Lecture 15: Dealing with Uncertainty
All maps are distortions

• “It’s not easy to lie with maps, it’s essential...to present a useful and truthful picture, an accurate map must tell white lies.” -- Mark Monmonier

• “A map is a set of errors that have been agreed upon”

• Distort 3-D world into 2-D abstraction (projection: distorts scale, direction, area, shape)

• Convert real-world features into symbolic objects e.g. city to circle

• Maps can portray abstractions (e.g., gradients, contours) as distinct spatial objects
The limits to mapping

- Measurement errors e.g. sampling, missing data
- Methodological errors e.g. conflation
- Symbology errors
- Map use and interpretation errors
- Misuse, misinterpretation and belief
Attribute uncertainty

Presence/ absence of rabies in 2007

Disclaimer: © World Health Organization. The boundaries and names shown and the designations used on this map do not imply the expression of an opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.
What is Uncertainty?

• NIST data quality:
  • lineage
  • positional accuracy
  • attribute accuracy
  • completeness
  • logical consistency

• Also:
  • source
  • scale
  • methodology
  • Reliability
  • trust and confidence
Example

• 3 reports, two say a bridge exists one says it is destroyed
• Report 3 introduces uncertainty
• Bridge is 0.666 certain, >0.5
• Trust? Reliability of sources
• Method in OSM
Lineage

Produced by the United States Geological Survey
North American Datum of 1983 (NAD83)
World Geodetic System of 1984 (WGS84). Projection and
1000-meter grid: Universal Transverse Mercator, Zone 11S
10 000-foot ticks: California Coordinate System of 1933
(zone 8)

Imagery.........................NAIP, July 2010 - August 2010
Roads..............................©2006-2011 TomTom
Roads within US Forest Service Lands..............FSTopo Data
with limited Forest Service updates, 2009
Names................................GNIS, 2011
Hydrography.......................National Hydrography Dataset, 2010
Contours............................National Elevation Dataset, 2005
Boundaries.........................Census, IBWC, IBC, USGS, 1972 - 2010
Positional accuracy e.g. NMAS

1. Horizontal accuracy. For maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch. These limits of accuracy shall apply in all cases to positions of well-defined points only. Well-defined points are those that are easily visible or recoverable on the ground, such as the following: monuments or markers, such as bench marks, property boundary monuments; intersections of roads, railroads, etc.; corners of large buildings or structures (or center points of small buildings); etc. In general what is well defined will be determined by what is plottable on the scale of the map within 1/100 inch.

2. Vertical accuracy, as applied to contour maps on all publication scales, shall be such that not more than 10 percent of the elevations tested shall be in error more than one-half the contour interval. In checking elevations taken from the map, the apparent vertical error may be decreased by assuming a horizontal displacement within the permissible horizontal error for a map of that scale.

3. The accuracy of any map may be tested by comparing the positions of points whose locations or elevations are shown upon it with corresponding positions as determined by surveys of a higher accuracy. Tests shall be made by the producing agency, which shall also determine which of its maps are to be tested, and the extent of the testing.

4. Published maps meeting these accuracy requirements shall note this fact on their legends, as follows: “This map complies with National Map accuracy Standards.”

5. Published maps whose errors exceed those aforesaid shall omit from their legends all mention of standard accuracy.

6. When a published map is a considerable enlargement of a map drawing (manuscript) or of a published map, that fact shall be stated in the legend. For example, “This map is an enlargement of a 1:20,000-scale map drawing,” or “This map is an enlargement of a 1:24,000-scale published map.”

7. To facilitate ready interchange and use of basic information for map construction among all Federal mapmaking agencies, manuscript maps and published maps, wherever economically feasible and consistent with the uses to which the map is to be put, shall conform to latitude and longitude boundaries, being 15 minutes of latitude and longitude, or 7.5 minutes, or 3-3/4 minutes in size.

Issued June 10, 1941 U.S. BUREAU OF THE BUDGET
Revised April 26, 1943
Revised June 17, 1947
Attribute accuracy/definition and scale
Temporal and differential completeness

Coverage and currency of the original 7.5-minute topographic map series in 1970
Logical completeness
Nick Chrisman’s View
(www.wiley.com/college/chrisman/define.html)
Tile/Merge

Source: Longley et al.
Spot the tiles
Scale-induced, temporal error
Generalization uncertainty

• Measurements not perfectly accurate
• Maps distorted when generalized
  • Selection
  • Simplification
  • Combination
  • Displacement
• Objects at scale can be far less than 0.1mm
• Definitions vague, ambiguous, subjective
• Landscape has changed over time
Classification inconsistency
Classification purpose

Example of National Wetland Inventory (NWI) Map

NWI maps show many different wetland types.
Dealing with uncertainty

• Simple quantification
• Showing missing data
• Conflation
• Symbolizing uncertainty
Map showing the discoveries of Abel Tasman in 1642-43 and 1644. The map includes the track of Tasman’s first voyage 1642-43 from Mauritius. It was included in Vol. 3, part 2 of his Francois Valentijn’s history, Oud en nieuw Oost-Indien (Old and new East Indies).
A Cartographic Trap

