Spatial Interpolation & Geostatistics

Kriging – Step 1

- Describe spatial variation with Semivariogram
- "Point cloud"

Kriging – Step 2

- Summarize spatial variation with a function
- Several choices possible; curve fitting defines different types of Kriging (circular, spherical, exponential, gaussian, etc.)

Key features of fitted variogram:
- Nugget: semivariance at d = 0
- Range: d at which semivariance is constant
- Sill: constant semivariance beyond the range
Kriging – Part II

- **Goal**: predict values where no data have been collected
  - Relies on first establishing:
    - **DEPENDENCY**: $z$ is, in fact, correlated with distance
    - **STATIONARITY**: $z$ values are stochastic (except for spatial dependency they are randomly distributed) and have no other dependence – use "detrending" or transformation tools if not Gaussian
    - **DISTRIBUTION**: works best if data are Gaussian. If not they have to first be made close to Gaussian.

ESRI Geostatistical Analyst Products

- **Map types**:
  - **Prediction** – contours of interpolated values
  - **Prediction Standard Errors** – show error distribution, as quantified by minimized RMS error (see below)
  - **Probability** – show probability of values exceeding a specified threshold
  - **Quantile** – show where thresholds overestimate or underestimate predictions

Some Kriging Products

- **Prediction map** – interpolated values
- **Probability map** – showing where critical values exceeded

Figures from ESRI "Using Geostatistical Analyst"

Dark orange and red indicate a probability greater than 62.5% that radiocesium contamination exceeds the critical threshold in forest berries.

Prediction map of radioceasium soil contamination levels in the country of Belarus after the Chernobyl nuclear power plant accident.
Kriging – Part II

Goal: predict values where no data have been collected
- Relies on first establishing:
  - DEPENDENCY – z is, in fact, correlated with distance
  - STATIONARITY – z values are stochastic (except for spatial dependency they are randomly distributed) and have no other dependence – use "detrending" or transformation tools if not Gaussian
  - DISTRIBUTION – works best if data are Gaussian. If not they have to first be made close to Gaussian.

1. SPATIAL DEPENDENCY
- Test with semivariogram & cross-validation plots
  - Violates Stationarity – detrend first

Spatial Dependence:
Semivariogram and Semivariogram Surface

Spatial Dependence:
Cross-Validation Diagnostic
- Use a subset of the data to test measured vs. predicted values
Kriging – Part II

- **Goal:** predict values where no data have been collected
- **Relies on first establishing:**
  - **DEPENDENCY** – z is, in fact, correlated with distance
  - **STATIONARITY** – z values are stochastic (except for spatial dependency they are randomly distributed) and have no other dependence – use "detrending" or transformation tools if not Gaussian
  - **DISTRIBUTION** – works best if data are Gaussian. If not they have to first be made close to Gaussian.

2. **STATIONARITY - Randomness**

- Data variance and mean is the same at all localities (or within a neighborhood of nearest points); data variance is constant in the neighborhood of investigation
- Correlation (covariance) depends only on the vector that separates localities, not exact locations, number of measurement or direction

California Ozone Demo.

- Data in “Geostat_demo” folder

ArcGIS Kriging Processing Steps

1. Add and display the data
2. Explore the data’s statistical properties
3. Select a model to create a surface – make a prediction map!
4. Assess the result
5. Compare to other models
Data Exploration

1. Examine the distribution – normal (Gaussian)?
   Transformation to normal required?
   - Histograms and QQPlots
2. Identify trends, if any
   - Trend Analysis
3. Understand spatial autocorrelation and directional influences
   - Semivariogram analysis

Data Exploration: Examine the Distribution

- Normal (Gaussian) distribution? (central value, spread, symmetry; mean and median the same?)
  Transformation to normal required?
  - Histogram tool, QQPlot tool (compare real and standard normal distributions)

Data Exploration: Identify Trends, If Any

- Underlying trends affect Kriging assumption of randomness – remove and work with “residuals”
  - Trend Analysis tool

Data Exploration: Spatial Autocorrelation & Directional Influences

- Variogram Analysis:
  - Look for correlation with distance
  - Look for directional trends among pairs of points
    - Semivariogram/Covariance Cloud tool

Exhibits Spatial Autocorrelation

Directional Influence shown here – pairs in NW-SE direction show largest covariance over shortest distances
ArcGIS Kriging Processing Steps

1. Add and display the data
2. Explore the data's statistical properties
3. Select a model to create a surface – make a prediction map!
4. Assess the result
5. Compare to other models

Mapping Ozone Concentration

1. Incorporate results of Data Exploration into Model selection
   - This example:
     - Remove underlying trends discovered during data exploration that have a rational explanation. (Analysis is then performed on residuals and trend surface is added back into final surface) = "Detrending"
     - Remove directional trends between pairs of points in certain directions closer points are more alike than in other directions = "anisotropy removal"

ArcGIS Kriging Processing Steps

1. Add and display the data
2. Explore the data's statistical properties
3. Select a model to create a surface – make a prediction map!
4. Assess the result – Cross Validation Plots
5. Compare to other models

Mapping Ozone Concentration – Interpolation & Cross Validation

2. Define search neighborhood for interpolation (c.f. I.D.W.)
   - Use a search ellipse (or circle) to find nearest neighbors; specify radii of ellipse, min. & max. number of points per sectors
3. Examine Cross Validation plot
   - Predicted vs. Measured for subset(s) of the data
   - "Mean error" should be close to zero
   - "RMS error" and "mean standardized error" should be small
   - "RMS standardized error" should be close to one.
Comparing Model Results

- Cross validation comparisons:
  - "Mean error" should be close to zero
  - "RMS error" and "mean standardized error" should be small
  - "RMS standardized error" should be close to one.

Probability Mapping with Indicator Kriging

- Task: Make a map that show the probability of exceeding a critical threshold, e.g. 0.12 ppm ozone for an 8 hr. period
- Technique:
  - Transform data to a series of 0s and 1s according to whether they are above or below the threshold
  - Use a semivariogram on transformed data; interpret indicator prediction values as probabilities