GIS in Hydrology and Water Management

Leibniz Universität Hannover

GIS in Hydrology and Water Management

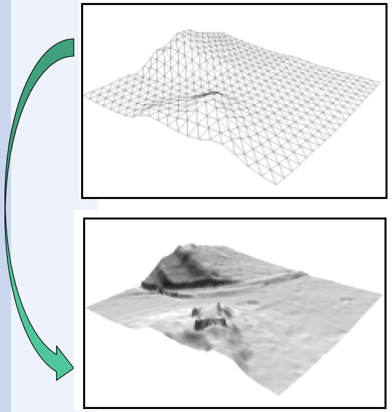
Data Interpolation

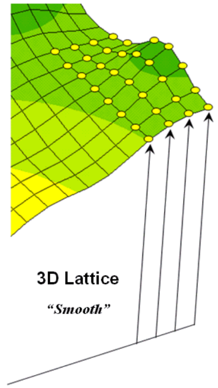
Sagi Dalyot
Anna Maria Walter
Institute of Cartography and Geoinformatics
Leibniz Universität Hannover, Germany

Interpolation

- Motivation:
  Continuous phenomena can only be measured using samples at specific points, e.g. discrete data
- Goal:
  - Derive (predict, approximate) unknown\unsampled value(s) for phenomena at every point in space (raster) via statistical approximation based on a set of known values (observations)
  - Convert data from point observations to continuous fields
- Example:
  - Terrain is continuous phenomenon
  - Measured at specific data points, e.g. samples (observations)
  - Interpolation to generate a continuous surface description
  - elevation, rainfall, chemical concentrations, noise levels,...

Interpolation



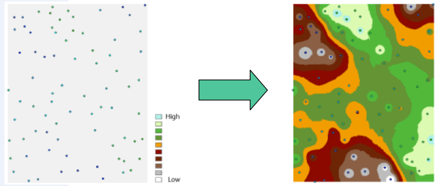


Interpolation

- Determine unknown values from known measurement values
- Idea:
  - Values at locations close to measurements will be similar to those of the unknown measurements
  - Different interpolation methods will almost always produce different results
  - Spatially distributed objects are spatially correlated; things that are close together tend to have similar characteristics, i.e., values close to sampled points are more likely to be similar than those that are farther apart = geo-statistics
  - Crossvalidation and validation – remove one (or more) data and then predict their values using the rest of data; this helps choosing the correct model to use for the phenomenon
- Tobler's First Law of Geography:
  - "Everything is related to everything else, but near things are more related than distant things"

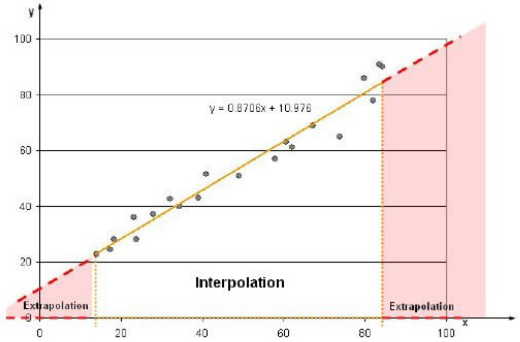
Interpolation

13	14	16	20	23
14	14	16	19	24
18	16	16	18	22
24	22	19	19	21
30	27	23	20	20



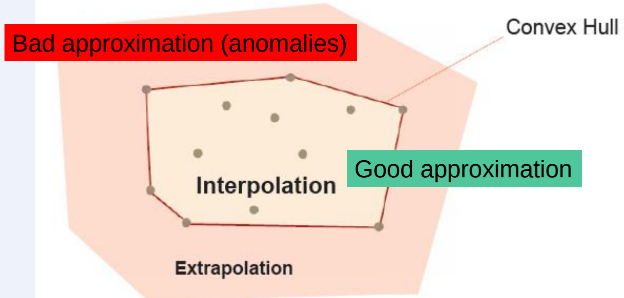
Interpolation / Extrapolation

- 'Inside' measurement (observation) range (set): interpolation
- 'Outside' measurement (observation) range (set): extrapolation



