# GEOG574G Introduction to Geostatistics

Spatial Interpolation

**Daoqin Tong**

School of Geography & Development
408 Harvill Building
Email: daoqin@email.arizona.edu

## Geostatistical Data

- Observations of a spatial stochastic process $\{Y(s), s \in \Re\}$ , where $Y(s)$ varies in R continuously, has been sampled at a finite number of fixed locations $s_1, \ldots, s_n$
- Objective: to infer the nature of spatial variation in an attribute over the whole study area R based on sampled point values
- Notations

$$(Y(s_1),\ldots,Y(s_n))^T \quad (y(s_1),\ldots,y(s_n))^T \Rightarrow (y_1,\ldots,y_n)^T$$

## Properties

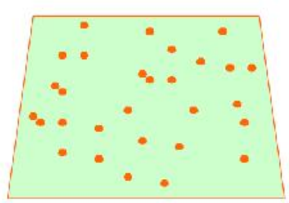
- First order effects
  - The mean value $E(Y(s)) = \mu(s)$
  - Global trend, large scale variation
- Second order effects
  - Spatial dependence between two points

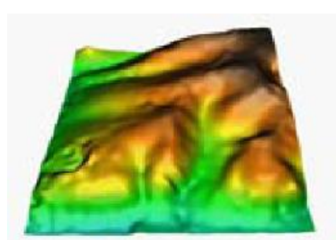
$$COV(Y(s_i), Y(s_j))$$

  - Local or small scale variation

## Introduction to Interpolation

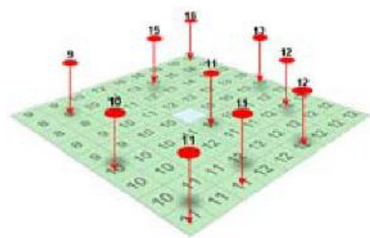
- We can't measure phenomena at every place in an area, so we sample
- Using samples, we can make inferences about the non-data locations and hence the nature of a phenomena
- With spatial interpolation, the goal is to create a surface that models the sampled phenomenon in the best way possible





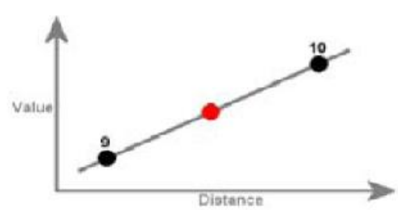
## Introduction to Interpolation

- The primary assumption of spatial interpolation is that points near each other are more alike than those farther away; therefore, any location's values should be estimated based on the values of points nearby



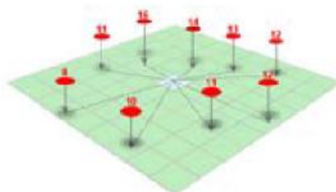
## What is Interpolation

- Interpolation is the process of estimating unknown values that fall between known values



## Spatial Interpolation

- Spatial interpolation calculates an unknown value from a set of sample points with known values that are distributed across an area
- The distance from the cell with unknown value to the sample cells contributes to its final value estimation



## Comments

- You can use spatial interpolation to create an entire surface from just a small number of sample points. However, more sample points are better if you want a detailed surface
- In general, sample points should be well-distributed throughout the study area
- Some areas, however, may require a cluster of sample points because the phenomenon is transitioning or concentrating in the location
- For example, trying to determine the size and shape of a hill might require a cluster of samples, whereas the relatively flat surface of the surrounding plain might require only a few

## Types of Spatial Interpolation

- To create a surface from a set of points, you can use an interpolator:
  - Exact vs. Inexact Interpolators
  - Global vs. Local Interpolators
  - Deterministic vs. Stochastic Interpolators
  - Gradual vs. Abrupt Interpolators

