Watershed Sciences 4930 & 6920
GEOGRAPHIC INFORMATION SYSTEMS

WEEK EIGHT– Lecture

INTERPOLATION & SPATIAL ESTIMATION

Joe Wheaton

HOUSEKEEPING

- Quizzes
- Lab 8?

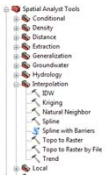
READING FOR TODAY

TODAY'S PLAN

I. **Sampling**

II. Spatial Interpolation
   I. TIN
   II. Nearest Neighbor
   III. Fixed Radius – Local Averaging
   IV. Inverse Distance Weighted
   V. Natural Neighbor
   VI. Spline
   VII. TOPO
   VIII. Kriging

III. Interpolations Compared

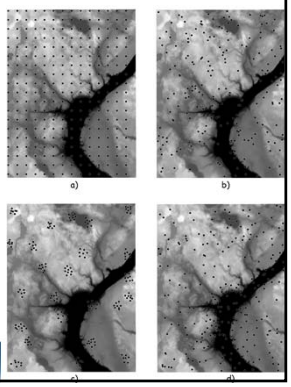


WHAT CAN WE COLLECT AT POINTS?

- Think of everything we can....

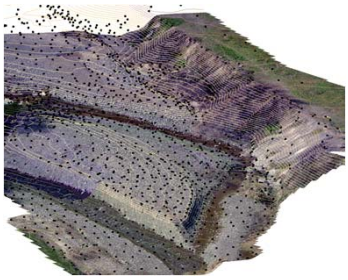
DIFFERENT SAMPLING STRATEGIES

- A) Regular Grid (systematic)
- B) Random
- C) Cluster
- D) Adaptive



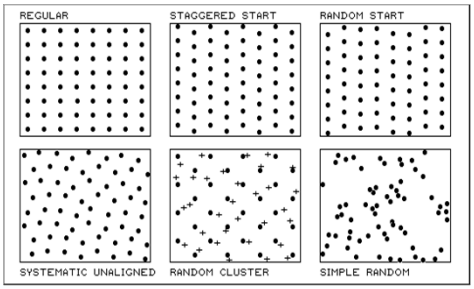
SAMPLING PATTERNS WE'VE DISCUSSED

- Point Based Sampling
- Sampling Patterns?
  - Uniform grid?
  - Sections?
  - Irregular
  - Random
  - Feature-based



A FEW OTHER VARIANTS…

Sampling Pattern: appropriate arrangement of samples considering both spatial interpolation and statistical inference



WHAT'S WRONG WITH OUR SAMPLING?

- NOT spatially continuous over area of interest
- Can't afford (time and money) to have samples everywhere
- Even if finite number of objects in study area, usually too many to measure all…
- Some sites impossible or unfeasible to visit

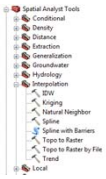
TODAY'S PLAN

I. Sampling

II. Spatial Interpolation

I. TIN
II. Nearest Neighbor
III. Fixed Radius – Local Averaging
IV. Inverse Distance Weighted
V. Natural Neighbor
VI. Spline
VII. TOPO
VIII. Kriging

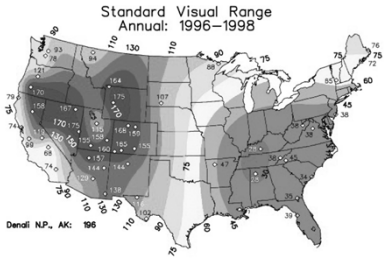
III. Interpolations Compared



INTERPOLATION DEFINED

- 'Interpolation predicts values for cells in a raster from a limited number of sample data points. It can be used to predict unknown values for any geographic point data, such as elevation, rainfall, chemical concentrations, noise levels, and so on.'

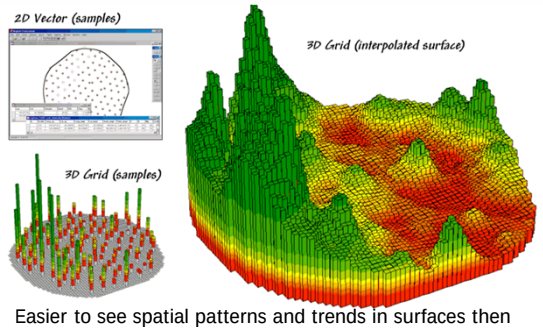
NATIONAL AIR QUALITY…



Interpolated in space from 50ish stations…
Interpolated in time through temporal averaging

From Bolstad (2008) Chapter 12

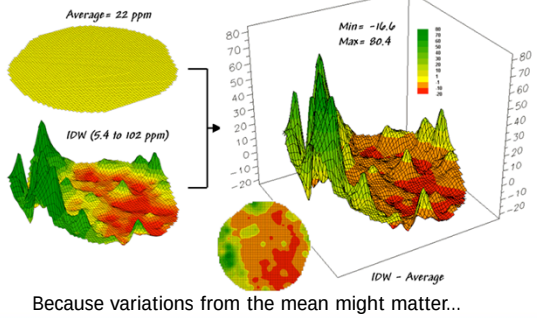
WHY INTERPOLATE?



Easier to see spatial patterns and trends in surfaces then raw discrete data

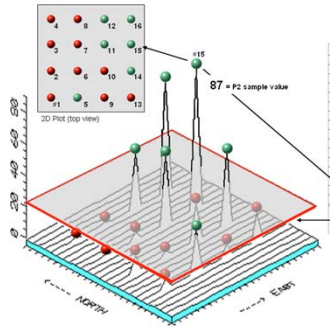
WHY INTERPOLATE?