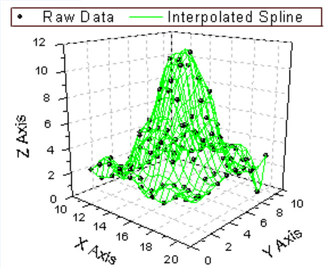
F. Hebeler

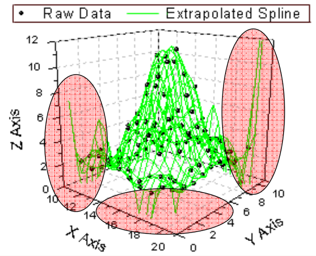
Interpolation / Extrapolation



F. Hebeler

Interpolation / Extrapolation





Interpolation / approximation

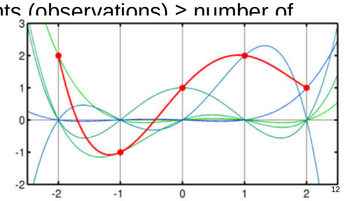
- Different methods:
  - Global vs. local
  - Exact vs. approximated
  - …
- E.g.:
  - Inverse Distance to Weighted (IDW)
  - Polynomial interpolation (mathematical)
  - Spline interpolation
  - Trend surfaces (local neighborhood)
  - Statistical approaches (Kriging)

Interpolation / approximation

- The interpolation tools can be divided into deterministic and geostatistical methods.
  - Deterministic interpolation methods assign values to locations based on the surrounding measured values and on specified mathematical formulas that determine the smoothness of the resulting surface - IDW (Inverse Distance Weighting), NN (Natural Neighbor), Trend, and Spline.
  - Geostatistical methods are based on statistical models that include autocorrelation (the statistical relationship among the measured points). Because of this, geostatistical techniques not only have the capability of producing a prediction surface but also provide some measure of the certainty or accuracy of the predictions - Kriging.

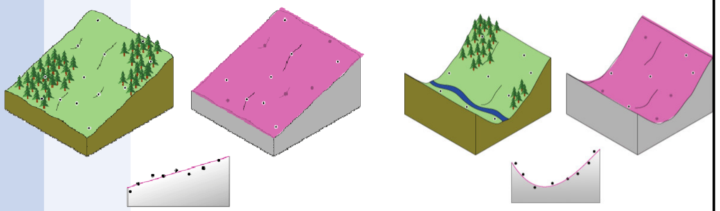
Polynomial interpolation (Trend)

- Given: *n* measured points (observations)
- Goal: compute a function that
  - Goes (exist) through all measured points (exact and global interpolation), or
  - Approximates points best (global approximation, LSM)
- Depending on the number of coefficients of polynomial function (degree), different goals can be achieved
- Number of measurements (observations) > number of polynomial coefficients



Polynomial interpolation (Trend)

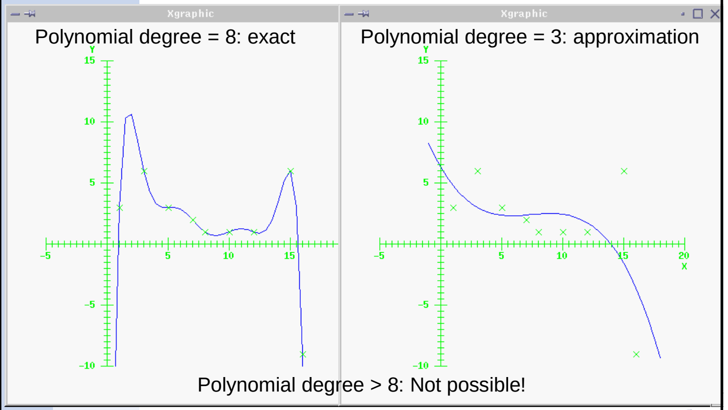
- Order of the polynomial is increased -> fitted surface is more complex. A higher-order polynomial will not always generate the most accurate surface; it is dependent on the data.
- RMS (residuals) can be used to determine the best value to use for the polynomial order (trying to achieve lowest RMS error).
- Trying to fit a piece of paper between observations.



