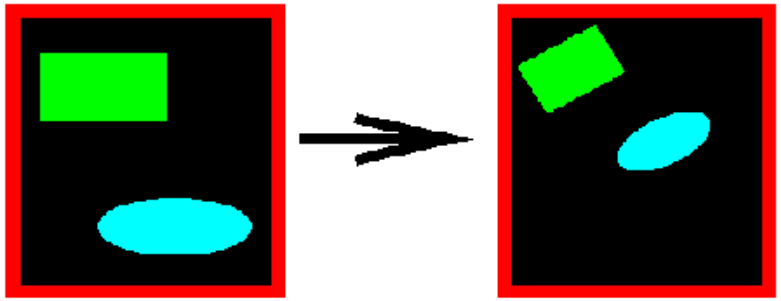
Question: Is RMS error consistent throughout the map or image?



Link	Source X	Source Y	X Map	YMap	Residual X	Residual Y	Residual
1	3.430459	-4.713265	-11,888,998....	4,747,325.49...	-37.371599	36.793688	52.444369
2	4.277447	-3.932807	-11,888,447....	4,747,022.68...	17.284938	-17.017645	24.256325
3	1.253048	-6.070465	-11,889,914....	4,748,279.88...	15.320981	-15.084059	21.500263
4	7.358529	-3.180755	-11,887,637....	4,744,952.10...	4.765680	-4.691984	6.687781

Recap: Resampling of pixel values

$$x = Au + Bv + C$$
$$y = Du + Ev + F$$



- The affine transformation equations build the mapping relations between the locations of pixels on the original and new images
- However, the new image has no pixel values (blank image)
- **Resampling** means filling each pixel of the new image with a value or derived value from the original image

Recap: Resampling of pixel values

- Three commonly used resampling methods are:
    1. Nearest neighbor resampling
    2. Bilinear interpolation (distance-weighted)
    3. Cubic convolution (distance-weighted)
- The above three methods are listed in order of increasing complexity and accuracy

Outline of this lecture

- Spatial data quality
  - Data quality problems
  - Reflection of spatial data quality
  - Causes of spatial data quality problem
  - Characteristics of spatial data quality problem
- Spatial data editing
  - Topological errors and editing
  - Non-topological editing

Spatial data quality

The proposal of spatial data quality problem

- Spatial data have the characteristics of complexity, ambiguity (e.g., living areas of African elephants)
- Human errors will be introduced during data processing
- Differences between the measuring value and the real value are inevitable
- These differences are **impossible** to be fully eliminated, but can be controlled
- These are the spatial data quality problems that are to be solved

The quality of spatial data is the key point for the success of GIS applications (e.g., Navigation)

Reflection of spatial data quality

- Location error

The positioning accuracy determines the success of the rocket launch

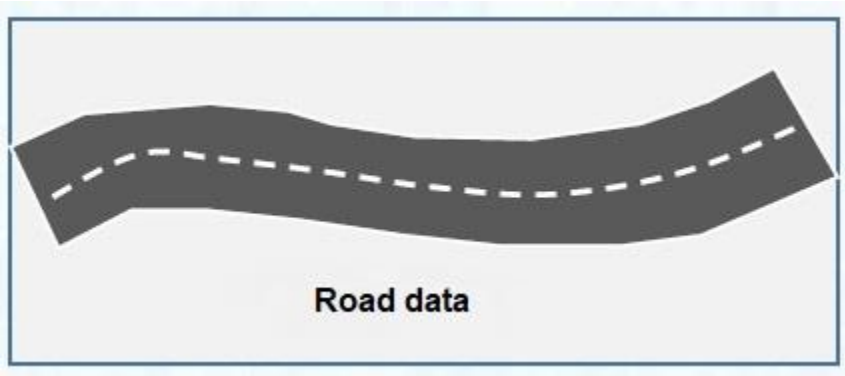
- Time error

Before disaster

After disaster

Reflection of spatial data quality

- Attribute error



Road name	Clementi road
Road type	Highway
Lanes	4
Max speed	70
Bicycle	Yes

- How do we know if attribute errors exist in a database?
  - Manual checking
  - Logical range (e.g., max speed between 30 and 50) and type checking (e.g., Boolean variable)

Causes of spatial data quality problem

1. Multiple data sources

Remote sensing

Survey vehicle

Onsite surveying

2. Data entry by someone is not strict

Surveyor	First measur byement	Second measurement	Average
John	100	110	105
Adam	112	96	104
Lily	106	108	107

But what if the measuring time is missing?

3. Too much data

Which one is more suitable?

Causes of spatial data quality problem

4. Data standards by different users

Government

Enterprise

Individual

Same data quality requirement?

5. Different encoding methods

0	2	2	5	5
2	2	2	5	5
2	2	2	3	3

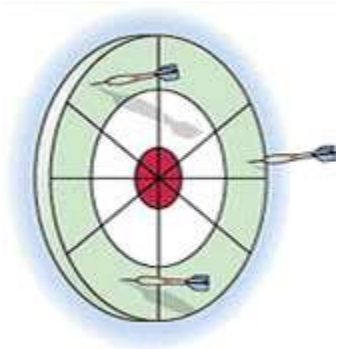
Different encoding methods will result in different data compression

6. Different data access limit

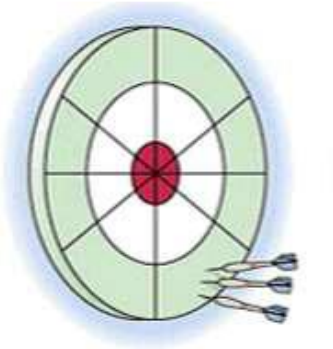
- US commercial satellite image (> 50cm)
- Due to sensitive information

Characteristics of spatial data quality

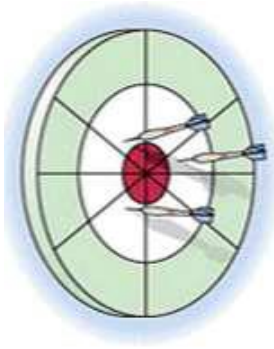
- **Data accuracy and precision**
  - Accuracy measures how close the recorded location of a spatial feature is to its real location (reference)
  - Precision measures how exactly the location is recorded



