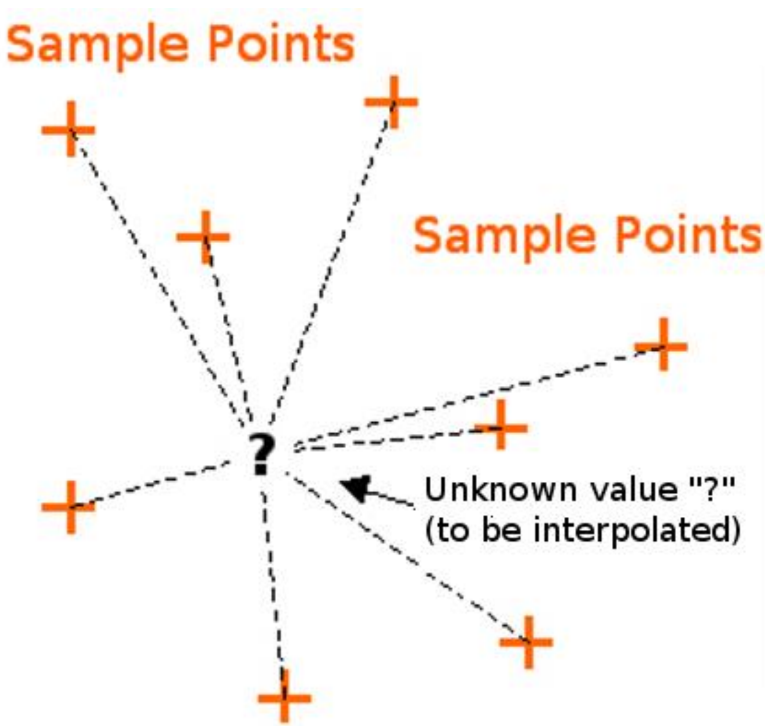
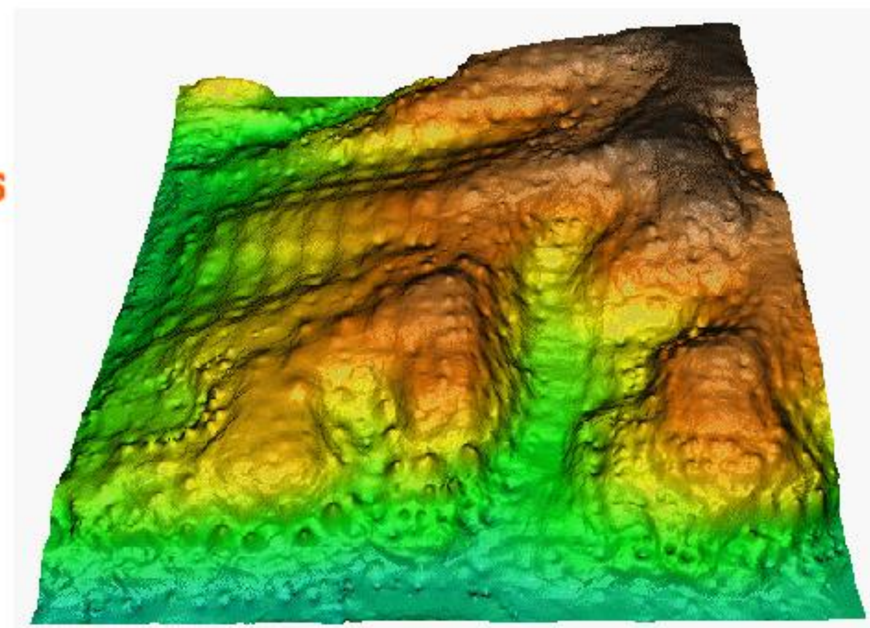
Geostatistics and Interpolation

SPATIAL INTERPOLATION

- **Things that are close together tend to have similar characteristics.**
- **Spatial Interpolation** is the process of **estimating the value of unknown locations across space from a set of observations.**
  - such as elevation, rainfall, chemical concentrations, noise levels, and so on.
- An **observation** is a location on the surface at which measurements of an attribute have been made





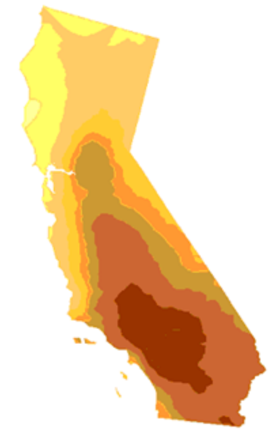
Spatial Interpolation Examples

Input rainfall point data

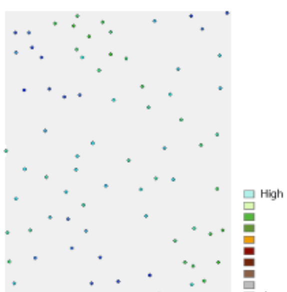
13	14	16	20	23
14	14	16	19	24
18	16	16	18	22
24	22	19	19	21
30	27	23	20	20

Interpolated rainfall surface

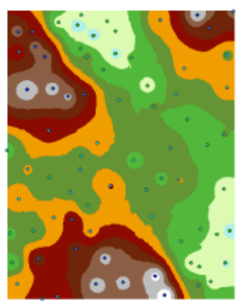
Point locations of ozone monitoring stations



Interpolated prediction surface



Input elevation point data



Interpolated elevation surface

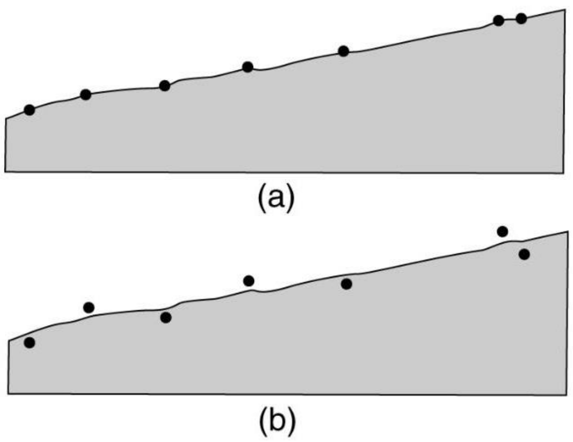
https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/understanding-interpolation-analysis.htm

Methods of Interpolation

- *Deterministic* methods
  - Use **mathematical functions** to calculate the values at unknown locations based either on the **degree of similarity** (e.g. IDW) or the **degree of smoothing** (e.g. RBF) in relation with neighboring data points.
  - Examples include:
    - Inverse Distance Weighted (IDW)
    - Radial Basis Functions (RBF)
- *Geostatistical* methods
  - Use both **mathematical and statistical methods** to predict values at all locations within region of interest and to provide **probabilistic estimates** of the quality of the interpolation based on the **spatial autocorrelation** among data points.
    - Include a deterministic component and errors (uncertainty of prediction)
  - Examples include:
    - Kriging
    - Co-Kriging

Exact vs. Inexact Interpolation

- Interpolators can be either *exact* or *inexact*
  - At sampled locations, ***exact interpolators*** yield values **identical** to the measurements.
  - At sampled locations, ***inexact interpolators*** predict values that are **different from** the measured values.
    - The **resulting surface** will **not pass through** the original point
    - Can be used to **avoid sharp spikes or dips** in the output surface
    - **Model quality** can be assessed by the statistics of the **differences between predicted and measured values**
  - The IDW is exact. Kriging can be exact or inexact.