What difference does it make?



Because variations from the mean might matter...

THE AVERAGE IS HARDLY ANYWHERE!



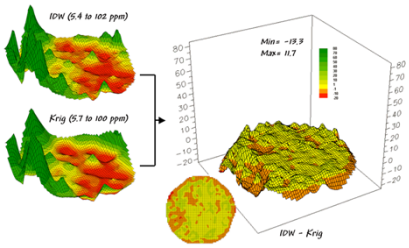
Field Collected Data

Sample	X	Y	P1	P2
#1	1000	1000	11	4
2	1000	1500	19	9
3	1000	2000	8	0
4	1000	2500	0	0
5	1500	1000	27	25
6	1500	1500	12	2
7	1500	2000	14	4
8	1500	2500	2	0
9	2000	1000	10	6
10	2000	1500	17	23
11	2000	2000	34	42
12	2000	2500	22	33
13	2500	1000	20	16
14	2500	1500	28	43
15	2500	2000	42	87
16	2500	2500	34	68
		Total	300	361
		Average	19	23
		StDev	12	26

Arithmetic Average – plot of the data average is a horizontal plane in 3-dimensional geographic space with some of the data points balanced above (green) and some below (red) the "typical" value

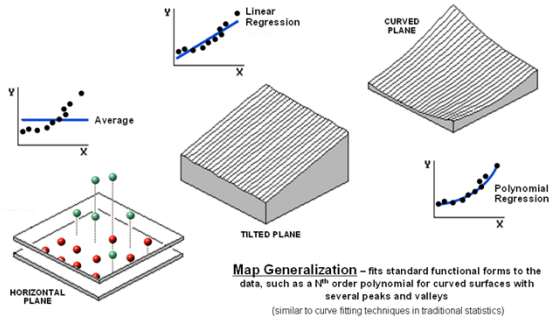
(uniform estimate of the spatial distribution)

DIFFERENT INTERPOLATIONS ALSO GIVE DIFFERENT RESULTS....



These differences are less problematic...
But worth considering...

INTERPOLATION TYPES VARY IN SOPHISTICATION



Map Generalization – fits standard functional forms to the data, such as a Nth order polynomial for curved surfaces with several peaks and valleys

(similar to curve fitting techniques in traditional statistics)

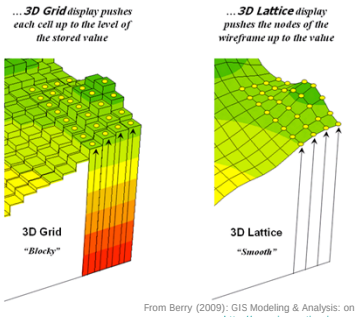
WHAT DO WE USE INTERPOLATION FOR?

- Interpolation from point observations to:
  - A continuous surface, which can be:
    - Vector Polygons (e.g. TIN or Theisan Polygons)
    - Raster Grid
- Interpolation from Vector Polygons (e.g. TIN or Theisan Polygons) to:
  - Raster Grid
- Extraction (by interpolation) from raster to points

Because we want to know approximate values at areas where we don't have observations

IS YOUR RASTER A GRID OR LATTICE?

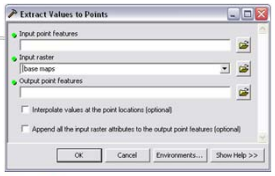
- Is this just a display thing... or is it algorithmic?
- Grid centered or node centered?



From Berry (2009): GIS Modeling & Analysis: on http://www.innovativegis.com

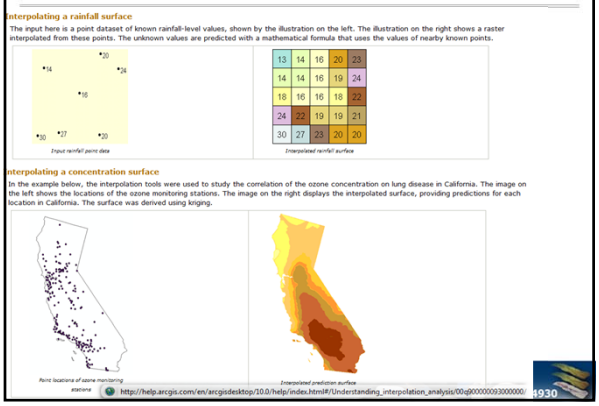
POINT EXTRACTION

- Points to be extracted won't line up exactly on a cell center... so what is the value?



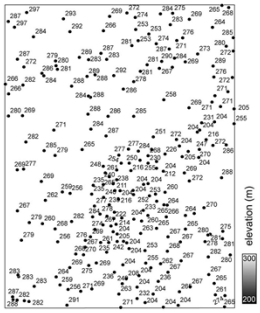
- If 'interpolation' checked, it uses a bilinear interpolation?

SOME INTERPOLATION EXAMPLES

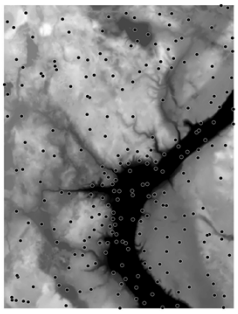


A DATASET WE'LL KEEP LOOKING AT...

Elevation points....