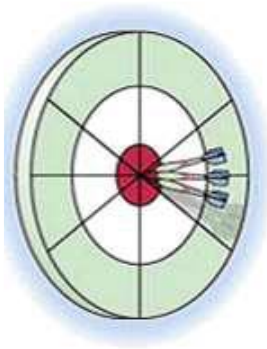
Low accuracy
Low precision



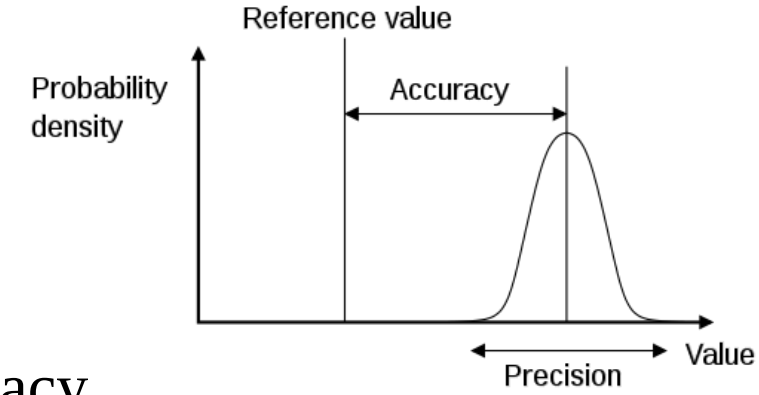
Low accuracy
High precision



High accuracy
Low precision

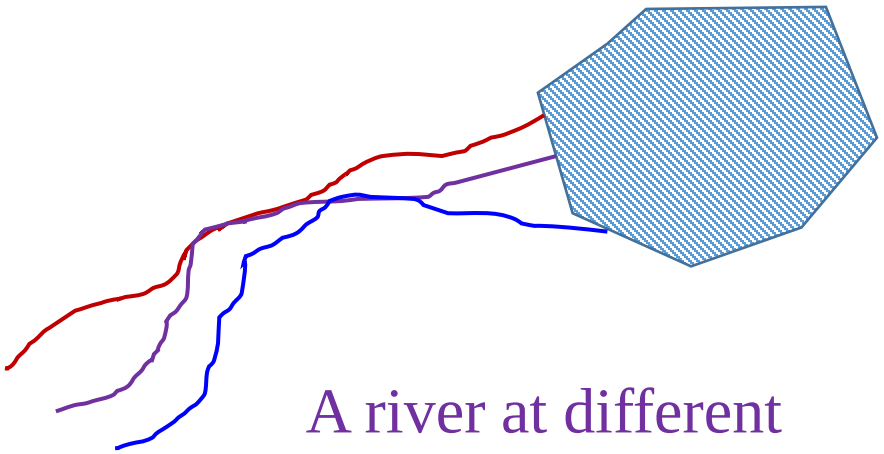


High accuracy
High precision



Characteristics of spatial data quality

- **Data uncertainty**
  - The real values of some phenomenon in the real world are difficult to know or immeasurable
- **Uncertainty of the real world itself**



A river at different times in the history

Characteristics of spatial data quality

- **Uncertainty of the human cognition**

A mountain, when viewed in face, may look like a range; when viewed from the side, it may look like a peak.

Characteristics of spatial data quality

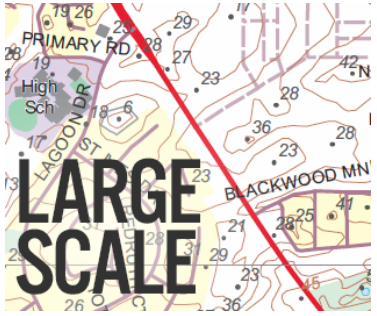
- **Data compatibility**

Are 3D and 2D maps compatible?

3D Map

2D Map

Overlay



Large-scale Map

How do we combine them?

Small-scale Map

Spatial data accuracy measurement

- Map Accuracy measurement (**horizontal** accuracy)
  - Can be obtained by **comparing** the correct location to the plotted location on a map (standard map)
  - Can be obtained by **testing** well-defined points (e.g., road intersection, center points of a building, using GPS)

Outline of this lecture

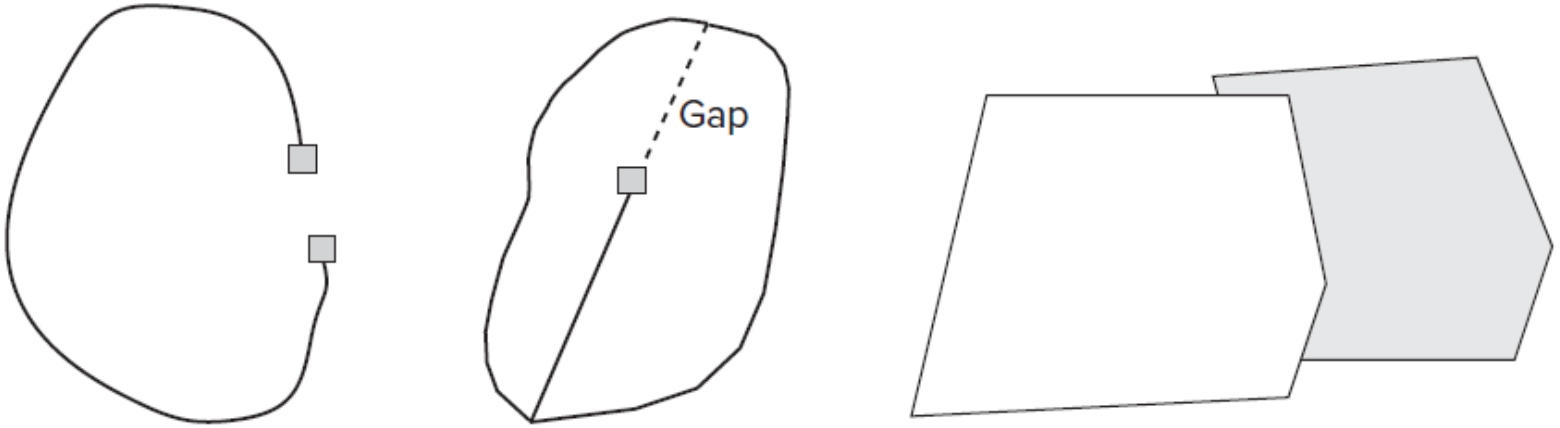
- Spatial data quality
  - Data quality problems
  - Reflection of spatial data quality
  - Causes of spatial data quality problem
  - Characteristics of spatial data quality problem
- Spatial data editing
  - Topological errors and editing
  - Non-topological editing

Topological errors

- **Topological errors** violate topological relationship rules
- **Topological relationship rules** can be defined by:
  - The data model (e.g., a polygon must be closed)
  - The users (e.g., highways across two states must connect perfectly)
- **Topological relationship rules** can be defined:
  - Within a feature class (point, polyline or polygon)
  - Between feature layers

Topological errors within a feature class

- Topological errors within a spatial feature class can be classified by polygon, line, and point
- Common topological errors with polygon features