## Interpolators

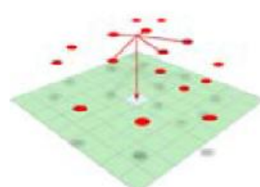
- Exact interpolators
  - Return the exact value of an interpolated surface at sampled sites
- Inexact interpolators
  - Seek to fit smooth functions that may fit some criteria of the surface, such as minimizing the squared deviations of the interpolated values to sampled values
- Global interpolators
  - Fit a function and use information for the full spatial domain: trend and surfaces (polynomial fits)
- Local interpolators
  - Fit a function drawn from a local neighborhood, specified by a radius or number of nearest neighbors. Moving averages are the most common local interpolators

## Interpolators

- Deterministic interpolators
  - Same result each time when the model is run
- Stochastic interpolators
  - Incorporate some kind of randomness in the model
- Gradual interpolators
  - Create smoothed surfaces overall
- Abrupt interpolators
  - Include barriers in the model

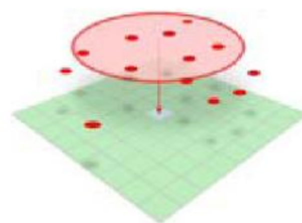
## Sample Size

- Some interpolation methods allow you to control the number of sample points used to estimate cell values
  - For example, if you limit your sample points to five, the interpolator will use the five nearest points to estimate unknown cell values
- The distance to each sample point will vary depending on the distribution of the points
- If you have a lot of sample points, reducing the size of the sample you use will speed up the interpolation process because a smaller set of numbers will be used to estimate each cell value



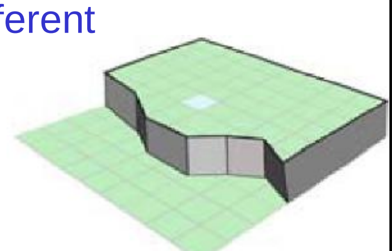
## Sample Size Radius

- You can also control your sample size by defining a search radius
- The number of sample points found within a search radius can vary depending on how the points are distributed
- You can choose to use some or all of the samples that fall within this radius to calculate the cell value.
- A variable search radius will continue to expand until the specified sample size is found
- A fixed search radius will use only the samples contained within it, regardless of how many or how few that might be



## Interpolation Barriers

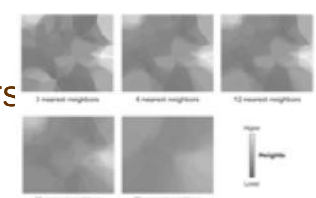
- The physical, geographic barriers that exist in the landscape, like cliffs or rivers, present a particular challenge when trying to model a surface using interpolation
- The values on either side of a barrier that represents a sudden interruption in the landscape are drastically different



## Interpolation Methods

- Spatial Moving Averages
  - Unweighted averages
  - Weighted averages
    - ⇒ Inverse Distance Weighting (IDW)
    - ⇒ Kernel Estimation
- Interpolation based on Tessellations
  - Piecewise polynomial functions
  - Spline functions

## Spatial Moving Averages



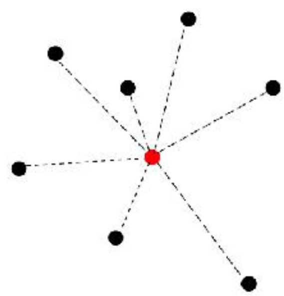
- Unweighted Averages
  - Average *k* nearest neighbors of location *s*
  - Equal weights
- Weighted Averages
  - Weighted average of data points around location s
  - Weights vary with distances

## Inverse Distance Weighting (IDW)

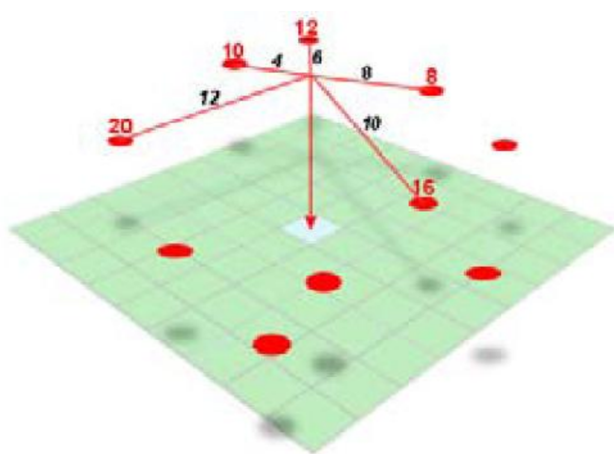
- Each input point has local influence that diminishes with distance
- Estimates are weighted averages of values at *n* known points within window