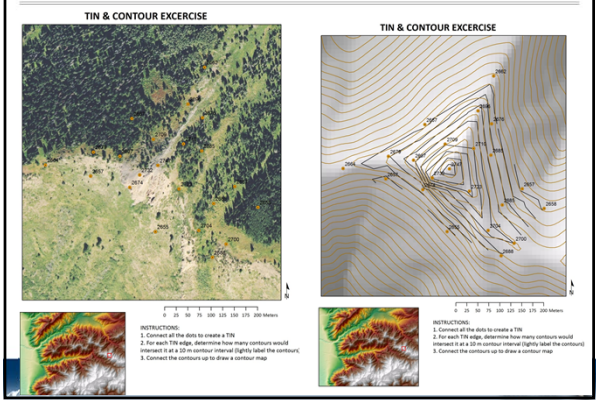


DEM



From Bolstad (2008) Chapter 12

ANOTHER DATASET WE'LL KEEP LOOKING AT



INTERPOLATION IN ARCGIS

- Three primary places:
  1. 3D Analyst Tools
  2. Geostatistical Analyst Tools
  3. Spatial Analyst Tools

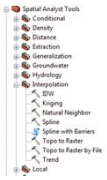
TODAY'S PLAN

I. Sampling

II. Spatial Interpolation

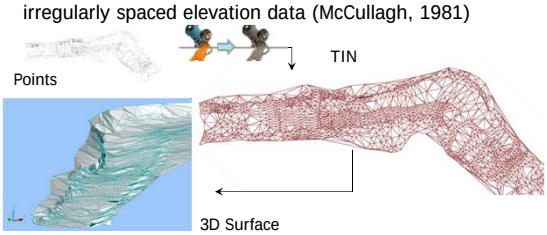
- **I. TIN**
- II. Nearest Neighbor
- III. Fixed Radius – Local Averaging
- IV. Inverse Distance Weighted
- V. Natural Neighbor
- VI. Spline
- VII. TOPO
- VIII. Kriging

III. Interpolations Compared



BUILDING A TIN FROM VECTOR DATA

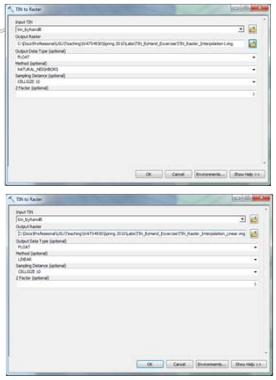
A Triangular Irregular Network (TIN) is the simplest and most common interpolation technique for building surfaces with irregularly spaced elevation data (McCullagh, 1981)



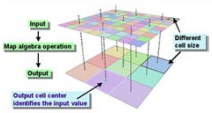
McCullagh MJ. 1981. Creation of smooth contours over irregularly distributed data using local surface patches. Geophysical Analysis. 13: 51-63.

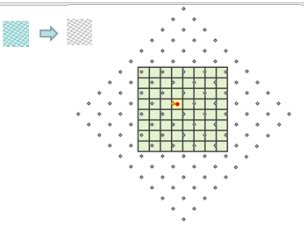
TIN to RASTER

- Natural neighbors vs. Linear
- An interpolation choice when going from vector to raster…



RESAMPLING (A FORM OF INTERPOLATION)

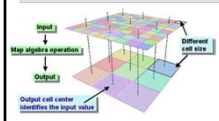


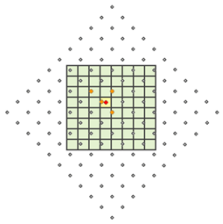


1. **Nearest Neighbor Assignment**
2. Bilinear Interpolation
3. Cubic Convolution

- Good for discrete (categorical) data since it does not alter value of input
- Once location of cell's center on output raster located on input raster, nearest neighbor assignment determines location of the closest cell center

RESAMPLING (A FORM OF INTERPOLATION)

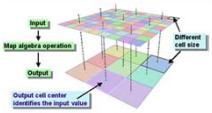


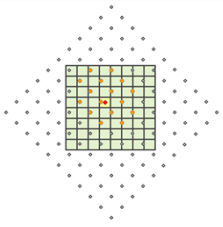


1. Nearest Neighbor Assignment
2. **Bilinear Interpolations**
3. Cubic Convolution

- Uses values of **4** nearest input cell centers to determine the value of output
- New value is a distance-weighted average
- Results in smoother-looking surface then nearest neighbor…

RESAMPLING (A FORM OF INTERPOLATION)

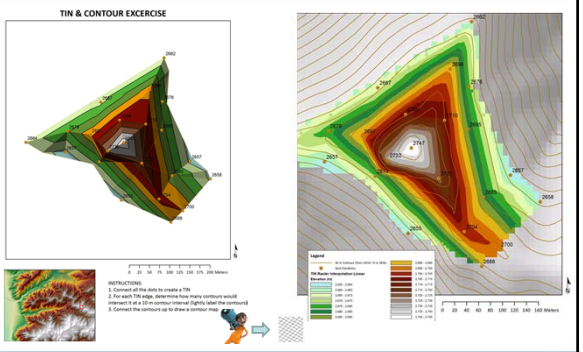




1. Nearest Neighbor Assignment
2. Bilinear Interpolation
3. **Cubic Convolution**

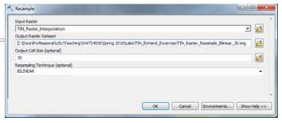
- Similar to bilinear interpolation except that weighted average is calculated from the 16 nearest input cell centers and their values

