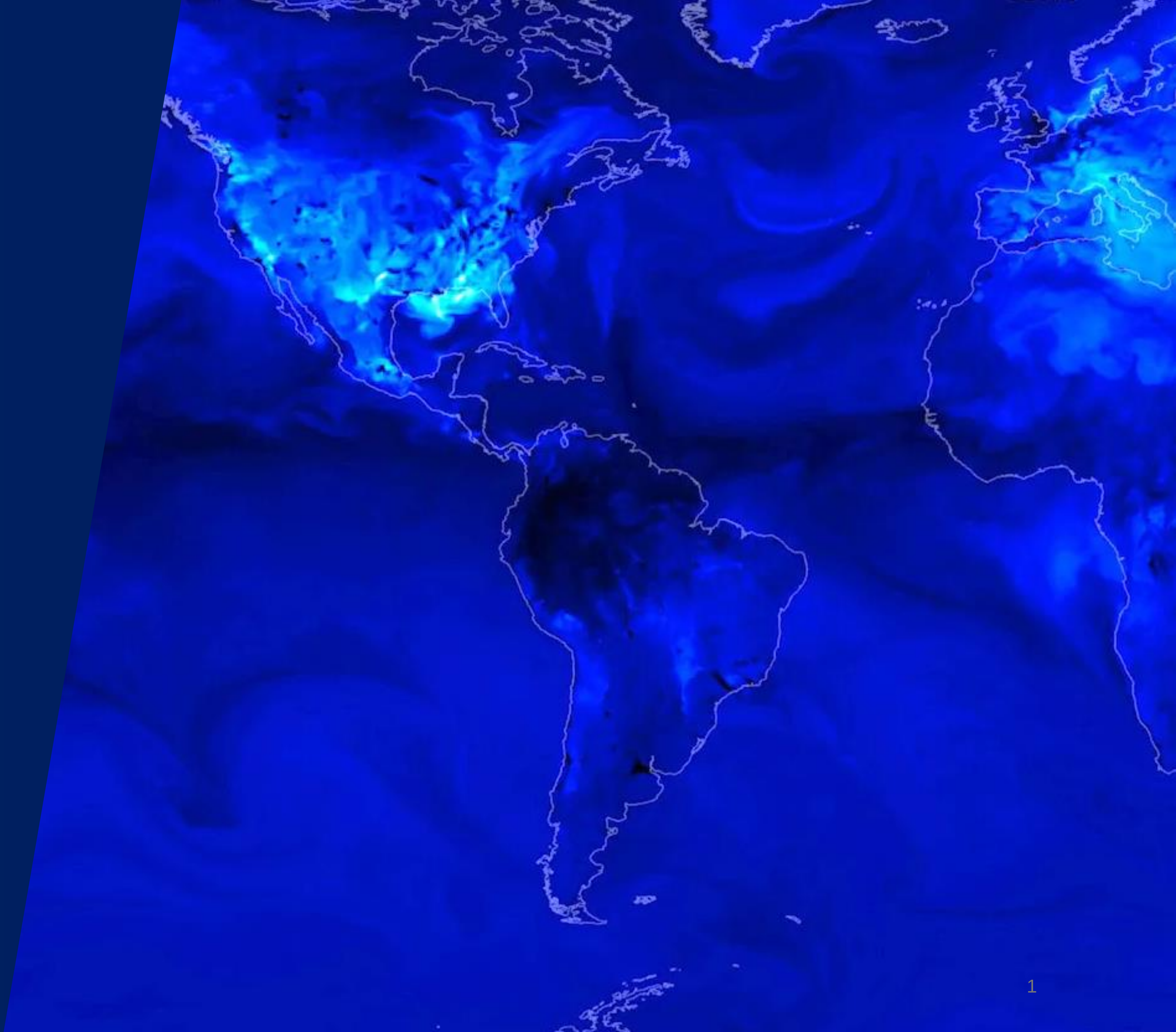
SAINT LOUIS
UNIVERSITY.

Surface and Field Analysis

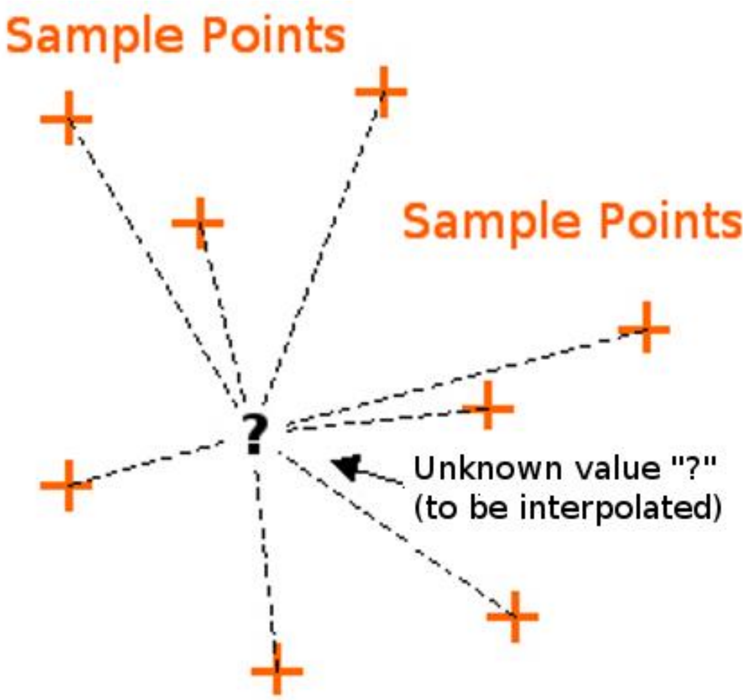
Part 3: Spatial Interpolation

Sourav Bhadra, Ph. D.

GIS4120/5120 Geospatial Analytics
10/8/2024

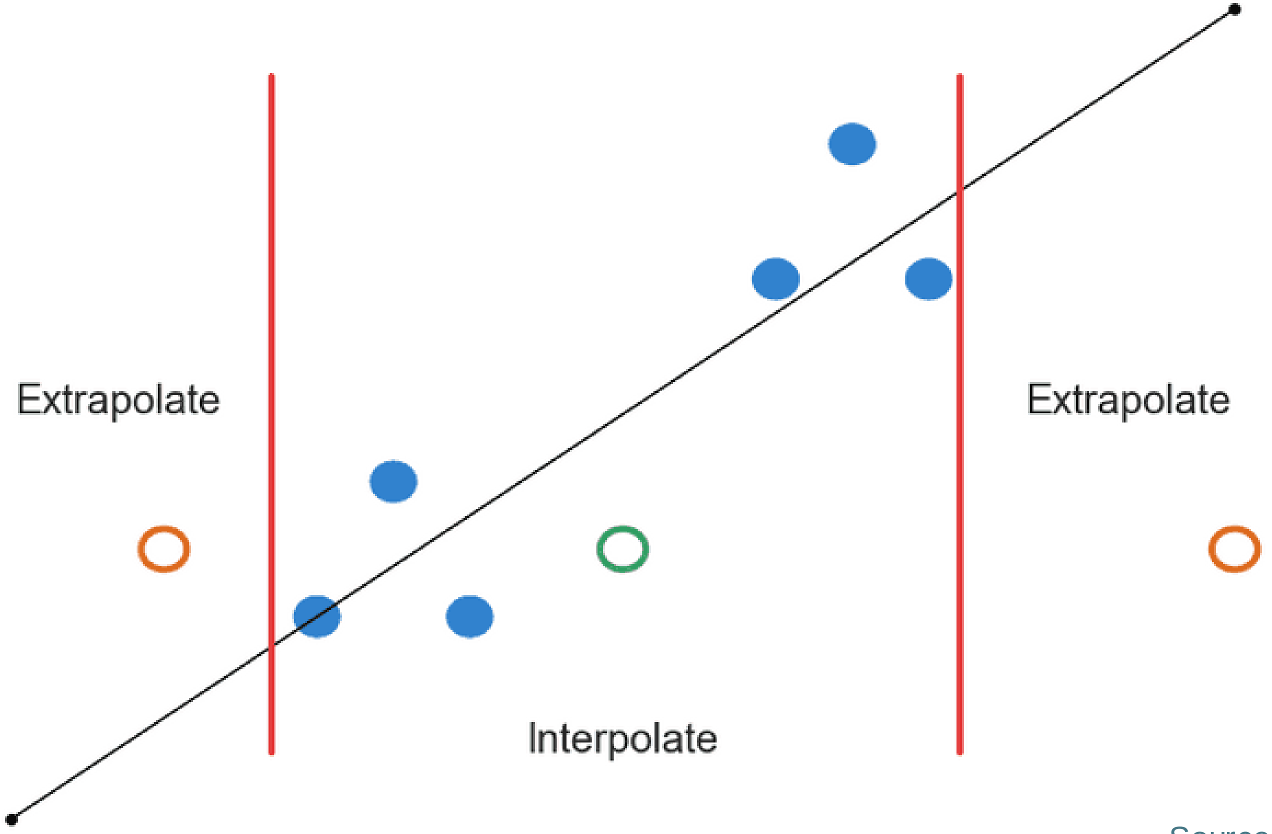


Interpolation: Knowing values of unknown points



- Many natural phenomena occur continuously across a landscape, for example, the distribution of nutrients in soil, dissolved oxygen in seawater, or rainfall.
- It is unrealistic to take measurements of natural phenomena everywhere.
- Interpolation comes into play in these situations.

Interpolation vs Extrapolation



Source

Remembering Tobler's first law of geography

Dr. Waldo R. Tobler
(1930 – 2018)

"Everything is related to everything else, but near things are more related than distant things"

Types of spatial interpolation

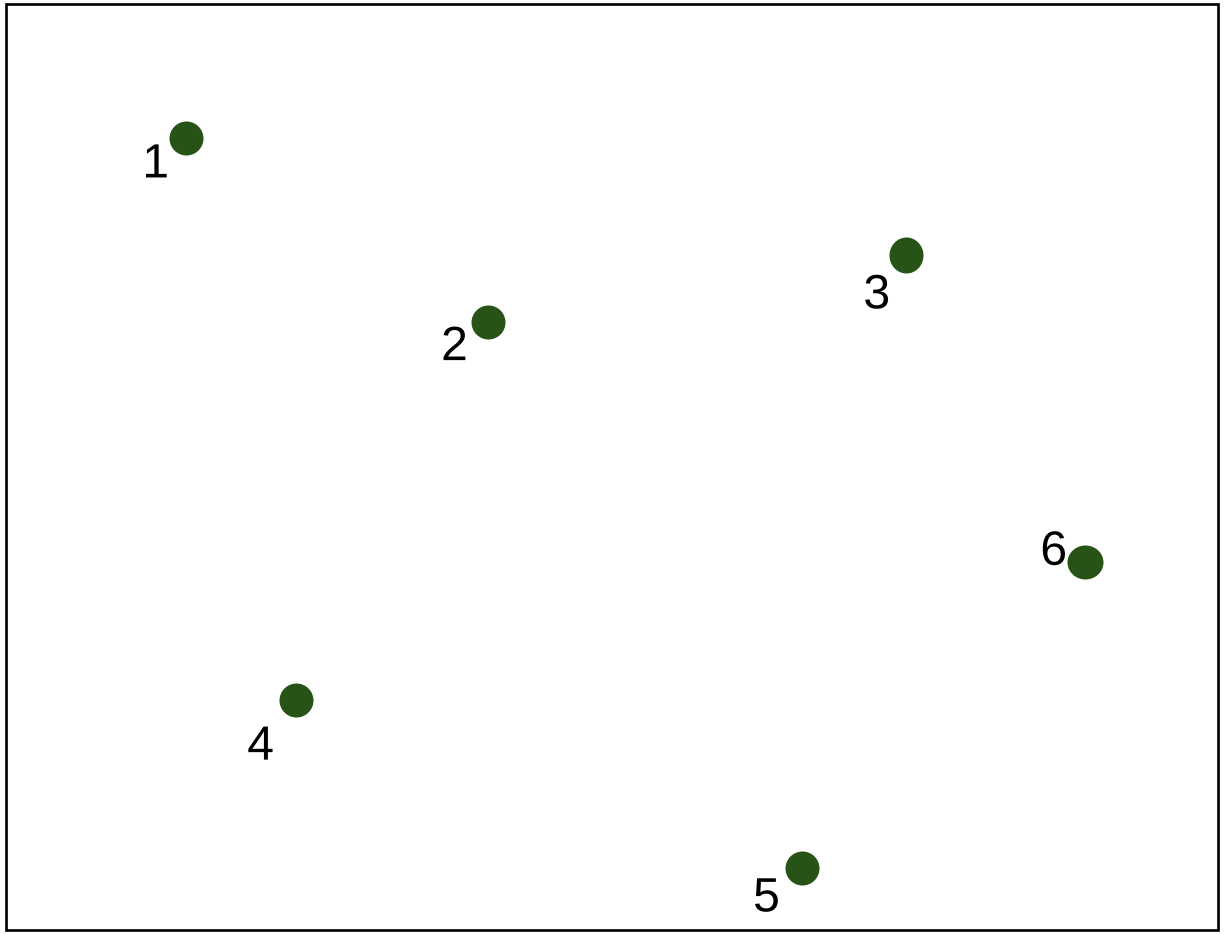
Deterministic Interpolation

- Use direct mathematical formulas to estimate values at unknown locations.
- Create smooth surface.
- Straightforward to use.
- Do not provide a measure of how uncertain the prediction might be.
- Examples: Theissen Polygon, Inverse distance weighting (IDW), Trend Analysis.

