Concepts and Applications of Kriging

Eric Krause

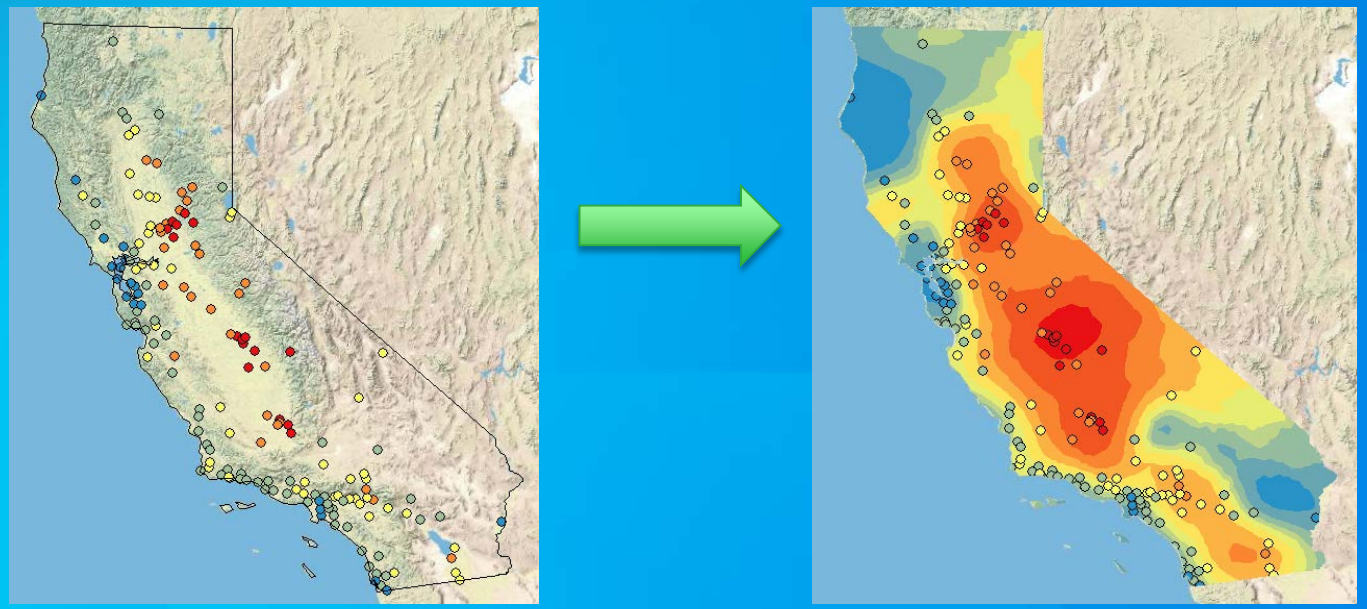
Konstantin Krivoruchko

Outline

- Introduction to interpolation
- Exploratory spatial data analysis (ESDA)
- Using the Geostatistical Wizard
- Validating interpolation results
- Empirical Bayesian Kriging
- Areal Interpolation
- Questions

What is interpolation?

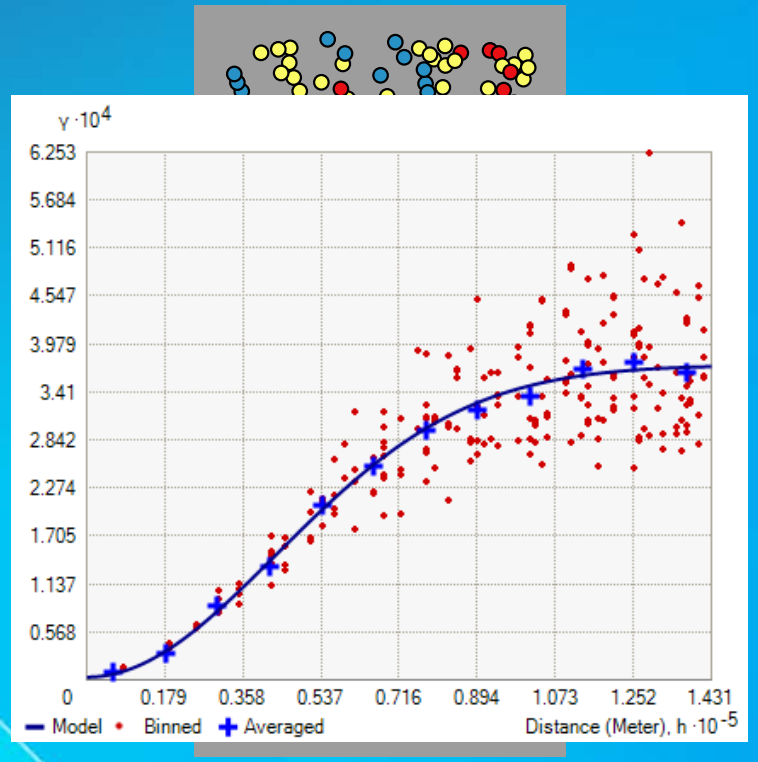
- Predict values at unknown locations using values at measured locations
- Many interpolation methods: kriging, IDW, LPI, etc

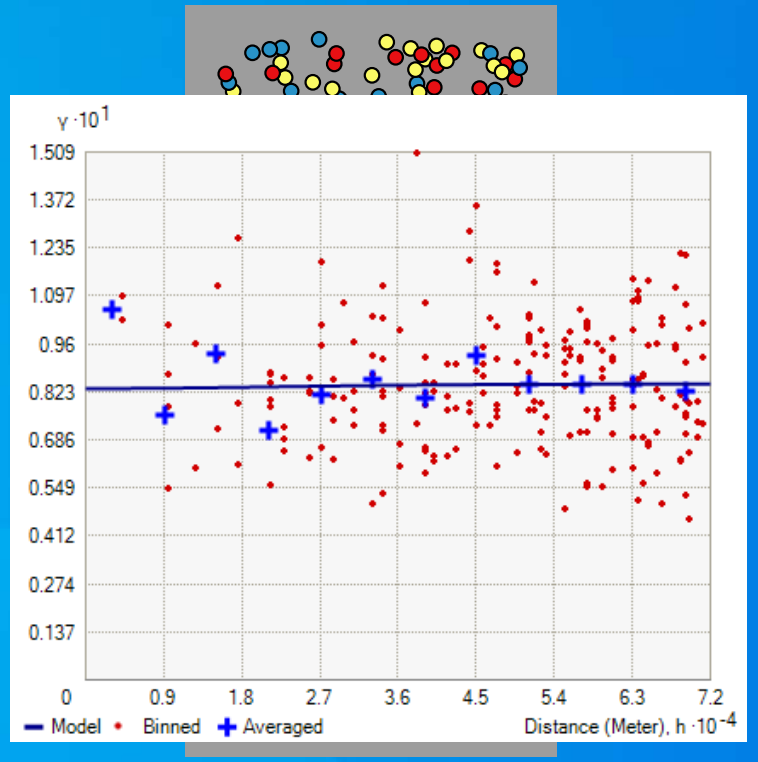


What is autocorrelation?