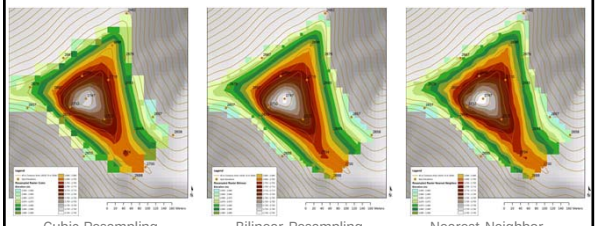
RECALL OUR TIN -> DEM



RESAMPLE EXAMPLE

- Resampling choices...
- 10 m to 30 m Raster





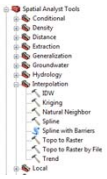
Cubic Resampling | Bilinear Resampling | Nearest Neighbor Resampling

TODAY'S PLAN

I. Sampling

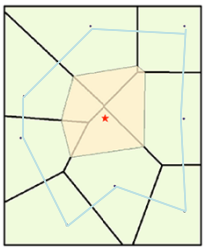
II. Spatial Interpolation
   I. TIN
   II. Nearest Neighbor
   III. Fixed Radius – Local Averaging
   IV. Inverse Distance Weighted
   V. Natural Neighbor
   VI. Spline
   VII. TOPO
   VIII. Kriging

III. Interpolations Compared



THIESSEN (VORONI) POLYGONS

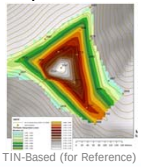
- First build the olive polygons by bisecting lines of perimeter polygon
- Then build the central peach polygon by bisecting lines to each of surrounding points and connecting midpoints
- Every polygon has ONLY one point to be interpolated!



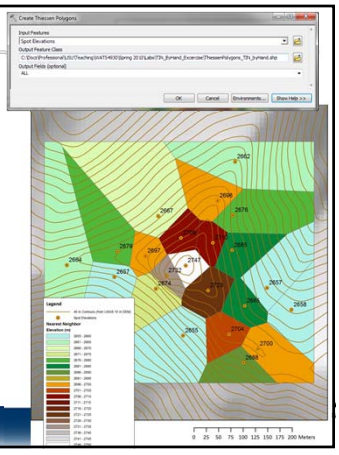
This is AKA Nearest Neighbor.... (not to be confused with Natural Neighbors)

IN ARC...

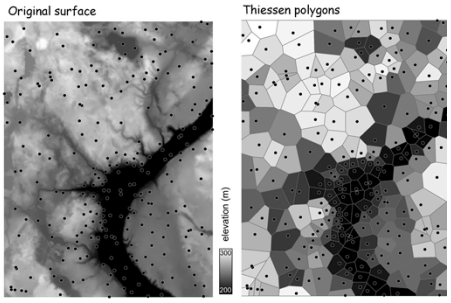
- Use Create Thiessen Polygons
- Notice
  - Influence of point spacing
  - Extension outside of sampled area



TIN-Based (for Reference)



NEAREST NEIGHBOR (AKA NATURAL NEIGHBOR)

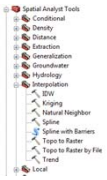


The points are not interpolated in this example, the region (Thiessen polygon) is...

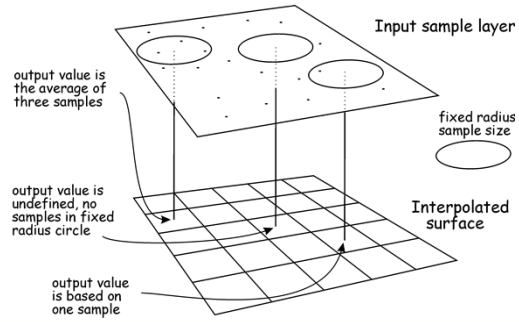
From Bolstad (2008) Chapter 12

TODAY'S PLAN

I. Sampling
II. Spatial Interpolation
  I. TIN
  II. Nearest Neighbor
  III. Fixed Radius – Local Averaging
  IV. Inverse Distance Weighted
  V. Natural Neighbor
  VI. Spline
  VII. TOPO
  VIII. Kriging
III. Interpolations Compared

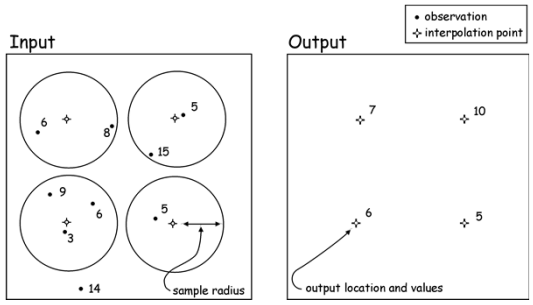


FIXED RADIUS INTERPOLATION



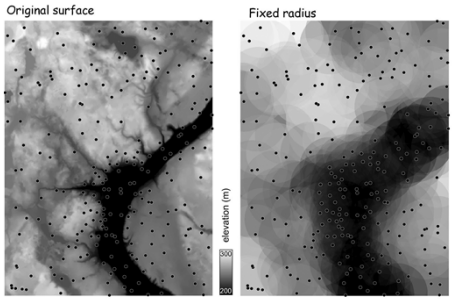
From Bolstad (2008) Chapter 12

FIXED RADIUS INTERPOLATION



From Bolstad (2008) Chapter 12

FIXED RADIUS APPLIED

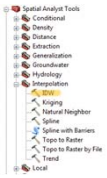


Not an exact interpolator.... Why?