Stochastic Interpolation

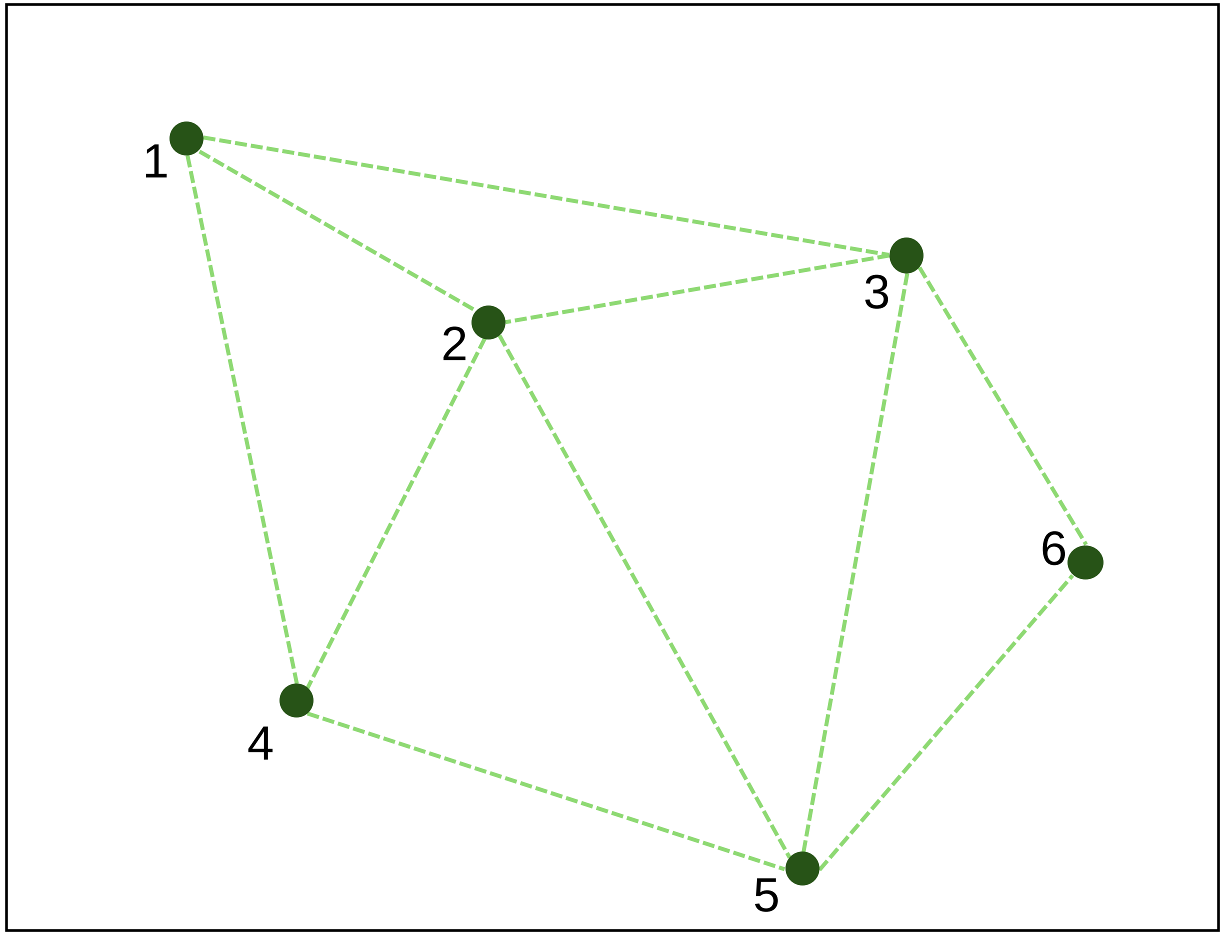
- Use statistical models to generate predictions as well as confidence.
- Incorporates randomness and provide and estimate of the uncertainty.
- Complex to implement.
- Kriging is a common stochastic method that uses the spatial correlation between data points to make predictions and provides variance estimates to understand the reliability of the prediction.

Theissen Polygon



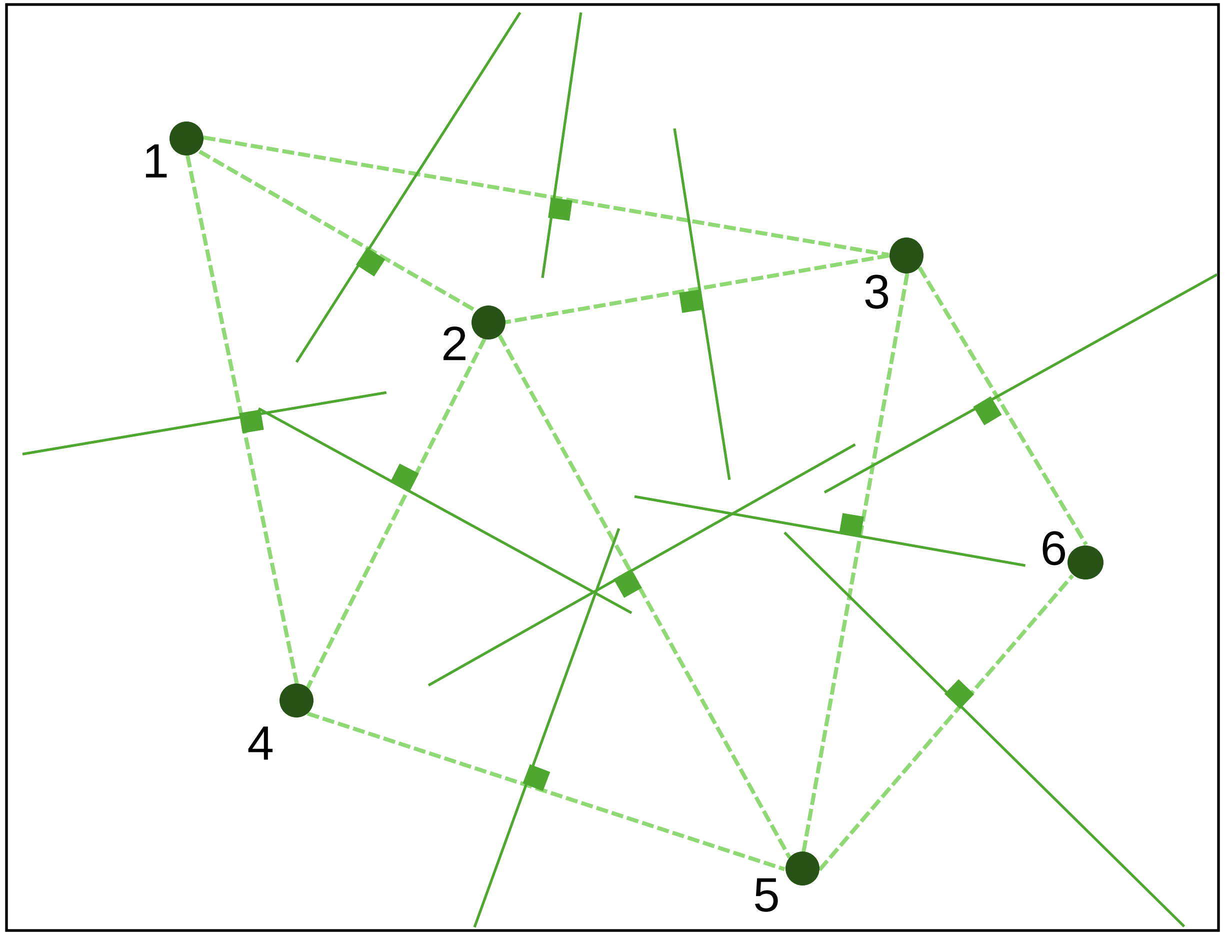
Station #	Annual Precip (mm)
1	985
2	1015
3	963
4	1101
5	1057
6	1078

Theissen Polygon



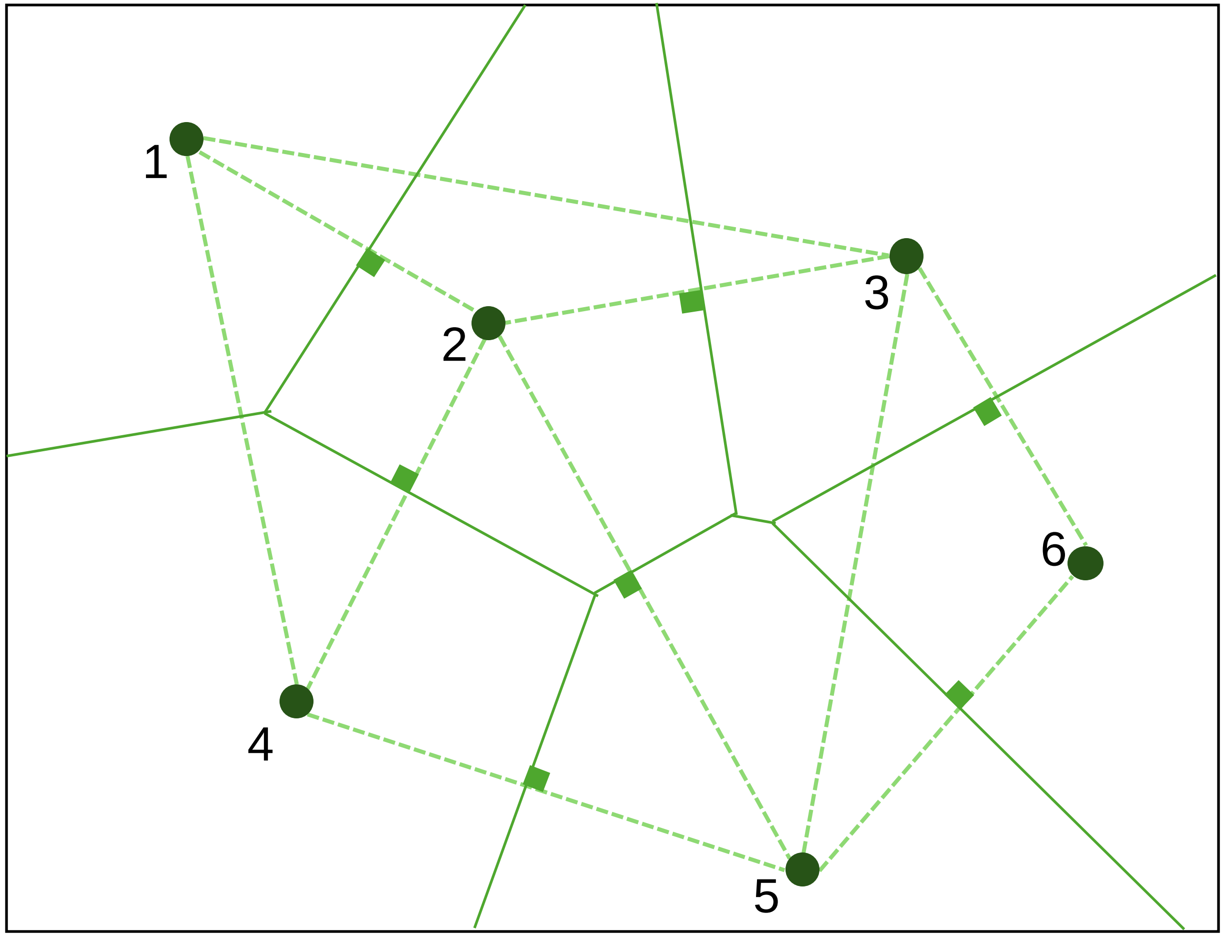
- Draw straight line from point to nearest neighbors.
- Should create a map of triangles.