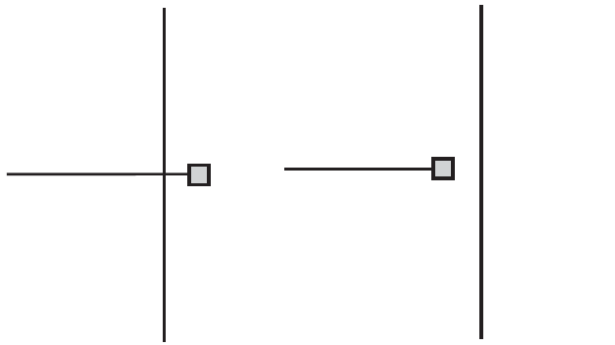


(a) An unclosed polygon (b) A gap between two polygons (c) Overlapped polygons

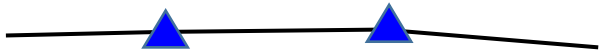
Note that it is possible to have overlapped polygons in some cases, e.g., activity spaces of two different people

Topological errors within a feature class

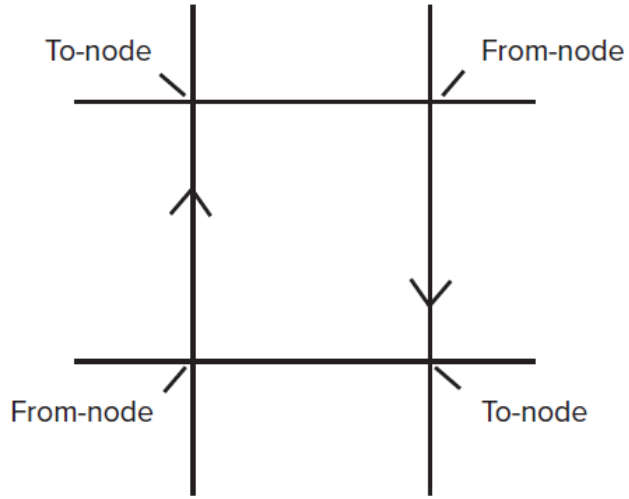
- Common topological errors with polylines:
  - **Dangling nodes**: line features do not meet perfectly at end points
    - **Overshoot**: a line is overextended
    - **Undershoot**: a gap between two lines
  - **Pseudo nodes:** divide a line feature unnecessarily into separate ones
  - **Wrong direction** of lines



Overshoot Undershoot



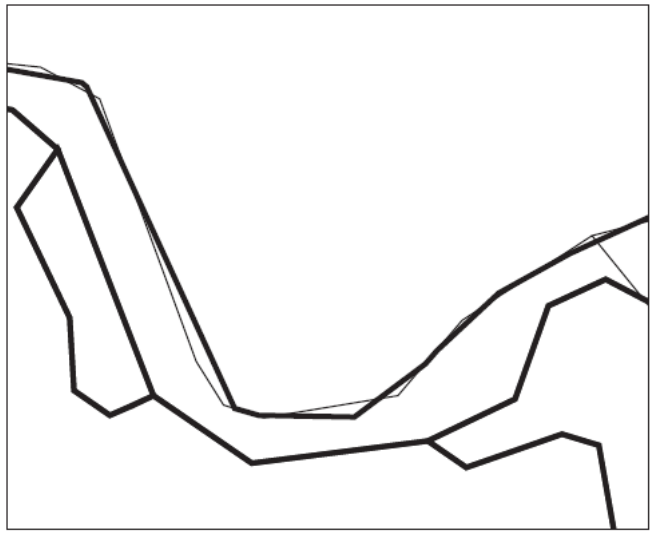
The pseudo nodes unnecessarily divide the same road into multiple road segments



Topological errors between layers

- Many operations and analysis in GIS require two or more layers
- Common errors between two **polygon** layers
  - The outline boundaries in the two polygon layers are not coincident
    - E.g., a GIS project uses a forest layer and a non-forest layer

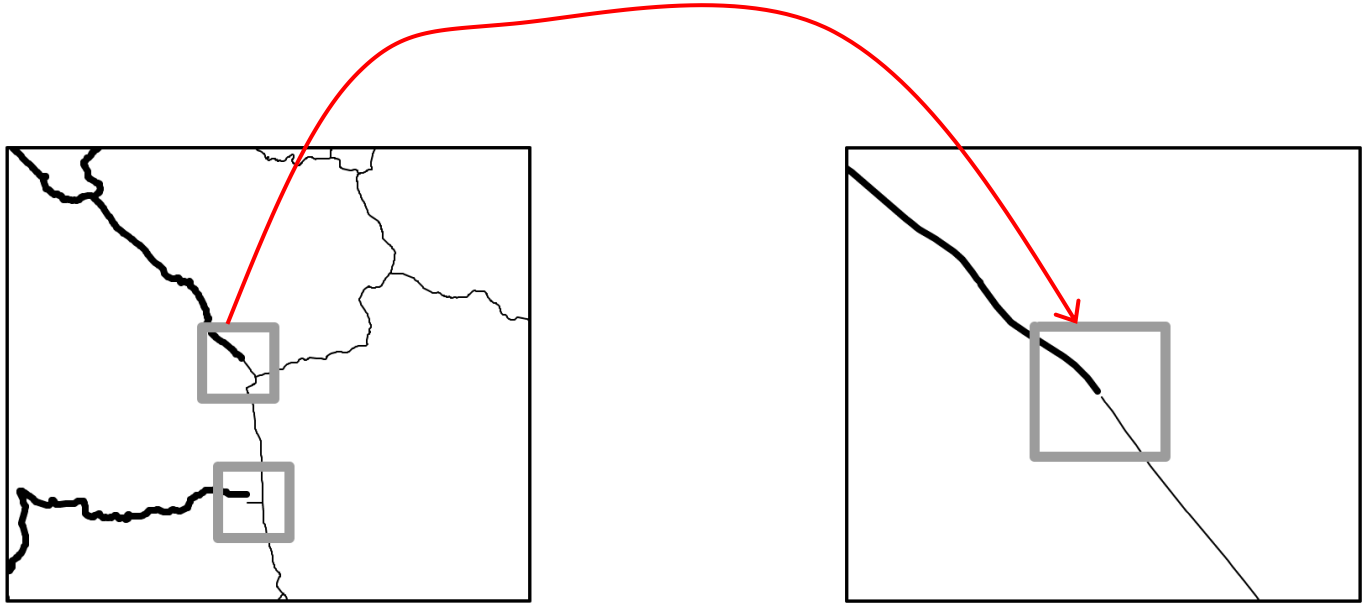
This error occurs when the two layers are digitized separately or from different data sources



The outline boundaries of the two layers, one shown as the **thicker line** and the other as the **thinner line**, do not coincide.