From Bolstad (2008) Chapter 12

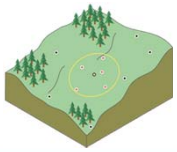
TODAY'S PLAN

I. Sampling
II. Spatial Interpolation
  I. TIN
  II. Nearest Neighbor
  III. Fixed Radius – Local Averaging
  IV. Inverse Distance Weighted
  V. Natural Neighbor
  VI. Spline
  VII. TOPO
  VIII. Kriging
III. Interpolations Compared



INVERSE DISTANCE WEIGHTED

- Estimates a value for a location as the weighted-average of the nearby data values within the roving window
- Average weighted so influence of surrounding values decrease with increasing distance from location being estimated



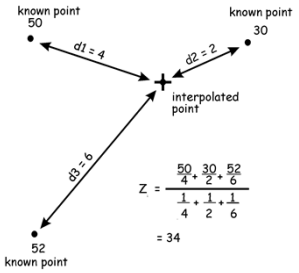
$$z_j = \frac{\sum_i \frac{z_i}{d^n{}_{ij}}}{\sum_i \frac{1}{d^n{}_{ij}}}$$

← Inverse Distance

INVERSE DISTANCE WEIGHTED

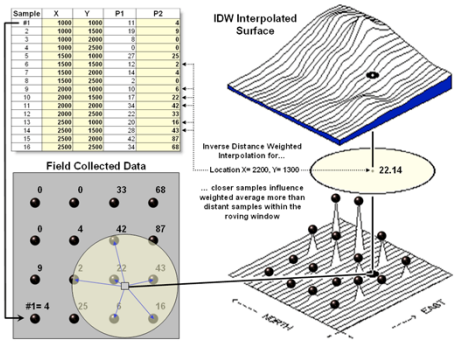
- Value of unknown points estimated using sampled values **and** distance to those samples
- The weight (or importance) of each sample point is determined by its distance away

$$z_j = \frac{\sum_i \frac{z_i}{d^n_{ij}}}{\sum_i \frac{1}{d^n_{ij}}}$$



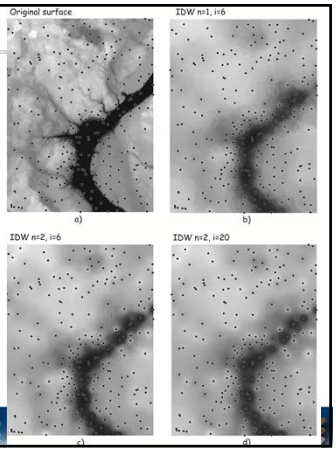
From Bolstad (2008) Chapter 12

IDW EXAMPLE...



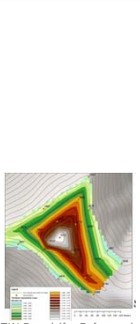
IDW SENSITVITY

- Exponent *n* ast a point's influence deteriorates with distance
- Number of points *i* controls search distance
- Local influences are stronger as both *n* and *i* increase

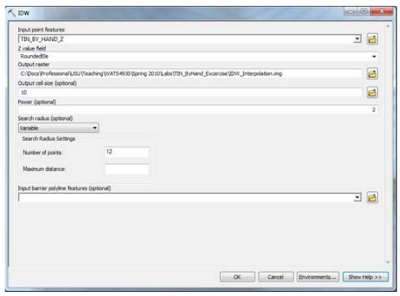


From Bolstad (2008) Chapter 12

INVERSE DISTANCE WEIGHTED

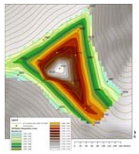


TIN-Based (for Reference)

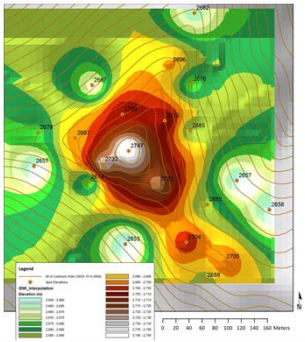


INVERSE DISTANCE WEIGHTED

- Method assumes variable being mapped decreases in influence with distance from sampled location...
- Notice Interpolation extent



TIN-Based (for Reference)

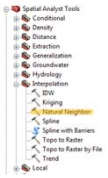


TODAY'S PLAN

I. Sampling

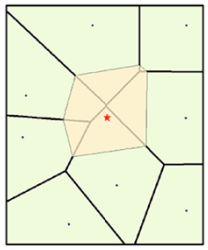
II. Spatial Interpolation
   - I. TIN
   - II. Nearest Neighbor
   - III. Fixed Radius – Local Averaging
   - IV. Inverse Distance Weighted
   - **V. Natural Neighbor**
   - VI. Spline
   - VII. TOPO
   - VIII. Kriging

III. Interpolations Compared

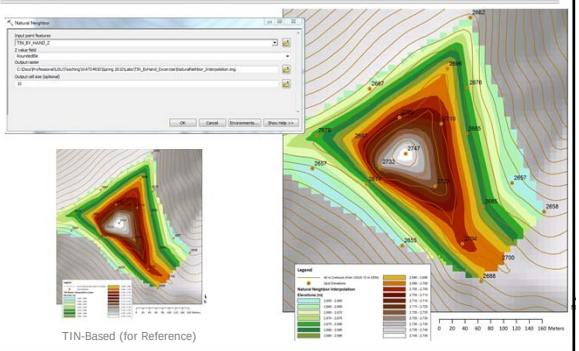


NATURAL NEIGHBOR

- Not to be confused with nearest neighbor...