Theissen Polygon



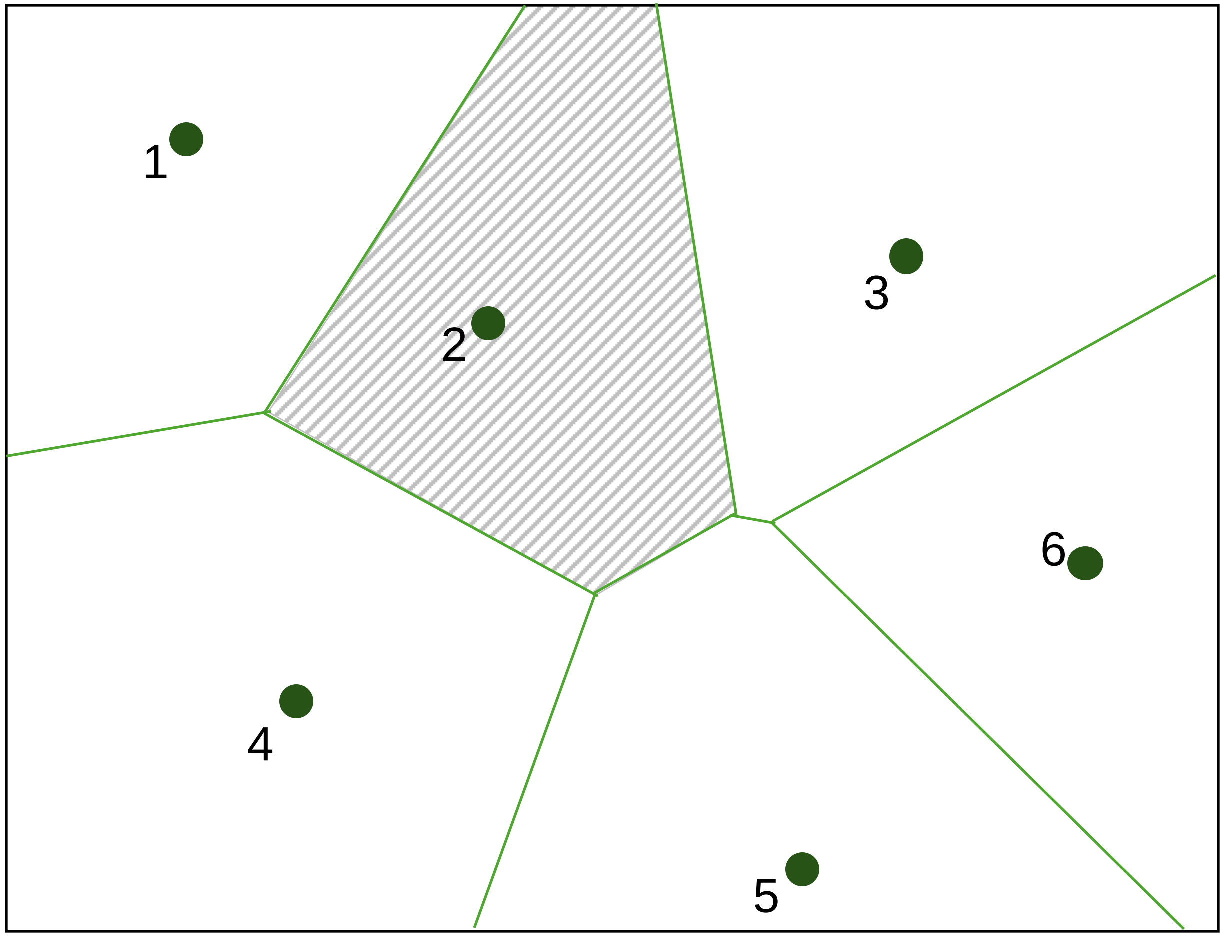
- Draw straight line from point to nearest neighbors.
- Should create a map of triangles.
- Draw perpendicular lines in the middle of each line between points (bisectors)

Theissen Polygon



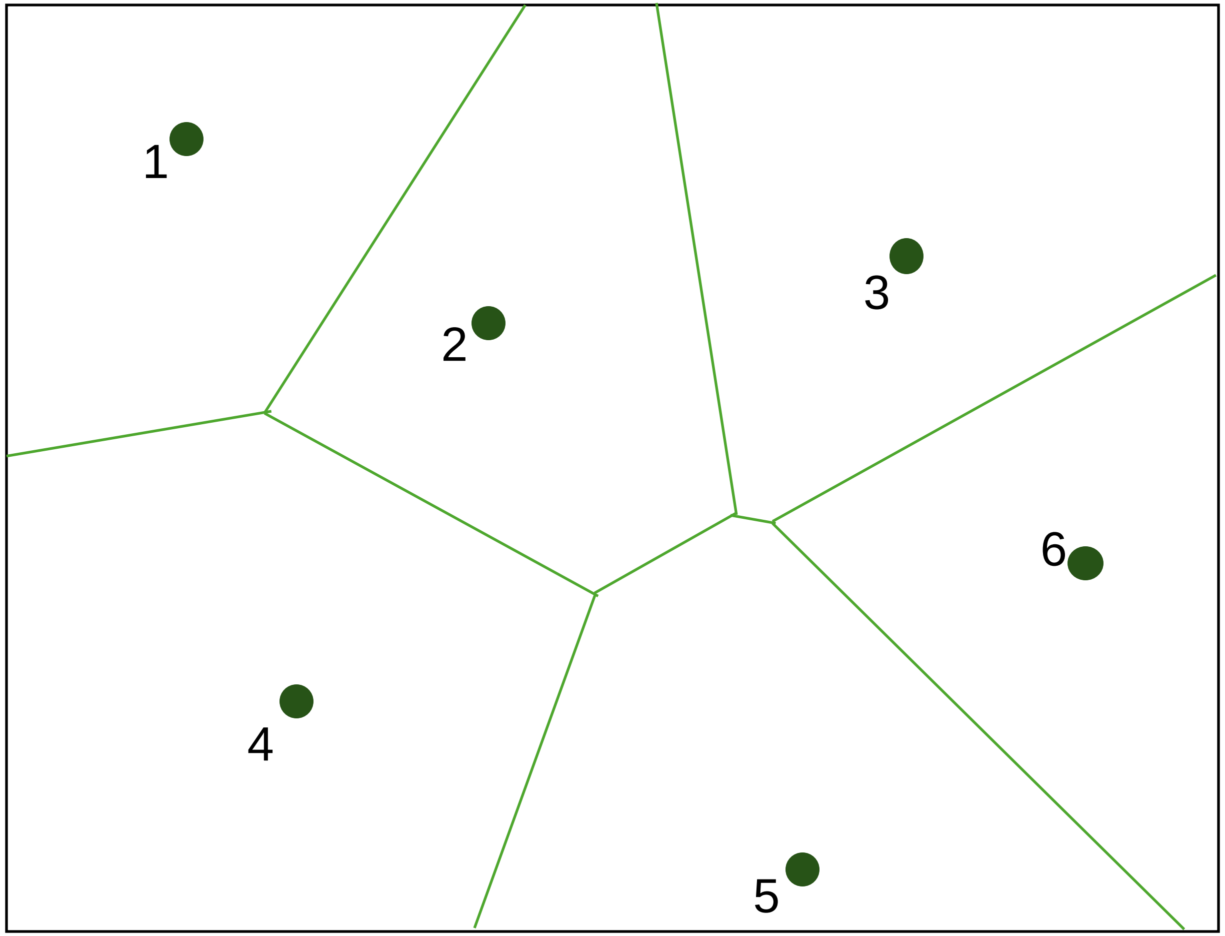
- Draw straight line from point to nearest neighbors.
- Should create a map of triangles.
- Draw perpendicular lines in the middle of each line between points (bisectors)
- Lengthen or shorten bisectors to create polygons. All lines may not be used. But # polygons must = # points

Theissen Polygon



- Draw straight line from point to nearest neighbors.
- Should create a map of triangles.
- Draw perpendicular lines in the middle of each line between points (bisectors)
- Lengthen or shorten bisectors to create polygons. All lines may not be used. But # polygons must = # points
- Shows area of influence

Theissen Polygon

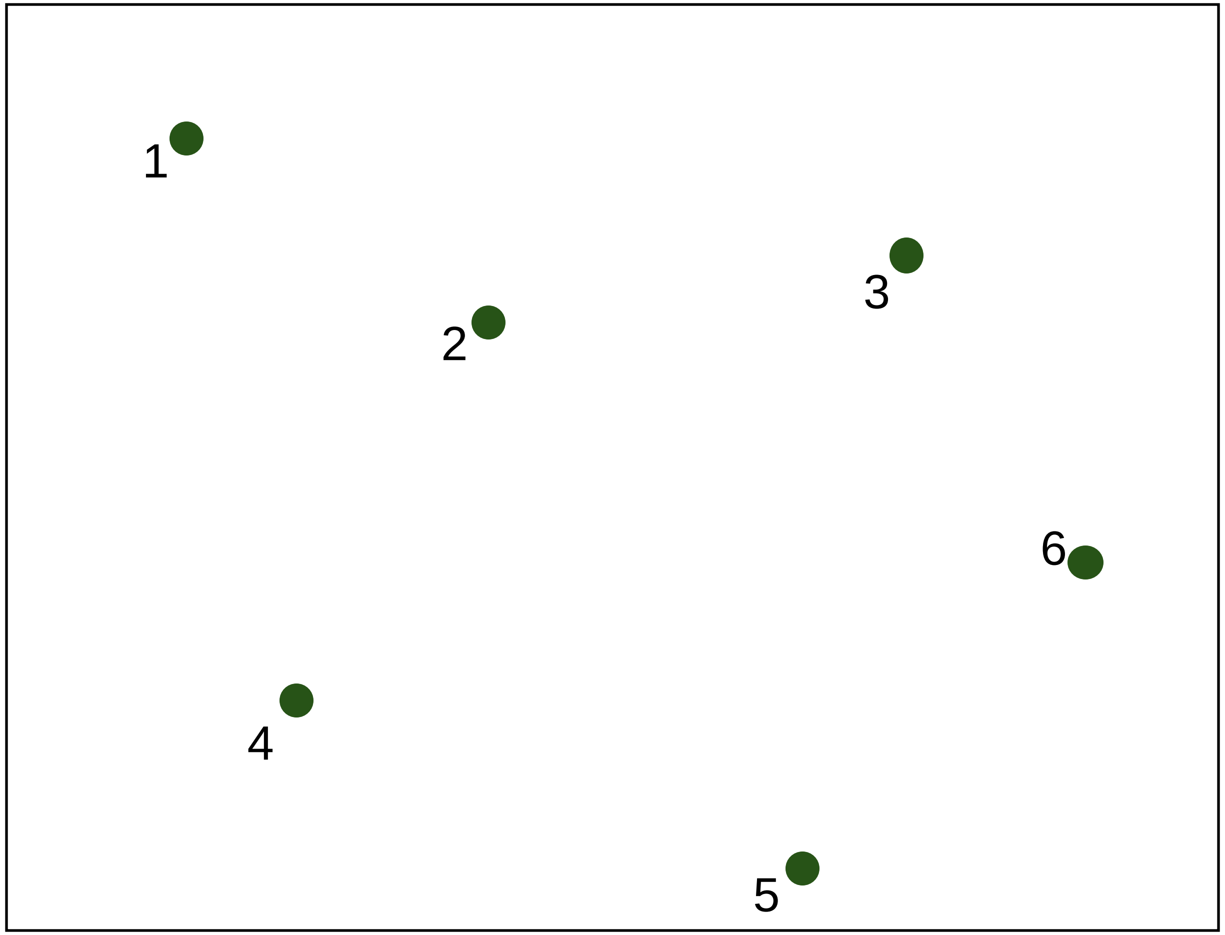


- Deterministic interpolation as a fixed mathematical approach is taken into account without knowing its confidence.
- Problems with Theissen polygon:
  - Might work when a dense network of samples are available.
  - Not creating a uniform grid-like surface.
  - Hard to perform any spatial statistics on top of theissen polygons.

Inverse Distance Weighting (IDW)

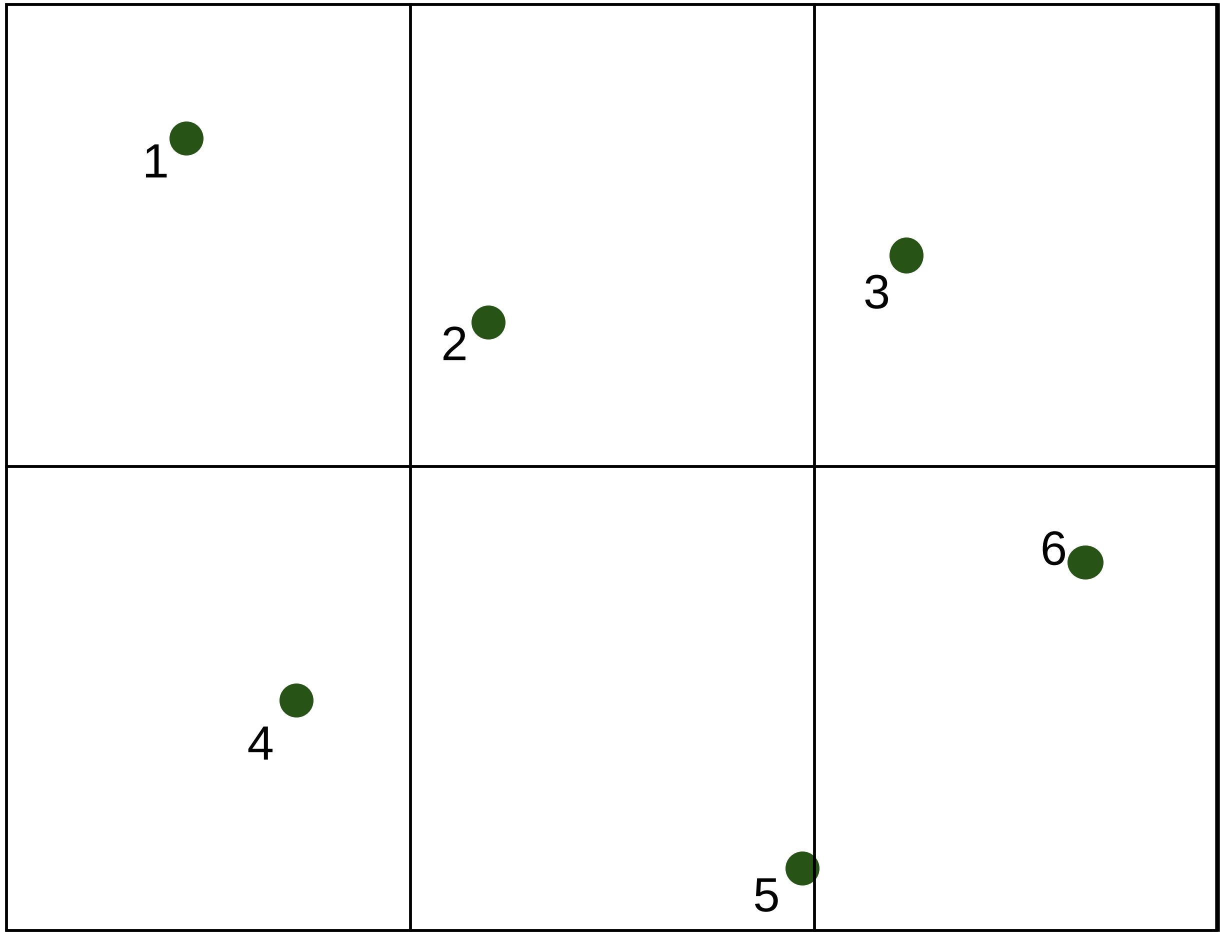
- IDW assumes that points closer to the unknown location have a greater influence on the predicted value than points farther away.
- Each known point is assigned a weight inversely proportional to its distance from the unknown point.
- The formula for the weight is $w_i = \frac{1}{d_i^p}$, where p is the power parameter that controls the influence of distance
- The predicted value is a weighted average of the known values
- $Z^*(x_o) = \frac{\sum w_i Z(x_i)}{\sum w_i}$