Topological errors between layers

- Common errors between two **line** layers
  - Lines from one layer do not connect with those from another layer at end points



Two highway layers from two adjacent states are not perfectly connected

Topological errors between layers

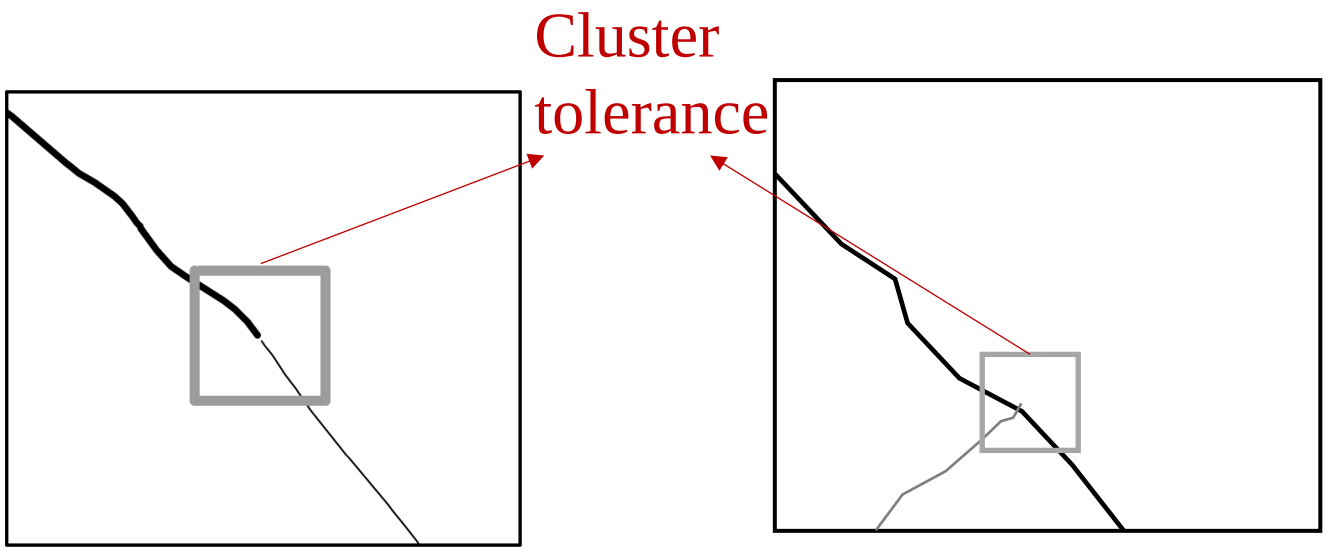
- Common errors between a point layer and a line layer
  - Points do not fall along line features in another layer
    - E.g., gauge stations for measuring streamflow do not fall along streams
- Common errors between a point layer and a polygon layer
  - Points do not fall within polygon features in another layer
    - E.g., police stations do not fall within their correct administration divisions

Topological editing

- **Topological editing** ensures that topological errors are removed
- Three steps to perform topological editing
  - Step 1: GIS can detect topological errors
  - Step 2: GIS can display topological errors
  - Step 3: GIS can fix topological errors

Topological editing – Cluster Tolerance

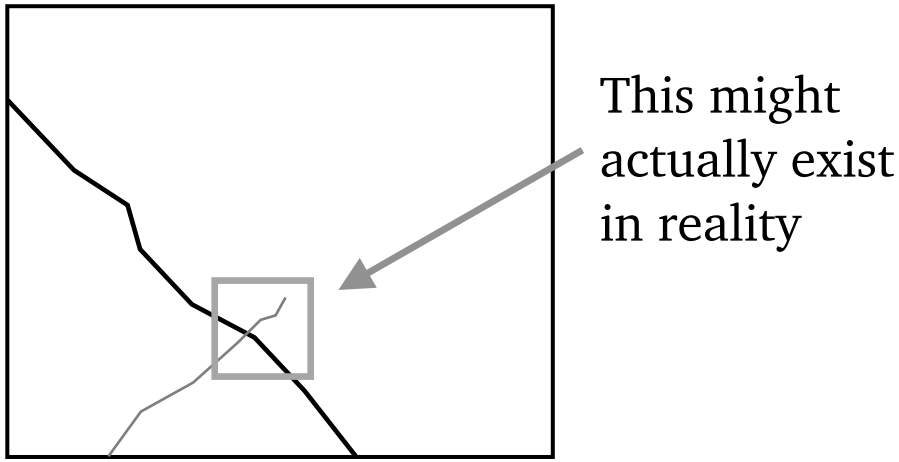
- Cluster tolerance is powerful for topological editing
- Cluster tolerance, also called XY tolerance, is used to snap vertices if they fall within a square area specified by the tolerance



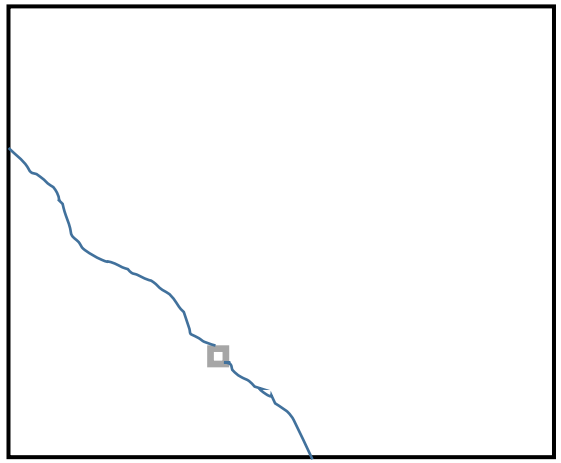
Cluster tolerance is powerful in handling undershoot and overshoot

Topological editing – Cluster Tolerance

- The cluster tolerance should not be set **too large** or **too small**



Cluster tolerance is set too large as it over-detects topological errors



Cluster tolerance is set too small as it under-detects topological errors

- The default cluster tolerance is 0.001 meter, but it can be changed depending on the accuracy of digitized data

Topological editing – Topology Rules

- **Topology rules** can be defined based on only one feature class or between spatial layers
- The geodatabase has more than 30 topology rules

Topological editing – Topology Rules

- How to use topology rules to correct topological errors
  1. **Create** a new topology by defining the participating feature classes, the topology rules, and a cluster tolerance
  2. **Validation** of topology. Identify errors that have violated the topology rules
  3. **Fix** topological errors or accept them as exceptions (e.g., acceptable dangling nodes)

Topology rules – Polygon

Polygon

Must not overlap

Polygons must not overlap within a feature class or subtype. Polygons can be disconnected or touch at a point or touch along an edge.

Polygon errors are created from areas where polygons overlap.

A voting district map cannot have any overlaps in its coverage.