Left: From A-Z streetmap of central Bristol (redrafted).
Above: Open Street Map
Occasionally TIGER Data are Inaccurate or Imprecise
Quantification

- Horizontal and vertical RMSE
- Recognized by GPS as DOP
- Categorical: by Confusion matrix, User/Producer, Kappa, ROC

![Image of ROC Curves and confusion matrix]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Water body [L1]</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>100.0%</td>
</tr>
<tr>
<td>Built up [L1]</td>
<td>20</td>
<td>50</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>18</td>
<td>95</td>
<td>100.0%</td>
</tr>
<tr>
<td>Local Play area [L1]</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>80.0%</td>
</tr>
<tr>
<td>Tree area [L1]</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>3</td>
<td>24</td>
<td>79.2%</td>
</tr>
<tr>
<td>Grass land [L1]</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>3</td>
<td>38</td>
<td>100.0%</td>
</tr>
<tr>
<td>Shrub &amp; mix [L1]</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>7</td>
<td>29</td>
<td>76</td>
<td>58.0%</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
<td>6</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Producer's Accuracy</td>
<td>36.0%</td>
<td>100.0%</td>
<td>66.7%</td>
<td>38.0%</td>
<td>76.0%</td>
<td>58.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIA Per Class</td>
<td>31.2%</td>
<td>100.0%</td>
<td>66.0%</td>
<td>31.6%</td>
<td>71.8%</td>
<td>40.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Classification Accuracy</td>
<td>61.7%</td>
<td>Kappa statistic</td>
<td>52.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Missing data

Age-adjusted suicide rates 200-206 per 100,000

Land-Only Temperature Departure from Average Apr 2014
(with respect to a 1981–2010 base period)

Data Source: GHCN-M version 3.2.2

NOAA’s National Climatic Data Center
Mon May 12 07:37:52 EDT 2014

Please Note: Gray areas represent missing data
Map Projection: Robinson
Rubber sheeting
Temporal conflation
Conflation
Interpolation method uncertainty
Top: IDW
Bottom: Spline
Fuzzy attributes and positions

- Positions assumed accurate
- But really, just best guess “Best Available Data”
- Differentiate best guesses from “truth”
- “Shadow map of certainty”
  - where an estimate is likely to be the most accurate
- Tracking error propagation
Source diagram: NOAA Chart (Kotzebue, AK)
Fig. 2. Two-way fuzzy comparison between simulated and observed changes. Based upon a "hard" per pixel comparison, there is no coincidence between the observed and the simulated maps. Using increasing fuzziness (longer tolerance distance), coincidence is higher. Note that coincidence is different depending if fuzziness is applied to the observed or the simulated map.
More Strategies

• Simulation
  • Complex models
  • Describing uncertainty as “a spatially autoregressive model with parameter rho” not helpful
  • How to get message across

• Many models out there
  • Research on modeling uncertainty (NCGIA Initiative 1)
  • Users can’t understand them all
  • Given choice, most users do not want uncertainty information
Probabilities

Historical Probability of a White Christmas (greater than 1 inch of snow on the ground)

<10% 10%-25% 25%-40% 40%-50% 50%-60% 60%-75% 75%-90% >90%

Based on 1981-2010 Climate Normals National Climatic Data Center

Digital Flood Insurance Rate Map (DFIRM) Database

Base Flood Elevation Lines
Base Flood Elevation Lines
Floodway
Floodway
Flood Hazard Boundaries
1% annual chance floodplain boundary
0.2% annual chance floodplain boundary
Floodway Boundary
Zone Break
Limit of Floodway, Study, or Delineated Study
Flood Hazard Area
Special Flood Hazard Areas
Other Flood Areas
Coastal Barrier Resource System Areas
Outliers

SRTM elevation errors of over 100m vertical
Options for dealing with map uncertainty

1. Ignore the issue completely
2. Describe uncertainty with measures (shadow map or RMSE)
3. Simulate equally probable versions of data
4. Be uncertainty-aware (trust, but verify)
Monte Carlo Simulation

A Review of Uncertainty in Data Visualization
Ken Brodlie, Rodolfo Allendes Osorio and Adriano Lopes
http://www.comp.leeds.ac.uk/kwb/publication_repository/2012/uncert.pdf

Fig. 5 Contour band: This shows the extent of the 95% confidence interval for the zero contour, with the zero contour for the average data at each point shown for comparison.

Fig. 6 Fuzzy contour: This shows the value of the t-statistic from the hypothesis test with a colour mapping from sea-blue to black based on the size of the t-statistic.

Fig. 7 Spaghetti Plot: zero contours from each of the eight individual models, superimposed on the 95% confidence band.
Kriging uncertainty in R

Figure: Visualisation of uncertainty for a quantitative variable (topsoil thickness in cm) interpolated using regression kriging: uncertainty included with whiteness and the accompanying two-dimensional legend.
Visualizing Uncertainty
Summary

• All data are uncertain
• Uncertainty types include lineage, positional accuracy, attribute accuracy, completeness and logical consistency
• Error due to scale, method, interpretation, source, bias
• Methods for dealing with error include ignoring it, quantification, simulation
• Few methods explore integrating error and uncertainty directly
• People have a hard time dealing with probabilities, let alone expected error information