Outline

- What is geostatistics?
- What is Geostatistical Analyst?
- Spatial autocorrelation
- Geostatistical Wizard and geoprocessing tools
- Where is it used?
- Demonstrations
- Tips and Tricks
- Conclusion
- Questions
Sessions of note…

Tuesday
- **ArcGIS for Geostatistical Analyst: An Introduction** (Tues 8:30-9:45 SDCC Rm17B)
- Creating Surfaces from Various Data Sources (Tues 3:15-4:30 SDCC Rm09)
- Concepts and Applications of Kriging (Tues 3:15-4:30 SDCC Rm17B)
- *Empirical Bayesian Kriging and EBK Regression Prediction – Robust Kriging as GP Tools* (Tues 5:30-6:15 Th07)

Wednesday
- *Choosing the Best Kriging Model for Your Data* (Wed 11:30-12:15 SDCC Th07)
- **ArcGIS for Geostatistical Analyst: An Introduction** (Wed 1:30-2:45 SDCC Rm17B)

Thursday
- Geostatistics in Practice: Learning Kriging Through Examples  (Thurs 8:30-9:45 SDCC Rm10)
- *Surface Interpolation in ArcGIS* (Thurs 10:30-11:15 SDCC Th07)
- *Performing Polygon-to-Polygon Predictions using Areal Interpolation* (Thurs 11:30-12:15 Th07)
- Creating Surfaces from Various Data Sources (Thurs 3:15-4:30 SDCC Rm09)
What is geostatistics?

- is a class of statistics used to analyze and predict values associated with spatial phenomena.
- it incorporates the spatial coordinates of the data
- Has evolved to not only provide
  - interpolated values, but also
  - measures of uncertainty
ArcGIS for Geostatistical Analyst

Too expensive to measure everywhere, however, we want to know values everywhere.
What is a semivariogram?

Semivariogram(distance $h$) = $0.5 \times$ average $(\text{value}_i - \text{value}_j)^2$

- Range = separation distance between pairs
- Sill = plateau the variogram reaches at the range
- Nugget = sampling error and short scale variability
Spatial autocorrelation
Geostatistical Analyst – What is it?

Provides a complete set of spatial analytical tools that range from techniques to explore the original data to post-processing evaluation of data and predictions uncertainties.

- **Geoprocessing tools**
  - Use within ArcMap / Pro / Server
  - Modelbuilder
  - Scripting
GP tool

Kernel Interpolation with Barriers

Eric Krause
Geostatistical Analyst – Geoprocessing tools
Geostatistical Analyst – What is it?

Provides a complete set of spatial analytical tools that range from techniques to explore the original data to post-processing evaluation of data and predictions uncertainties.

- **Wizard**

  - is a dynamic set of pages that is designed to guide you through the process of constructing and evaluating the performance of an interpolation model.
Geostatistical Wizard

Kernel Interpolation with Barriers

Eric Krause
ESDA
Exploratory Spatial Data Analysis

• Where is the data located?
• What are the values at the data points?
• How does the location of a point relate to its value?
Exploratory Spatial Data Analysis (ESDA)
What is kriging?
- It is a geostatistical interpolation technique
- that models the spatial correlation of point measurements
- to estimate values at unmeasured locations.
- Associates uncertainty with the predictions
What is kriging?

Tuesday
- Concepts and Applications of Kriging (Tues 3:15-4:30 SDCC Rm17B)
- *EBK and EBK Regression Prediction – Robust Kriging as GP Tools* (Tues 5:30-6:15 Th07)

Wednesday
- *Choosing the Best Kriging Model for Your Data* (Wed 11:30-12:15 SDCC Th07)

Thursday
- Geostatistics in Practice: Learning Kriging Through Examples (Thurs 8:30-9:45 SDCC Rm10)
- *Performing Polygon-to-Polygon Predictions using Areal Interpolation* (Thurs 11:30-12:15 Th07)
Geostatistical Wizard - Empirical Bayesian Kriging

General Properties
- Subset Size: 100
- Overlap Factor: 1
- Number of Simulations: 500
- Output Surface Type: Prediction
- Transformation Type: Empirical
- Semivariogram Type: Whittle

Semivariograms
- Partial Sill
- Range

Density (10^1) Simulations at (-1980920, 156542)

Nugget
- Transformation

Neighborhood Type
- Standard Circular
- Maximum Neighbors: 15
- Minimum Neighbors: 10
- Sector Type: 1 Sector
- Angle: 0
- Radius: 90085.7106446951

Identify Result
- X: -1980920
- Y: 156542

< Back  Next >  Finish

Geostatistical Wizard
Eric Krause
More ESDA
Interpolation workflow

- ESDA
- Interpolate
- Goodness of fit
Why use ESRI’s Geostatistical Analyst?