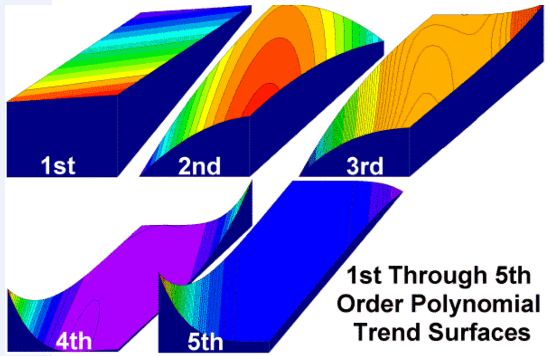
Polynomial interpolation (Trend)

- Using trend interpolation:
  - Normally will result in a smooth surface that represents gradual trends in the surface over the area of interest.
  - Fitting a surface to the sample points when the surface varies gradually from region to region over the area of interest—for example, pollution over an industrial area.
  - Examining or removing the effects of long-range or global trends.
  - The more complex the polynomial, the more difficult it is to ascribe physical meaning to it. Furthermore, the calculated surfaces are highly susceptible to outliers (extremely high and low values), especially at the edges.

2D Polynomial interpolation: 9 points (observations)

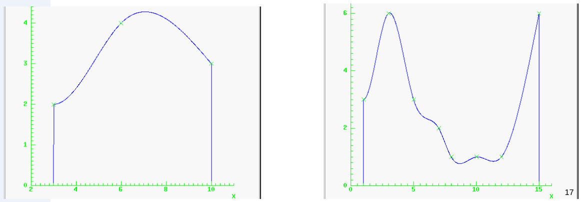


2.5D (Surface) Polynomial interpolation



Splines

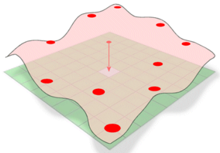
- Adjacent piecewise low-degree polynomial functions between two measurement points (e.g., interval) fit smoothly together
- Typical: cubic polynomial
- Additional constraints: smooth transitions in measured points
- Interpolates using a two-dimensional minimum curvature spline technique; resulting smooth surface passes exactly through the input points.
- Example: 2D spline through 3, and 8 points, respectively



17

Splines

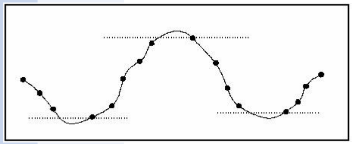
- Spline bends a sheet of rubber that passes through the input points while minimizing the total curvature of the surface. It fits a mathematical function to a specified number of nearest input points while passing through the sample points.
- Method is best for generating gently varying surfaces, such as elevation, water table heights, or pollution concentrations.
  - The surface must pass exactly through the data points.
  - The surface must have minimum curvature (thin plate)
- Ensuring a smooth (continuous and differentiable) surface, together with continuous first-derivative surfaces.

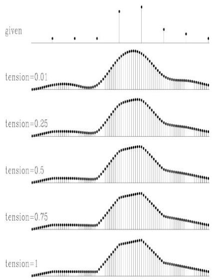


18

Splines

- There are two Spline methods:
  - Regularized - creates a smooth, gradually changing surface with values that may lie outside the sample data range.
  - Tension - controls the stiffness of the surface according to the character of the modeled phenomenon. It creates a less smooth surface with values more closely constrained by the sample data range.

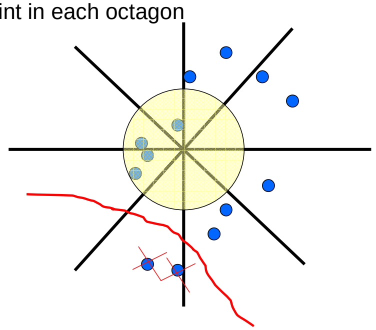




19

Local Interpolation: selection of points in neighborhood

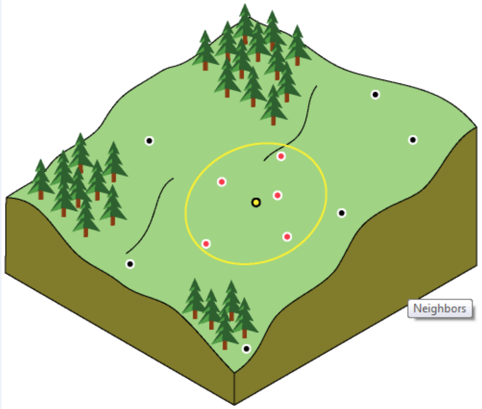
- Well distributed in neighborhood (i.e., not only the closet ones!)
- E.g. 1 point in each octagon



- No points that lie behind break lines

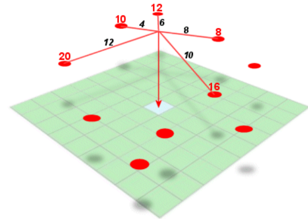
20

Local Interpolation: selection of points in neighborhood



Distance Function: IDW

- Inverse distance weighted (IDW) interpolation determines cell values using a linearly weighted combination of a set of sample points. The weight is a function of inverse distance. The surface being interpolated should be that of a locationally dependent variable.
- This method assumes that the variable being mapped decreases in influence with distance from its sampled location.



