Spatial Analysis and Modeling (GIST 4302/5302)

Guofeng Cao

Department of Geosciences

Texas Tech University

Outline of This Week

- Last topic, we learned:
  - Variogram/Covariogram
  - Spatial interpolation/Kriging
- This topic, we will learn:
  - Concept of spatial uncertainty
  - How should we do with uncertainty

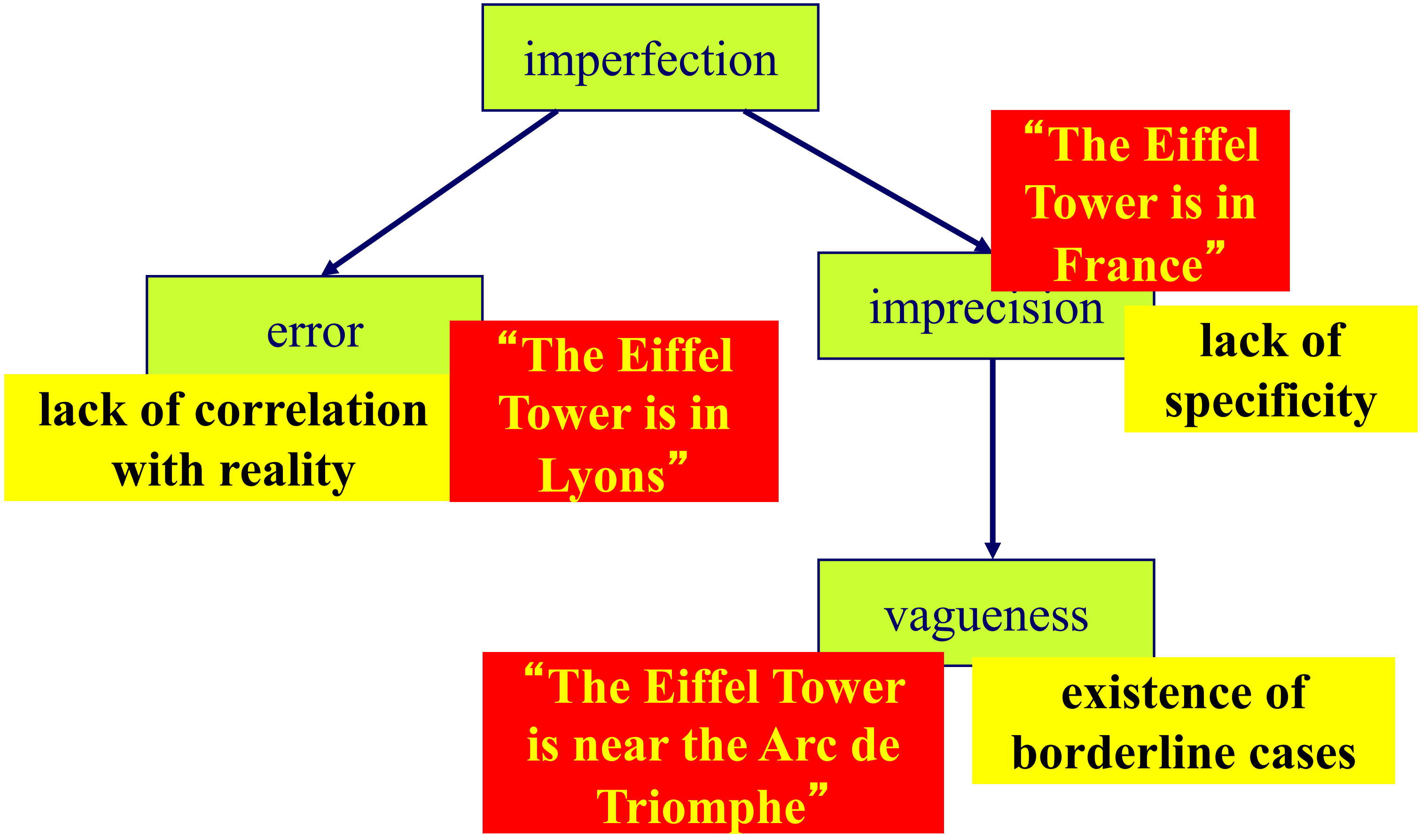
Uncertainty

- Uncertainty is an *unavoidable* property of the world, information about the world, and our cognition of the world
- Uncertainty
  - May refer to state of mind: “I am unsure where the meeting will take place”
  - May be applied directly to data or information about the world: “The depth of the sea at a particular location is uncertain”