Reference: Burrough, P. A., and R. A. McDonnell. 1998. Principles of geographical information systems. Oxford University Press, Oxford. 333pp.

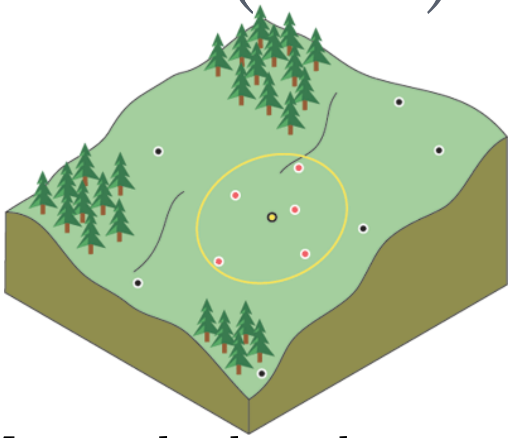
Inverse Distance Weighted (IDW)

Inverse Distance Weighted (IDW)

- **Method:** value at each grid cell location is a distance weighted average of the values at the **nearby** observations

$$z' = \frac{\sum_{i}^{k} \left(z_i \times \frac{1}{d_i^p}\right)}{\sum_{i}^{k} \frac{1}{d_i^p}}$$

- where
  z' = interpolated value
  $\boldsymbol{z_i}$ = observed value at location *i*
  $\boldsymbol{d_i}$ = Euclidean distance between the observations *i* and location of unknown value
  p = exponent modeling friction of distance
  k = number of observations in neighborhood



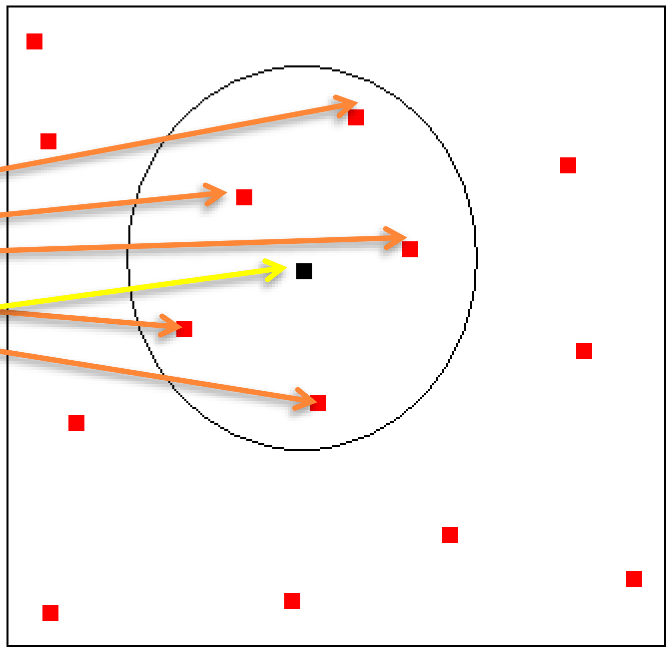
Measured values that are **nearest to the prediction location** will have **greater influence** (i.e., *weight*) on the predicted value at that unknown point **than those that are farther away**.

Weights of each measured point are **proportional to the inverse distance raised to the power value *p***. As a result, as the distance increases, the weights decrease rapidly. **How fast the weights decrease is dependent on the value for *p*.**

Search Neighborhood Specification

- Specify the k
  - Number of Neighbors

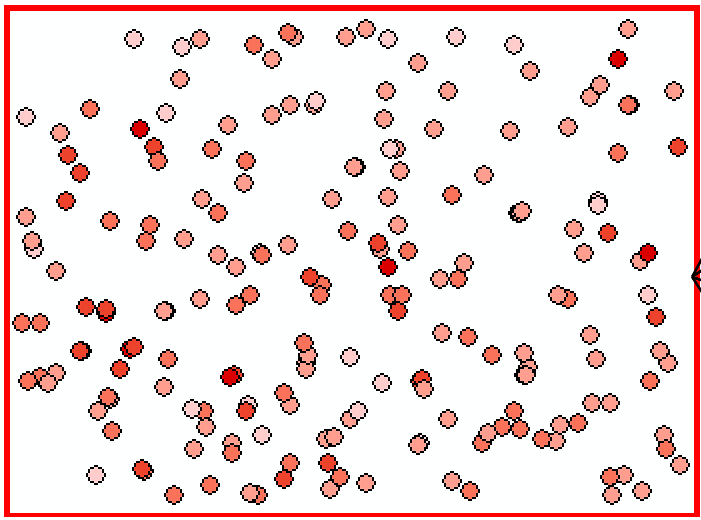
e.g. 5 nearest neighbors with known values (shown in red) of the unknown point (shown in black) will be used to determine its value