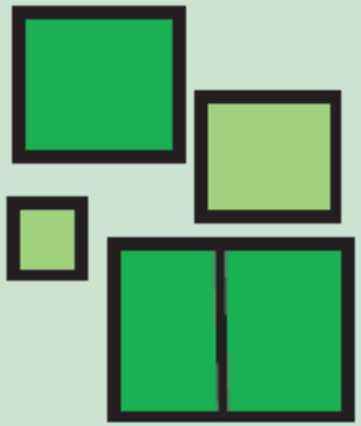
Use this rule to make sure that no polygon overlaps another polygon in the same feature class or subtype.

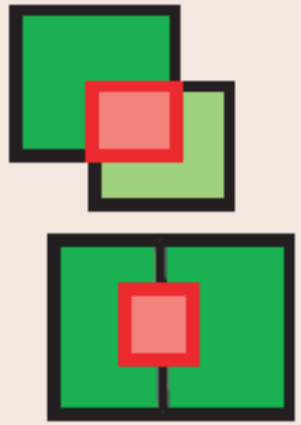
Topology rules – Polygon

Polygon

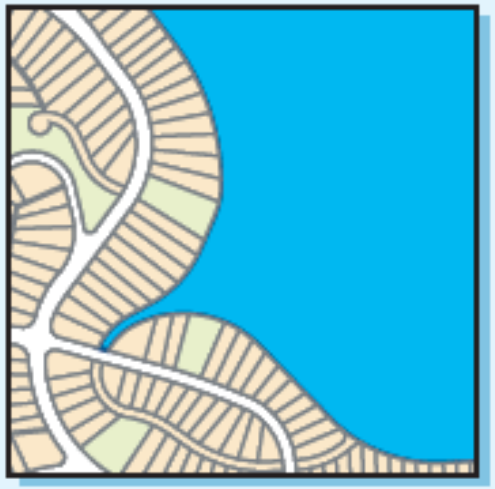
Must not overlap with

Polygons of the first feature class or subtype must not overlap polygons of the second feature class or subtype.





Polygon errors are created where polygons from the two feature classes or subtypes overlap.



Use this rule when polygons from one feature class or subtype should not overlap polygons of another feature class or subtype.

Lakes and land parcels from two different feature classes must not overlap.

Topology rules – Polygon

Polygon

Contains point

Each polygon of the first feature class or subtype must contain within its boundaries at least one point of the second feature class or subtype.

Polygon errors are created from the polygons that do not contain at least one point. A point on the boundary of a polygon is not contained in that polygon.

School district boundaries must contain at least one school.

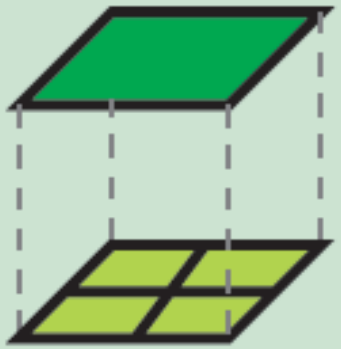
Use this rule to make sure that all polygons have at least one point within their boundaries. Overlapping polygons can share a point in that overlapping area.

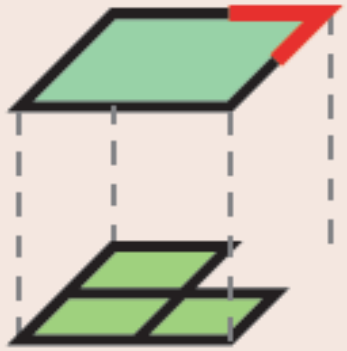
Topology rules – Polygon

Polygon

Area boundary must be covered by boundary of

The boundaries of polygons in one feature class or subtype must be covered by the boundaries of polygons in another feature class or subtype.





Line errors are created where polygon boundaries in the first feature class or subtype are not covered by the boundaries of polygons in another feature class or subtype.

Subdivision boundaries are coincident with parcel boundaries, but do not cover all parcels.

Use this rule when the boundaries of polygons in one feature class or subtype should align with the boundaries of polygons in another feature class or subtype.

Topology rules – Line

Line

Must not have dangles

The end of a line must touch any part of one other line or any part of itself within a feature class or subtype.

Point errors are created at the end of a line that does not touch at least one other line or itself.

A street network has line segments that connect. If segments end for dead-end roads or cul-de-sacs, you could choose to set as exceptions during an edit session.

Use this rule when you want lines in a feature class or subtype to connect to one another.

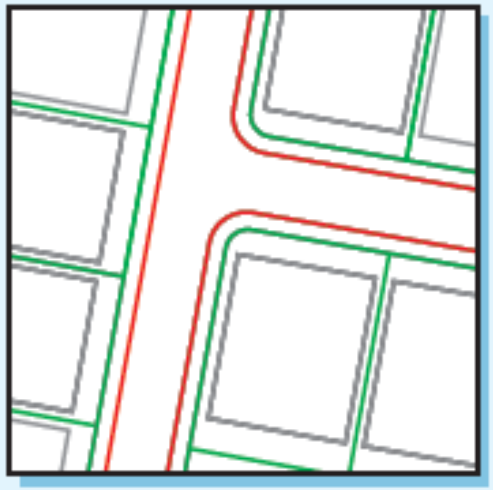
Topology rules – Line

Line

Must not overlap

Lines must not overlap any part of another line within a feature class or subtype. Lines can touch, intersect, and overlap themselves.

Line errors are created where lines overlap.



Lot lines cannot overlap one another.

Use this rule with lines that should never occupy the same space with other lines.

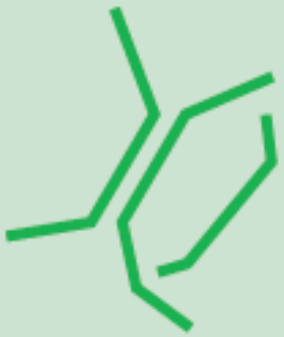
Note that intersection is allowed in this topology rule

Topology rules – Line

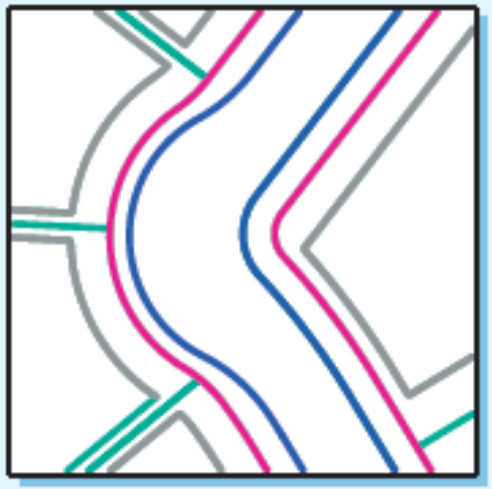
Line

Must not intersect

Lines must not cross or overlap any part of another line within the same feature class or subtype.



Line errors are created where lines overlap, and point errors are created where lines cross.



Use this rule with lines whose segments should never cross or occupy the same space with other lines.