Spatial uncertainty example

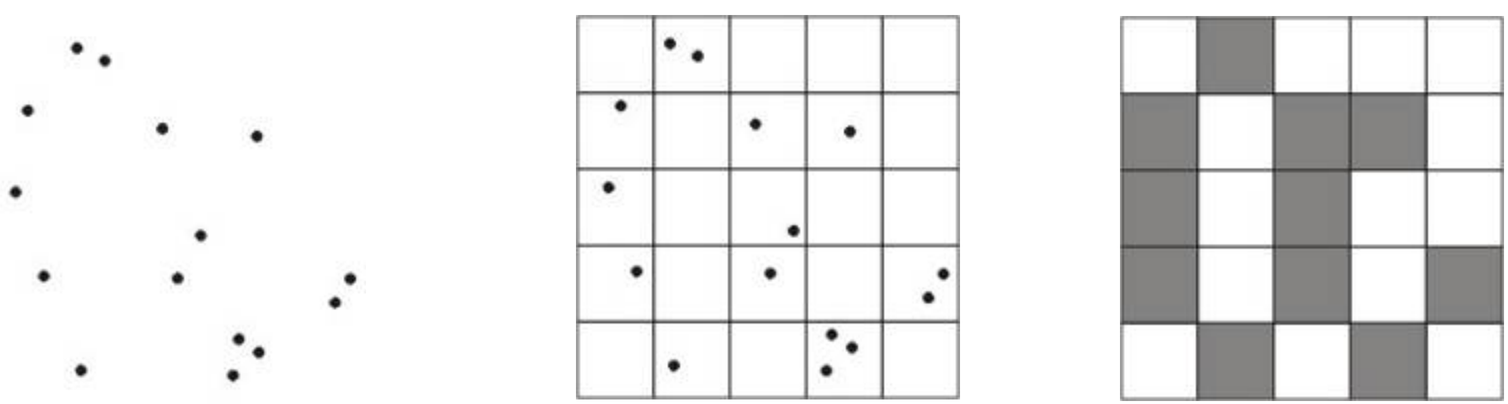
- Consider the capture of data about the boundary of a lake
  - ***Uncertain specifications***: The lake's boundary may not be completely specified, e.g.,
    - temporal variation in water's edge
    - lack of clarity in definition of lake (*vagueness*)
  - ***Uncertain measurements***: The location of the lake's boundary may be difficult to capture, e.g.,
    - Incorrect instrument calibration (*inaccuracy*)
    - Mistakes in using the instruments
    - Lack of detail in measurement (*imprecision*)
  - ***Uncertain transformations***: Transformation of the data may introduce further uncertainty, e.g.,
    - Measured points may be interpolated between to produce complete boundary

Typology of imperfection



Granularity and indiscernibility

- Granularity concerns the existence of “clumps” or “grains” in data, where individual element cannot be *discerned apart*
- Indiscernibility is often assumed to be an equivalence relation (reflexive, symmetric, and transitive)



Dimensions of data quality

- Data quality refers to the characteristics of a data set that may influence the decision based on that data set