Tobler’s first law of geography:

"Everything is related to everything else,
but near things are more related than distant things."





Demo

Geostatistical Wizard

Eric Krause

What is kriging?

- Kriging is the optimal interpolation method if the data meets certain conditions.
- What are these conditions?
  - Normally distributed
  - Stationary
  - No trends
- How do I check these conditions?
  - Exploratory Spatial Data Analysis (ESDA)

What is an “optimal” interpolator?

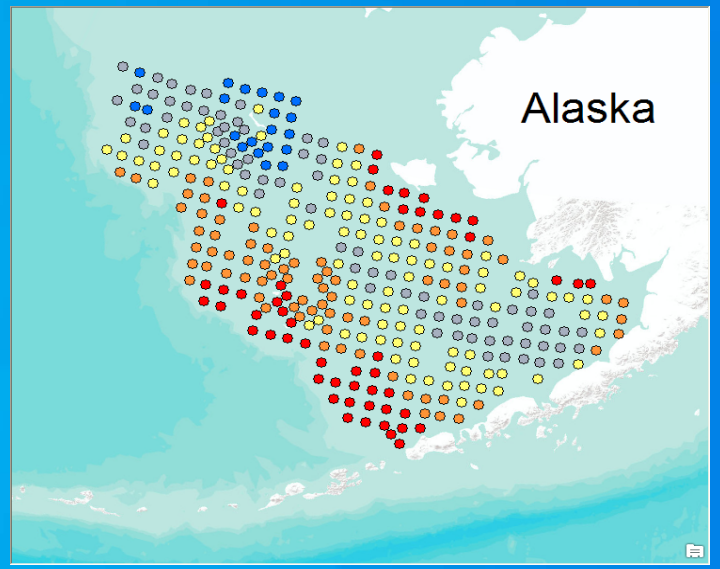
- Estimates the true value, on average
- Lowest expected prediction error
- Able to use extra information, such as covariates
- Filters measurement error
- Can be generalized to polygons (Areal interpolation, Geostatistical simulations)
- Estimates probability of exceeding a critical threshold

Geostatistical workflow

1. Explore the data
2. Choose an interpolation method
3. Fit the interpolation model
4. Validate the results
5. Repeat steps 2-4 as necessary
6. Map the data for decision-making

Exploratory Spatial Data Analysis

1. Where is the data located?
2. What are the values of the data points?
3. How does the location of a point relate to its value?



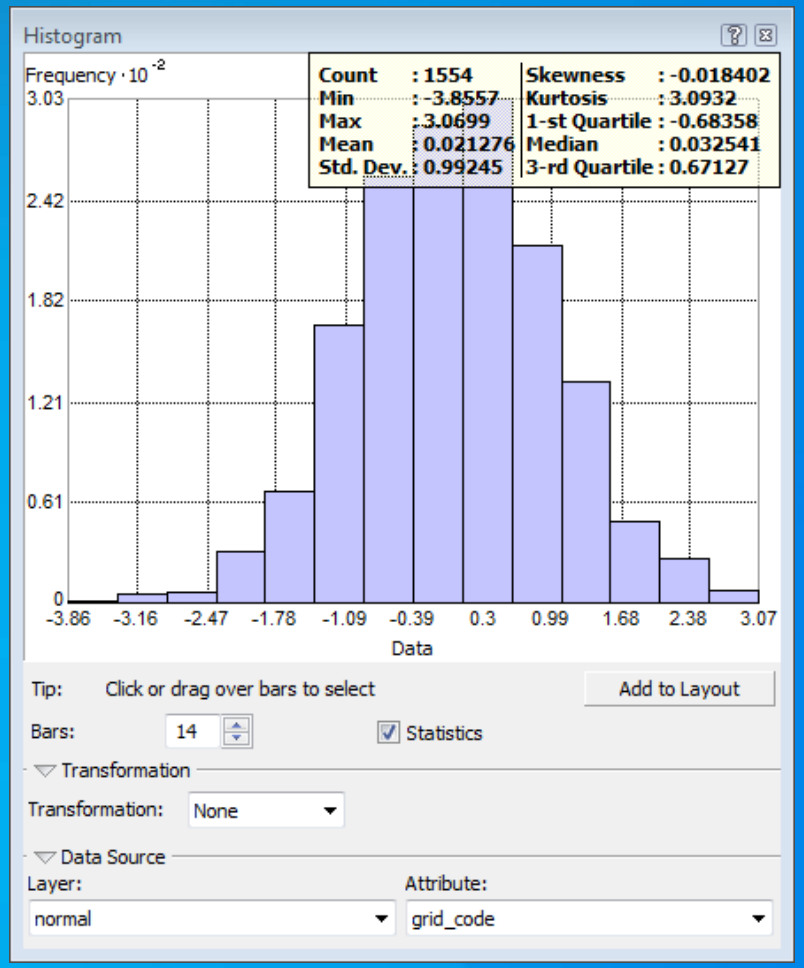
Does my data follow a normal distribution?

- How do I check?
  1. Histogram
     - Check for bell-shaped distribution
     - Look for outliers
  2. Normal QQPlot
     - Check if data follows 1:1 line