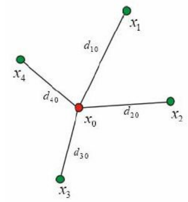
Distance Function: IDW

- Rely mainly on the inverse of the distance raised to a mathematical power. The Power parameter controls the significance of known points on the interpolated values based on their distance from the output point.
- By defining a higher power value, more emphasis can be put on the nearest points (less smooth surface). Specifying a lower value for power will give more influence to surrounding points that are farther away, resulting in a smoother surface.
- An optimal value for the power can be considered to be where the minimum mean absolute error (RMS) is at its lowest.
- The characteristics of the interpolated surface can also be controlled by limiting the input points. Input points far away from the prediction may have poor or no spatial correlation, so there may be reason to eliminate them from the calculation.
- Thus: specifying the number of points or specifying a fixed radius within which points will be included in the interpolation.

Distance Function: IDW

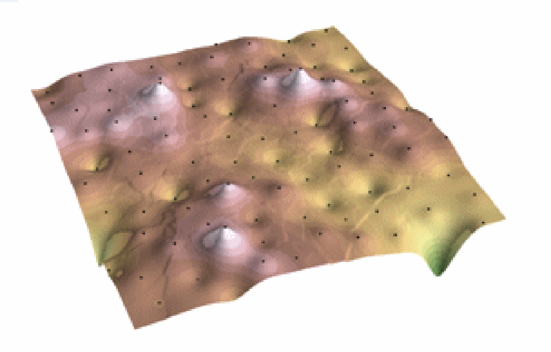
- Idea: Tobler's law
- Height as weighted mean
  - Weight: inverse proportional to distance
  - i.e., close points have highest influence

$$z(x,y) = \frac{\sum_{p=1}^{R} Z_p d_p^{-n}}{\sum_{p=1}^{R} d_p^{-n}}$$



- With
  - z: unknown height of new point (x_0)
  - R: number of points in neighborhood (=4)
  - z_p: height of neighbor point (x_1 – x_4)
  - d: distance between unknown point and neighbor (d_{10} – d_{40})
  - n: power (typical: n = 1 or 2)

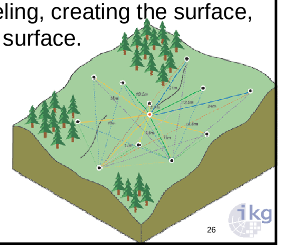
Distance Function: IDW



25

Kringing - Geostatistical

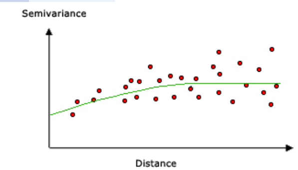
- Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface.
- Kriging fits a mathematical function to a specified number of points (local), or all points (global), to determine the output value for each location.
- Kriging is a multistep process; it includes exploratory statistical analysis of the data, variogram modeling, creating the surface, and (optionally) exploring a variance surface.



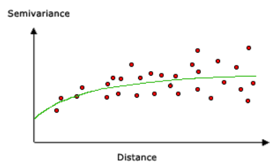
26

Kringing - Geostatistical

- Kriging is most appropriate when knowing that there exists a spatially correlated distance or directional bias in the data. It is often used in soil science and geology.
- Based on statistical models that include autocorrelation—that is, the statistical relationships among the measured points. Because of this, geostatistical techniques not only have the capability of producing a prediction surface but also provide some measure of the certainty or accuracy of the predictions.



Spherical



Exponential

27

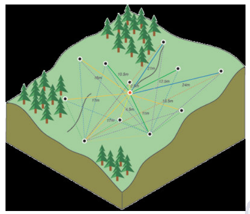
Kringing - Geostatistical

- Idea: statistical models that include autocorrelation (correspondence between observations)
- Providing measures of the certainty or accuracy of predictions, e.g., heights.
- Distance (or direction) between samples corresponds to spatial correlation explained by surface variations.