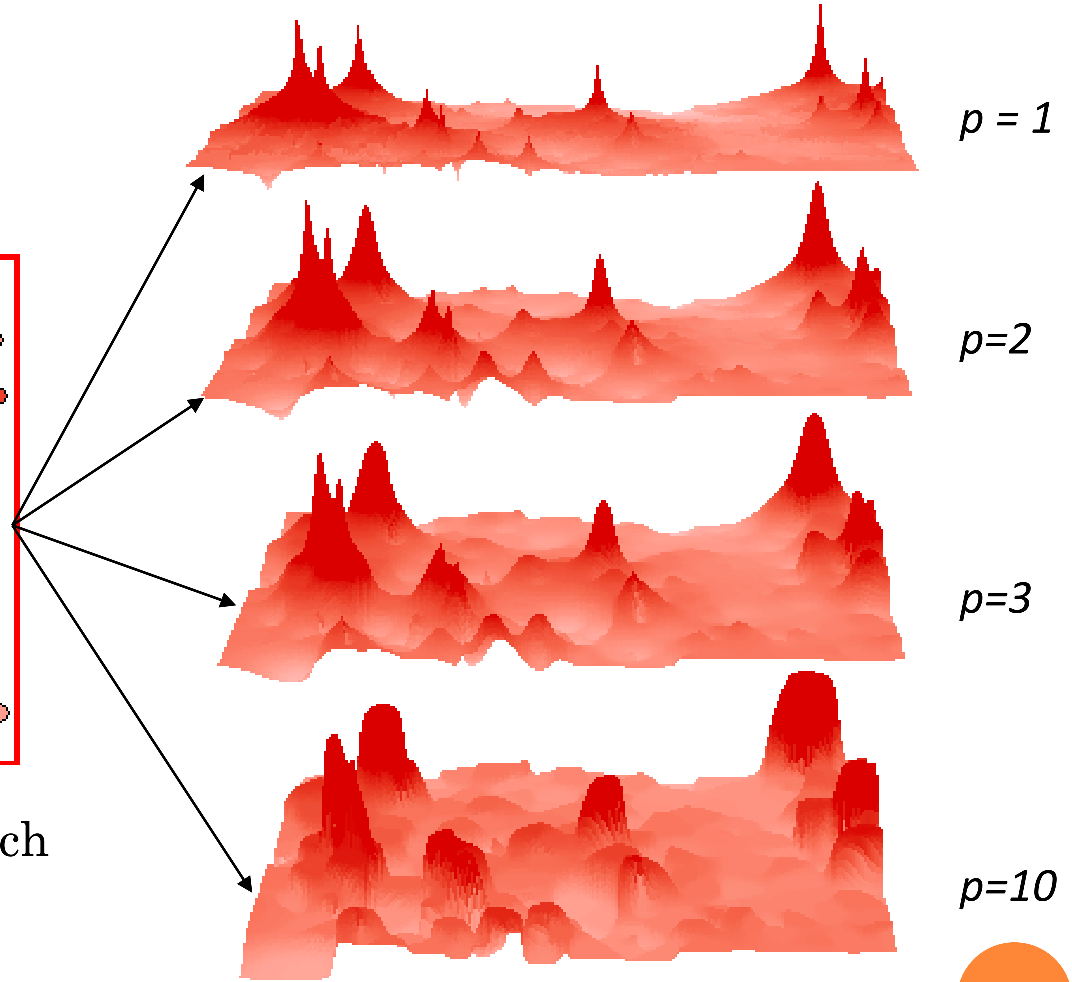


  - Fixed search radius

Examples of IDW with Different *p*'s

Gold concentrations at locations in western PA





- Larger *p*'s (i.e., power to which distance is raised) yield smoother surfaces
- Food for thought: What happens when *p* is set to 0?

Inverse Distance Weighted (IDW)

- **Advantages**
  - Results in a continuous and smooth surface
  - Is an exact interpolator (i.e. derived surface passed through observed values)

- **Disadvantages**
  - Requires subjective selection of parameters (*k* and *p*)
  - Does not extrapolate beyond the minimum and maximum of observed values
  - Does not consider direction of observations

- **Application**
  - most appropriate when the phenomenon presents local variability

The Accuracy of the Results

One way to assess the accuracy of the interpolation is known as ***cross-validation***

- Remember the initial goal: use *all* the measured points to create a surface
- However, assume we **remove *one*** of the measured points from our input, and **re-create the surface** using all the remaining points.
- Now, we can look at the ***predicted value*** **at that removed point** and **compare it to the point's *actual value***!
- We do the same thing **for all the points**
- If the average (squared) difference between the actual value and the prediction is small, then our model is doing a good job at predicting values at unknown points. If this average squared difference is large, then the model isn't that great.

Kriging

Kriging

Mathematical methods of interpolation (e.g. local spatial average, IDW) determine the distance weighting function and neighborhood definition based on expert knowledge, not from the data

Kriging estimates the choice of function, weights, and neighborhood from the sampling data, and interpolate the data with these choices.

Kriging

Kriging is a statistical interpolation method that is **optimal** in the sense that it makes best use of what can be inferred about the spatial structure in the surface to be interpolated **from an analysis of the control point data**.

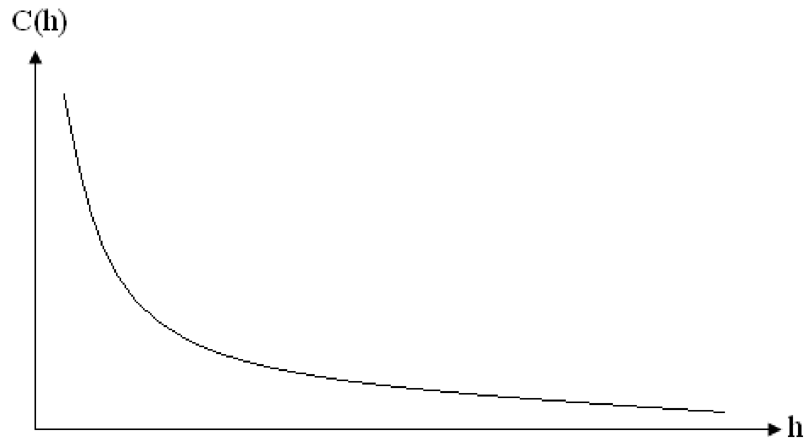
✓Methods used in the South African mining industry by Danie Krige

Three steps

1) Produce a description of the spatial variation (**Semi-variogram**) in the sample control point data
2) Summarizing the spatial variation by a regular **mathematical function**
3) Using this model to determine the **interpolation weights**
4) Create a **prediction surface**
5) Use **cross-validation** to evaluate the model performance.