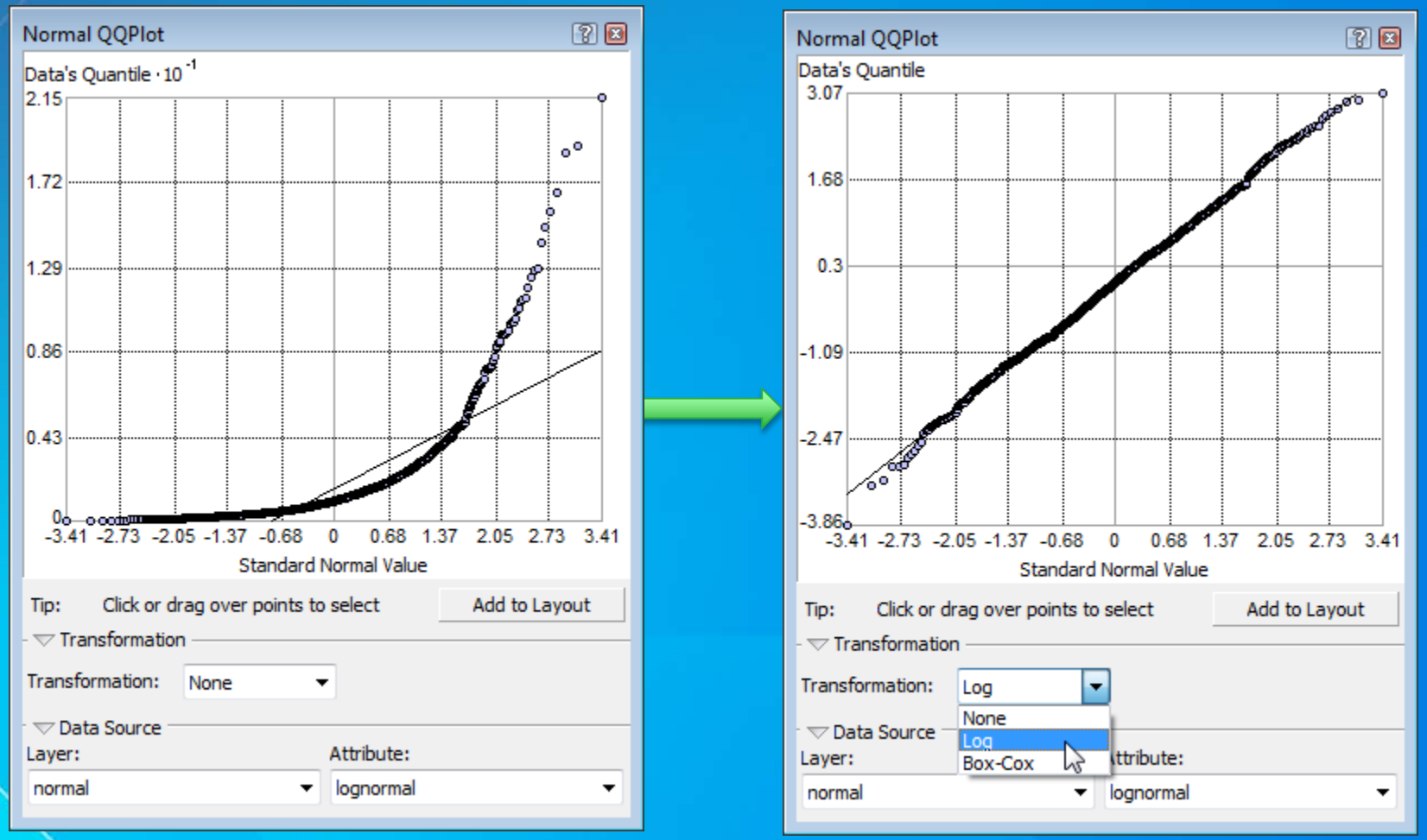
- What can I do if my data is not normally distributed?
  - Apply a transformation
    - Log, Box Cox, Arcsin, Normal Score Transformation

Does my data follow a normal distribution?

- What should I look for?
  - Bell-shaped
  - No outliers
  - Mean ≈ Median
  - Skewness ≈ 0
  - Kurtosis ≈ 3

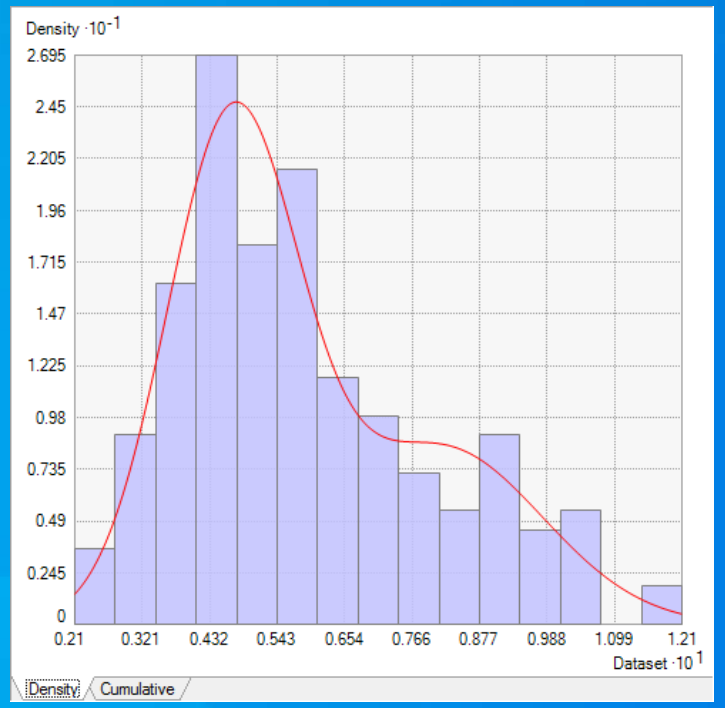


Does my data follow a normal distribution?



Normal Score Transformation

- Fits a smooth curve to the data
- Performs a quantile transformation to the normal distribution
- Performs calculations with transformed data, then transforms back at the end
- Simple kriging with normal score transformation is default in ArcGIS 10.1 and beyond

