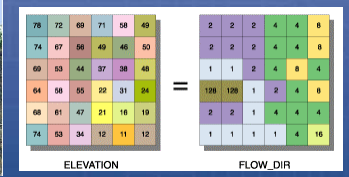
Geostatistical Analyst: An Introduction

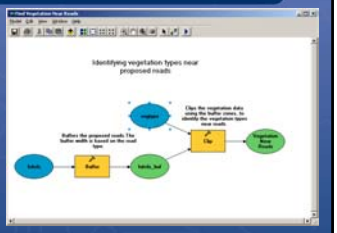
Katja Krivoruchko, ESRI

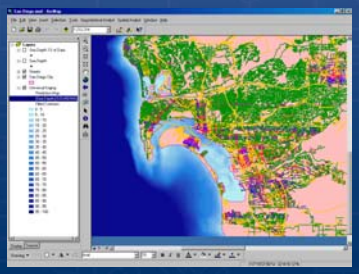
Outline

- Introduction
- Uncertainty
- About spatial data
- Where is geostatistics used
- Short Demo
- Geostatistical Analyst: Features
- A bit about Spatial Analyst's kriging

Introduction: Modeling







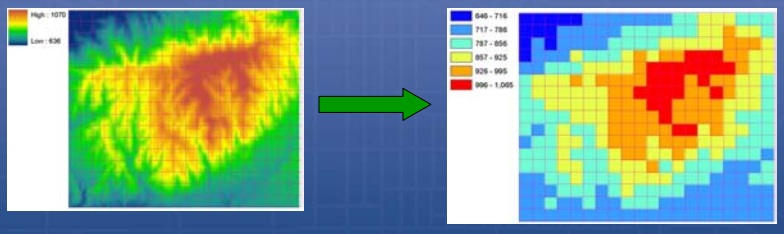
Uncertainty

- "All models are wrong. We make tentative assumptions about the real world which we know are false but which we believe may be useful" - George Box, 1976
- Uncertainty arises from attempting to represent real-world in a digital database
  - (in)accuracy
  - unknown phenomena
  - uncertain nature of the attribute
  - prediction uncertainty

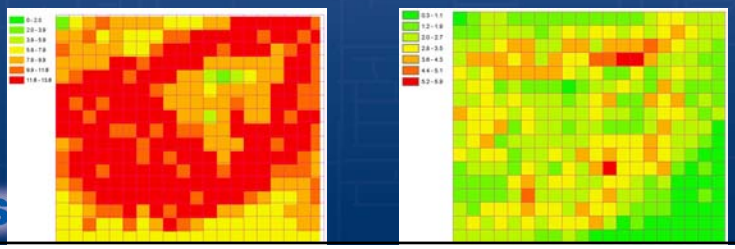
Uncertainty: Example

- It is safe to live with air pollutant concentration of 15 units or less
- My town has a measured concentration of 10 units
- I am ok, right?
- Yes: if error is +/- 3 units
- No: if error is +/- 6 units