NATURAL NEIGHBOR



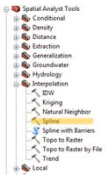
TIN-Based (for Reference)

TODAY'S PLAN

I. Sampling

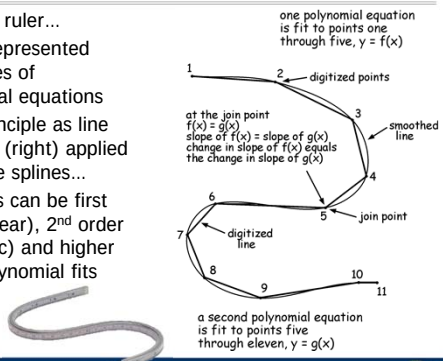
II. Spatial Interpolation
   I. TIN
   II. Nearest Neighbor
   III. Fixed Radius – Local Averaging
   IV. Inverse Distance Weighted
   V. Natural Neighbor
   VI. Spline
   VII. TOPO
   VIII. Kriging

III. Interpolations Compared

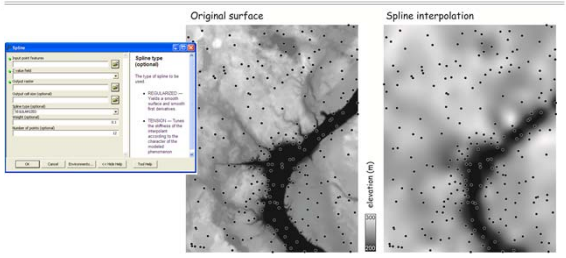


SPLINE INTERPOLATION 1D

- A flexible ruler...
- Can be represented with series of polynomial equations
- Same principle as line functions (right) applied to surface splines...
- Equations can be first order (linear), 2nd order (quadratic) and higher order polynomial fits



SPLINE INTERPOLATION 2D

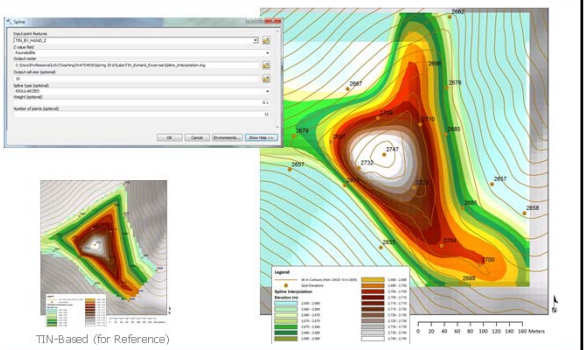


- Designed to minimize surface curvature
- Creates smooth surface that passes exactly through input points

http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/How_Spline_works/009z000000078000000/

From Bolstad (2008) Chapter 12

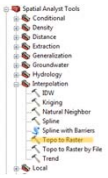
SPLINE



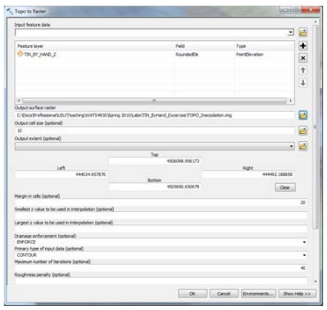
TIN-Based (for Reference)

TODAY'S PLAN

I. Sampling
II. Spatial Interpolation
  I. TIN
  II. Nearest Neighbor
  III. Fixed Radius – Local Averaging
  IV. Inverse Distance Weighted
  V. Natural Neighbor
  VI. Spline
  VII. TOPO
  VIII. Kriging
III. Interpolations Compared

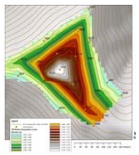


TOPO TO RASTER

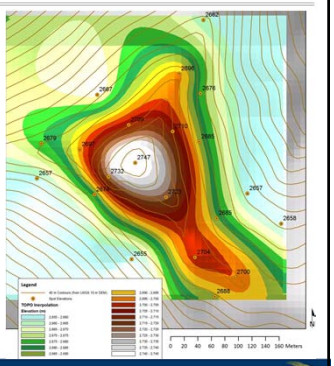


TOPO TO RASTER

Variance Estimate

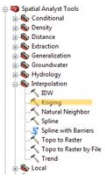


TIN-Based (for Reference)



TODAY'S PLAN

I. Sampling
II. Spatial Interpolation
  I. TIN
  II. Nearest Neighbor
  III. Fixed Radius – Local Averaging
  IV. Inverse Distance Weighted
  V. Natural Neighbor
  VI. Spline
  VII. TOPO
  VIII.Kriging
III. Interpolations Compared



WHAT IS KRIGING?

- An advanced geostatistical procedure that generates an estimated surface from a scattered set of points with z-values
- Unlike other interpolation methods, Kriging involves an interactive investigation of the spatial behavior of the phenomenon represented by the z-values before you select the best estimation method for generating the output surface

The kriging formula

Kriging is similar to IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location. The general formula for both interpolators is formed as a weighted sum of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

- where:
  $Z(s_i)$ = the measured value at the ith location
  λ_i = an unknown weight for the measured value at the ith location
  s_0 = the prediction location
  N = the number of measured values

In IDW, the weight, λ_i, depends solely on the distance to the prediction location. However, with the kriging method, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement of the measured points. To use the spatial arrangement in the weights, the spatial autocorrelation must be quantified. Thus, in ordinary kriging, the weight, λ_i, depends on a fitted model to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location. The following sections discuss how the general kriging formula is used to create a map of the prediction surface and a map of the accuracy of the predictions.

http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z0000007t000000.htm

TWO STEPS....

Creating a prediction surface map with kriging

To make a prediction with the kriging interpolation method, two tasks are necessary:

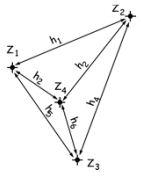
- Uncover the dependency rules.
- Make the predictions.