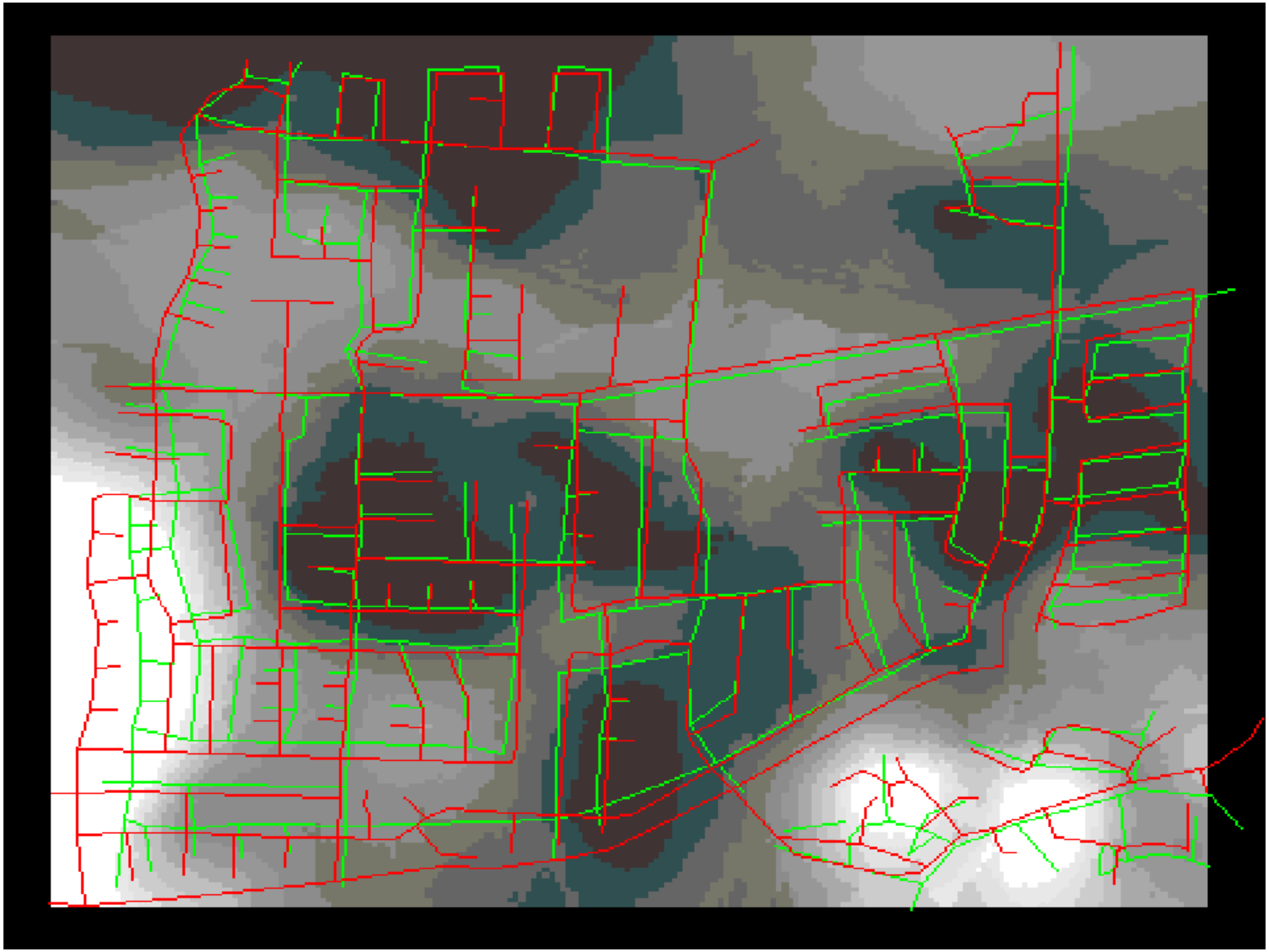
Element	**Concise definition**
accuracy	Closeness of the match between data and the things to which data refers
bias	Existence of systematic distortions within data
completeness	Exhaustiveness of data, in terms of the types of features that are represented in data
consistency	Level of logical contradictions within data
currency	How “up-to-date” data is
format	Structure and syntax used to encode data
granularity	Existence of clumps or grains within data
lineage	Provenance of data, including source, age, and intended use
precision	Level of detail or specificity of data
reliability	Trustworthiness of degree of confidence a user may have in data
timeliness	How relevant data is to the current needs of a user