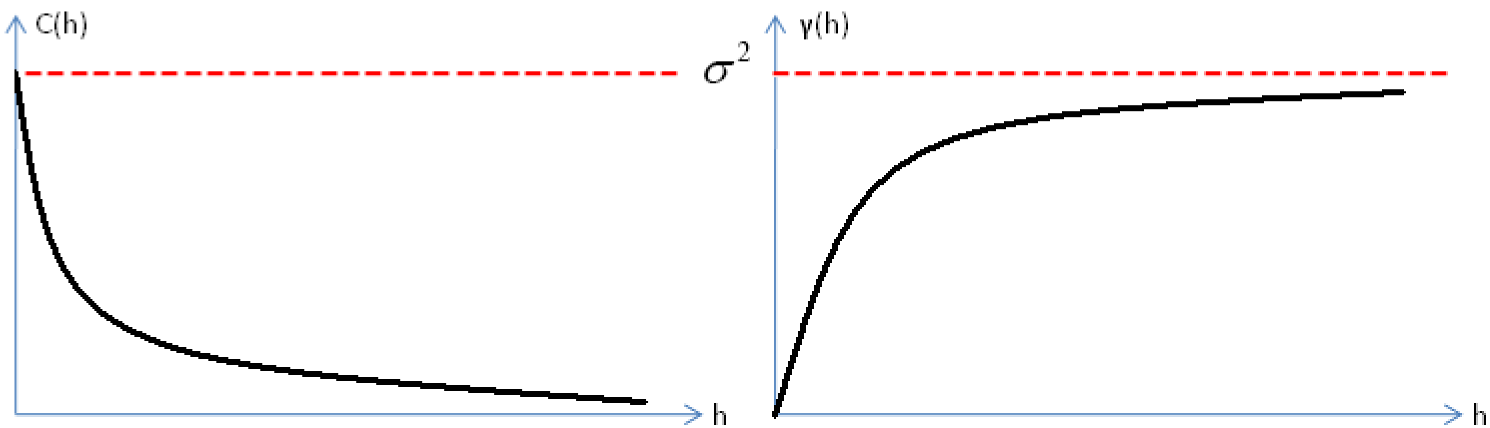
Covariance and Distance

- From the First Law of Geography it would then follow that **as distance between points increases**, the **similarity** (i.e., covariance or correlation) between the values at these points **decreases**
- If we plot this out, with **inter-point distance *h*** on the x-axis, and **covariance *C(h)*** on the y-axis, we get a graph that looks something like the one below. This representation of *covariance* as a function of distance is called as the *covariogram*



But…

- Unfortunately, it so happens that one generally cannot estimate covariograms directly
- For that purpose, a related function of distance (*h*) called the *semi-variogram* (or simply the *variogram*) is calculated
  - The variogram is denoted by γ(h)
  - One can easily obtain the *covariogram* (*C(h)*) from the *variogram*(γ(h)) (but not the other way around)
- Covariograms and variograms tell us the spatial structure of the data



Covariogram C(h)

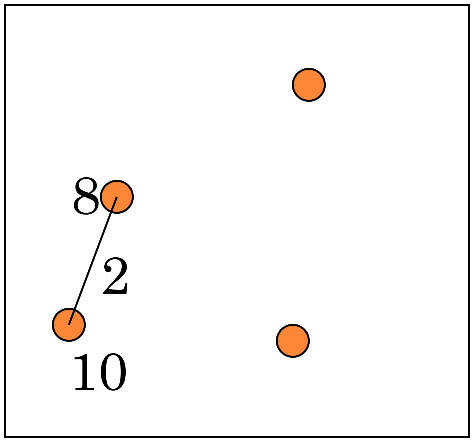
Variogram γ(h)

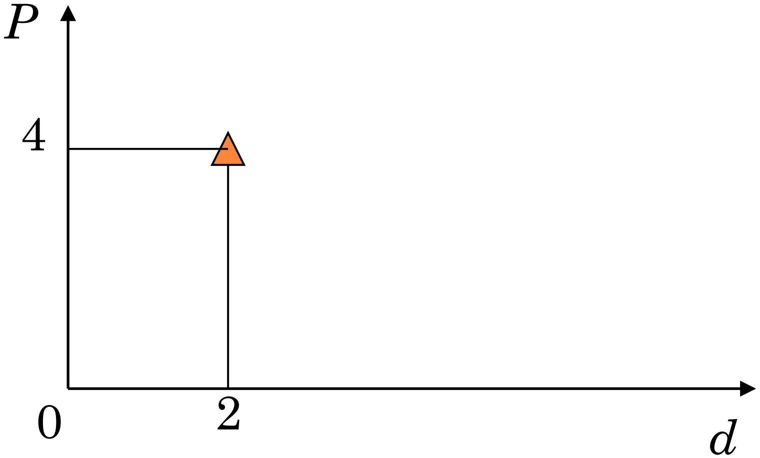
Kriging

- Describing the spatial variation: the semi-variogram

Variogram cloud: a plot of a measure of differences against the distance d_{ij} between the control points for all possible pairs of points.

$$P_{ij}(d) = (z_i - z_j)^2$$

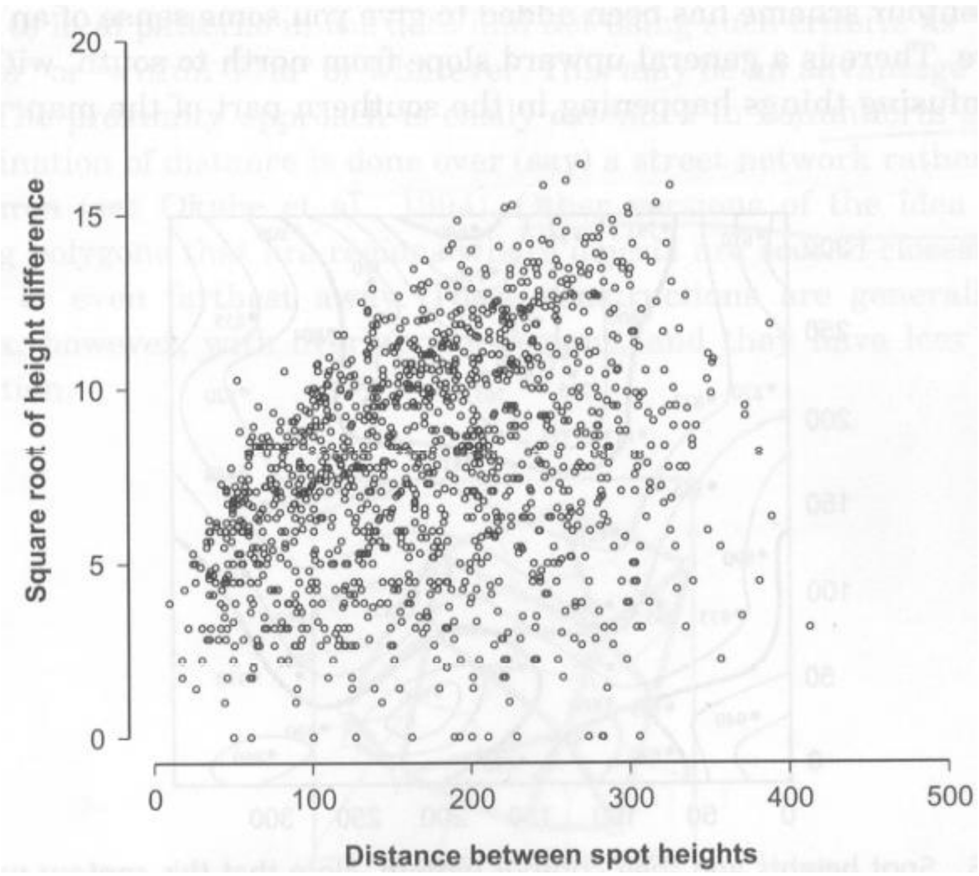




KRIGING

- Describing the spatial variation: the semi-variogram

Example of variogram cloud



There is a trend such that height differences increase as the separation distance increases

Indicating the farther apart two control points are, the greater is the likely difference in their value.