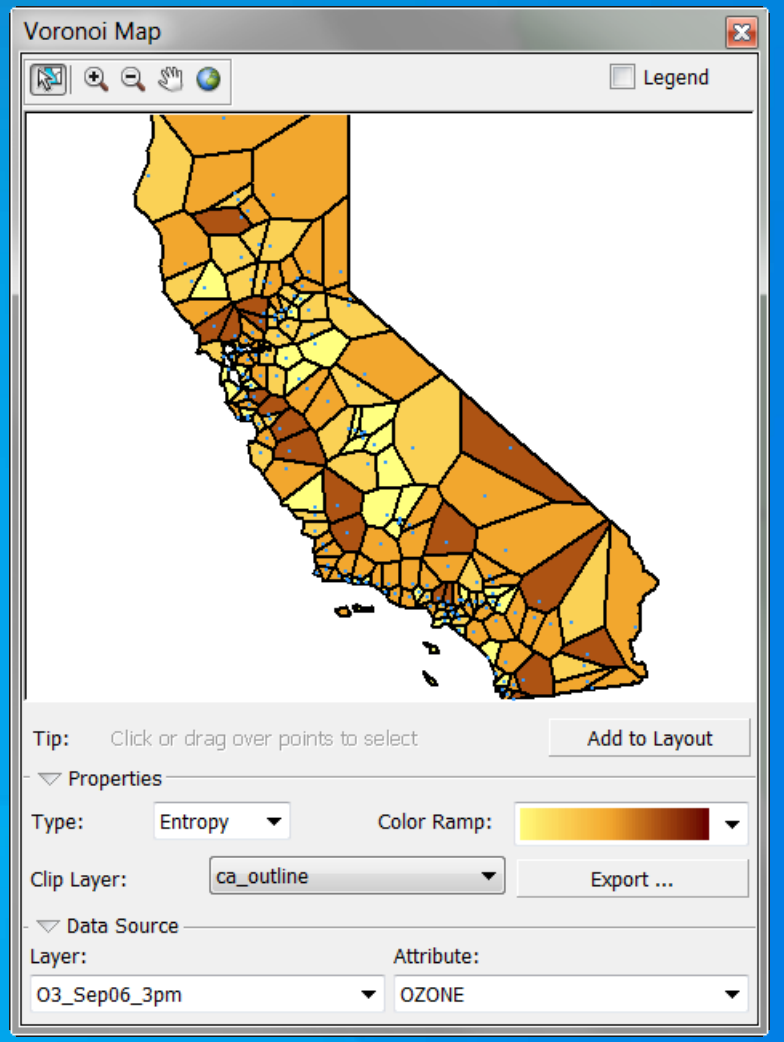


Is my data stationary?

- What is stationarity?
  - The statistical relationship between two points depends only on the distance between them.
  - The variance of the data is constant (after trends have been removed)
- How do I check for stationarity?
  - Voronoi Map symbolized by Entropy or Standard Deviation
- What can I do if my data is nonstationary?
  - Transformations can stabilize variances
  - Empirical Bayesian Kriging

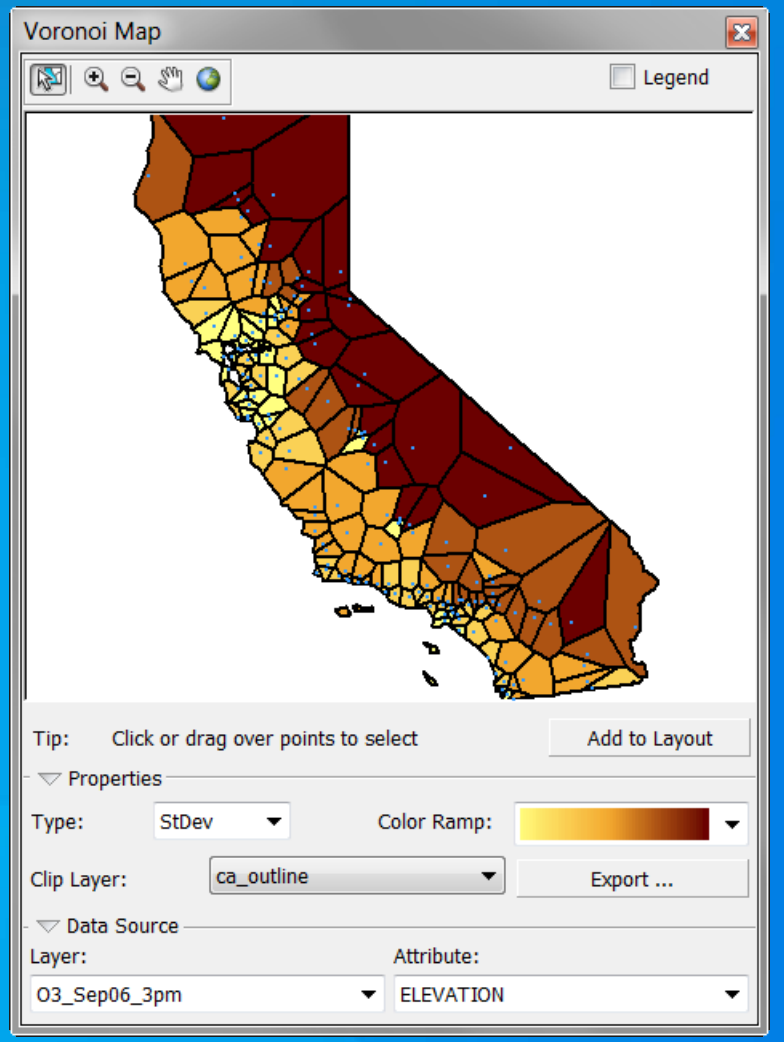
Is my data stationary?

- When symbolized by Entropy or StDev, look for randomness in the symbolized Thiessen Polygons.



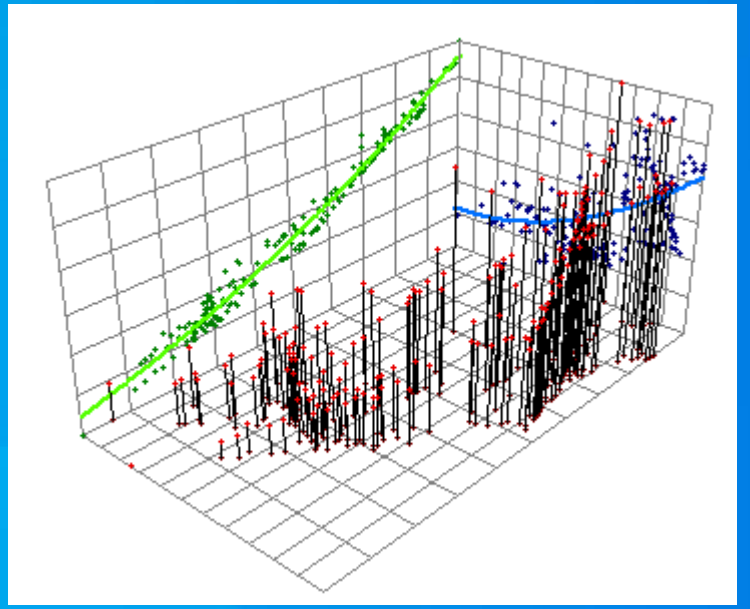
Is my data stationary?

- When symbolized by Entropy or StDev, look for randomness in the symbolized Thiessen Polygons.



Does my data have trends?

