ENGRG 59910: Introduction to GIS

Lecture 06: Spatial Interpolation and Statistics

Source: NASA

Acknowledgment: Michael Piasecki
Outline Today

• Concepts
  • What is spatial interpolation
  • Why is necessary
  • Sample of interpolation (size and pattern)

• Interpolation methods
  • Thiessen polygon
  • Local average
  • Inverse Distance Weighted
  • Spline
  • Kriging
Spatial Interpolation

- **Spatial interpolation** is the process of using points with known values to estimate values at other points. These points with known values are called known points, control points, sampled points, or observations.

- Interpolation only works where values are spatially dependant, or spatially auto-correlated, that is, where nearby locations tend to have similar Z values.

Examples of spatially auto-correlated features:
- elevation,
- property value,
- crime levels,
- precipitation

- **Cannot** use interpolation where values are not spatially auto-correlated
• Positive **spatial autocorrelation** is when similar values cluster together in a map. Negative **spatial autocorrelation** is when dissimilar values cluster together in a map.
Why Spatial Interpolation?

• GIS require data at all points (subject to spatial resolution of question).

• Impossible to measure continuous variables at all points.
  • Infinite number of points
  • Finite amount of time and other resources
  • Inaccessible points

• It is necessary to evaluate new values
  • Evaluate new value if cell size is changed
  • Replace missing or erroneous data (noise data)
Principles of Interpolation

• Problem: Impossible ever to measure any continuous variable at all points

• Objective: Estimate unknown points by similarity and proximity with known points

• Basic Principle: Points close together in space are more likely to have similar values than points far apart (Tobler's First Law of Geography, 1970)

• Basis:
  • Sample of Known Points
  • Unknown points
Interpolation is prediction within the range of our data

- E.g., having temperature values for a bunch of locations all throughout PA, predict the temperature values at all other locations within PA

Extrapolation is prediction outside the range of our data

- E.g., having temperature values for a bunch of locations throughout PA, predict the temperature values in Kazakhstan
Sample Patterns

- Systematic pattern
- Random pattern
- Cluster pattern
- Adaptive pattern
Systematic Sampling

• Easy
• Samples spaced uniformly at fixed X, Y intervals
• Parallel lines

Advantages
Easy to understand

Disadvantages
• All receive same attention
• Difficult to stay on lines
• May be biases
Random Sample

• Select point based on random number process

Advantages
Less biased (unlikely to match pattern in landscape)

Disadvantages
Does nothing to distribute samples in areas of high
Difficult to explain, location of points may be a problem
Cluster Sampling

- Group samples around centers
- Centers can be located *systematically* or *randomly* (or *for convenience*, if statistically valid)
- Samples within a cluster can be located randomly, systematically, or otherwise.

**Advantage**
- Reduced travel time
- Take advantage of existing access

**Disadvantages:**
For *cluster center placement*, same as for *systematic* and *random point placement*
Adaptive Patterns

- **Adaptive Sampling** – higher sampling density where important variable(s) have greatest variation.

**Advantages**
- Sampling efficiency optimized
- Logistics can be planned in many cases
- More efficient, homogeneous areas have few samples, better representation of variable areas.

**Disadvantages**
- Preliminary sampling may be necessary, e.g. soil chemistry
- Significant prep time
- Subjective
Interpolation: *The Unknown Points*

- All interpolation methods use sampled values and positions to model un-sampled points.

- *Deterministic* interpolation techniques create surfaces from measured points, based on either the extent of similarity (e.g., Inverse Distance Weighted) or a degree of smoothing (e.g., Trend Surface Analysis).

- *Geostatistical* interpolation techniques (e.g., Kriging) are based on statistics and are used for more advanced prediction surface modeling, which also includes error or uncertainty of predictions.

- NOTE: Different methods will (almost always) produce different results.
Classification of Interpolation

• Local or global
• Exact or approximate (inexact)
• Deterministic or stochastic
• Gradual or abrupt
Global vs. Local

A **global interpolation** method uses every known point available to estimate an unknown value.

A **local interpolation** method, on the other hand, uses a sample of known points to estimate an unknown value.
Exact vs. Inexact

**Exact interpolation** predicts a value at the point location that is the same as its known value. In other words, exact interpolation generates a surface that passes through the control points.

**Inexact interpolation** or approximate interpolation predicts a value at the point location that differs from its known value.
A deterministic interpolation method provides no assessment of errors with predicted values.

A stochastic interpolation method, on the other hand, offers assessment of prediction errors with estimated variances.
Gradual or Abrupt

• Gradual methods:
  • produce smooth surface between data points
  • appropriate for interpolating data of low local variability

• Abrupt methods:
  • produce surfaces with a stepped appearance
  • appropriate for interpolating data of high local variability or data with discontinuities
Interpolation Methods

• Different methods will produce different results with same input data.

• No single method is more accurate than others for all situations.

• Accuracy – may be determined by comparison with a second set of “withheld” samples for accuracy checking.