$$\hat{y}(s) = \frac{\sum_{i=1}^{n} w(s_i, s)\, y(s_i)}{\sum_{i=1}^{n} w(s_i, s)}$$

$$w(s_i, s) = d(s_i, s)^{-k}, \quad w(s_i, s) = e^{-kd(s_i, s)}$$



## Inverse Distance Weighting

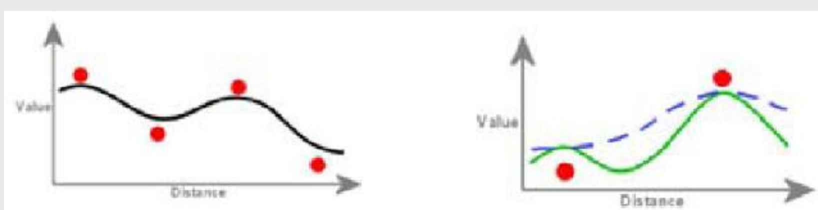



Each of the five sample points have different values and distances from the estimated cell

# Inverse Distance Weighting

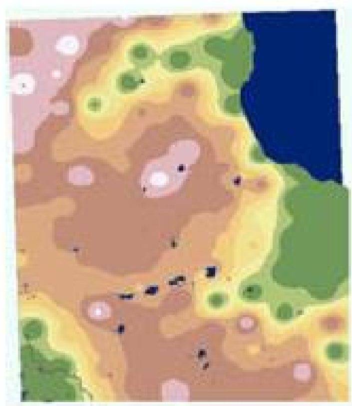
- IDW works best for dense, evenly-spaced sample point sets
  - It does not consider any trends in the data, so, for example, if actual surface values change more in the north-south direction than they do in the east-west direction (because of slope, wind, or some other factor), the interpolated surface will average out this bias rather than preserve it
- IDW interpolation considers the values of the sample points and the distance separating them from the estimated cell
  - Sample points closer to the cell have a greater influence on the cell's estimated value than sample points that are further away.
- Inverse Distance Weighting cannot make estimates above the maximum or below the minimum sample values
  - For an elevation surface, this has the effect of flattening peaks and valleys (unless their high and low points are part of the sample). Because the estimated values are averages, the resulting surface will not pass through the sample points

# IDW: Adjusted Power Settings

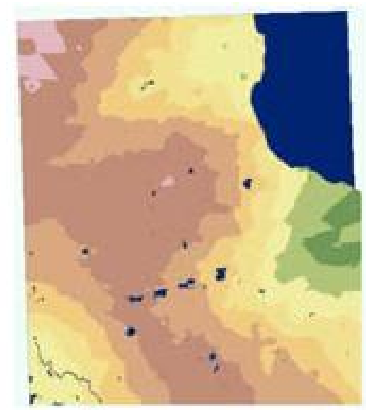



- You can adjust the relative influence of sample points. In other words, you can increase how much power the values of sample points have over the interpolation process. Increased power means that the output cell values become more localized and less averaged. Their influence, however, drops off rapidly with distance
- Lowering the power that sample point values have provides a more averaged output because sample points farther away become more and more influential until all of the sample points have the same influence.

## IDW: Adjusting the Power Settings



Power of 2



Power of 0.5

## Kernel Estimation

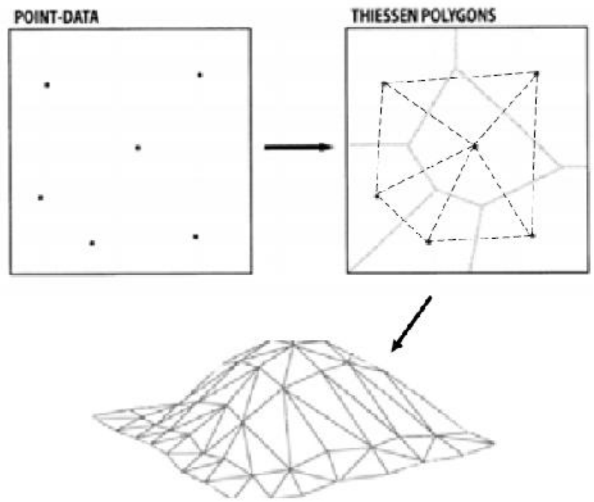
- Recall, in point patterns, the intensity is estimated by