KRIGING

- Describing the spatial variation: the semi-variogram

Spatial dependence can be described more concisely by the experimental semivariogram function as follows

$$2\hat{\gamma}(d) = \frac{1}{n(d)} \sum_{d_{ij}=d} (z_i - z_j)^2$$

n(d) is the number of pair of points at separation *d*
$\hat{\gamma}$ is the estimated semi-variogram

$$2\hat{\gamma}(d) = \frac{1}{n(d \pm \Delta/2)} \sum_{d \pm \Delta/2} \left(z_i - z_j\right)^2$$

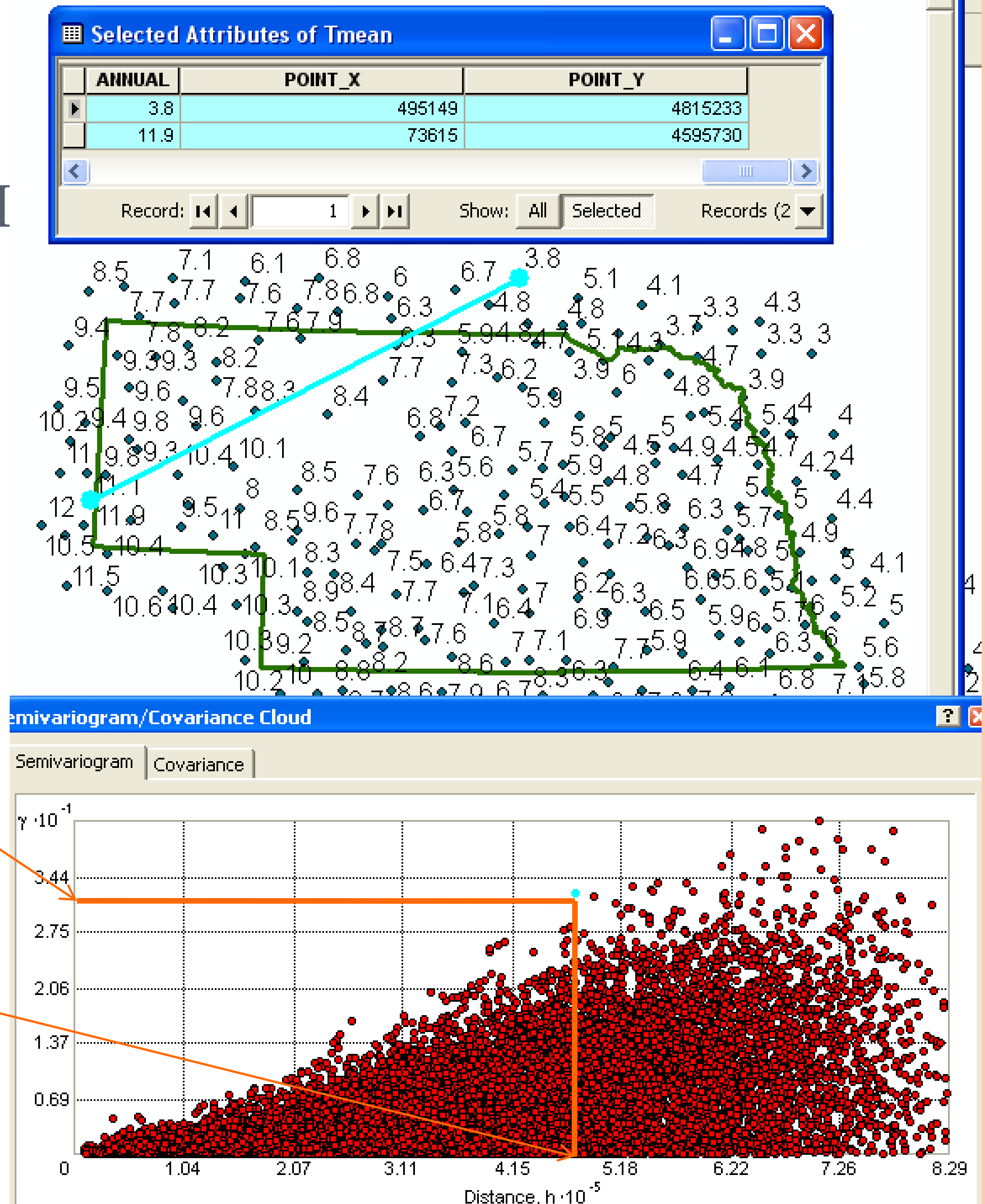
SEMIVARIOGRAM

$$\gamma = \frac{1}{2} \sum_{i,j} \left[z_i - z_j \right]^2$$

$$\gamma = \frac{1}{2} \sum_{i,j} \left[3.8 - 11.9 \right]^2$$

$$\gamma = 32.805$$

Dist = 4.75 x 10^5m



Kriging

The estimate to be calculated, i.e. an output pixel value $\hat{z}$, is a linear combination of weight factors (w_i) and known input point values (Z_i):

$\hat{Z} = \Sigma(w_i * Z_i)$

In case the value of an output pixel would only depend on 3 input points, this would read:

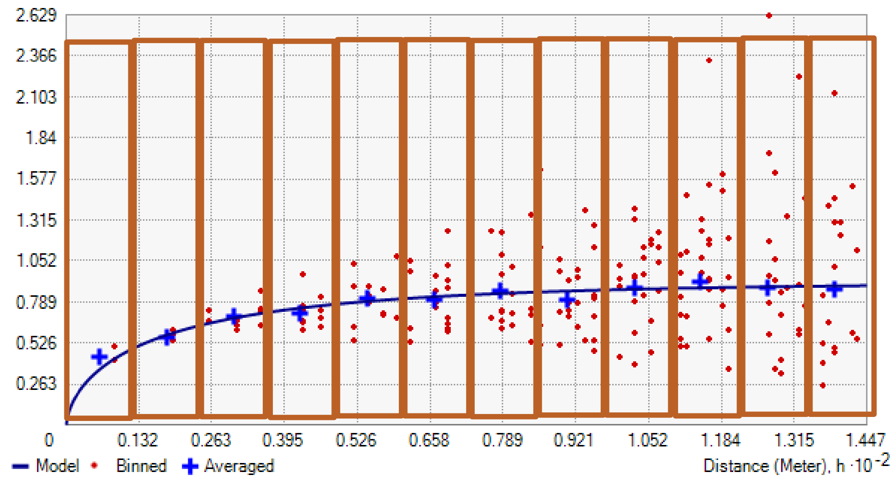
$\hat{Z} = w_1 * Z_1 + w_2 * Z_2 + w_3 * Z_3$