- What are trends?
  - Trends are systematic changes in the values of the data across the study area.
- How do I check for trends?
  - Trend Analysis ESDA tool
    - What can I do if my data has trends?
      - Use trend removal options
      - Potential problem – Trends are often indistinguishable from autocorrelation and anisotropy
      - EBK

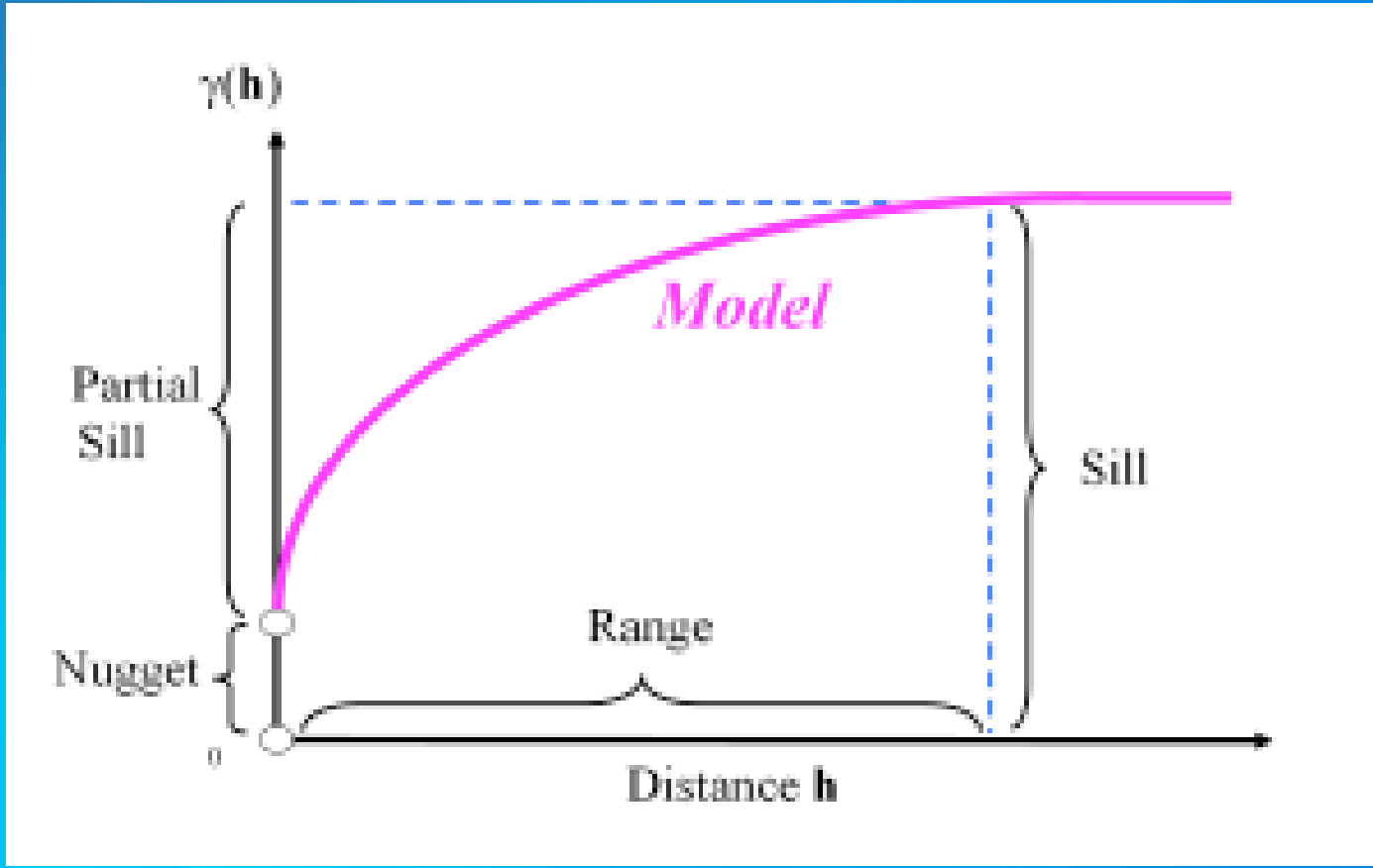


Demo

ESDA

Eric Krause

Semivariogram/Covariance Modeling



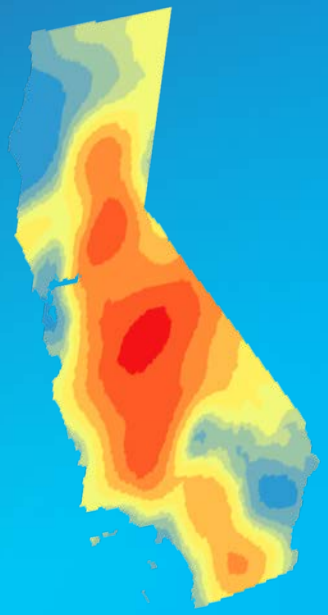
Cross-validation

- **Used to determine the quality of the model**
  - **Iteratively discard each sample**
  - **Use remaining points to estimate value at measured location**
  - **Compare predicted versus measured value**

Prediction Errors	
Samples	75049 of 75049
Mean	-3.308263e-7
Root-Mean-Square	0.002382183
Mean Standardized	-0.0001263315
Root-Mean-Square Standardized	0.8544209
Average Standard Error	0.002788466
Export Result Table	

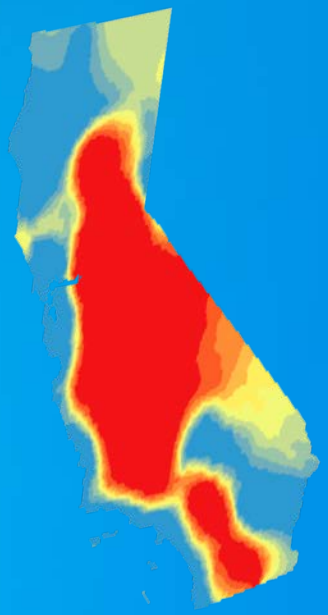
Kriging output surface types

Prediction

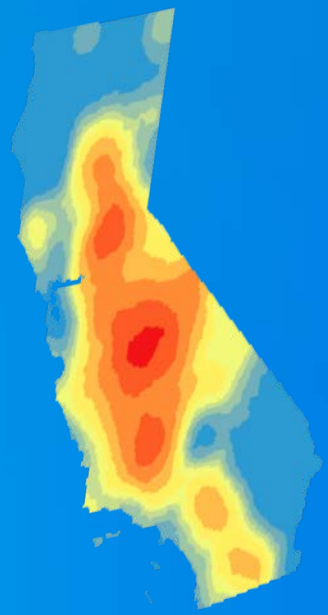


Error of Predictions

Probability



Quantile



Demo

Kriging

Eric Krause

Empirical Bayesian Kriging

- Advantages
  - Requires minimal interactive modeling, spatial relationships are modeled automatically
  - Usually more accurate, especially for small or nonstationary datasets
  - Uses local models to capture small scale effects
    - Doesn’t assume one model fits the entire data
  - Standard errors of prediction are more accurate than other kriging methods
- Disadvantages
  - Processing is slower than other kriging methods
  - Limited customization

How does EBK work?