• The researcher should select the method based on knowledge of the study area, phenomena of interest, and available resources.
Choose Interpolation Methods

• The best interpolation method for any given application depends on
  • The characteristics of the variable to be estimated
  • the cost of sampling
  • available resources
  • accuracy requirements of the user

• Relative performance of interpolators has been determined for some variables in some locations

• This is truly a matter of trying out different methods, once you have some data.
The Basic Equation for Spatial Interpolation

\[ Z_0' = \sum_{i=1}^{n} w_i \cdot Z_i \]

Where:
- \( Z_0' \) is the attribute value to be predicted at the unvisited-site
- \( Z_i \) is the attribute value at the \( i \) point of the nearby locations
- \( w_i \) is the weight assigned to the attribute at point \( i \), \( w_i \) should sum up to 1 (to be unbiased)
- \( n \) is the total number of nearby locations involved

Key issues:
1) How many nearby points should we include for a given un-visited site?
2) How to select these nearby points?
3) How to allocate the weight for each nearby point?
Interpolation methods

• Most GIS packages offer a number of methods

• Typical methods are:
  • Thiessen polygons
  • Spatial moving average
  • Trend Surfaces
  • Inverse Distance Weighted
  • Kriging
Nearest Neighbor Interpolation (Thiessen polygons)

• Value assigned from nearest sample

• Value assigned across unusually large cells may be questionable; watch for them

• Edge effects
Thiessen polygon construction
Polygon Models: Thiessen (Voronoi polygons, Dirichlet cells) Polygons

Thiessen Polygons: Areas of Influence
An “Exact Interpolator”: Original Values are Preserved
Local Average Method (Spatial Moving Average)

- Calculates new value of each location based on range of values associated with neighbouring points

- Neighbourhood determined by a filter
  - size, shape and character of filter

- Avoids shaky assignments at extreme distances

- Potential problem: Assigns zero to sites with no samples within radius

- Not exact interpolator
Spatial Moving Average

\[
\frac{3 + 7 + 5}{3} = 5
\]
Local Average: Fixed Radius

- Raster grid specified
- Sample points within a fixed, “search” radius with center in each cell are averaged.
Inverse-Distance Weighting (IDW)

Exact Interpolator:
Value calculated directly from three (usually) nearest samples

Near one sample point, its influence greatly predominates.

Must know!
Exponent Value: Degree of Influence
Interpolation with Splines

- Force line to pass through known points.
- Joined polynomial functions
- Can be 1-3 order
- Segments meet at join points
- Iterative process

One polynomial equation is fit to points one through five, $y = f(x)$.

At the join point:
- $f(x) = g(x)$
- slope of $f(x)$ = slope of $g(x)$
- change in slope of $f(x)$ equals the change in slope of $g(x)$

A second polynomial equation is fit to points five through eleven, $y = g(x)$. 
Nearest Neighbor “Thiessen” Polygon Interpolation

Spline Interpolation
Effects of data uncertainty

- Original surface
- Interpolation based on 100 points
- Error map
- Interpolation based on 10 points
- Error map
Edge effects

- Original surface with sample points
- Interpolated surface
- Error map and extract

Legend:
- Low
- Medium
- High
Spatial Interpolation in ArcGIS – Geostatistical Analyst

• The ArcGIS GeoStatistical Analyst extension provides a variety of interpolation methods for the creation of an optimal interpolated surface from your data.

• A friendly wizard helps you through the interpolation process.
Interpolation with Geostatistics

• Several options; best known is:

• Kriging
  • Developed by D.G. Krige and G. Matheron.
  • Statistically-based estimator
  • Depends on Variogram analysis: defined as “... a fct describing the degree of spatial dependence of a spatial random field or stochastic process ... the variance of the difference between field values at two locations \( x \) and \( y \) across realizations in the field ...”

• Weighting to interpolate variable value similar to IDW, but weights are based on spatial autocorrelation in the sample, not merely distance.
Kriging Steps

1. Calculate semivariance: spatial autocorrelation
2. Identify a model to estimate the weights
3. Calculate covariance within the data
4. Make the estimation
5. Check for error
Geostatistics with the ArcGIS Geostatistical Analyst Extension
Again, use of this methodology deserves an entire course...
Cross Validation

- The validation technique compares the interpolation methods by first dividing known points into two samples: one sample for developing the models for each interpolation method to be compared and the other sample for testing the accuracy of the models.

- Steps
  - Remove a known point from the data set.
  - Use the remaining points to estimate the value at the point previously removed.
  - Calculate the predicted error of the estimation by comparing the estimated with the known value.
Root mean square (RMS) error

\[
RMS = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Z_{i,\text{act}} - Z_{i,\text{est}})^2}
\]

Standardized RMS = 

\[
\sqrt{\frac{1}{n} \sum_{i=1}^{n} \frac{(Z_{i,\text{act}} - Z_{i,\text{est}})^2}{s^2}}
\]
Summary

• Concepts
  • What is spatial interpolation and why is necessary
  • Sample pattern: systematic; random; cluster; adaptive

• Classification of interpolation methods
  • Global vs Local
  • Exact vs Inexact
  • Graduate vs abrupt
  • Deterministic vs Stochastic

• Interpolation methods
  • Thiessen polygon
  • Local average
  • Inverse Distance Weighted
  • Spline
  • Kriging