- With

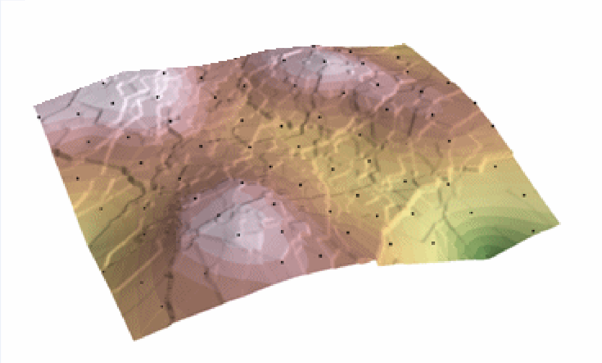
$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

  - $Z(s_i)$: measured va
  - λ_i: unknown weight for position i
  - s_0: prediction position
  - N: number of observations



28

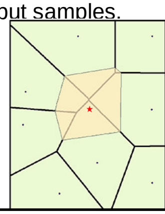
Kringing - Geostatistical



29

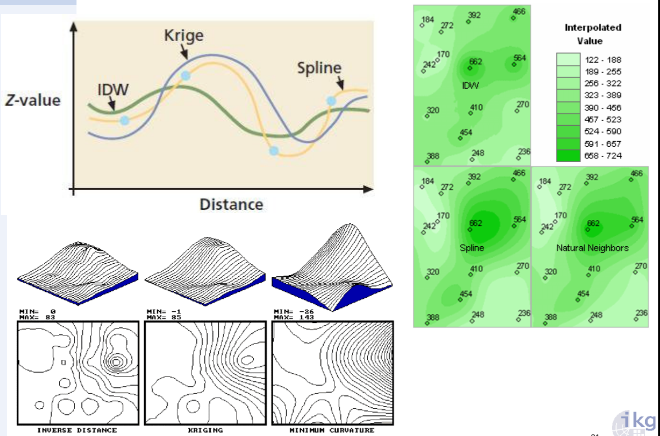
Natural Neighbor

- Finding the closest (local) subset of input samples to a query point and applies weights to them based on proportionate areas to interpolate a value ("area-stealing" interpolation).
- Interpolated heights are guaranteed to be within the range of the samples used.
- It does not infer trends and will not produce peaks, pits, ridges, or valleys that are not already represented in data.
- Surface passes through the input samples and is smooth everywhere except at locations of the input samples.



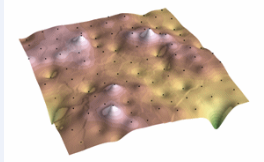
30

Comparison

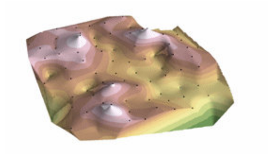


31

Comparison

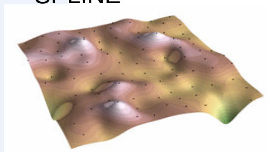


IDW2

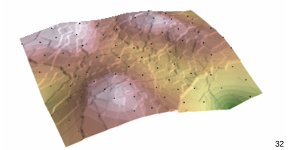


NN

SPLINE

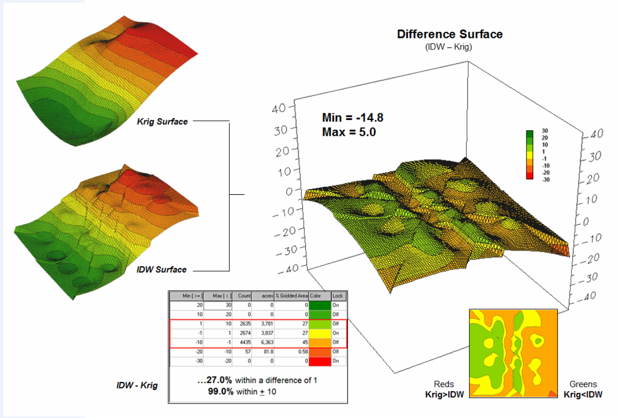


KRIGING



32

Comparison



Summary

- Geometric algorithms and statistics are essential for the automatic analysis of spatial data (continuous phenomena).
- Algorithms have to cope with large volumes of data.
- Discipline in Computer Science dealing with these problems
  - Computational Geometry
- Ongoing development:
  - precise definition of spatial operators
  - implementation in database system (e.g. set operations, buffer, distance, polygon overlay, …) -> e.g. realized in Database Management System Oracle Spatial
  - Analysis-derived (specified) algorithms