Inverse Distance Weighting (IDW)



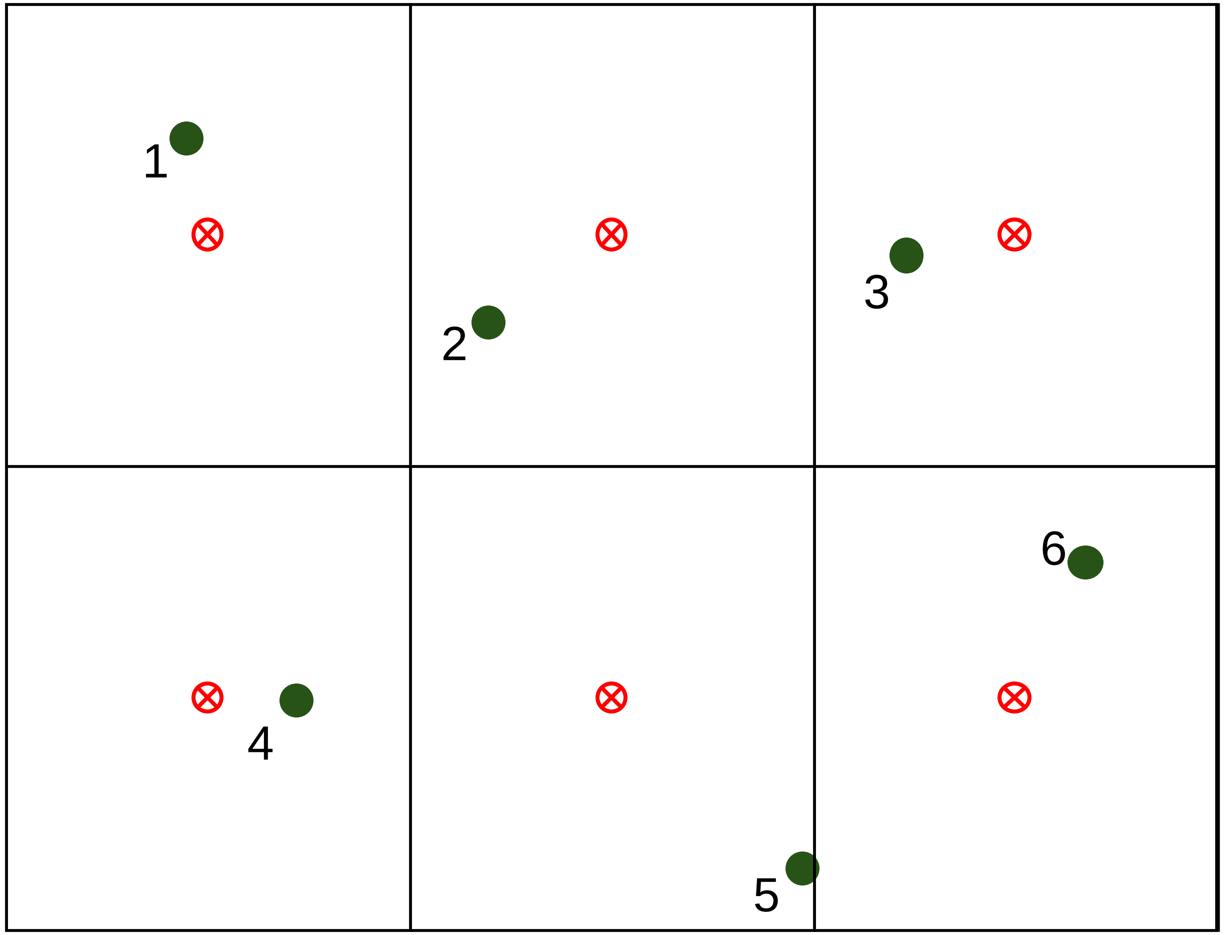
- Same problem as before, 6 weather stations with precipitation values

Inverse Distance Weighting (IDW)



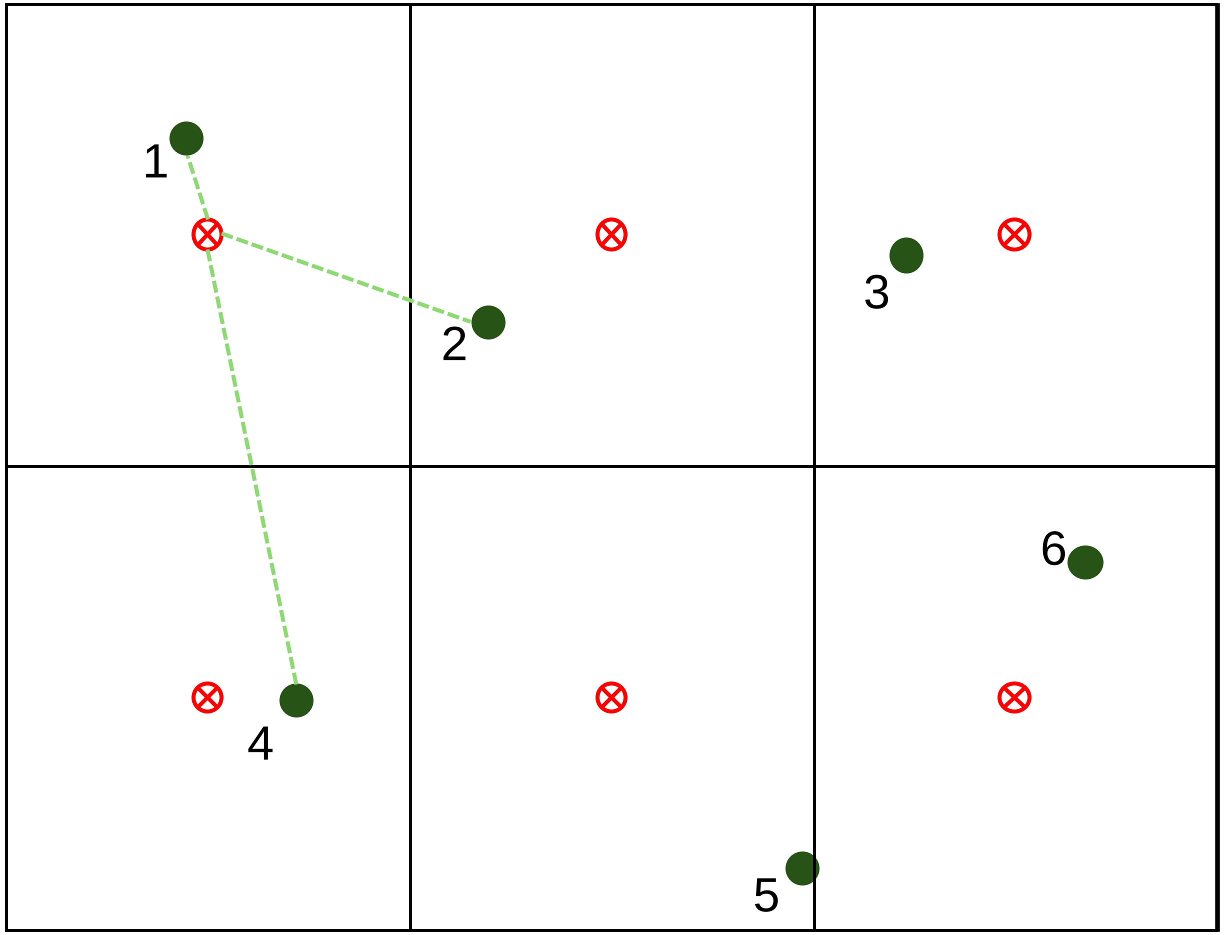
- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels

Inverse Distance Weighting (IDW)



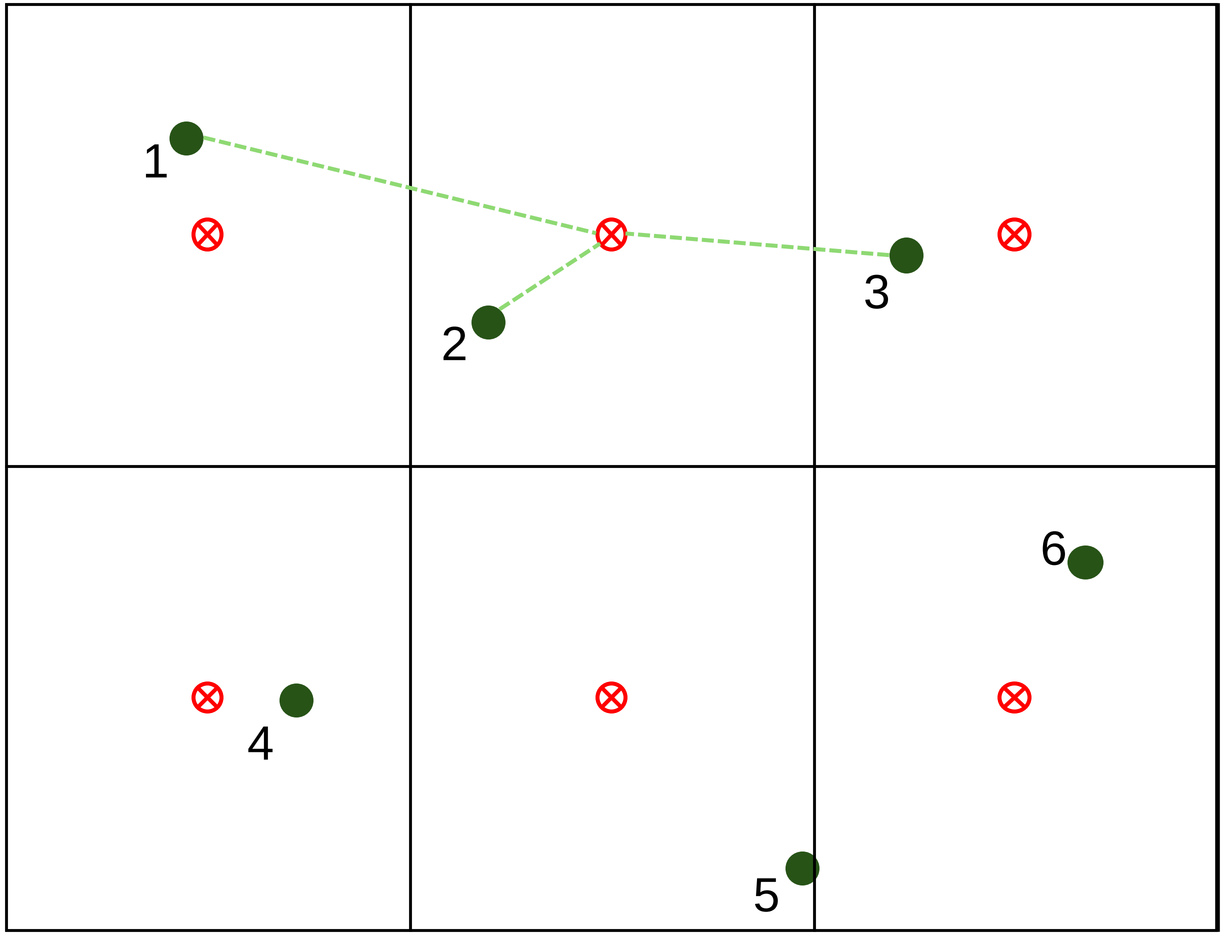
- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels
- Create centroid for each pixel

Inverse Distance Weighting (IDW)