1. Producing a description of the spatial variation in the sample control point data
2. Summarizing this spatial variation by a regular mathematical function
3. Using this model to determine interpolation weights

Kriging

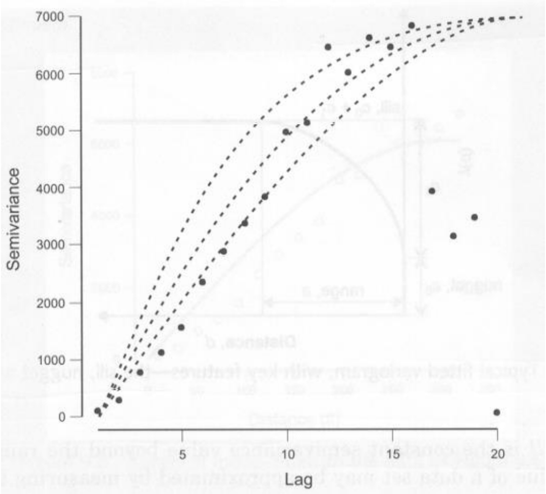
- Describing the spatial variation: the semi-variogram

- There will generally be only one pair of points that are exactly *h* units apart, unless we're dealing with regularly spaced samples. Therefore, we create "bins", or **distance ranges**, into which we place point pairs with similar distances, and estimate γ only for midpoints of these bins rather than at all individual distances.
  - These bins are generally of the same size
  - It's a rule of thumb to have at least 30 point pairs per bin
- We call these estimates of γ(h) at the bin midpoints the *empirical variogram*



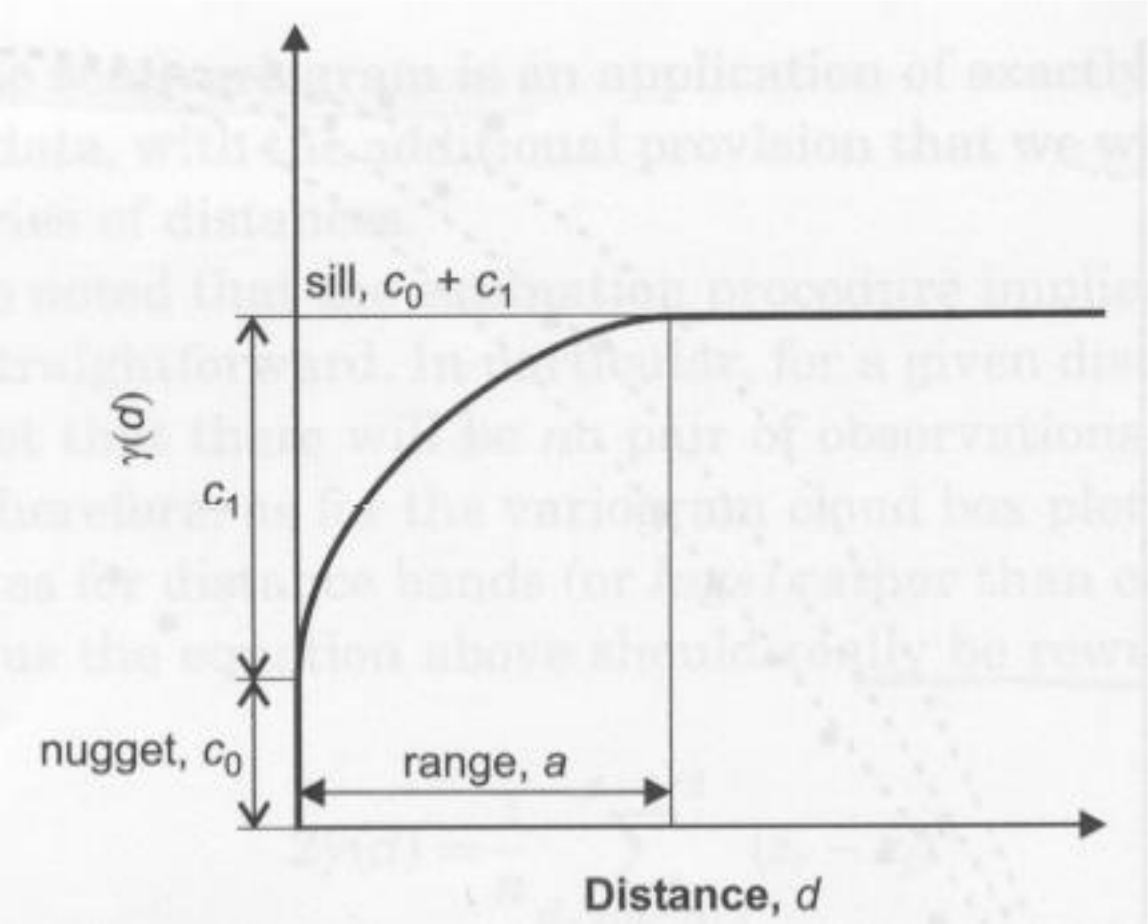
Kriging

- Summarize the spatial variation by a regular mathematical function
  - Now, we're going to **fit a variogram model (i.e., curve)** to the empirical variogram
  - That is, based on the shape of the empirical variogram, **different variogram curves might be fit**
  - The curve fitting generally employs the method of **least squares** – the same method that's used in regression analysis



KRIGING

- Summarize the spatial variation by a regular mathematical function



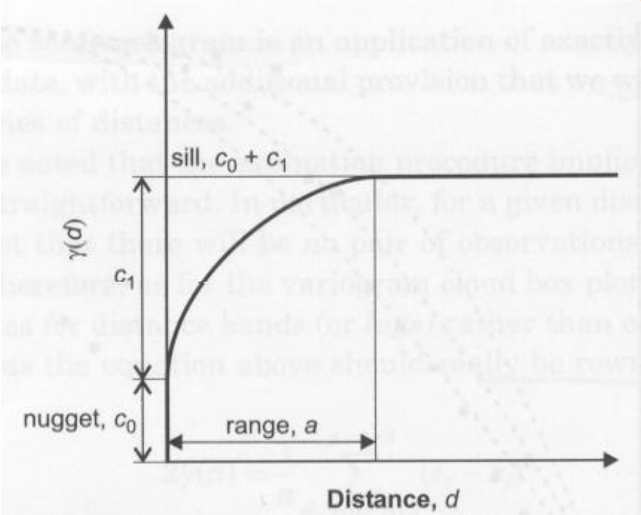
Nugget (c_0): variance at zero distance

Range (a): the distance at which the semivariogram levels off and beyond which the semivariance is constant (correlation is approaching 0)

Sill (c_0+c_1): the constant semivariance value beyond the range

Spatial Independence at Small Distances – Nugget (c_0)

- Even though we assume that values at points that are very near each other are correlated, **points that are separated by very, very small values might be considerably less correlated**.
  - E.g.: you might find a gold nugget and no more gold in the vicinity
- In other words, even though $\gamma(0)$ is always 0, however γ at very, very small distances will be equal to a value c_0 that is considerably greater than 0.
- This value denoted by c_0 is called the ***nugget***.