To realize these two tasks, kriging goes through a two-step process:

1. It creates the variograms and covariance functions to estimate the statistical dependence (called spatial autocorrelation) values that depend on the model of autocorrelation (fitting a model).
2. It predicts the unknown values (making a prediction).

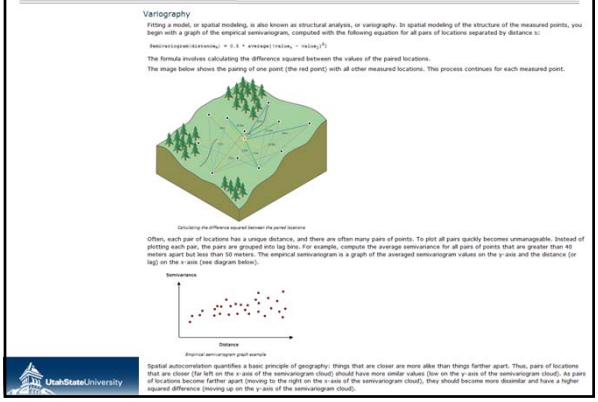
It is because of these two distinct tasks that it has been said that kriging uses the data twice: the first time to estimate the spatial autocorrelation of the data and the second to make the predictions.

- Statistically based estimator of spatial variables
- Relies on:
  - Spatial Trend – increase or decrease in a variable in some direction
  - Local spatial autocorrelation
  - Random stochastic variation

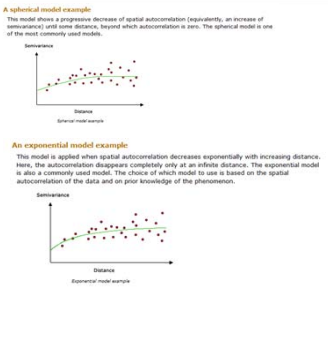


http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z0000007t000000.htm

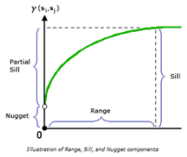
THE FIRST STEP



FIT EMPIRICAL DATA TO SEMIVARIOGRAM



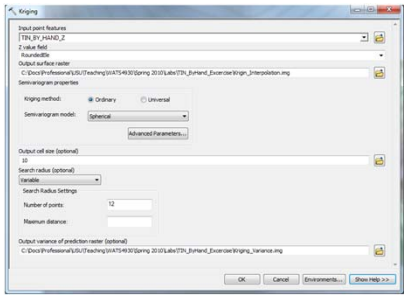
- Models:
  - Circular
  - Spherical
  - Exponential
  - Gaussian
  - Linear



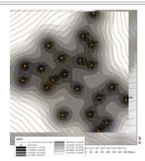
STEP 2: MAKING THE PREDICITON

- Now you've uncovered dependence (autocorrelation) in your data, we can make a prediction using fitted model (set aside empirical semivariogram)
- Like IDW, kriging forms weights from surrounding measured values to predict unmeasured locations… Instead of just using distance, kriging uses semivariogram

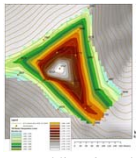
KRIGING



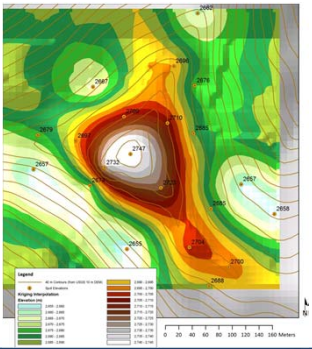
KRIGING



Variance Estimate



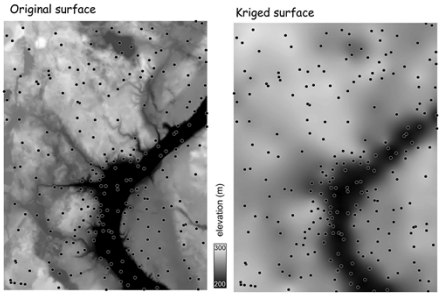
TIN-Based (for Reference)



KRIGING

Original surface

Kriged surface

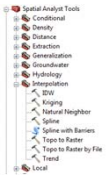


From Bolstad (2008) Chapter 12

TODAY'S PLAN

I. Sampling
II. Spatial Interpolation
  I. TIN
  II. Nearest Neighbor
  III. Fixed Radius – Local Averaging
  IV. Inverse Distance Weighted
  V. Natural Neighbor
  VI. Spline
  VII. TOPO
  VIII. Kriging

III. Interpolations Compared



INTERPOLATION METHODS COMPARED

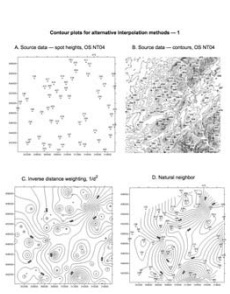


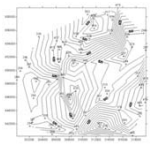
Table 6-2 Gridding and interpolation methods

Method	Speed	Type	Comments
Inverse distance weighting (IDW)	Fast	Exact, unless smoothing factor specified	Tends to generate bull's eye patterns. Simple and effective with dense data. No extrapolation. All interpolated values between data points lie within the range of the data point values and hence may not approximate valleys and peaks well
Natural neighbor	Fast	Exact	A weighted average of neighboring observations using weights determined by Voronoi polygon concepts. Good for dense datasets. Typically implementations do not provide extrapolation
Nearest neighbor	Fast	Exact	Most useful for almost complete datasets (e.g. grids with missing values). Does not provide extrapolation
Kriging -Geostatistical models (Stochastic)	Slow/ Medium	Exact if no nugget (assumed measurement error)	Very flexible range of methods based on modeling variograms. Can provide extrapolation and prediction error estimates. Some controversy over aspects of the statistical modeling and inference. Speed not substantially affected by increasing number of data points. Good results may be achieved with <250 data points.
Conditional simulation	Slow	Exact	Flexible range of techniques that use a fitted variogram as a starting point. Effective as a means of reproducing statistical variation across a surface and obtaining pseudo-confidence intervals