Uncertainty in GIS

- GIS databases built from maps are not necessarily objective, scientific measurements of the world it is impossible to create a perfect representation of the world in a GIS database therefore all GIS data are subject to uncertainty
  - Uncertainty arise in every state of map production processes
  - uncertainty regarding what the data tell us about the real world a range of possible truths
  - that uncertainty will affect the results of analysis
  - <u>all GIS results should have confidence limits, "plus or minus"</u>

Uncertain in GPS

Uncertain in Lines



Uncertain in Lines

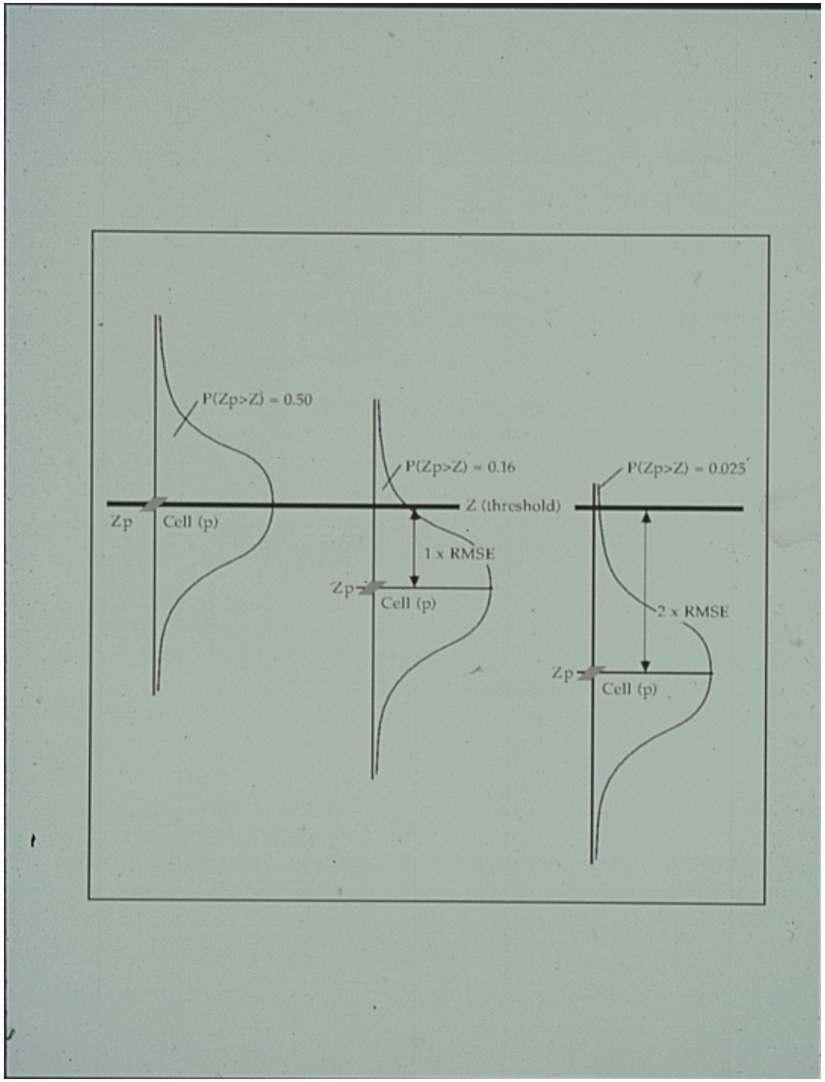
- It is an example of positional errors in two commercial street centerline databases of Goleta
  - the background fill is darkest where errors are smallest
    - note how the errors are often up to 100m
  - a problem if someone reports the location of a fire using one map, and a response is dispatched using the other map
  - the response vehicle could be sent to the wrong street note also how many streets are not in both databases
  - notice how errors persist over large areas
    - the error at one point is not independent of error at neighboring points
    - this is a general characteristic of error in GIS databases