1. Divide the data into subsets of a given size
   - Controlled by “Subset Size” parameter
   - Subsets can overlap, controlled by “Overlap Factor”
2. For each subset, estimate the semivariogram
3. Simulate data at input point locations and estimate new semivariogram
4. Repeat step 3 many times. This results in a distribution of semivariograms

   Controlled by “Number of Simulations”
5. Mix the local surfaces together to get the final surface.

What is coming in the future?

In ArcGIS Pro 1.2, EBK will support covariate rasters.

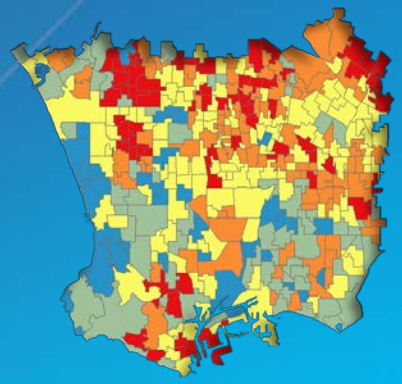
1. **Will allow EBK to use extra information not contained in the sample points**
2. **Expands use cases to non-continuous data, such as housing prices and predicted sales revenue.**

Demo

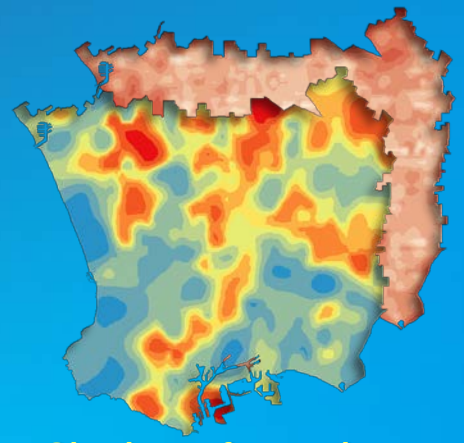
Empirical Bayesian Kriging

Eric Krause

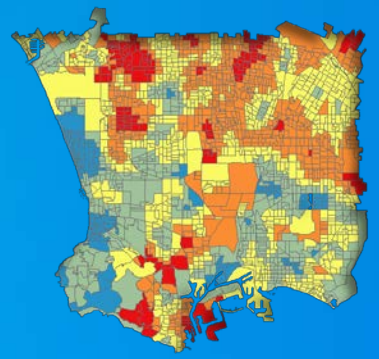
Areal Interpolation



Obesity by school zone



Obesity surface and error surface



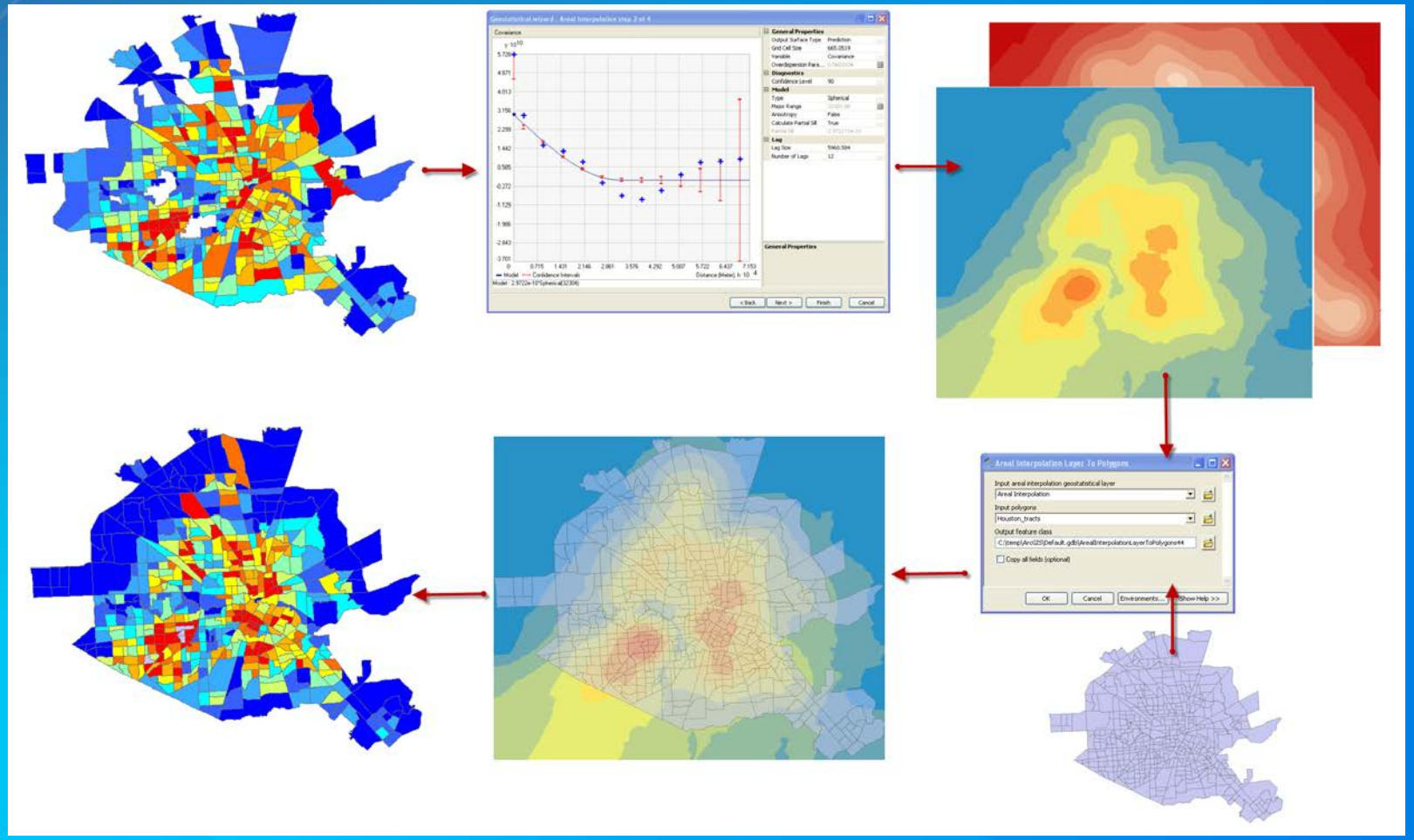
Obesity by census block

- Predict data in a different geometry
  - School zones to census tracts
- Estimate values for missing data