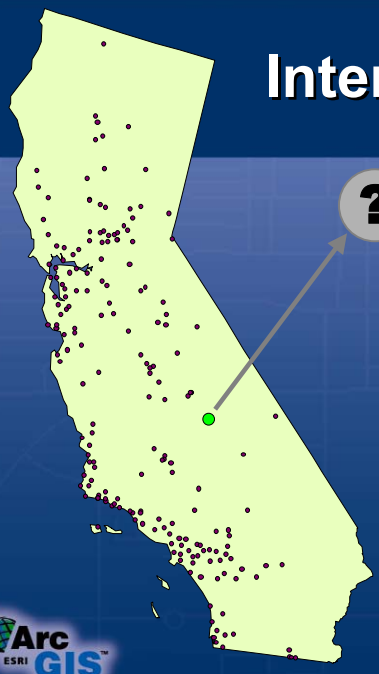
Uncertainty: Example



Uncertainty



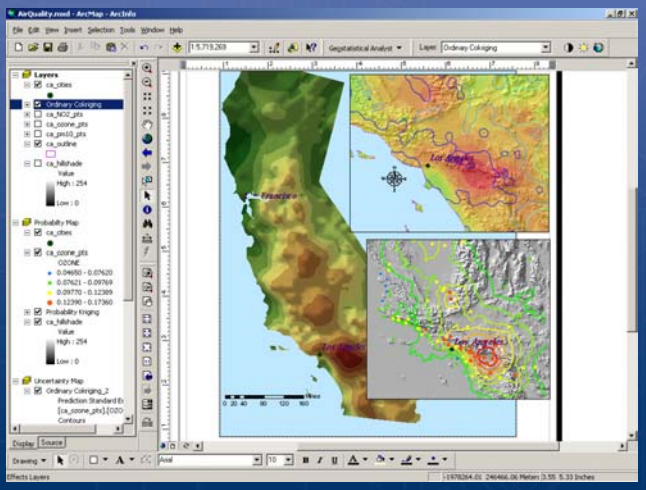
Interpolation



Latin verb *interpolare*

inter	meaning between,
polare	meaning to polish

Interpolation: Example



Interpolation: Data Reqs

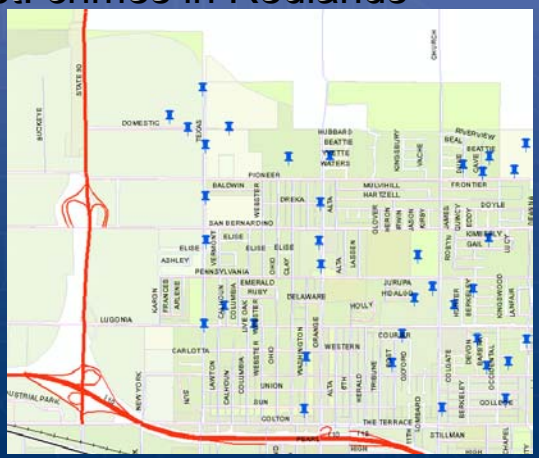
- Spatial correlation
  - data points closer together are more alike than those farther apart

if not, why interpolate?

- Continuous data
  - Examples include temperature, pollutant concentration, coded surface (1 for yes and 0 for no)

Interpolation: review

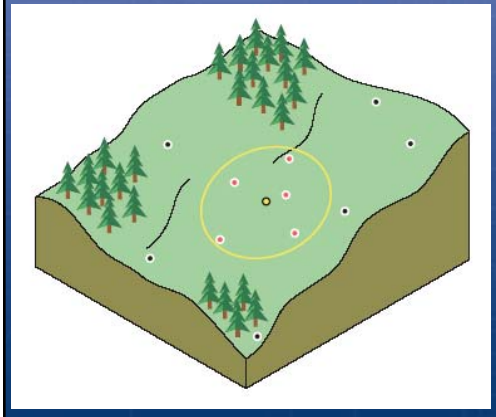
Dataset: crimes in Redlands



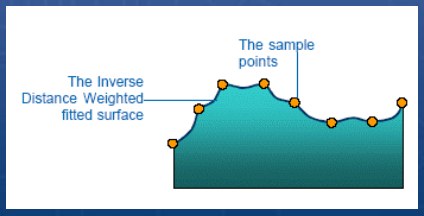
Interpolation - Methods

- Deterministic
  - Inverse Distance Weighted
  - Global polynomial
- Geostatistical
  - Kriging
  - Cokriging

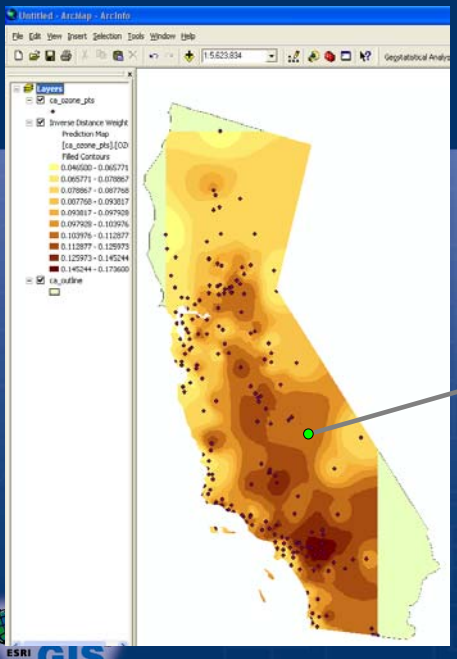
Inverse Distance Weighted



- Distance
- searching neighborhood



DEMO: Interpolation



Can we attach any certainty to this value predicted by IDW ?

?

Geostatistics

It is a set of models and tools developed for the statistical analysis of continuous data.

The predictions that are made by ***Geostatistical Analyst*** are accompanied by information on uncertainties.

Geostatistics: Example

- Ozone levels in California
  - Ozone measured in a few locations
  - What is the concentration in Redlands?
  - What is the probability of regulatory level being exceeded during year?
  - Where else would we place a monitoring station to increase prediction accuracy?