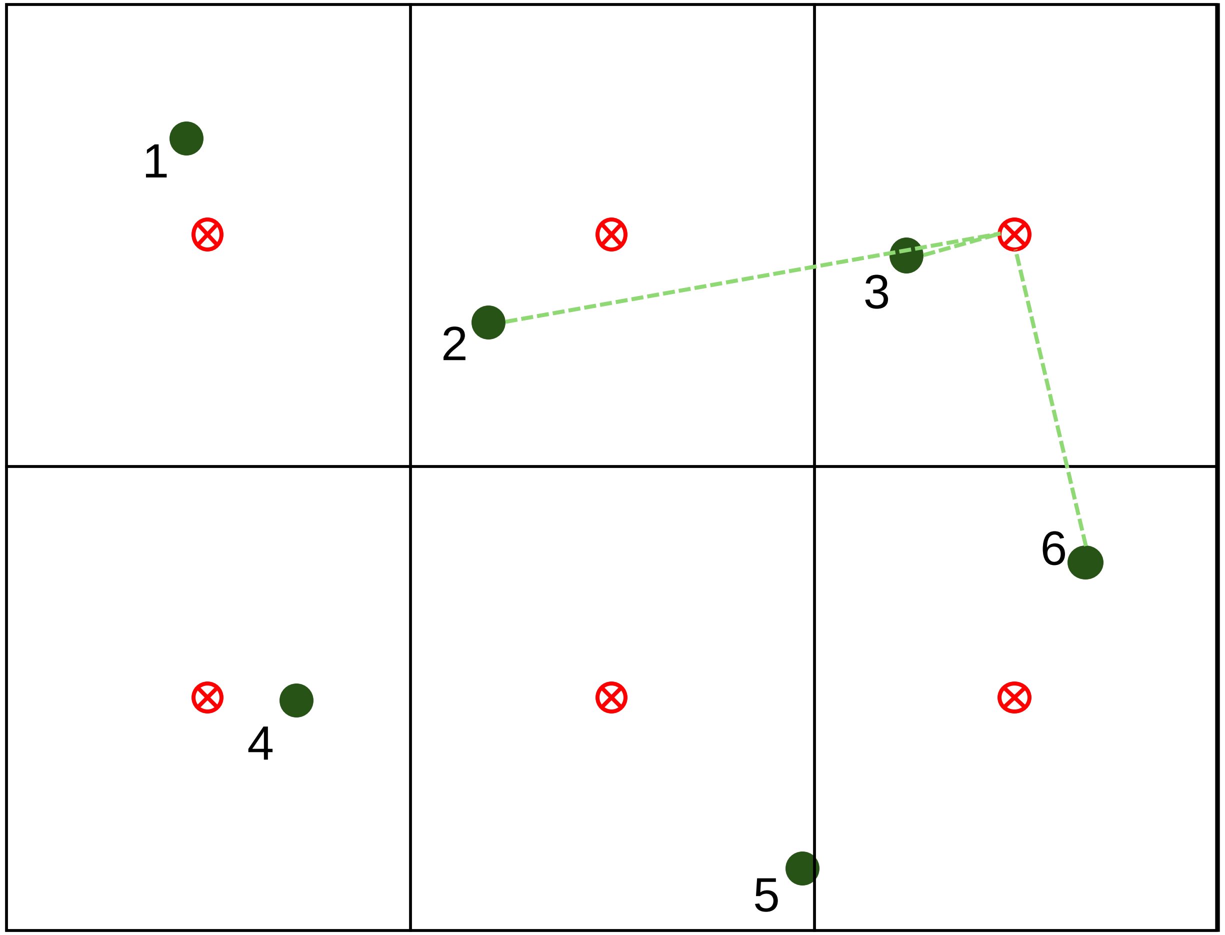
- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels
- Create centroid for each pixel
- Find closest points for each pixel, in this case, let’s consider 3 nearest neighbors

Inverse Distance Weighting (IDW)



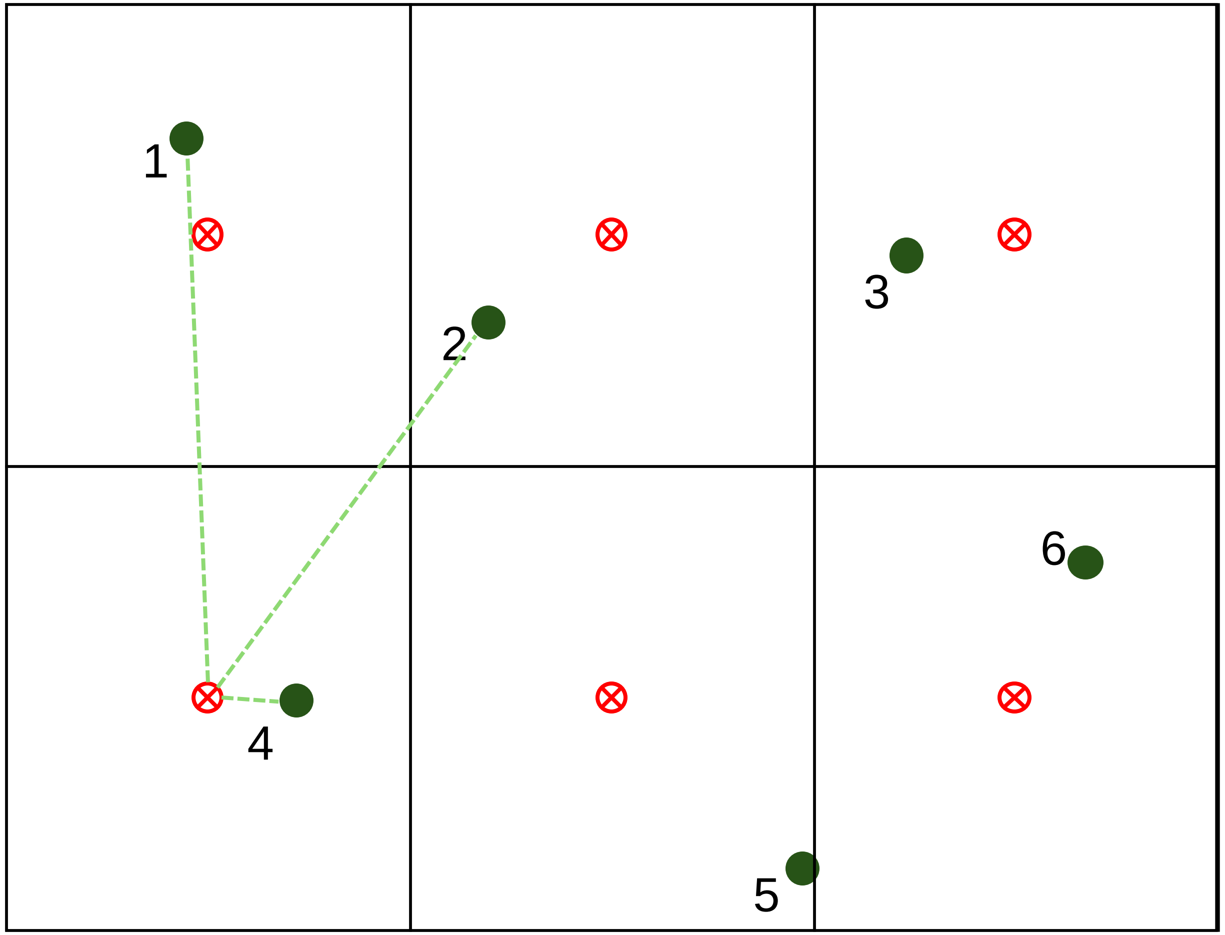
- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels
- Create centroid for each pixel
- Find closest points for each pixel, in this case, let’s consider 3 nearest neighbors

Inverse Distance Weighting (IDW)



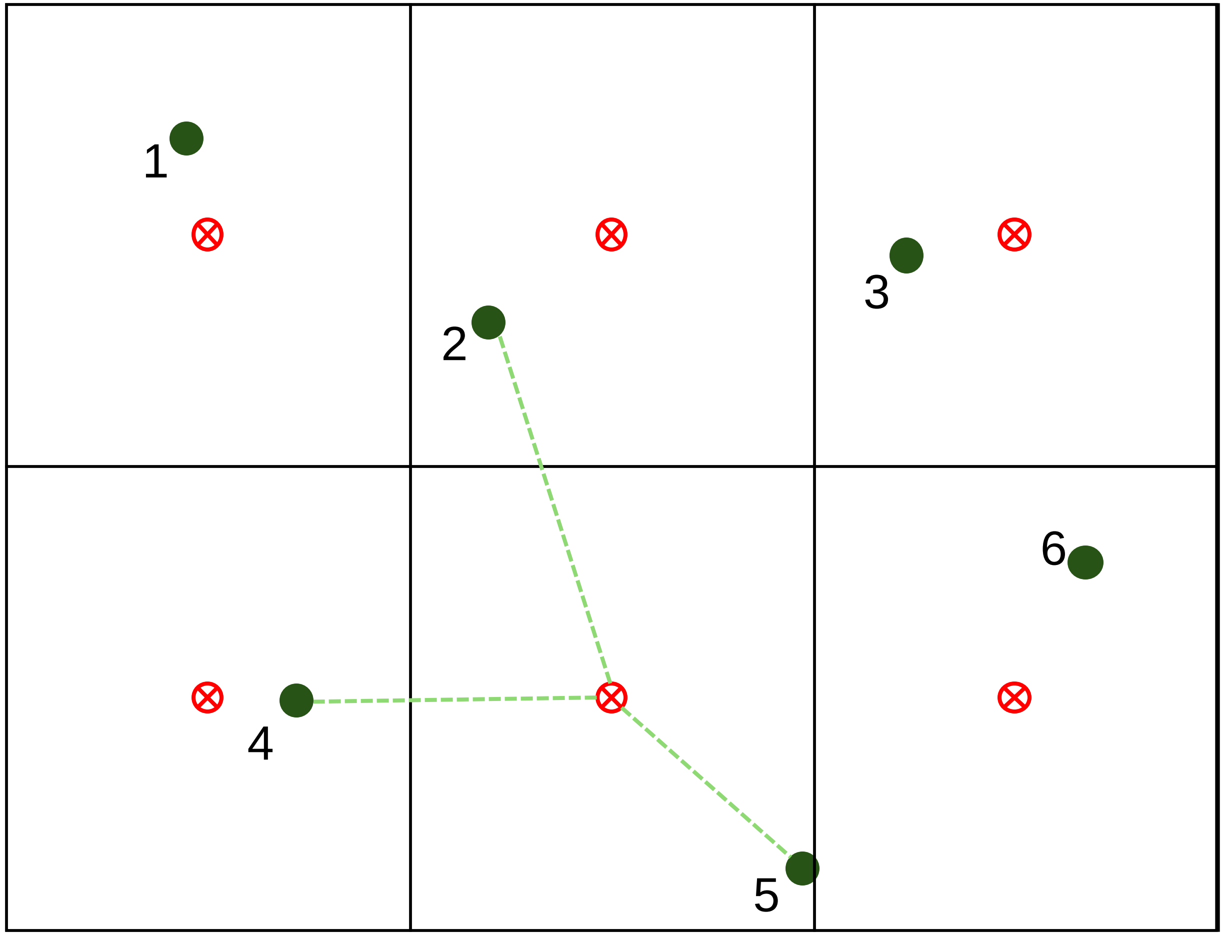
- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels
- Create centroid for each pixel
- Find closest points for each pixel, in this case, let's consider 3 nearest neighbors

Inverse Distance Weighting (IDW)