Uncertainty in Lines

Uncertainty in DEMs

- USGS quality description
  - a DEM provides measurements of the elevation of the land surface at each grid point
  - errors are due to:
    - measurement of the wrong elevation at the grid point
    - measurement of the right elevation at the wrong location
    - any combination of these
    - it is impossible to determine which case applies
- the USGS provides simple quality statements for its DEMs
  - given as "root mean square error“
  - this is the square root of the average squared difference between recorded elevation and the truth
  - roughly interpreted as the average difference
  - e.g. many DEMs have RMSE of 7m
  - an error of 7m is common and errors of 10m, even 20m occur sometimes

Uncertainty in DEMs

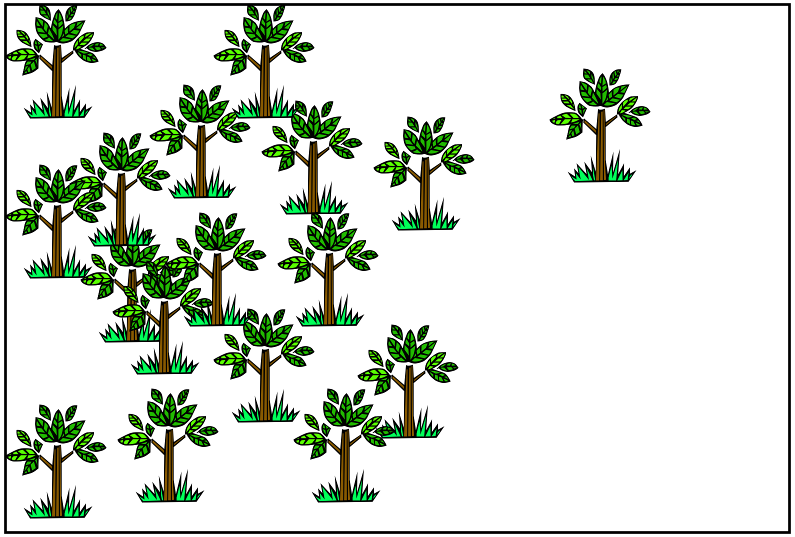


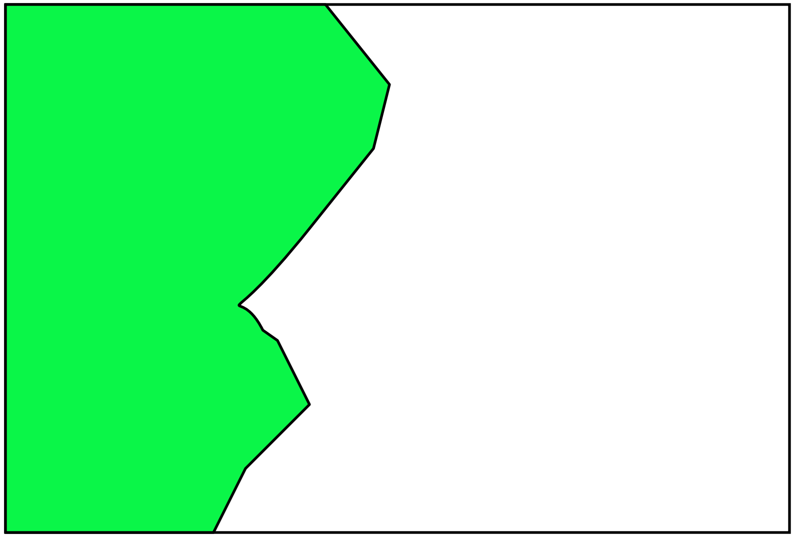
Uncertainty in Area-class Maps

- Nature of errors
  - area class maps show a class at every point
  - Typical examples include vegetation cover maps, soil maps, land use maps
  - they imply that class is uniform within areas, changes abruptly between areas
    - in fact both assumptions are not right
    - there should be variation within areas (heterogeneity)
    - there should be blurring across boundaries
  - area class maps have been described as "maps showing areas that have little in common, surrounded by lines that do not exist"

