Spatial Data Editing

- Why do we need to edit spatial data?
  - Correct location/topology errors
  - Update/modify spatial information
- Errors
  - Detection
  - Quantification
  - Spatial data accuracy standard
  - Correction

Sources of Errors in GIS Data

- Conception
- Measurement
- Representation
- Analysis

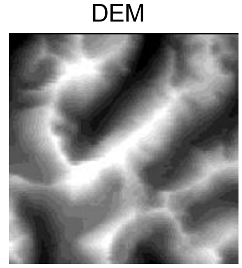
Detection of Errors

- External validation (using referencing data)
  - Remotely sensed imagery
  - Maps
  - GPS
    - HARN (High Accuracy Reference Network)
    - HPGN (High Precision Geodetic Network)
    - HARN and HPGN are both names for the same project that is focused on readjusting the NAD83 datum to a higher level of accuracy state by state.
- Internal validation
  - Simulation
  - Spatial analysis (derivatives)
  - Scientific visualization

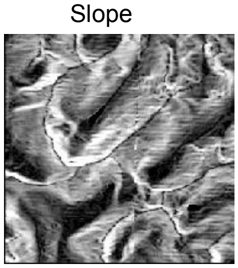
Geographic Visualization/Spatial Analysis

- Tools for GIS data verification

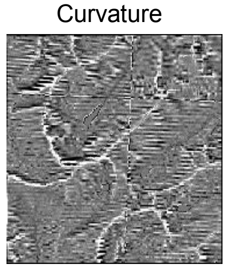
DEM



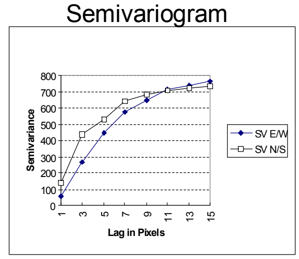
Slope



Curvature



Semivariogram



Quantifying Errors

- Overall magnitude
  - RMSE
  - Producer’s accuracy (probability of errors)
  - User’s accuracy (reliability of data)
- Spatial distribution of errors
  - Simulation
  - Error propagation models
- Qualitative / topological errors
  - Topology rules

Editing

- Coverage, shapefile, geodatabase feature class
- Topological/non-topological editing
- Single layer/multiple layers editing
- Other editing tools/procedures
  - Line simplification and smoothing
  - Rubbersheeting
  - Transformation (scaling, shifting, rotation)
  - Edgematching

Spatial Data Accuracy Standards

- 1947 US National Map Accuracy Standard
  - <10% sample points with > 1/30” displacement for scale > 1:20K, or >1/50” for scale <= 1:20K
- 1990 American Society for Photogrammetry & Remote Sensing Standard
  - RMSE 16.7 ft for 1:20K; 2ft for 1:2400 (1/100”)
- 1998 Federal Geographic Data Committee National Standard for Spatial Data Accuracy
  - Standard for < 1:20K maps and NSSDA statistic
  - Estimate accuracy: RMSE x 1.7308 (95% confidence level)

The Use of Accuracy Standards

- Interpretation
  - Confidence intervals (probability)
  - Scale
  - RMSE sample size (is 20 points enough?)
- Regulation
  - Quality control
  - Meta data
- Reporting measurements
  - Precision & accuracy

Use of Accuracy Information

- Fuzzy tolerance (ArcInfo)

  FuzzyT = Scale x Digitizer precision (in inches) / 12

  Digitizer precision ~ 0.002” (500 dpi)

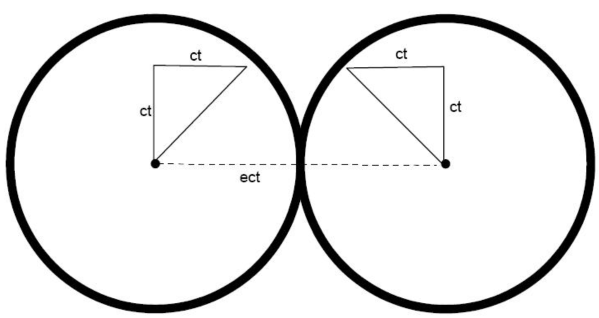
- Cluster (xy) tolerance (ArcGIS)

  An order of magnitude less then the highest accuracy of your data.

  1 inch data accuracy ~ ClusterT = 1/12 /10 = 0.00833 ft.

ex: if cluster tolerance is set to 1 unit then each point is allowed to move 1 unit in the X direction AND 1 unit in the Y direction for an effective cluster tolerance (ect) defined by the following:

$$ect = 2\sqrt{ct^2}$$



$$Radius = \sqrt{ct^2 + ct^2}$$

$$ect = 2 \times \sqrt{2 \times ct^2}$$

$$ct = \frac{ect}{2\sqrt{2}}$$

$$ct \approx \frac{ect}{2.83}$$

solve for ct to determine input parameter:

$$ct = \sqrt{ect^2} \Big/ 2$$

Cluster tolerance is specified in the units of the spatial reference in which the data participating in the topology is stored.

from ArcGIS Desktop Help: The cluster tolerance should be very small, so only very close vertices are snapped together. A typical cluster tolerance is at least an order of magnitude smaller than the accuracy of your data. For example, if your features are accurate to 2 meters, your cluster tolerance should be no more than 0.2 meters.

Line Simplification – Douglas-Peucker Algorithm

1. Connecting end points of a line to form a trend line
2. Calculate deviations of vertices from the trend line
3. If there are deviations larger than a user-specified tolerance, then the vertex with the largest deviation is selected as the new end point, otherwise stop
4. Repeat 1-3 until no deviation exceeds the tolerance