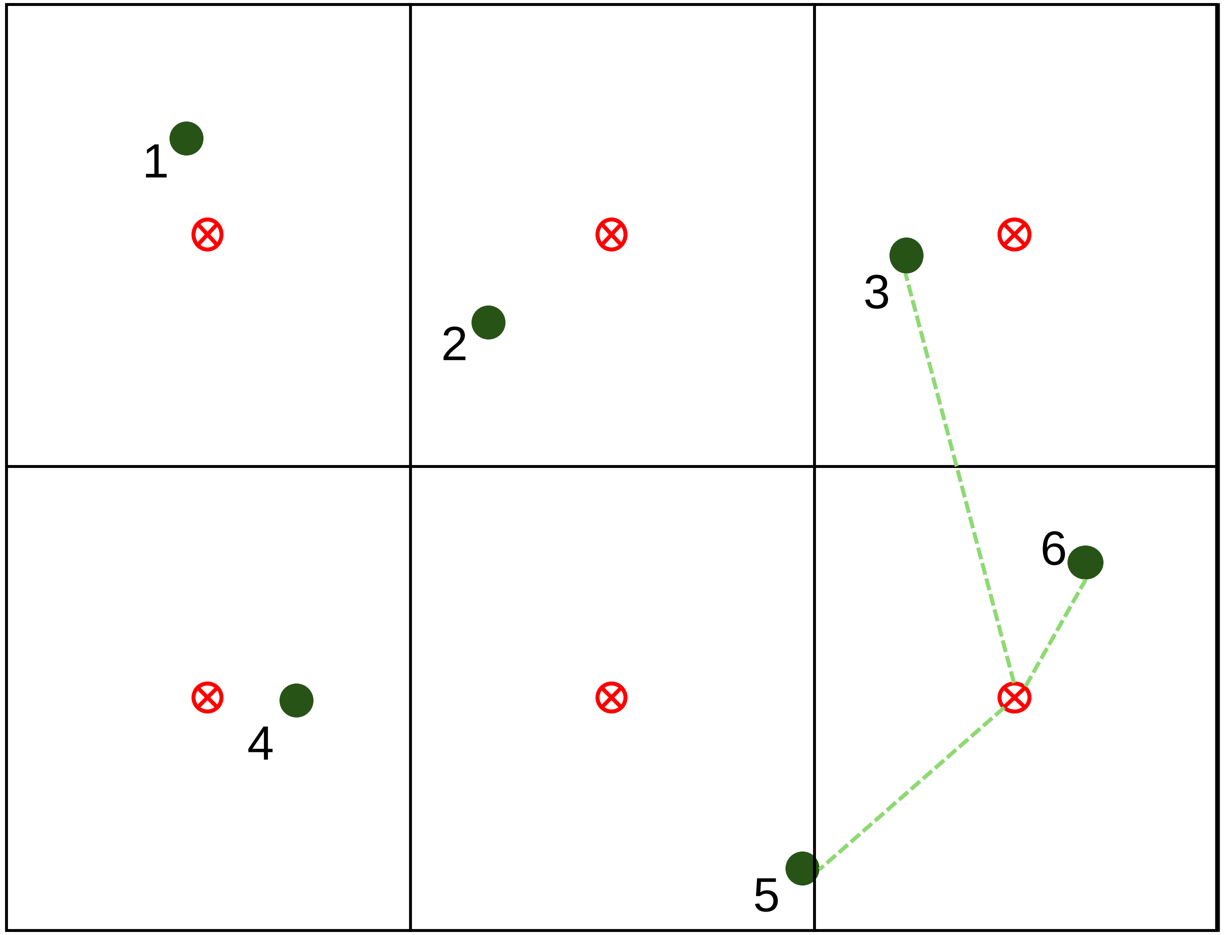
- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels
- Create centroid for each pixel
- Find closest points for each pixel, in this case, let’s consider 3 nearest neighbors

Inverse Distance Weighting (IDW)



- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels
- Create centroid for each pixel
- Find closest points for each pixel, in this case, let's consider 3 nearest neighbors

Inverse Distance Weighting (IDW)



- Same problem as before, 6 weather stations with precipitation values
- Decide number of pixels for the output raster, we can choose six pixels
- Create centroid for each pixel
- Find closest points for each pixel, in this case, let’s consider 3 nearest neighbors

Inverse Distance Weighting (IDW)

Station #	Nearby Stations	Distance (cm)	Annual Precip (mm)	Grid Annual Precip (mm)
1	1	1.2	985	
	2	3.8	1015	
	4	6.4	1101	
2				
3				
4				
5				
6				

$$\frac{\left(\frac{1}{1.2} * 985\right) + \left(\frac{1}{3.8} * 1015\right) + \left(\frac{1}{6.4} * 1101\right)}{\frac{1}{1.2} + \frac{1}{3.8} + \frac{1}{6.4}}$$

$$z_j = \frac{\sum w_i * z_i}{\sum w_i}$$

$$w_i = \frac{1}{d_i}$$

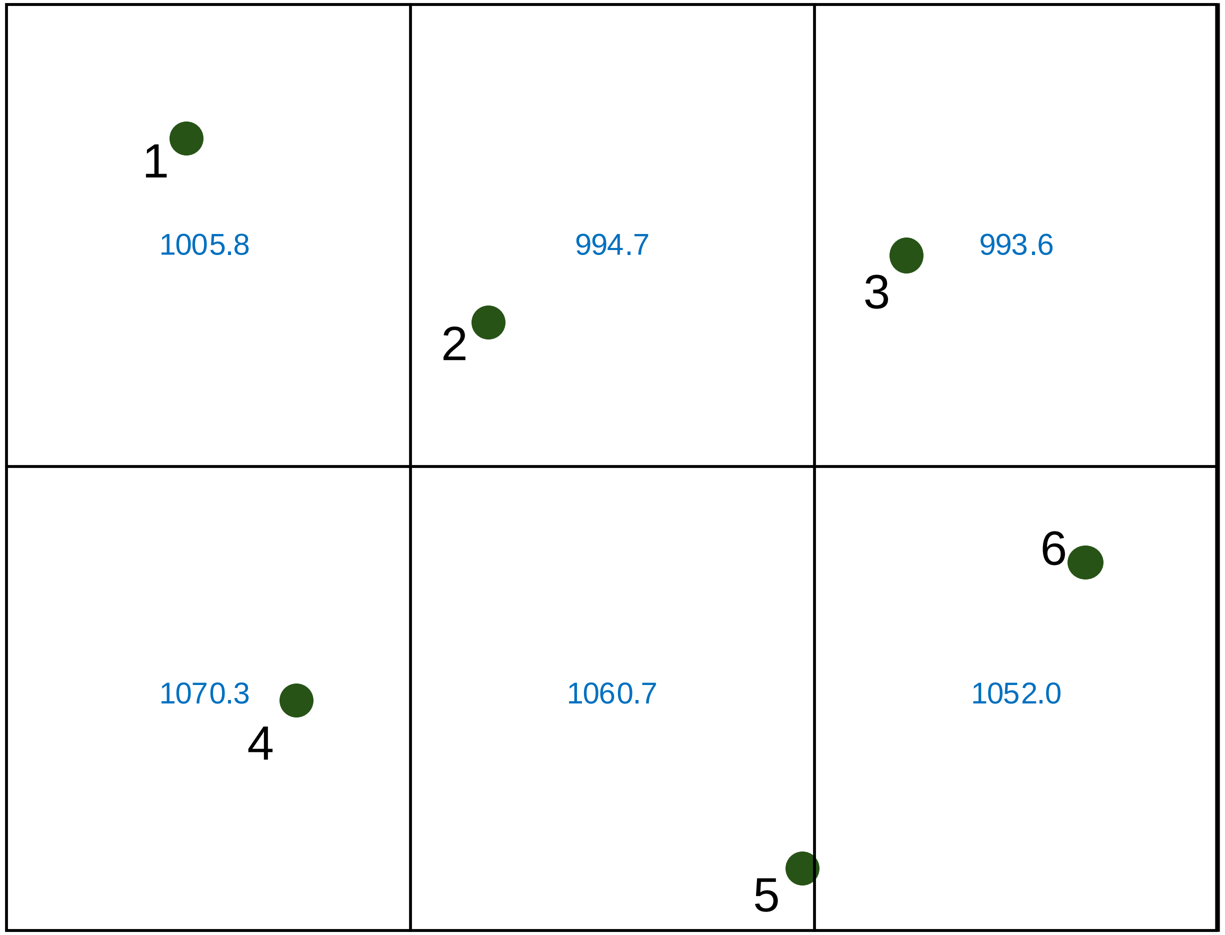
Inverse Distance Weighting (IDW)

Station #	Nearby Stations	Distance (cm)	Annual Precip (mm)	Grid Annual Precip (mm)
1	1	1.2	985	1005.8
	2	3.8	1015	
	4	6.4	1101	
2	1	5.5	985	994.7
	2	2.0	1015	
	3	3.8	963	
3	2	6.8	1015	993.6
	3	1.4	963	
	6	4.5	1078	
4	1	6.3	985	1070.3
	2	5.5	1015	
	4	1.3	1101	
5	2	5.1	1015	1060.7
	4	4.0	1101	
	5	3.4	1057	
6	3	5.9	963	1052.0
	5	3.5	1057	
	6	1.9	1078	

$$z_j = \frac{\sum w_i * z_i}{\sum w_i}$$

$$w_i = \frac{1}{d_i}$$

Inverse Distance Weighting (IDW)



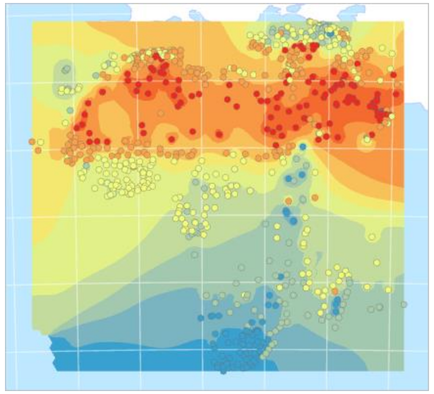
- Easy to understand
- Minimal parameters
- No need for statistical assumptions
- Flexible power parameters
- Efficient for dense data
- Quick computation
- Deterministic, always produce same results

Inverse Distance Weighting (IDW): Challenges

Assumption of Local Influence

IDW assumes that points closer to the unknown location have a greater influence on the estimate, which may not always be true, especially if spatial relationships are not purely distance-based. This can lead to inaccuracies in areas where there are local variations that are not related to distance.

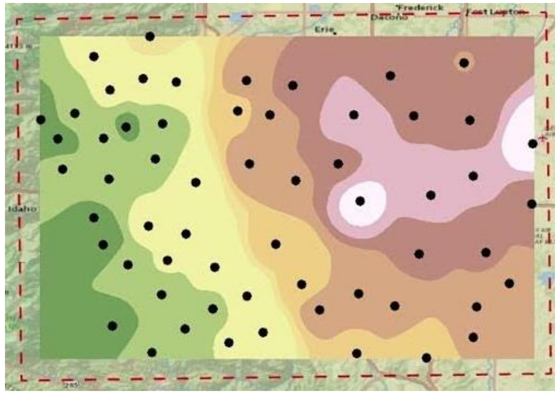
Sensitivity to Point Distribution



Source

IDW results depend heavily on the distribution of input points. If there is an uneven distribution or clustering of points, this can lead to biases in the results. Areas with sparse data may have large interpolation errors.

Edge Effects

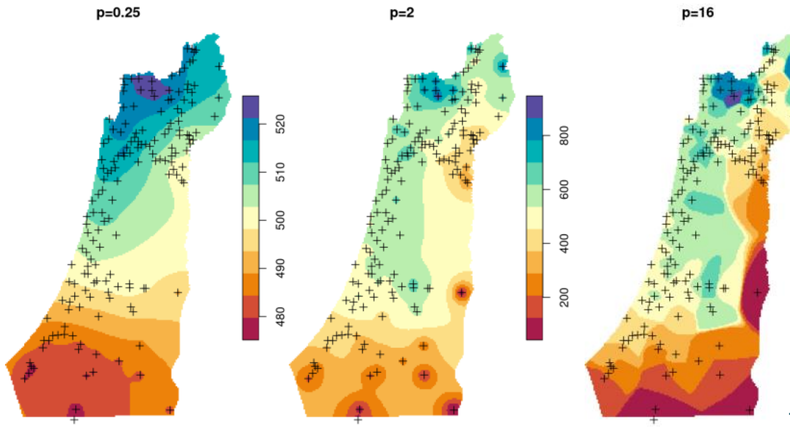


Source

At the boundaries of the study area, IDW can produce unreliable estimates due to a lack of neighboring data points outside the boundaries, which means those locations have less information to draw from, leading to less accurate interpolations.

Inverse Distance Weighting (IDW): Challenges

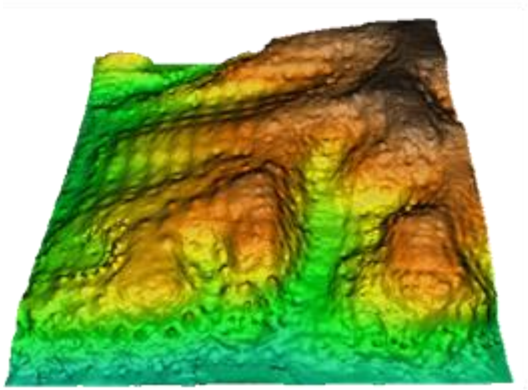
Choice of Parameters



Source

IDW involves parameters such as the power parameter and search radius, which can significantly affect the interpolation result. The choice of these parameters can be somewhat subjective and requires trial and error. An inappropriate selection can lead to over- or under-smoothing of the interpolated surface.

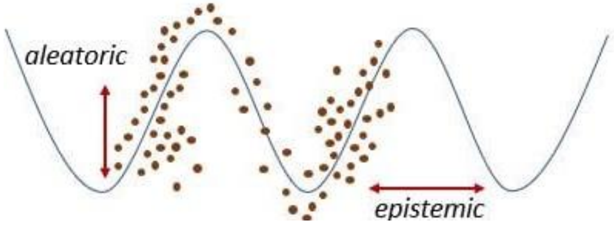
Limited to Local Patterns



Source

IDW only accounts for the local influence of neighboring points and ignores broader trends that may be present in the dataset. This limitation means that IDW may not accurately reflect regional trends or patterns beyond the local level.