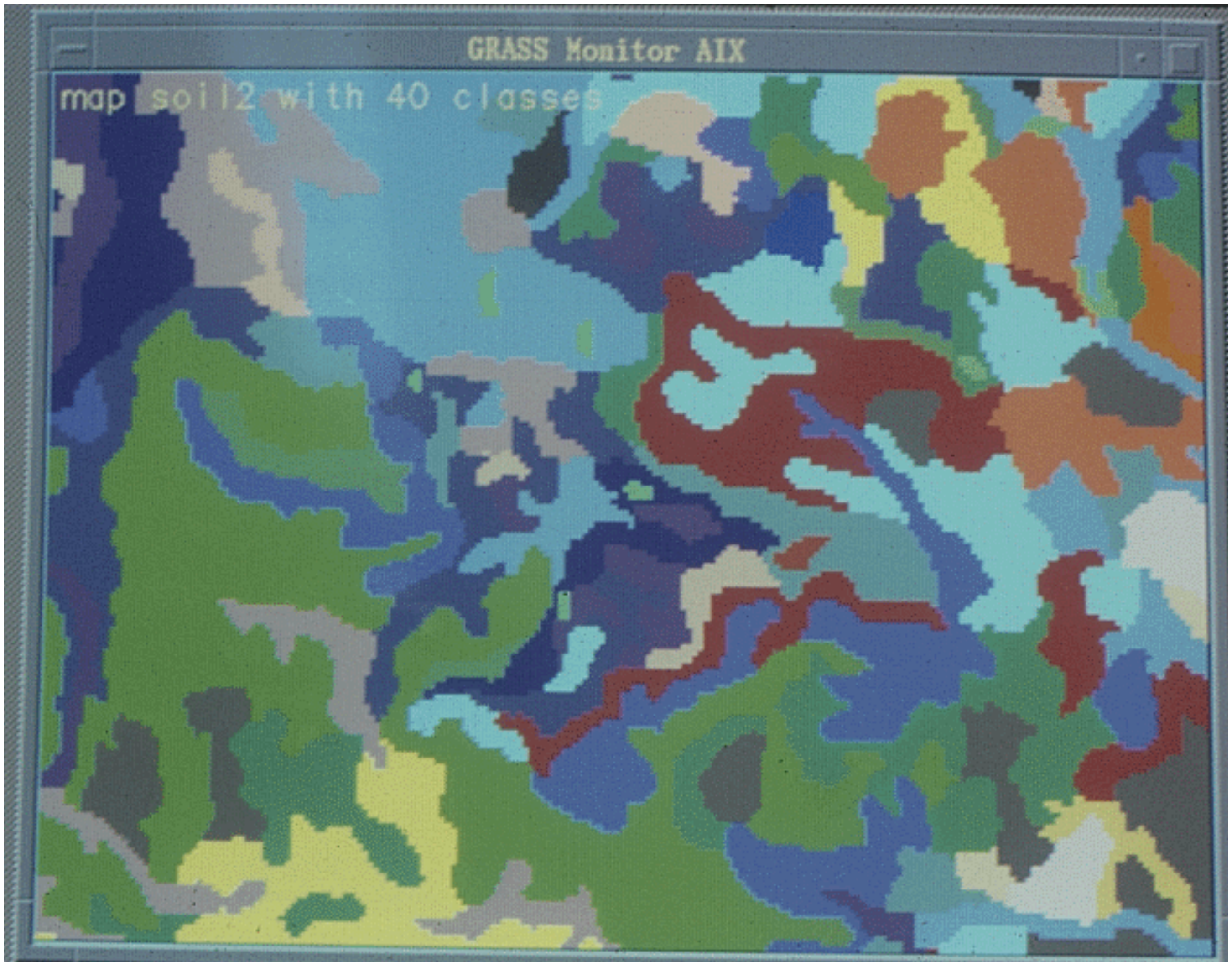
Uncertainty in Area-class Maps

- Attribute uncertainty (Forest vs. Ag)
- Positional uncertainty
- Definitional uncertainty
- Measurement uncertainty