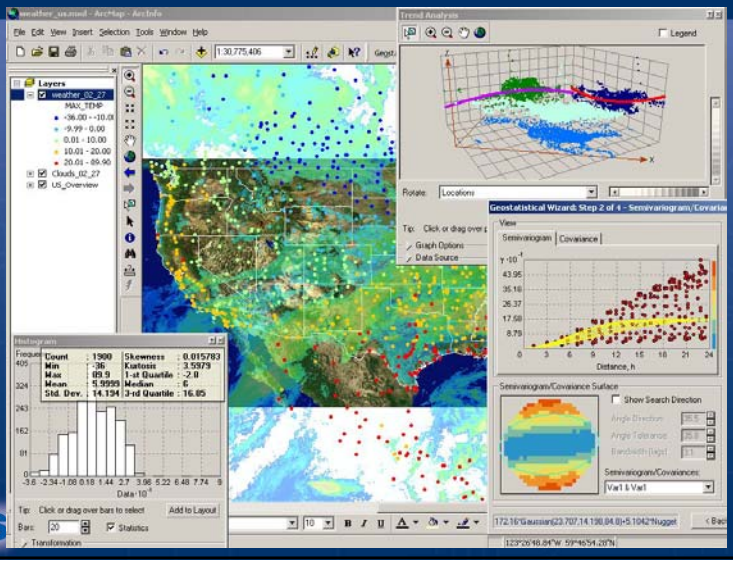
Geostatistics: Who Needs It?

- Meteorology
  - atmospheric data analysis
  - rainfall analysis
- Agriculture
- Environmental
  - fisheries
  - air/soil/groundwater pollutants
  - probability that concentration has been exceeded
- Geology
  - petroleum industry
  - mining industry
  - soil chemistry

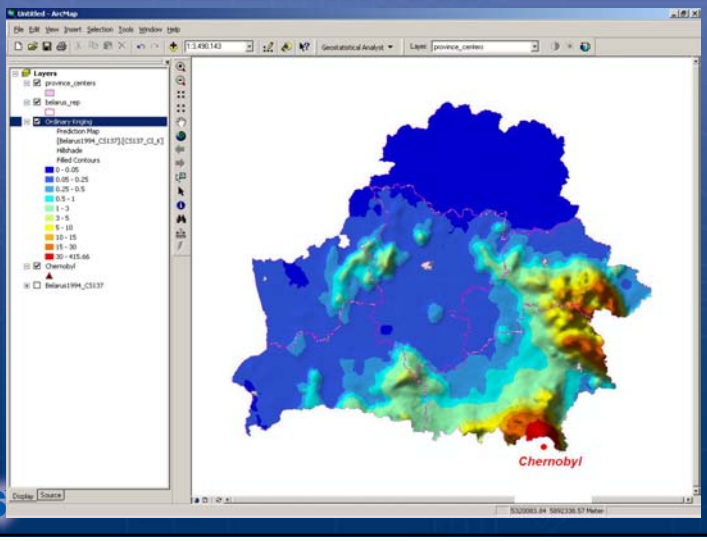
Who Uses GA?

- USGS/ Geological firms/agencies
- USDA/ Agricultural firms/agencies
- US EPA/ Environmental organizations
- US Fish and Wildlife Service
- CDC/ Public health organizations
- Over 700 universities
- Flood Control Districts
- Engineering Firms

