Interpolation
Procedure of estimating value of properties at unsampled sites within the area covered by existing observations
As opposed to extrapolation
Many, many methods exist
All rely upon assumptions about how property varies across the area
Spatial Autocorrelation (spatial dependence)
Any method only as good as the degree to which assumptions match reality

Why interpolate elevation data?
To build regular grid from scattered points
To build grid from contour lines
Turn contour lines into points & go from there
To identify elevation at off-grid points
To resample a raster grid
New resolution
Change of projection

Classifying Interpolation Methods
Points to Points – our focus today
Lines to Points
Contours
Also mix: use contours & spot heights to build grid
Areal
Transform values from one tiling scheme to another; not common for elevation data
Global
Uses all points in region to establish point height
Trend Surface fitting
Local
Uses nearby points (local neighborhood)
Many methods are effectively local
IDW, Kriging, Splines
Exact
Interpolated surfaces passes through known points
In geostats speak, it 'honors the data'.
Approximate
Interpolated surface not constrained to honor data
Typical for trend surfaces
Deterministic
A single value is estimated for each point
Stochastic
Height characterized as a random variable
Probabilities of different heights may be assessed
Gradual
Surface is relatively smooth
Abrupt
Discontinuities are employed in the interpolation
Fault lines
Univariate
Covariate
Additional variable(s) used to help with interpolation
Not typical for elevation interpolation, but...
Elevation makes a useful covariate

Proximal (Nearest Neighbor)
Value is assigned using the closest known point
\[ Z(u,v) = Z \text{ (nearest point)} \]
Default in raster resampling algorithms
Fast to calculate
Not Smooth
Conceptually, a Thiessen Polygon (Voronoi)

Splines
Piecewise (local) polynomial functions
Fit a curve to a subset of points
Higher order very smooth
No discontinuities in 1st or 2nd derivatives
Smooth in elevation, slope, curvature
Minimizes local surface change
Exact or data smoothing
Poorer for rough surfaces

Fitting Splines
Linear: order 1
Linear segments connect the dots
Quadratic & Cubic: orders 2 & 3
Enable curves
B-splines
Weights contributions of each point in the subset
Enables much greater local control
Greater smoothness

Smoothness and Piecewise
By using subsets, algorithms are suited for very large problems
Millions of points
Higher order = Smoother
Other factors important too
Tension – controls degree of smoothness of the overlap between subsets
Larger tension = more local control
Lower tension = greater generalization

Inverse Distance Weighting
aka IDW, distance weighted average
Assign weight to each nearby elevation based on distance from point to be interpolated
\[ Z = \frac{\sum (z_i \cdot w_i)}{\sum w_i} \]
i = 1 to # of points
w: weight based on distance: \[ w_i = 1/d_i^k \]
d is distance between point i and point t.b.i.
k is a real number, usually between 0-2.
Example
Unweighted Average: \( (10 + 14 + 6 + 9) / 4 = 9.75 \)
Proximity (N-N): 6

Effect of IDW k parameter
As k gets large, more distant points get very small weights
Influence of nearby points is very large
As k approaches 0, weights approach 1
  Simple unweighted average
k is very important! How to choose?
  2, because it's usually the default
  Try different ones, see how the surfaces look
  Cross-validation

Bilinear Interpolation
Variant of IDW that uses relative distance, not absolute, in a two-stage process
Identify a few (typically 4) surrounding points
  Good for interpolating with a grid
Estimates fall within range of the data
Continuous, but not smooth (jumps at edges)
Fast, stable
Step 1: linearly interpolate the blue dots
Step 2: linearly interpolate X

Kriging
Geostatistical approach for estimation
Actually a form of IDW, but w/ complex weights
Weights depend on a variogram model and on the configuration of nearby points
  Distance from nearby points is included
  Distance between nearby points is included
Kriging supplies an estimate and a standard error (uncertainty about the estimate)
Learn more in GEO 866!

Triangulated Irregular Network
TIN
x,y,z coordinates of important points stored
Points are joined w/ vertices to form triangles
Area between points interpolated as a plane

Why 'triangular'?
Why not squares, pentagons, etc?
Triangles fill the plane and avoid gaps in surface
All pieces are guaranteed to fit
Much terrain is effectively represented by triangular facets
  Fluvially eroded landscapes
  Sharp breaks in slope, where triangle sides can match the breaks
Why 'irregular'?  
Great flexibility in shape introduced  
  - Many points can be employed in rugged areas  
  - Few points used in flat areas  
Able to handle both regular and irregular constellations of points  
(x,y,z) must all be stored explicitly  
  - Unlike raster  
  - More on storage in a bit

Preferential Sampling in Rugged Terrain

Why 'network'?  
Connections between points comprise network  
Configuration of network critical to model  
  - As important as the (x,y,z) data

Creating TINs  
Assume a set of (x,y,z) data exists  
The big challenge is identifying triangles  
There are usually many tessellations  
  - Delaunay Triangulation is common

TINs to Rasters  
TINs interpolate very easily to rasters  
  - Default QGIS approach  
Overlay grid of locations (cell centers)  
Interpolate height at the location  
  - Planar: each point falls within a facet w/ constant slope  
  - Cubic: curved surface fit to each facet

Summary  
Typologies of Interpolation  
Methods  
  - Nearest Neighbor  
  - Splines / Local polynomials  
  - IDW  
  - Bilinear Interpolation  
  - Kriging  
TINs  
  - Structure  
  - Facet interpolation