Uncertainty in Area-class Maps



Uncertainty in Area-class Maps

- For yellow regions
  - let's assume the legend says this class is "80% sand, with 20% inclusions of clay“
  - this map is used for many purposes
  - some involve land use regulation
  - some involve taxation, compensation
  - in principle, all of these are uncertain if the map is uncertain
  - GIS applications are in deep trouble in court if it can be shown that regulations, taxes were based on uncertainty and that no effort was made to deal with that uncertainty

Strategies

- Producer of data must describe uncertainty
  - RMSE 7m
  - Metadata
  - Spatial Data Transfer Standard (STDS)
    - Positional accuracy
    - Attribute accuracy
    - Logical consistency (logical rules? polygons close?)
    - Completeness
    - Lineage
  - Not effective

- Monte Carlo Simulation strategy
  - Simulate equally probable versions of data

General Strategy for Uncertainty Evaluation: Integrated with Monte Carlo Simulation

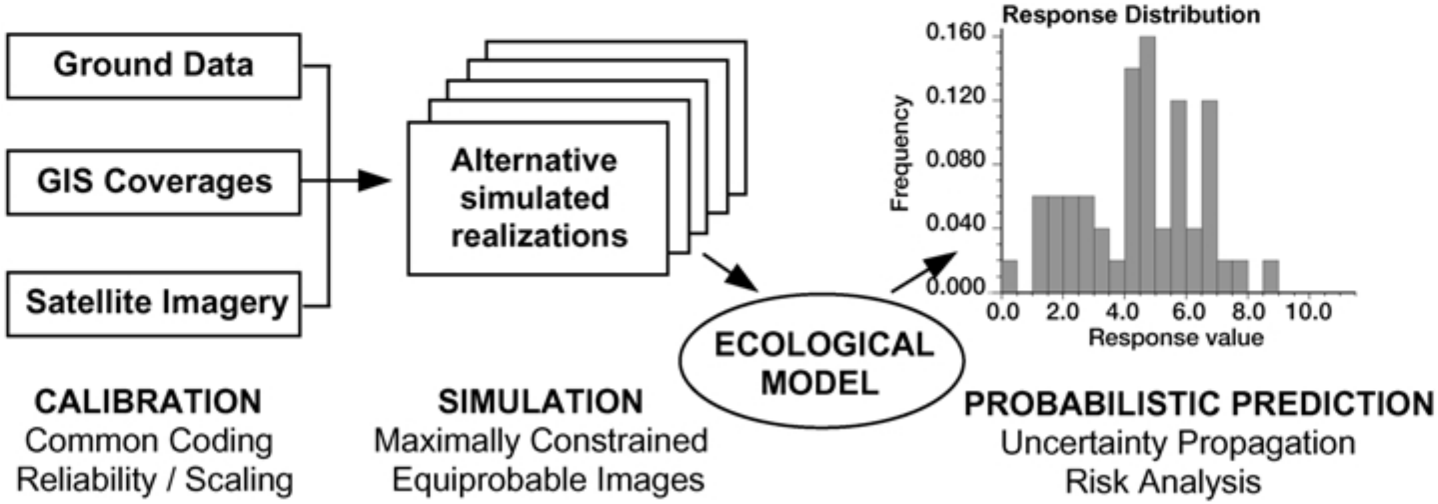
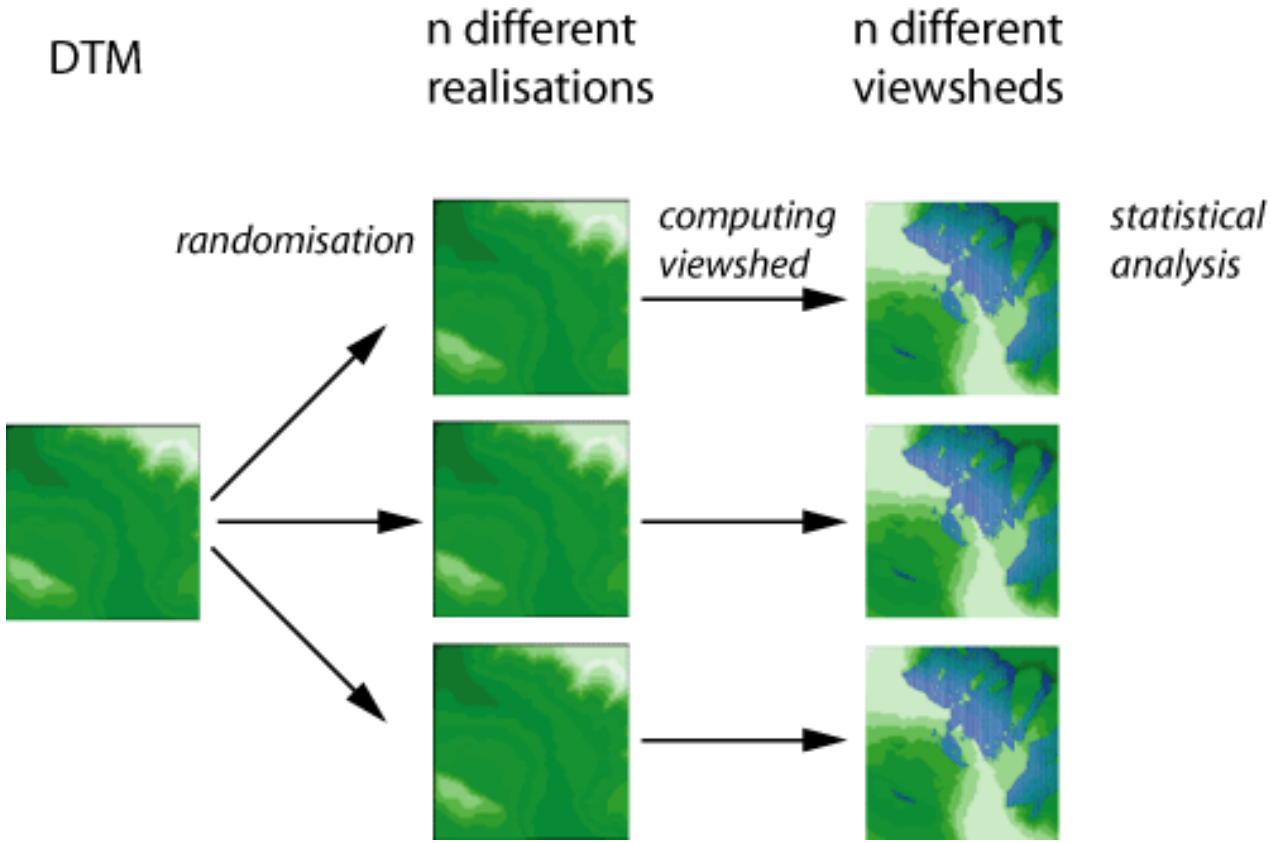


Image courtesy: Phaedon C. Kyriakidis and Jennifer L. Dungan

Uncertain Viewsheds

- ***Viewshed***: a region of terrain visible from a point or set of points
- ***Probable viewshed***:
  - Uncertainty arising through imprecision and inaccuracy in measurements of the elevation
  - Boundary will be crisp but its position uncertain
- ***Fuzzy viewshed***:
  - Uncertainty arising from atmospheric conditions, light refraction, and seasonal and vegetation effects
  - Boundary is broad and graded
  - Fuzzy regions are often used

Uncertain Viewsheds



- End of this topic