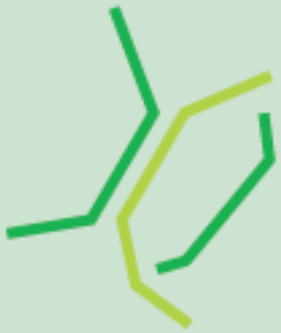
Lot lines cannot intersect or overlap, but the endpoint of one feature can touch the interior of another feature.

Topology rules – Line

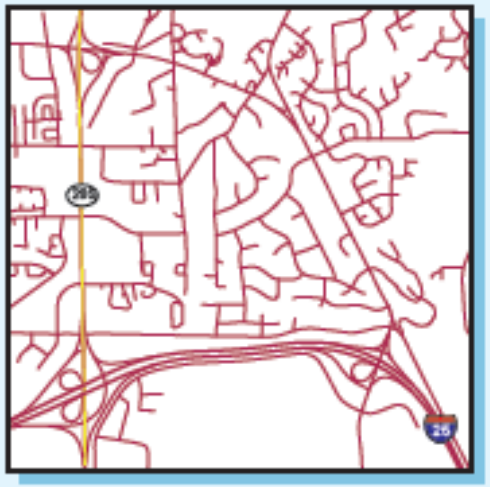
Line

Must not intersect with

Lines in one feature class or subtype must not cross or overlap any part of a line in another feature class or subtype.



Line errors are created where lines overlap, and point errors are created where lines cross.



Local roads cannot intersect or overlap major highways and must connect only at ramps.

Use this rule with lines whose segments should never cross or occupy the same space with lines in another feature class or subtype.

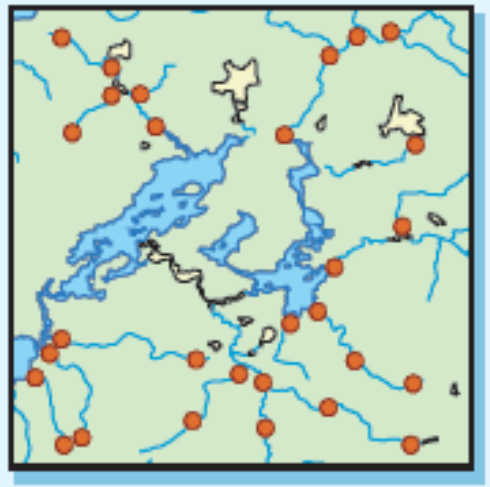
Topology rules – Points

Point

Point must be covered by line

Points in one feature class or subtype must be covered by lines in another feature class or subtype.

Point errors are created on the points that are not covered by lines.



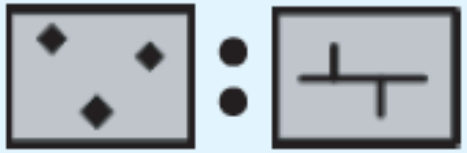
Monitoring stations must fall along streams.

Use this rule when you want to model points that are coincident with lines.

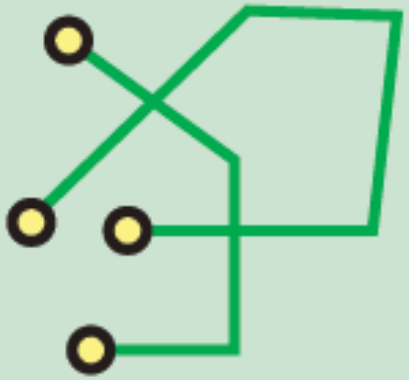
Topology rules – Points

Point

Must be covered by endpoint of

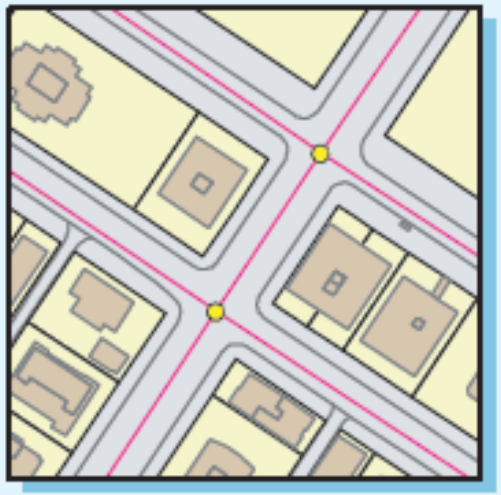


Points in one feature class or subtype must be covered by the ends of lines in another feature class or subtype.



Point errors are created on the points that are not covered by the ends of lines.

Use this rule when you want to model points that are coincident with the ends of lines.



Street intersections must be covered by the endpoints of street centerlines.

Topology rules – Points

Point

Must be covered by boundary of

Points in one feature class or subtype must touch boundaries of polygons from another feature class or subtype.

Point errors are created where points do not touch the boundaries of polygons.

Utility service points might be required to be on the boundary of a parcel.

Use this rule when you want points to align with the boundaries of polygons.

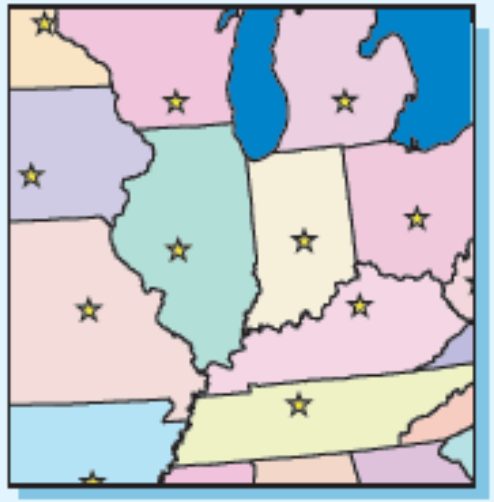
Topology rules – Points

Point

Must be properly inside polygons

Points in one feature class or subtype must be inside polygons of another feature class or subtype.

Point errors are created where the points are outside or touch the boundary of the polygons.



State capitals must be inside each state.

Use this rule when you want points to be completely within the boundaries of polygons.