No Uncertainty Measure

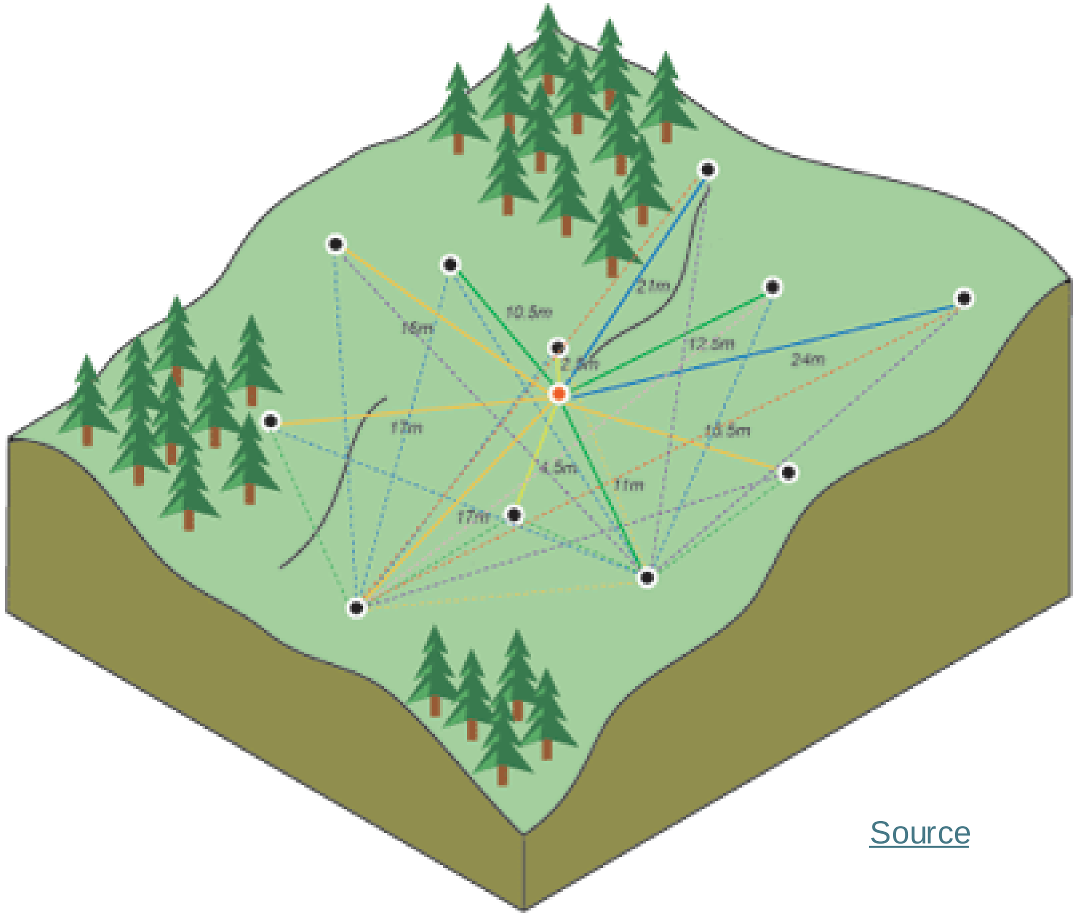


Source

IDW does not provide any measure of uncertainty or error with the predicted values. This makes it difficult to evaluate how reliable an estimate might be.

Kriging

- IDW takes spatial autocorrelation into account, but the weighting is arbitrary
- Kriging combines the approaches conceptually: it uses distance weighting approach, but lets the data speak for themselves to define the weights



Source

History of Kriging

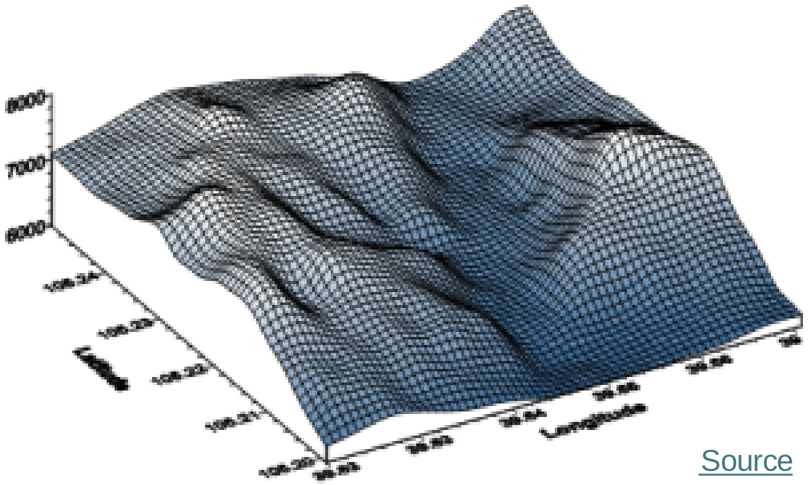
- Geostatistics, first developed by *Georges Matheron* (1930-2000), the French geomathematician. The major concepts and theory were discovered during 1954-1963 while he was working with the **French Geological Survey** in Algeria and France.
- In 1963, he defined the linear geostatistics and concepts of variography, variances of estimation and kriging (named after Danie Krige) in the *Traité de géostatistique appliquée*.
- Kriging was named in honor of Danie Krige (1919-2013), the South African mining engineer who developed the methods of interpolation.

Georges Matheron

Danie G. Krige

What is Geostatistics?

- Techniques which are used for mapping of surfaces from limited sample data and the estimation of values at unsampled locations
- Geostatistics is used for
  - spatial data modelling
  - characterizing the spatial variation
  - spatial interpolation
  - Simulation
  - optimization of sampling
  - characterizing the uncertainty



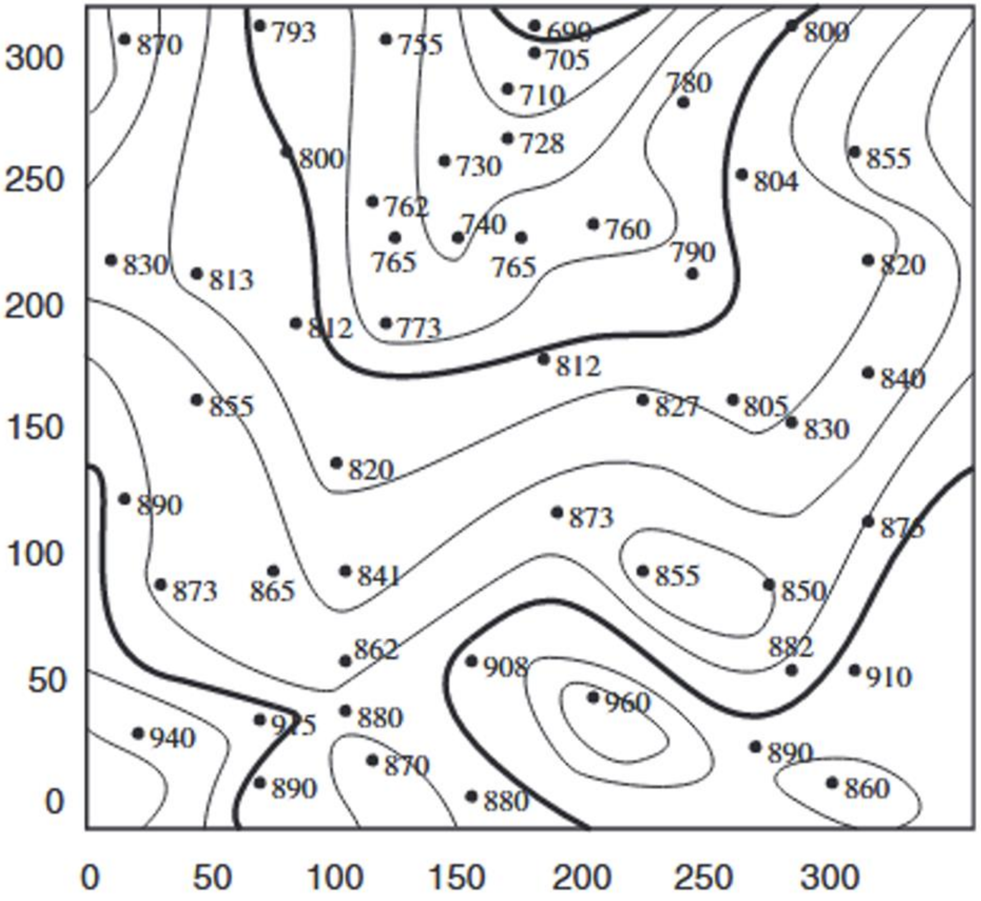
Source

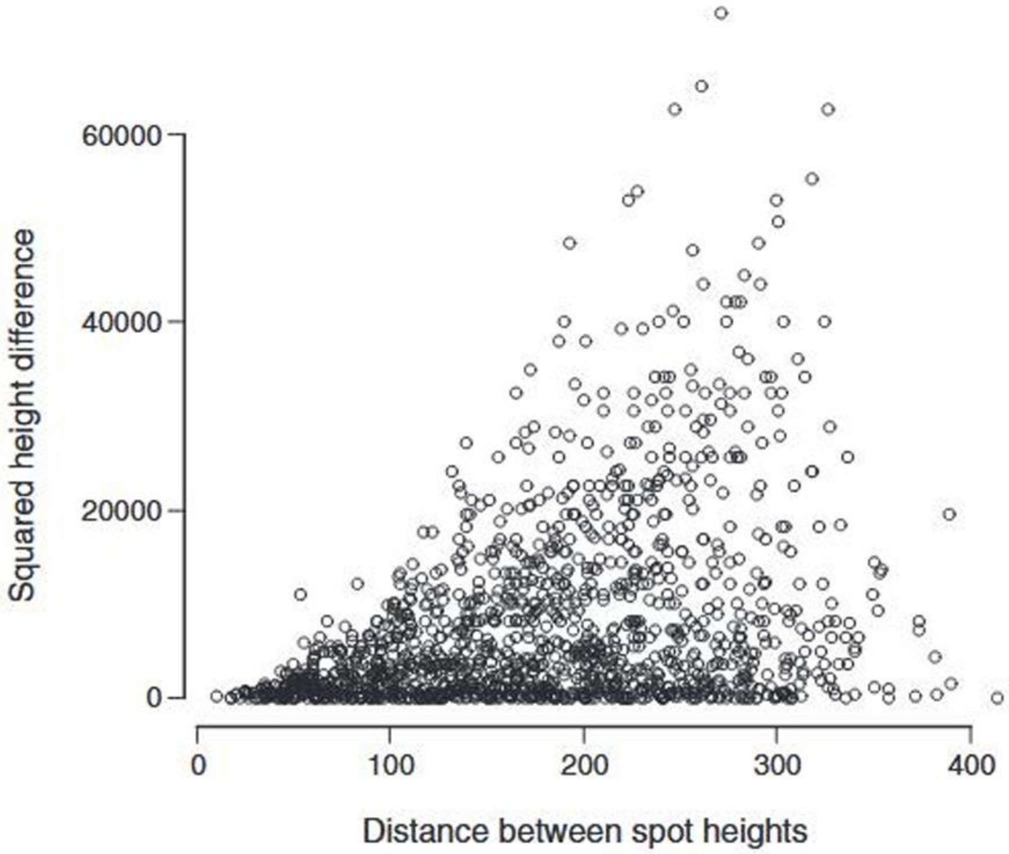
Source

Three major steps in kriging

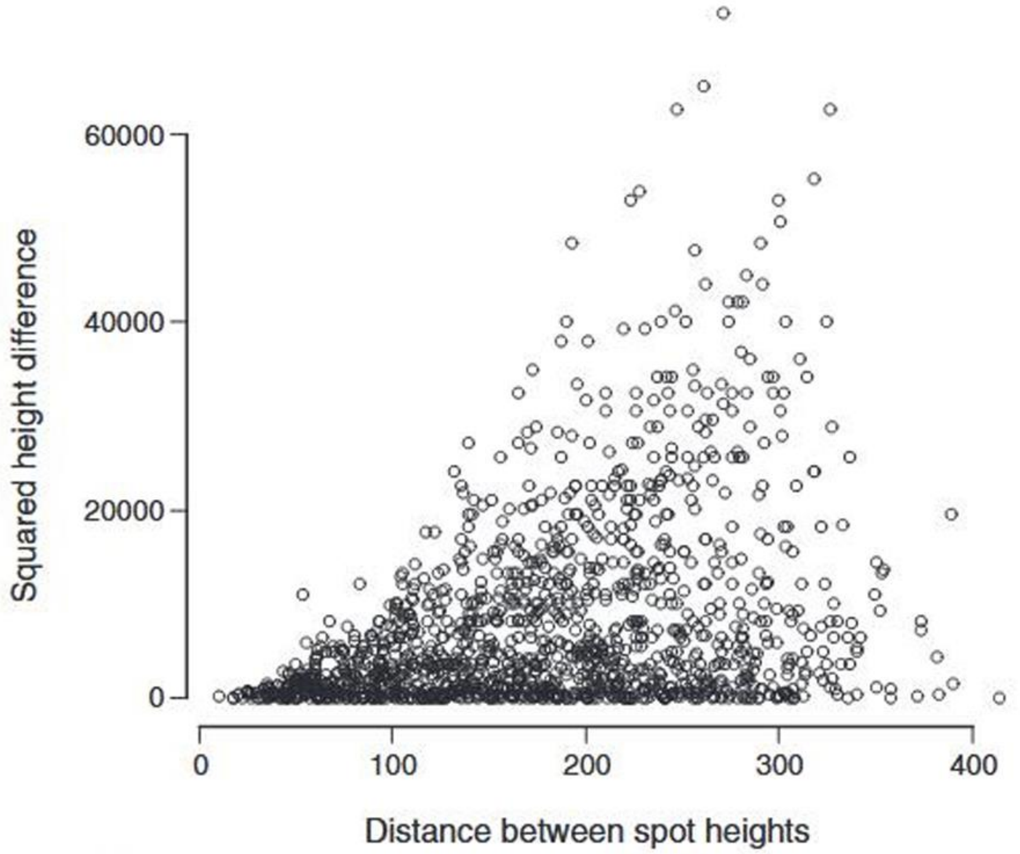
- Describe the spatial variation with variogram
- Summarize the variation with a mathematical function
- Use the function to determine interpolation weights