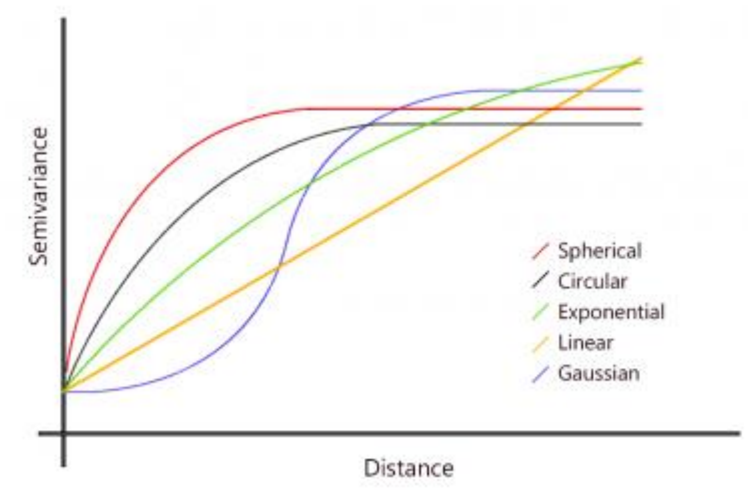


- The ratio of the nugget to the sill (c_0/(c_0+ c_1)) is known as the ***nugget effect,*** and may be interpreted as the **percentage of variation in the data that is not spatial**.

Mathematical Functions

- Summarize the spatial variation by a regular mathematical function

✓Nugget model
✓Linear model
✓Spherical model
✓Exponential model
✓Power model
✓Gaussian model
✓Others

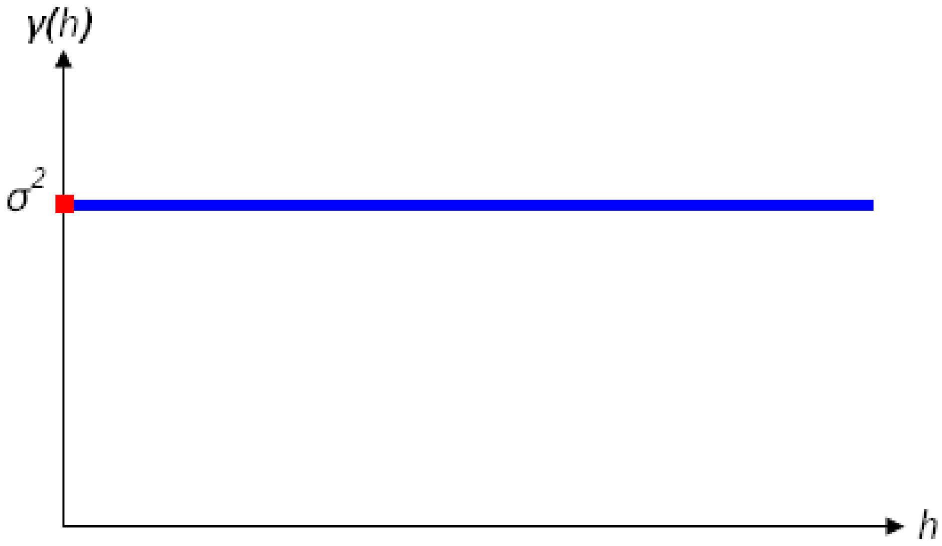


https://gisgeography.com/kriging-interpolation-prediction/

Nugget model

- Summarize the spatial variation by a regular mathematical function

Nugget model: A constant variance model

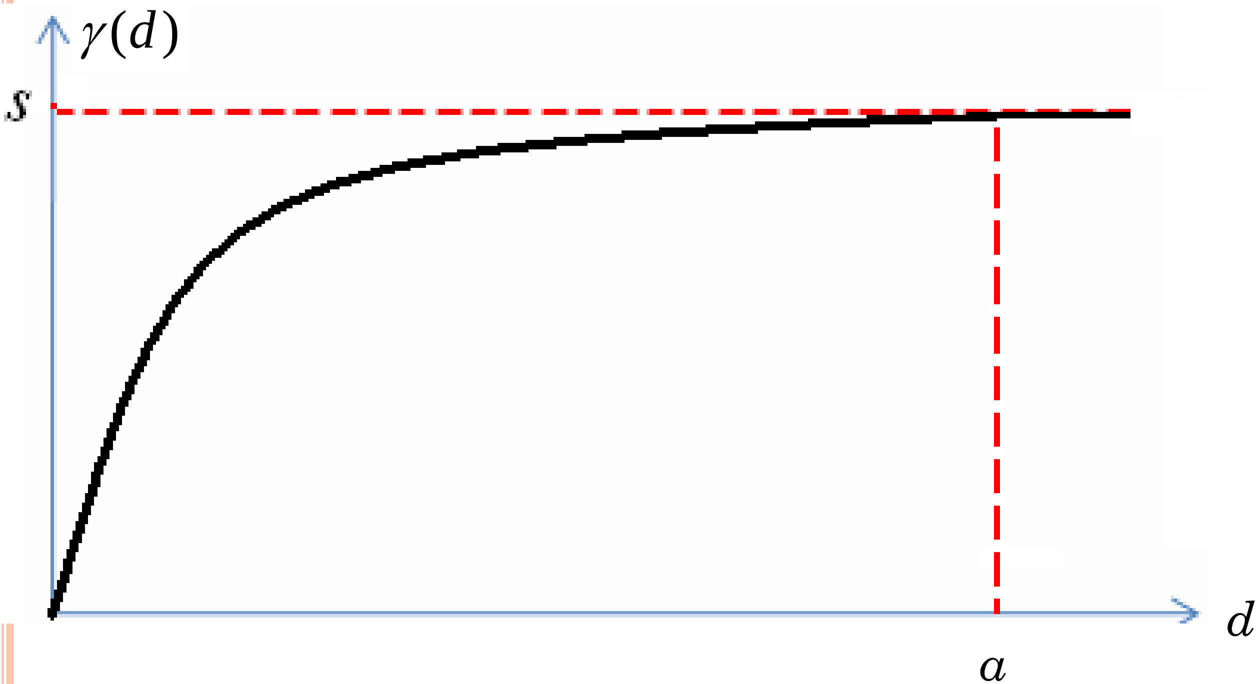


- That is, there is absolutely no spatial autocorrelation in the data (even at small distances)

Spherical model

The spherical model is the most widely used variogram model

Spherical model starts from a nonzero variance (c_0) and rise as an elliptical arc to a maximum value (c_0+c_1) at distance a.