Types of Areal Interpolation

- Average (Gaussian)
  - Median age, average temperature
- Rate (Binomial)
  - Cancer rates, obesity rates, percent of college graduates
- Event (Overdispersed Poisson)
  - Animal counts, crimes

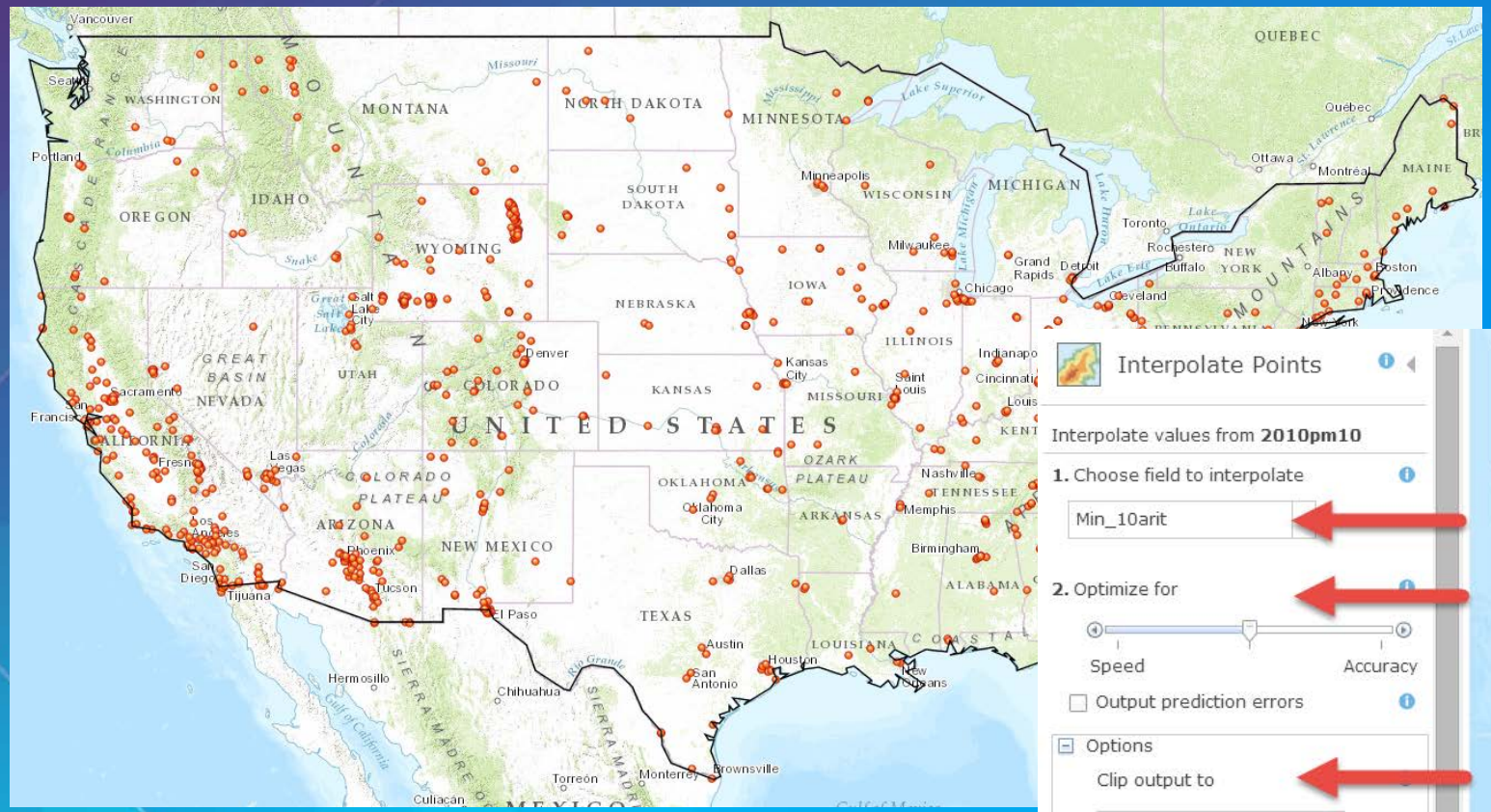
Polygon to Polygon Workflow

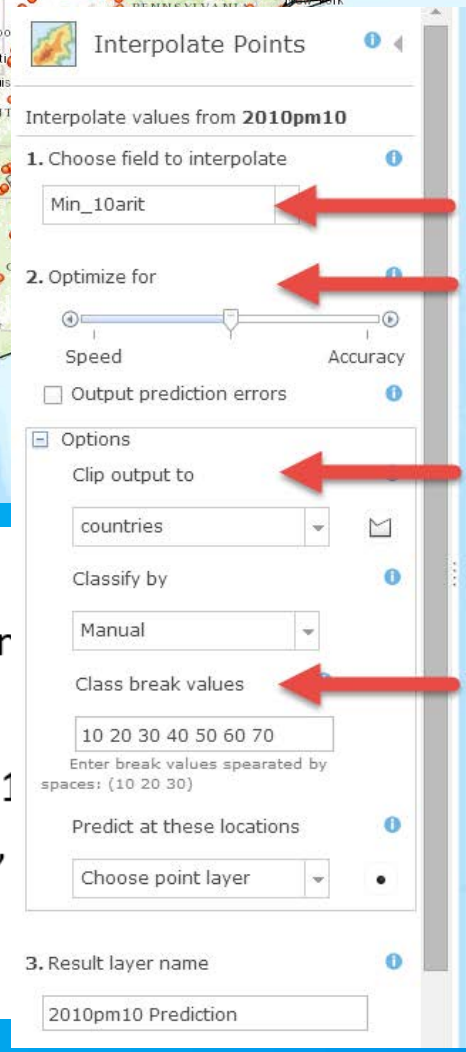


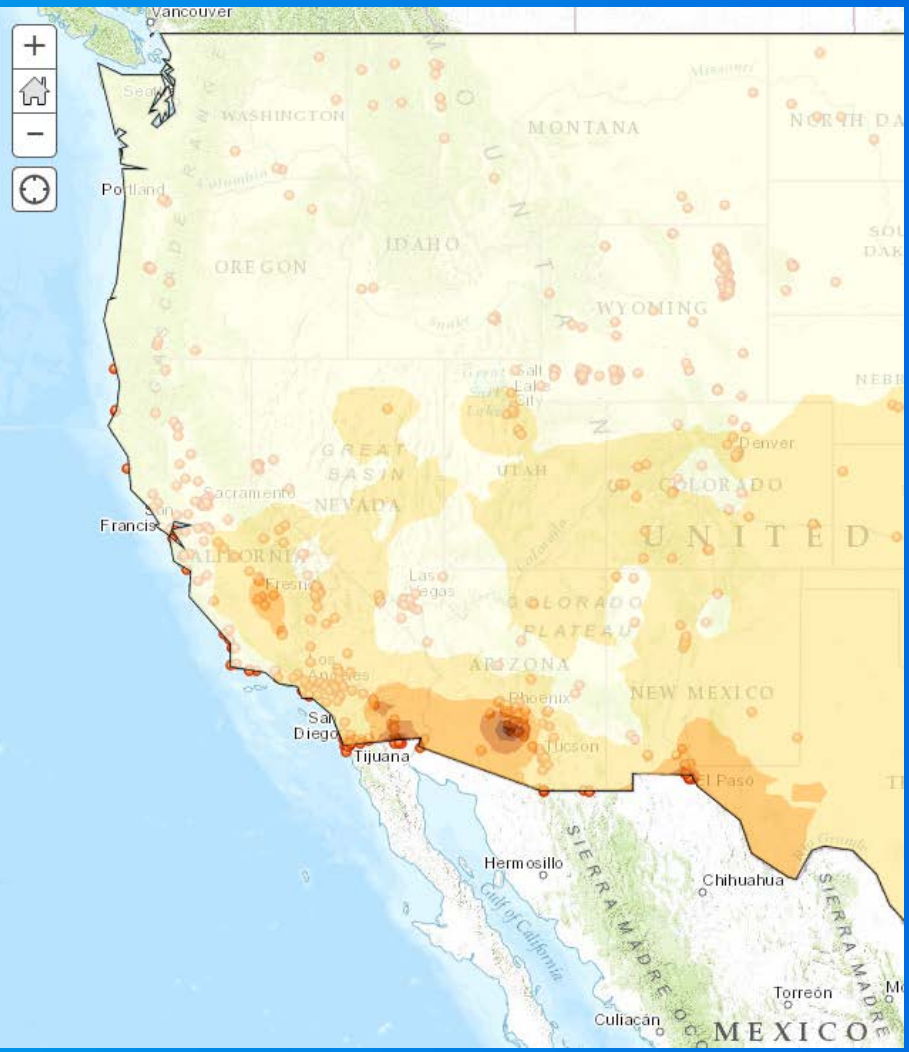
Demo

Areal Interpolation

Eric Krause







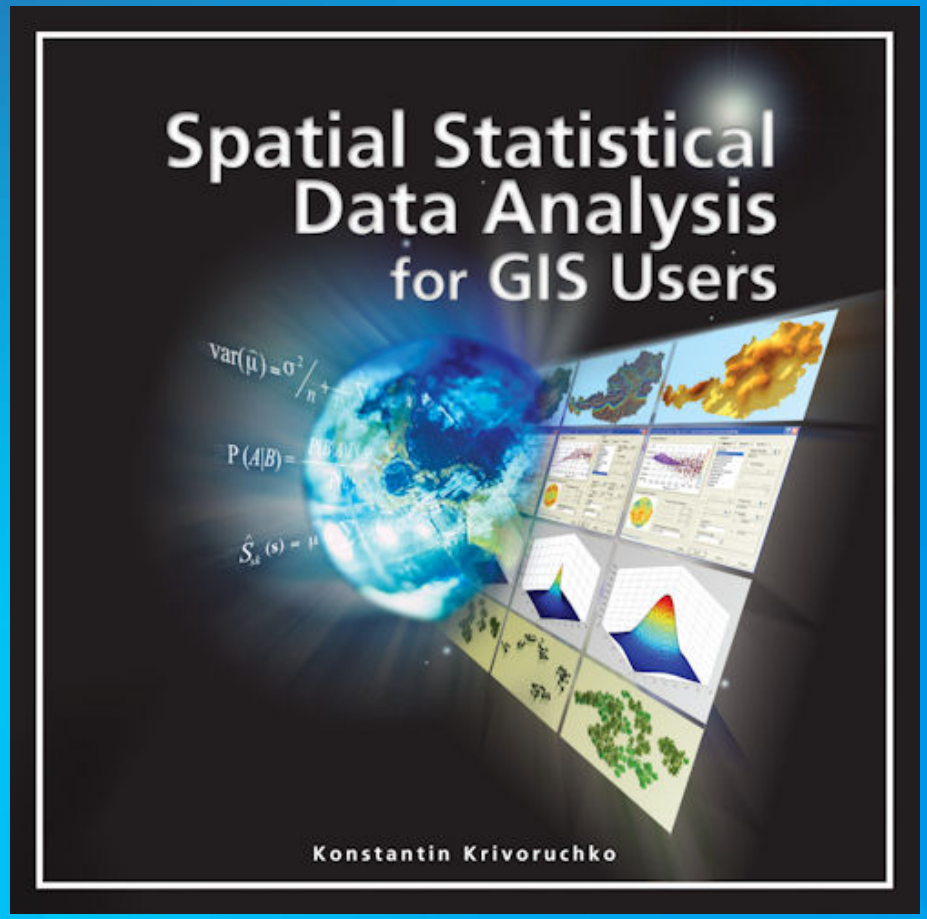
PM 10 Interpolation using ArcGIS Online

Particulate matter are a form of hazardous air pollution cor
small enough to get into the lungs.

In the United States, areas are not allowed to exceed a PM1
once a year, averaged over 3 years. In California, however,
at 50 µg/m^3.

Used daily arithmetic average PM10 value for 2010

Available in the bookstore and from Esri Press



Sessions of note…

Tuesday		
• ArcGIS Geostatistical Analyst - An Introduction	8:30-9:45	Room 14 A
• Concepts and Applications of Kriging	10:15-11:30	Room 14 A
• Creating Surfaces from Various Data Sources	10:15-11:30	Room 17 B
• EBK – Robust Kriging as a Geoprocessing Tool	3:30-4:15	Exhibit Hall B
Wednesday		
• ArcGIS Geostatistical Analyst - An Introduction	1:30-2:45	Room 05 A
• Concepts and Applications of Kriging	3:15-4:30	Room 05 A
• Surface Interpolation in ArcGIS	4:30-5:15	Exhibit Hall B
• Geostat. Simulations - Preparing for Worst Case Scenarios	5:30-6:15	Exhibit Hall B
Thursday		
• Polygon-to-Polygon Predictions using Areal Interpolation	9:30-10:15	Exhibit Hall B
• Creating Surfaces from Various Data Sources	10:15-11:30	Room 17 A

Understanding our world.