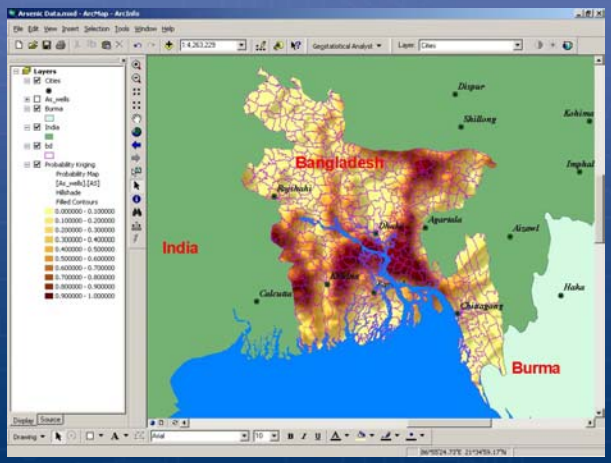
Weather Monitoring



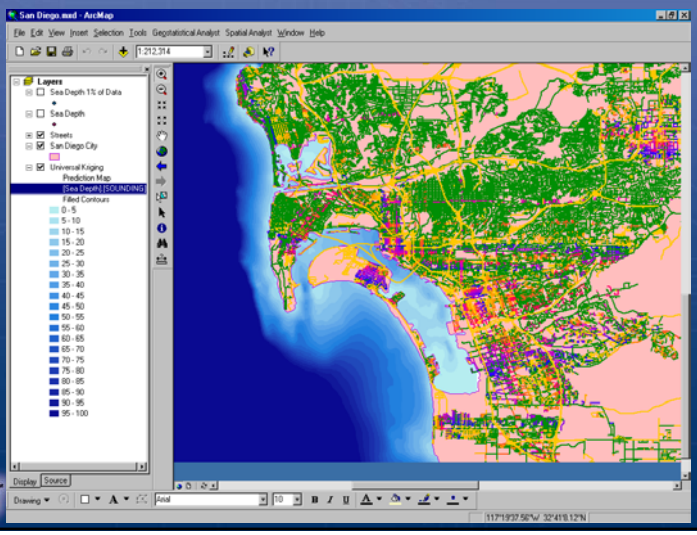
Radioactive Contamination



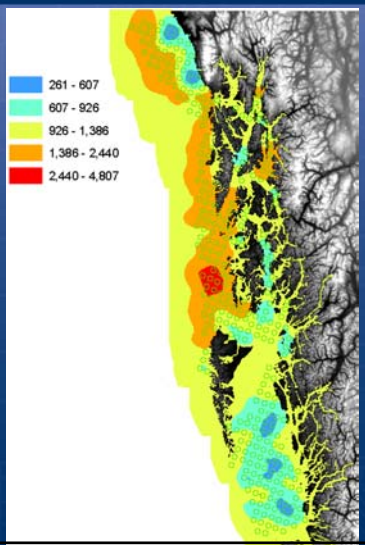
Groundwater Contamination



Bathymetry



Fish Abundance



What is Kriging?

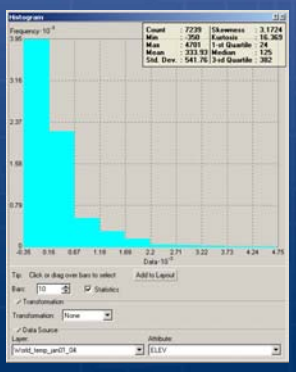
- A weighted, moving-average estimation technique based on Geostatistics that uses the spatial correlation of point measurements to estimate values at adjacent, unmeasured points
- Associates uncertainty with the predictions

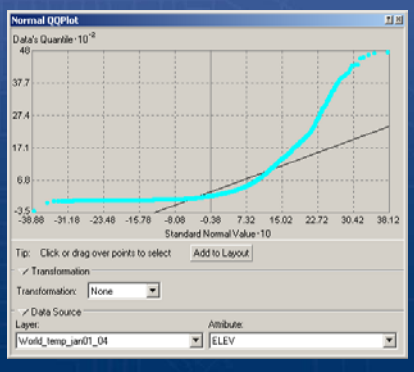
Why Is It Called Kriging?

- **DG Krige,** ore exploration, in 50-ies.
- **Matheron,** published his work in the 60-ies and coined the term *kriging*
- **Gandin,** first developed a technique for use in synoptic meteorology, before Krige and Matheron.
  - Gandin called this method, **optimum interpolation**.

Kriging: Data requirements

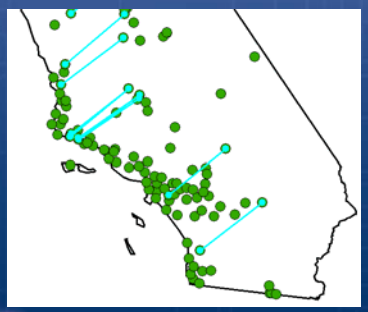
- Gaussian distribution
  - "Normal," or bell-shaped distribution