If $d <= a$ then

$$\gamma(d) = c_0 + c_1\left[\frac{3d}{2a} - 0.5\left(\frac{d}{a}\right)^3\right]$$

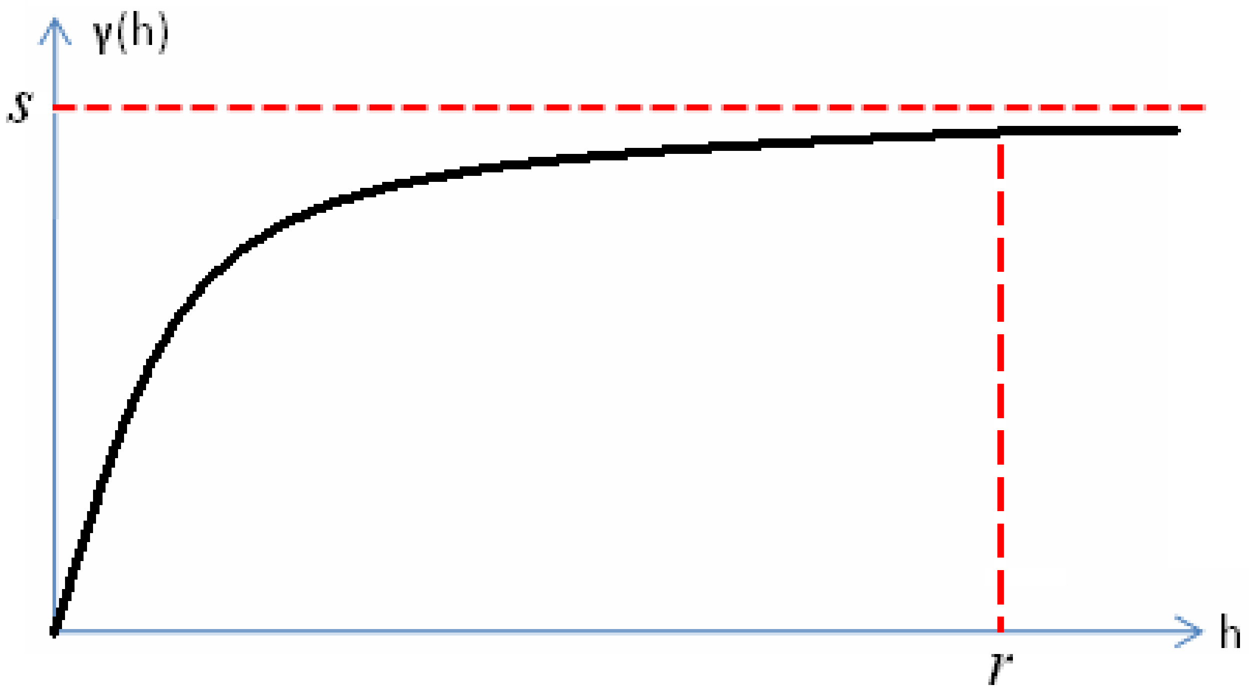
If $d > a$ then

$$\gamma(d) = c_0 + c_1$$

covariance is assumed to be exactly zero at distances $d \geq a$

Exponential Model

- The exponential variogram looks very similar to the spherical model, but assumes that the **correlation never reaches exactly zero**.
- The model is monotonically increasing
  - i.e., as *h* goes up, so does $\gamma(h)$



Using model to determine the interpolation weights

$$\begin{pmatrix} 0 & \gamma(h_{12}) & \gamma(h_{13}) & \dots & \gamma(h_{1n}) & 1 \\ \gamma(h_{21}) & 0 & \gamma(h_{23}) & \dots & \gamma(h_{2n}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & 0 & \dots & \gamma(h_{3n}) & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \gamma(h_{n3}) & \dots & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \bullet \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ \dots \\ W_n \\ \lambda \end{pmatrix} = \begin{pmatrix} \gamma(h_{p1}) \\ \gamma(h_{p2}) \\ \gamma(h_{p3}) \\ \dots \\ \gamma(h_{pn}) \\ 1 \end{pmatrix}$$

h_{ik}	is the distance between input point *i* and input point *k*
h_{pi}	is the distance between the output pixel *p* and input point *i*
$\gamma(h_{ik})$	is the value of the semi-variogram model for the distance h_{ik}, i.e. the semi-variogram value for the distance between input points *i* and input point *k*
$\gamma(h_{pi})$	is the value of the semi-variogram model for the distance h_{pi} , i.e. the semi-variogram value for the distance between the output pixel *p* and input point *i*
w_i	is a weight factor for input point *i*
λ	is a Lagrange multiplier, used to minimize possible estimation error

Kriging

- Types of Kriging
  - **Ordinary Kriging** – assumes the trend is an unknown constant
  - **Simple Kriging** – assumes the trend is a known constant
  - **Universal Kriging** – assumes the trend is a known deterministic function
  - **Cokriging** – uses the correlation of additional variable in calculating spatial autocorrelation. Assume the trend for variables are unknown constant

Trend/mean

$$z_{(s)} = \mu + \varepsilon_{(s)}$$

variable of interest at location (s)

Spatial dependent of an stationary process

http://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/kriging-in-geostatistical-analyst.htm

Isotropy vs. Anisotropy

- When we use *isotropic* covariograms, we assume that the **covariance** between the point values **depends *only* on distance**
- *Anisotropic* (or directional) covariograms are used when we have reason to believe that *direction* plays a role as well (i.e., covariance is a function of **both distance *and* direction**)
  - E.g., in some problems, accounting for direction is appropriate (e.g., when wind or water currents might be a factor)

Kriging

Conclusion

1) Kriging is computationally intensive
2) All the results depend on the model we fit to the estimated semi-variogram from the sample data
3) If the correct model is used, the methods used in kriging have an advantage over other interpolation procedures

Consideration

- Check for enough number of pairs at each lag distance (from 30 to 50).
- Removal of outliers
- Truncate at half the maximum lag distance to ensure enough pairs
- Use a larger lag tolerance to get more pairs and a smoother variogram
- Start with an omnidirectional variogram before trying directional variograms
- Use transforms of the data for skewed distributions (e.g. logarithmic transforms).

Next Class

- Statistical Inference for Geographical Processes
- (Fotheringham & Brunsdon, 2004; McLaughlin & Boscoe, 2007)