More topology rules

- http://resources.arcgis.com/en/help/main/10.2/01mm/pdf/topology_rules_poster.pdf

Outline of this lecture

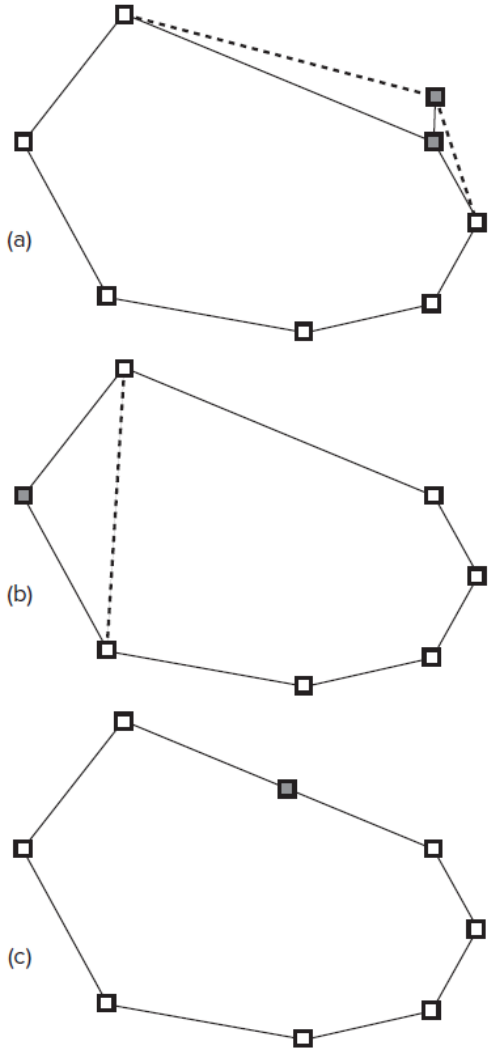
- Spatial data quality
  - Data quality problems
  - Reflection of spatial data quality
  - Causes of spatial data quality problem
  - Characteristics of spatial data quality problem
- Spatial data editing
  - Topological errors and editing
  - Non-topological editing

Non-topological editing

- Non-topological editing
  - Does not involve **topology** as defined in a **map topology** or a **topology rule**
  1. Modify existing features
  2. Create new features from existing features

Non-topological editing

- Modify existing features
  - Extend/trim lines
  - Delete/move features
  - Reshape features
  - Split lines and polygons



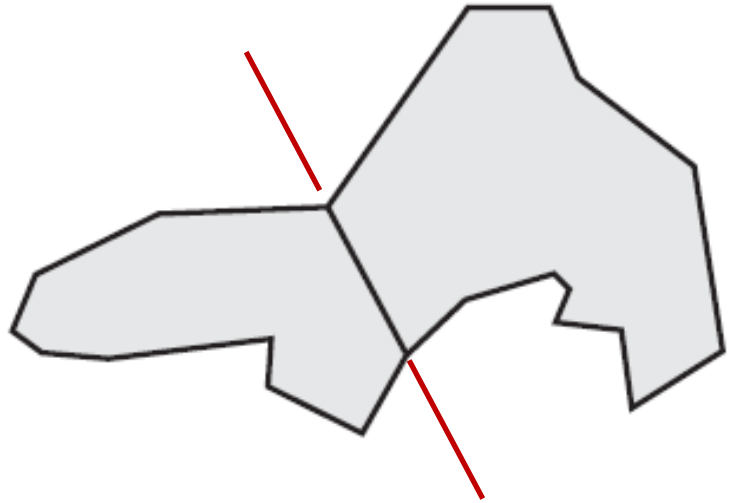
Reshape a polygon by **moving** a vertex

Reshape a polygon by **deleting** a vertex

Reshape a polygon by **adding** a vertex

Non-topological editing

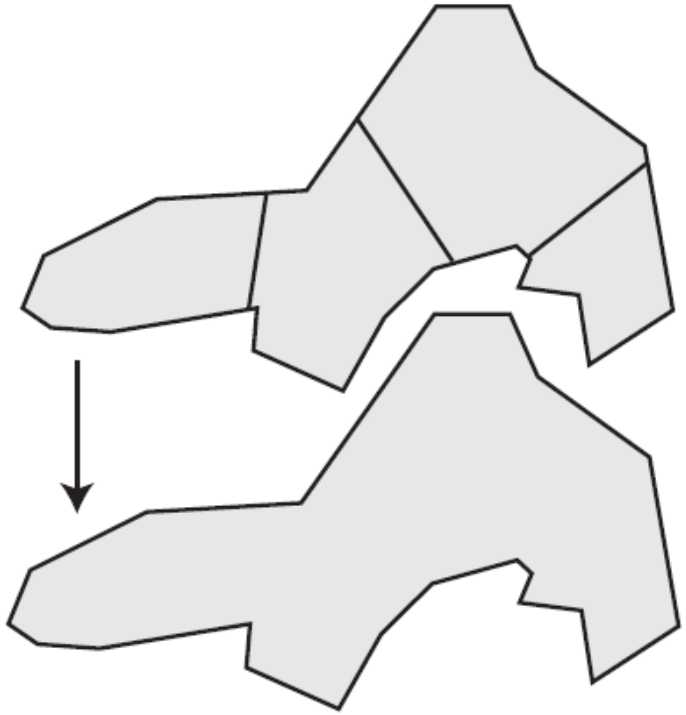
- Modify existing features
  - Extend/trim lines
  - Delete/move features
  - Reshape features
  - Split lines and polygons



Sketch a line across the polygon boundary to split the polygon into two

Non-topological editing

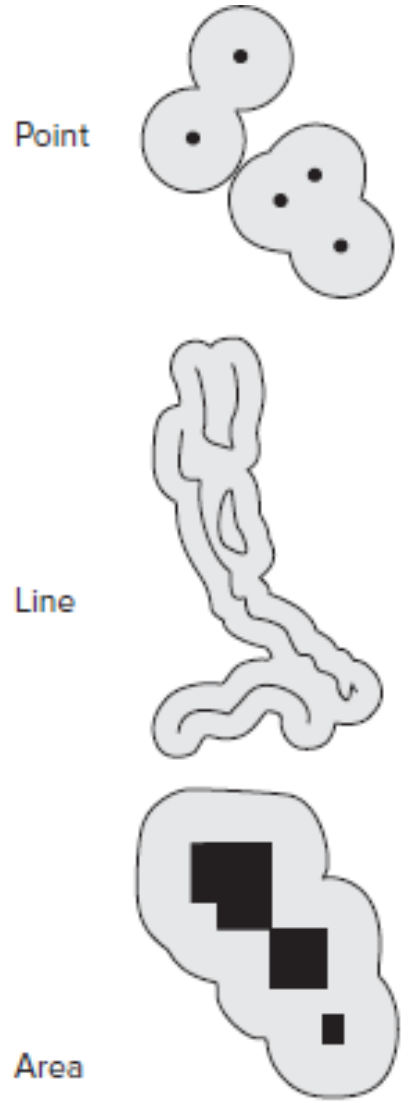
- Create features from existing features
  - Merge features into one feature
  - Buffer features to create a buffer feature
  - Union features to combine features from the same layer
  - Intersect features to create a new feature