- Search neighborhood
  - Sectors
  - Smooth
- Chordal distance
- Cross validation
- Error maps
- Interactive Variography
- Barriers
- Simulations
Search neighborhood - Smooth
Unlike smoothing the output, this method modifies the weights.
Search neighborhood - Standard

- 2 per sector
- 8 closest
Chordal distances
Only for EBK and EBK Regression Prediction

- Automatically kicks in when data are in GCS
- The chordal distance between any two points is the straight-line distance that connects the two points.
- This line will go through the earth rather than along its surface.

Distance between LA and New York
Geodesic  = 3,939.1 km
Chordal    = 3,877.0 km
Difference = 62.1 km (1.5%)
Chordal distances
Only for EBK and EBK Regression Prediction
Cross validation / Validation
Output surfaces

Prediction

Standard error of prediction

Probability that rainfall exceeds 900mm
Interactive Wizard

Geostatistical Wizard - Kriging - Semivariogram/Covariance Modeling

Geostatistical Wizard - Kriging - Cross validation

Summary

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.000020951329288017</td>
<td>0.005987875221249</td>
<td></td>
</tr>
<tr>
<td>Root-Mean-Square</td>
<td>0.0116929592502132</td>
<td>1.69736122980455</td>
<td></td>
</tr>
<tr>
<td>Mean Standardized</td>
<td>0.0046606084250038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root-Mean-Square Standardized</td>
<td>1.69736122980455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Standard Error</td>
<td>0.0046606084250038</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cross validation is a “leave one out” method that allows you to determine how well your interpolation model fits your data. Cross validation works by removing a single point from the dataset and using all remaining points to predict to the location of the point that was removed. The predicted value is then compared to the measured value, and many statistics are generated to determine the accuracy of the prediction.
Interactive Wizard

- Optimize model
- Function Type: Covariance
- Lag Size: 6.12540010014148
- Number of Lags: 12
- Nugget: Enable
- Nugget Value: 0.187500971494174

Buttons: < Back, Next, Finish
Barriers
Gaussian Geostatistical Simulations
Create multiple versions (realizations) of a surface to perform risk analysis.

- Any realization might be the “real” thing!
EBK Regression Prediction

Eric Krause
Geostatistical layers

Eric Krause
Where is GA used?

- Anyone who needs to statistically explore data and create surfaces for a number of variables will benefit from this statistical software package.

- Some of the various fields that use ArcGIS Geostatistical Analyst include:
  - agriculture,
  - geology,
  - meteorology,
  - hydrology,
  - archaeology,
  - forestry,
  - oceanography,
  - fishery,
  - health care, and
  - environmental studies.
Tips & Tricks

• Use Mask when creating a raster
  - 8700 pixels inside (55,000 outside) Japan (6 1/2 times)
Tips & Tricks

- Subset of the data
  - SubsetFeatures GP tool
  - Selection
Conclusions

https://geonet.esri.com/

GeoNet
The Esri Community

IDW interpolation method
Please Take Our Survey on the Esri Events App!

Download the Esri Events app and find your event

Select the session you attended

Scroll down to find the survey

Complete Answers and Select “Submit”
Find “optimal” script

```python
iterationsList = [30, 100, 500, 1000]
subsetSizeList = [100, 200, 500]
overlapList = [1, 3, 5]
semivariogram_list = [["POWER",'NONE'], ['LINEAR','NONE'], ['THIN_PLATE SPLINE','NONE'],
                      ['EXPONENTIAL_DETRENDED','EMPIRICAL'], ['EXPONENTIAL','EMPIRICAL'],
                      ['K_BESSSEL_DETRENDED','EMPIRICAL'], ['K_BESSSEL','EMPIRICAL'],
                      ['WHITTLE','EMPIRICAL'], ['WHITTLE_DETRENDED','EMPIRICAL']]

try:
    for iterations in iterationsList:
        for i in range(0, len(semivariogram_list)):
            for subsetSize in subsetSizeList:
                for overlap in overlapList:
                    t1 = time.time()
                    result = arcpy.EmpiricalBayesianKriging_ga(inPC, inField, outLyr, outRas, cellsize, 
                        transf, subsetSize, overlap, iterations, 
                        sn, 'PREDICTION', '', '', '', svg)
                    t2 = time.time()
                    tebk = t2-t1
                    cv = arcpy.CrossValidation_ga(outLyr)
                    rmseValue = cv.rootMeanSquare
                    rmsStd = cv.rootMeanSquareStandardized
                    toWriteList = [rmseValue, rmsStd, svg, transf, tebk, subsetSize, overlap, iterations]
                    outFile.writerow(toWriteList)
                    arcpy.Delete_management(outLyr)

except:
    print ('Tool execution FAILED')
    print (arcpy.GetMessages())
```