From http://www.spatialanalysisonline.com/output/

INTERPOLATION METHODS COMPARED

TIN with Linear Interpolation



Local Polynomial

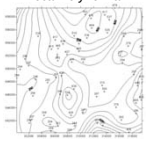
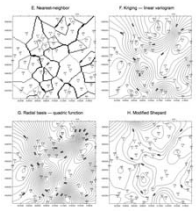


Table 6-2 Gridding and interpolation methods

Method	Speed	Type	Comments
Conditional simulation	Slow	Exact	Flexible range of techniques that use a fitted variogram as a starting point. Effective as a means of reproducing statistical variation across a surface and obtaining pseudo-confidence intervals
Radial basis	Slow/ Medium	Exact if no smoothing value specified	Uses a range of kernel functions, similar to variogram models in Kriging. Flexible, similar in results to Kriging but without addition assumptions regarding statistical properties of the input data points
Modified Shepard	Fast	Exact, unless smoothing factor specified	Similar to inverse distance, modified using local least squares estimation. Generates fewer artifacts and can provide extrapolation
Triangulation with linear interpolation	Fast	Exact	A Delaunay triangulation based procedure. Requires a medium-large number of data point to generate acceptable results
Triangulation with spline interpolation	Fast	Exact	A Delaunay triangulation based procedure, with bicubic spline fitting rather than plane surface fitting. Generates very smooth surfaces. Widely used in computer-aided design
Profiling	Fast	Exact	A procedure that converts contour data to grid format. Similar to linear interpolation methods used in generating contour data (see subsection 6.5.3). See also, Topogrid, subsection 6.6.14
Minimum curvature	Medium	Exact/Smoothing	Generates very smooth surfaces that exactly fit the dataset

INTERPOLATION METHODS COMPARED



Local Polynomial

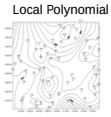
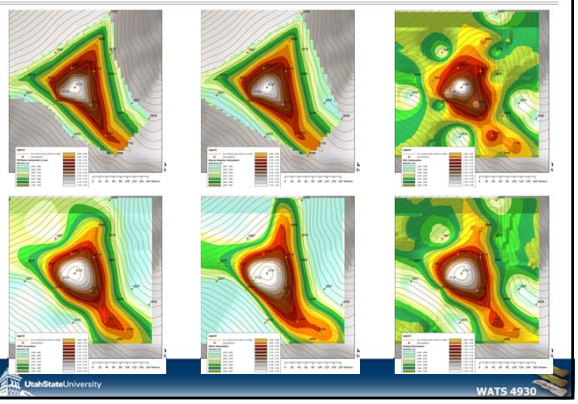


Table 6-2 Gridding and interpolation methods

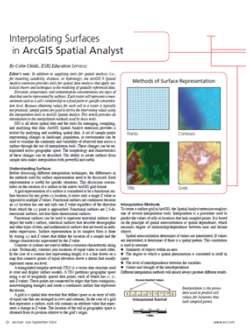
Method	Speed	Type	Comments
Spline functions	Fast	Exact (smoothing possible)	Available as a distinct procedure and incorporated into a number of other methods. Bicubic and biharmonic splines are commonly provided (see for example, Sandwell, 1987)
Local polynomial	Fast	Smoothing	Most applicable to datasets that are locally smooth
Polynomial regression	Fast	Smoothing	Provides a trend surface fit to the data points. Most effective for analyzing 1st-order (linear) and 2nd-order (quadratic) patterns, and residuals analysis/trend removal. Can suffer from edge effects, depending on the data. See also, Section 5.6
Moving average	Fast	Smoothing	Uses averages based on a user-defined search ellipse. Requires a medium-large number of data points to generate acceptable results
Topogrid/Topo to Raster	Slow/ Medium	Not specified	Based on iterative finite difference methods. Interpolates a hydrologically "correct" grid from a set of point, line and polygon data, based on procedures developed by Hutchinson (1988, 1989, 1996). Requires contour vector data as input. Available in ArcGIS based on Hutchinson's ANUDEM program

From http://www.spatialanalysisonline.com/output/

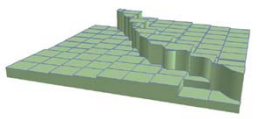
INTERPOLATION TECHNIQUES COMPARED



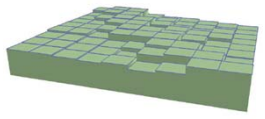
CONCISE ARTICLE SUMMARY



With Barrier

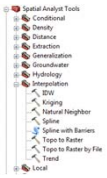


Without Barrier



Barriers, which reflect the presence of fault lines, cliffs, streams, and other features that create linear discontinuity in surfaces, also control how surfaces are generated. IDW and Kriging support the use of barriers.

http://www.esri.com/news/arcuser/0704/files/interpolating.pdf

TODAY'S PLAN

I. Sampling
II. Spatial Interpolation
   I. TIN
   II. Nearest Neighbor
   III. Fixed Radius – Local Averaging
   IV. Inverse Distance Weighted
   V. Natural Neighbor
   VI. Spline
   VII. TOPO
   VIII. Kriging
III. Interpolations Compared