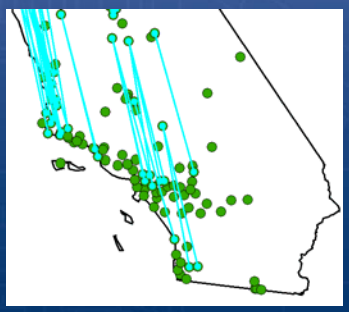




Kriging: Data requirements

- Stationarity
  - Statistical properties do NOT depend on exact locations





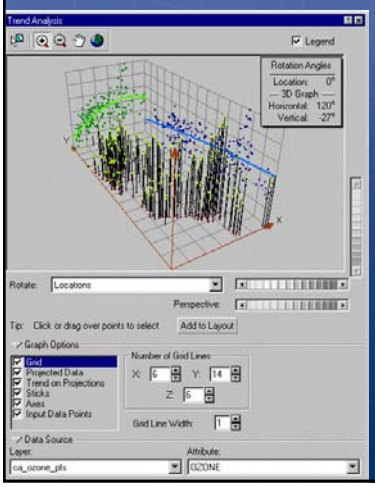
Geostatistical Analyst

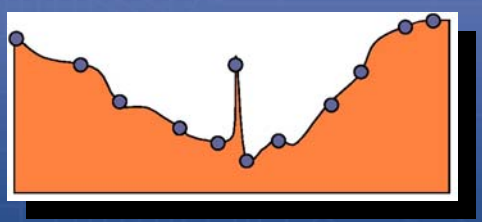
- Exploratory Spatial Data Analysis (ESDA)
- Geostatistical Methods
  - Kriging/cokriging
    - Provide prediction uncertainty
- Deterministic Methods
  - IDW, Polynomial, Radial Basis

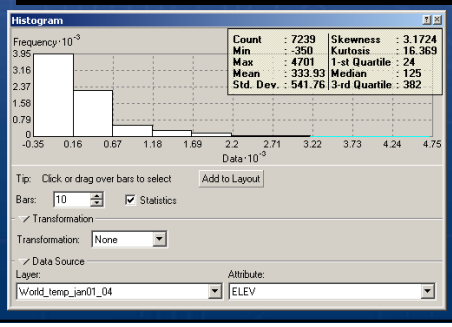
GA: ESDA Tools

- Geostatistical Analyst provides tools to explore the distribution of the data
  - stationarity
  - normality
  - local and global outliers
  - trends
  - spatial dependence

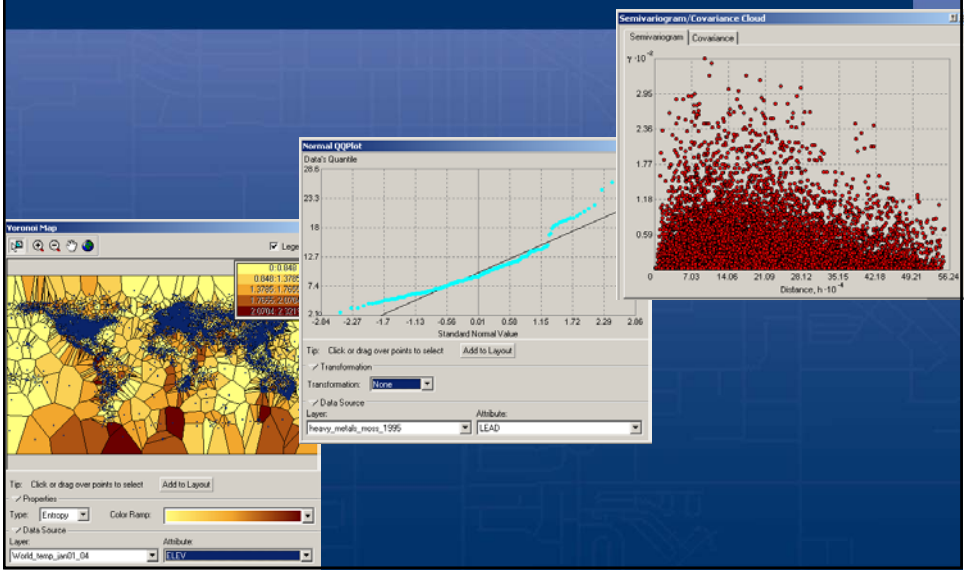
GA: ESDA Tools







GA: ESDA Tools



SHORT DEMO

- Create surface (default parameters)
- Explore data
  - Normality
  - Stationarity
  - Autocorrelation
- Compare Prediction Surfaces
- Create Prediction Standard Error Map

Who Reviewed GA?

- GA's implementation of geostatistics is based on Noel Cressie's *Statistics for Spatial Data*
- Dr. Noel Cressie was the consultant
- Dr. Cressie's PhD students were early beta testers

What's with Spatial Analyst?

