Uncertainty in Geospatial Data

Michael F. Goodchild
University of California
Santa Barbara
Traditional geospatial data

- Maps, atlases, globes
  - highly synthesized, compiled, abstracted
  - often rich quantitative attributes
  - no memory of observation provenance
General principles

• It is impossible to map perfectly
  – to determine position on the Earth’s surface
  – to assign locations to vaguely defined classes

• All geospatial data are therefore subject to uncertainty

• It can be as important to determine what geospatial data do not say
  – to quantify what is missing
  – to communicate uncertainty effectively

• Scientific results should be reported to a precision that reflects their accuracy
Line measurement (Geodesic)
Segment: 10,527,369.504217 Meters
Length: 10,527,369.504217 Meters
Outstanding issues

• Communicating uncertainty to the user
  – the user may not want to know
  – users expect maps to be perfect

• Spatial (and temporal) autocorrelation
  – geospatial data, errors are almost always positively autocorrelated
  – modeling, calibration, metadata difficult

• Ontological issues
  – choice of representation linked to uncertainty
  – raster > vector
Outstanding issues (2)

• Dependence on use case
  – fitness for use
  – can we address uncertainty generically?

• Importance of provenance
  – are two datasets interoperable?
    • how much provenance do they share?
    • *binary* metadata