$$\hat{\lambda}_\tau(s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{(s-s_i)}{\tau}\right)$$

- The weighted average uses kernel function to calculate weights

$$\hat{y}(s) = \frac{\sum_{i=1}^{n} w(s_i, s) y(s_i)}{\sum_{i=1}^{n} w(s_i, s)} \qquad w(s_i, s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{(s-s_i)}{\tau}\right)$$

## Interpolation Based on Tessellations

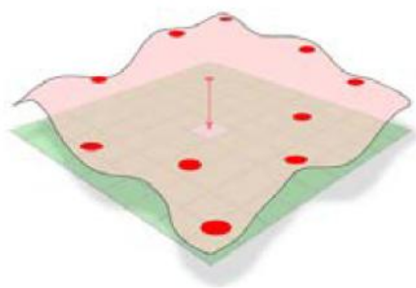



## Tessellations

Piecewise Polynomial Functions

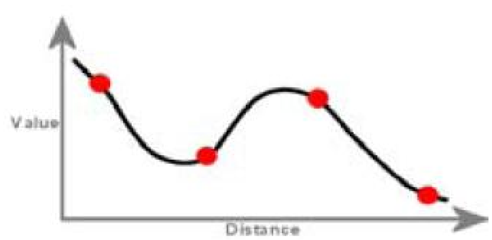
## Spline

- Instead of averaging values, like IDW does, the Spline interpolation method fits a flexible surface, as if it were stretching a rubber sheet across all the known points values



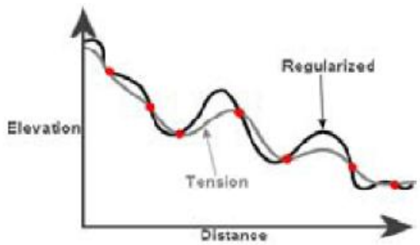
## Spline

- This stretching effect is useful if you want estimated values that are below the minimum or above the maximum values found in the sample data. This makes the Spline interpolation method good for estimating lows and highs where they are not included in the sample data

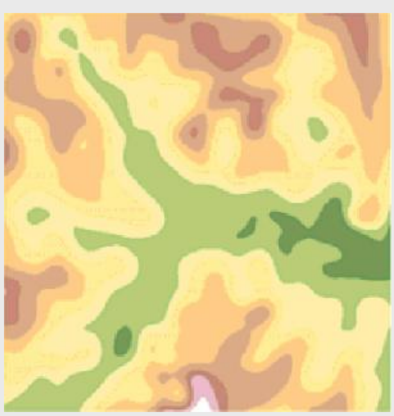


## Types of Spline

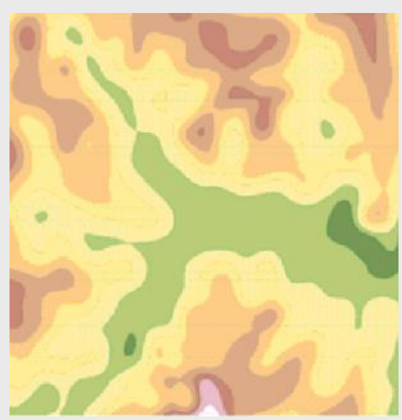
- Regularized and Tension
- A Tension Spline is flatter than a Regularized Spline of the same sample points, forcing the estimates to stay closer to the sample data
- Tension Spline method produces a surface more rigid in character, while the Regularized Spline method creates one that's more elastic.




## Spline Weights



Spline Regular 0.1



Spline Regular 1