GA vs SA

- SA created for spatial analysis
  - map algebra
  - cost analysis
  - watershed calculations
  - combinational operators
- GA created for spatial data analysis
  - statistical analysis of input data
  - statistical methods for surface creation
  - estimation of prediction uncertainty

Spatial Analyst's Kriging

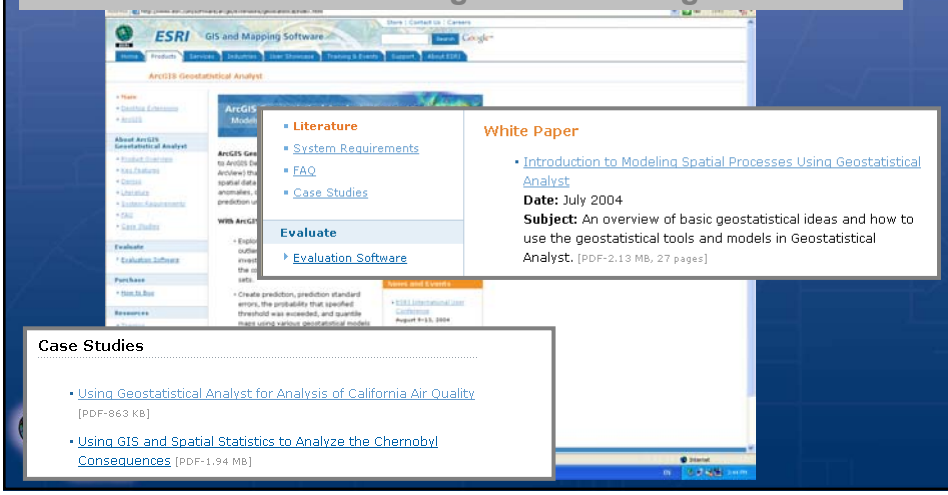
- No ESDA tools
- No detrending (to ensure stationarity)
- No transformation (for normality)
- "Black box"
  - few parameters
  - no interactive graphic interface
- No diagnostics

Spatial Analyst's Kriging

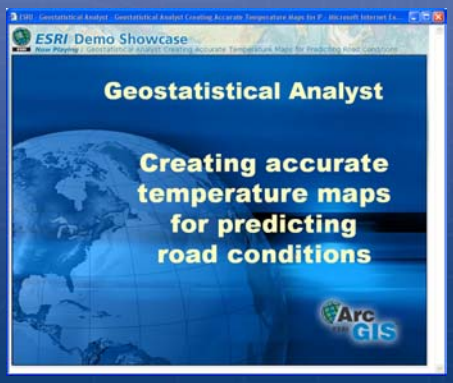
- Rule of Thumb:
  - Use SA when surface created for data display purposes
  - Use GA when surface created for decision-making

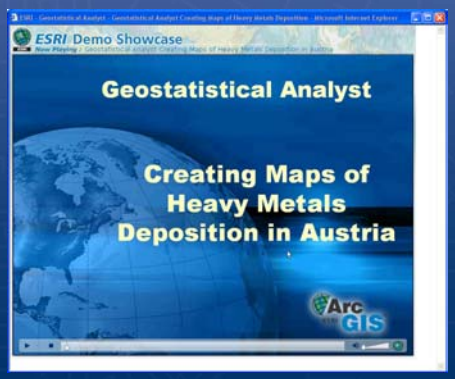
For More Details

www.esri.com/software/arcgis/extensions/geostatistical



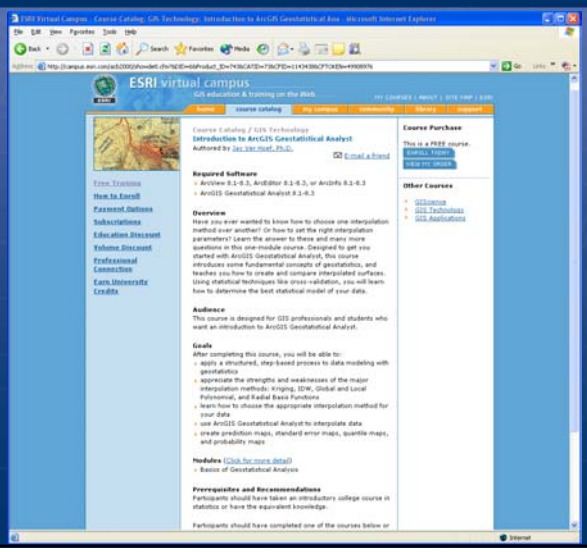
For More Details





ArcGIS

For More Details



ArcGIS

Points to ponder on ….

"Find out how useful the data are and not if they are completely free of errors"

Fotheringham et al.,

"Don't tell me how right you want to be, tell me how wrong you can be"

Ian Lerche