Merge four polygons into one

Non-topological editing

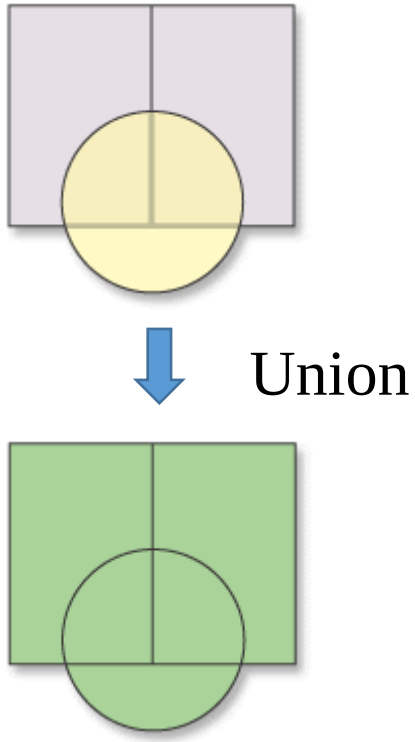
- Create features from existing features
  - Merge features into one feature
  - Buffer features to create a buffer feature
  - Union features to combine features from the same layer
  - Intersect features to create a new feature



Buffering around points, lines, and polygons to obtain new polygon features

Non-topological editing

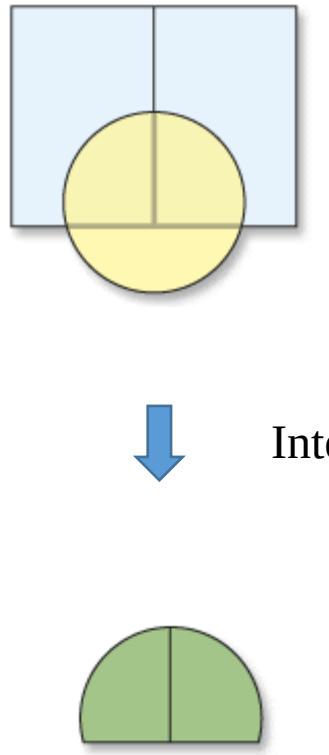
- Create features from existing features
  - Merge features into one feature
  - Buffer features to create a buffer feature
  - Union features to combine features from the same layer
  - Intersect features to create a new feature



The union of three polygon features to form a new feature class with five polygon features

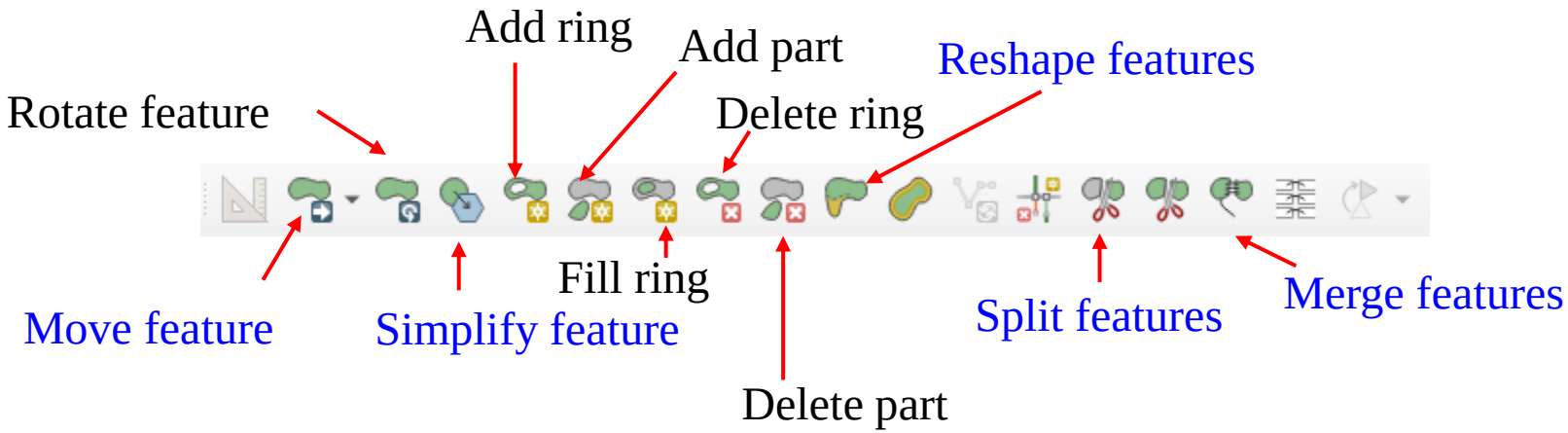
Non-topological editing

- Create features from existing features
  - Merge features into one feature
  - Buffer features to create a buffer feature
  - Union features to combine features from the same layer
  - Intersect features to create a new feature

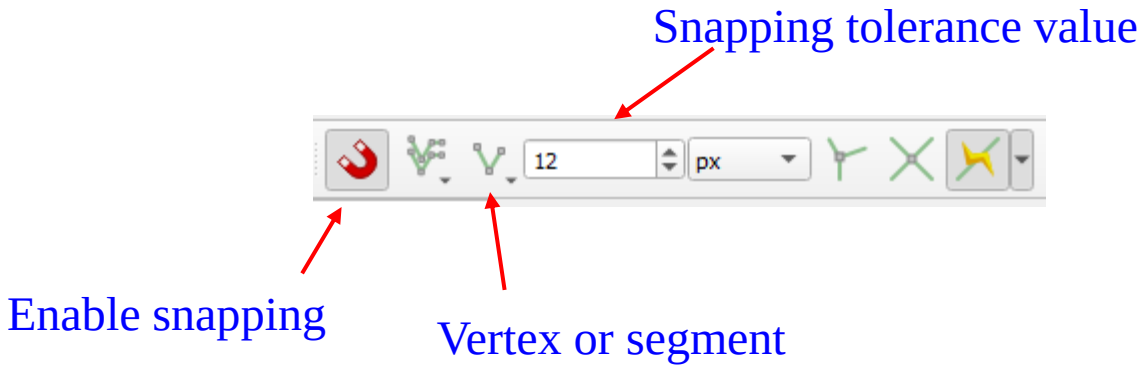


Intersect three polygon features to form a new feature class with two polygon features

Editing tools in QGIS



Snapping tools in QGIS



Summary

- Spatial data quality
  - Data quality problems
    - Spatial data quality – key point for success
  - Reflection of spatial data quality
    - Location error, time error and attribute error
  - Six causes of spatial data quality problem
  - Characteristics of spatial data quality problem
    - Data **accuracy** and **precision**, **uncertainty**, and **compatibility**

Summary

- Spatial data editing
  - Topological errors and editing
    - Topological errors **within** and **between** layers
    - Common topological errors in polygons, lines and points
    - Topological editing: cluster tolerance and topology rules
  - Non-topological editing
    - Modify existing features
      - Extend/trim, reshape, move, split features
    - Create features based on existing features
      - Merge, buffer, union, intersect features

THANK YOU