Step 1: Semivariogram cloud





Step 1: Semivariogram cloud

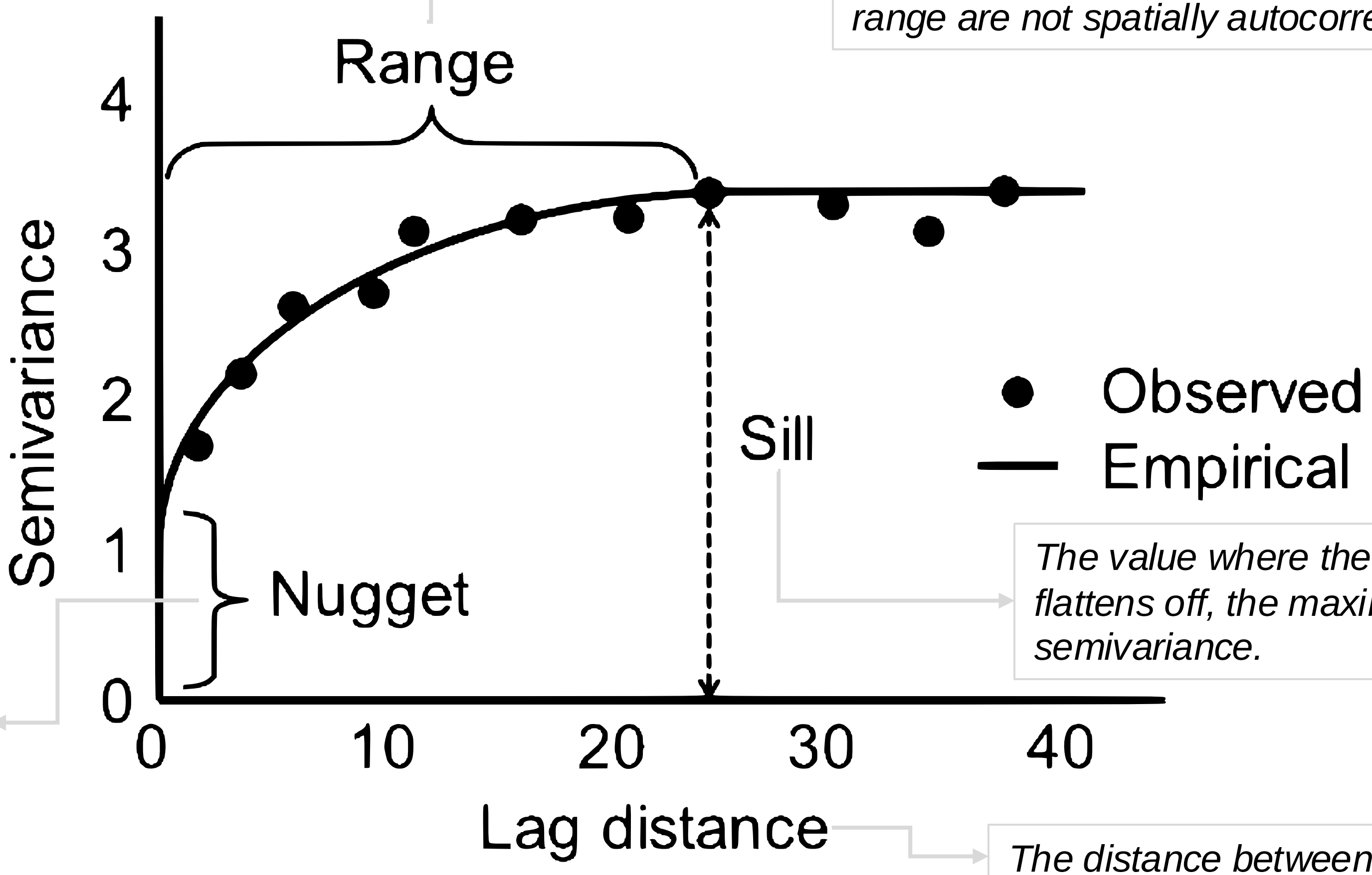


- To examine the spatial continuity of a regionalized variable and how this continuity changes as a function of distance.
- The computation of a variogram involves plotting the relationship between the semivariance and the lag distance
- Measure the strength of correlation as a function of distance
- Quantify the spatial autocorrelation

Step 1: Semivariogram cloud

The variability among the data

The point where the semivariogram reaches the sill on the lag-axis. Sample points that are farther apart than range are not spatially autocorrelated.



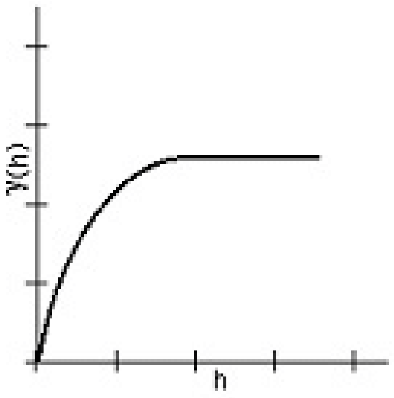
The value where the semivariogram first flattens off, the maximum level of semivariance.

The value of the variogram with 0 lag; errors in measurements

The distance between sampling pairs

Step 2: Mathematical model

Spherical

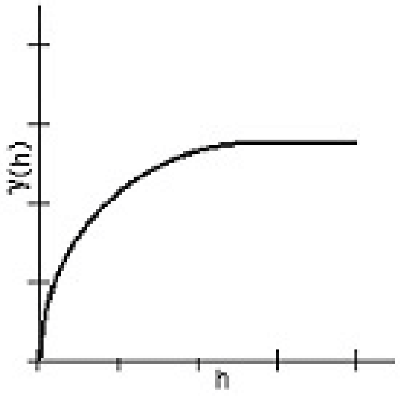


$$\gamma(h) = c_0 + c\left(\frac{3h}{2\alpha} - \frac{1}{2}\left(\frac{h}{\alpha}\right)^3\right)$$

$$\gamma(h) = c_0 + c \qquad h > \alpha$$

$$\gamma(0) = 0$$

Circular



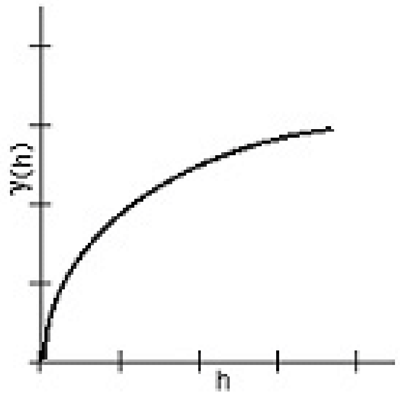
$$\gamma(h) = c_0 + c\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{\alpha}\right) + \sqrt{1 - \frac{h^2}{\alpha^2}}\right)$$

$$0 < h \le \alpha$$

$$\gamma(h) = c_0 + c \qquad h > \alpha$$

$$\gamma(0) = 0$$

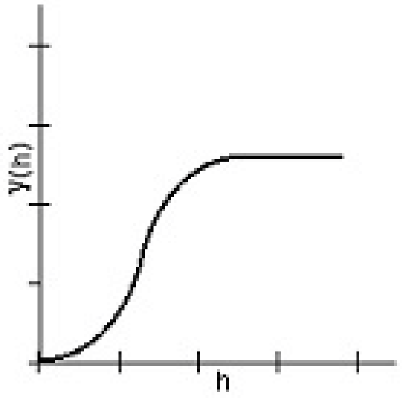
Exponential



$$\gamma(h) = c_0 + c\left(1 - exp\left(\frac{-h}{r}\right)\right) \qquad h > 0$$

$$\gamma(0) = 0$$

Gaussian



$$\gamma(h) = c_0 + c\left(1 - exp\left(-\frac{h^2}{r^2}\right)\right) \qquad h > 0$$

$$\gamma(0) = 0$$

Source

Step 3: Determine interpolation weights

Covariance matrix is built by using the variogram to determine how much each point is correlated

$$\begin{array}{ccccc} \mathbf{C} & \cdot & \mathbf{w} & = & \mathbf{D} \\ \underbrace{\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}_{(n+1)\times(n+1)} & \cdot & \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix}}_{(n+1)\times 1} & = & \underbrace{\begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}}_{(n+1)\times 1} \end{array}$$

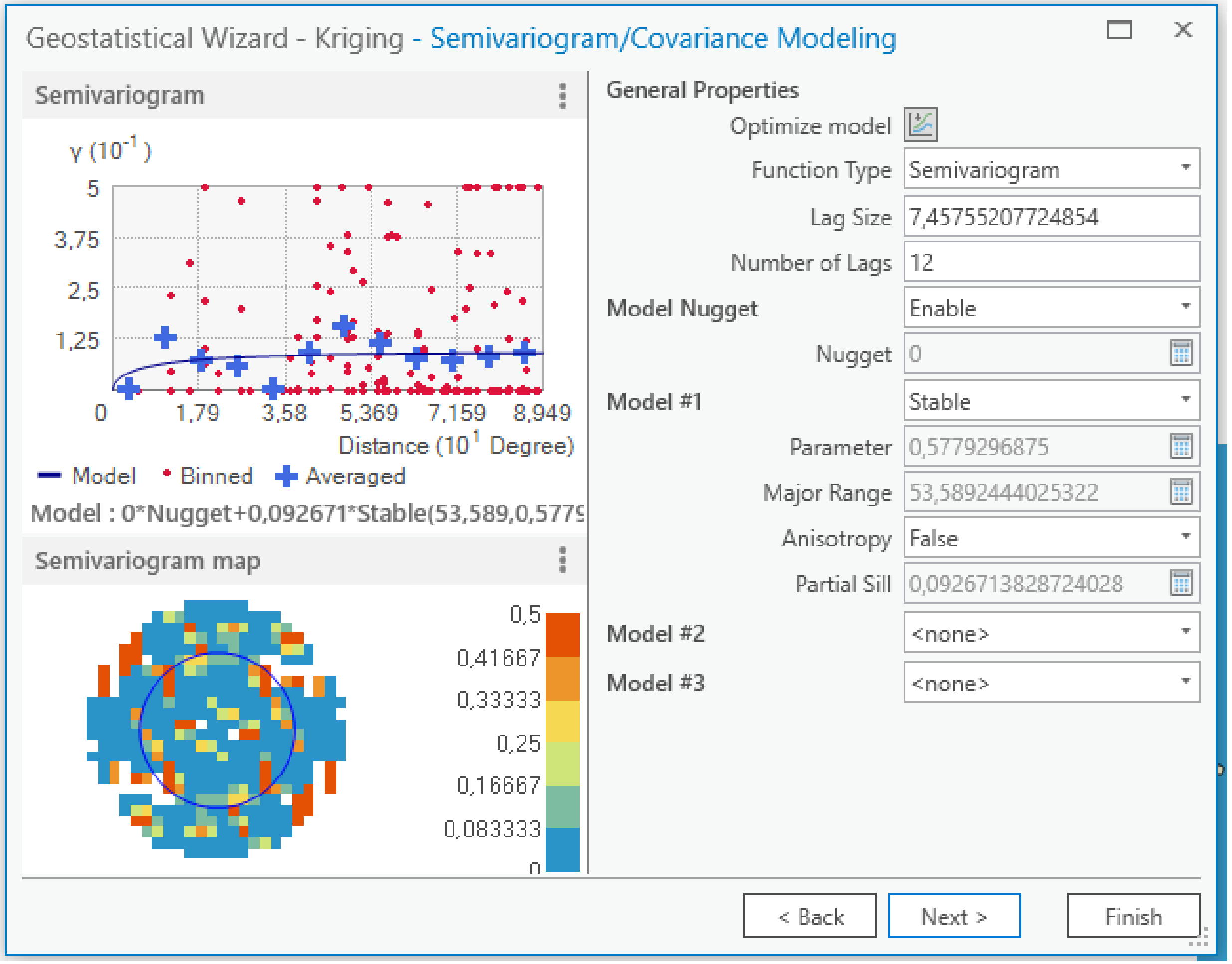
Another vector is created that represents the covariance between the unknown point and each known point.

Source

Solve for weights:

$$\begin{aligned} \mathbf{C} \cdot \mathbf{w} &= \mathbf{D} \\ \mathbf{C}^{-1} \cdot \mathbf{C} \cdot \mathbf{w} &= \mathbf{C}^{-1} \cdot \mathbf{D} \\ \mathbf{I} \cdot \mathbf{w} &= \mathbf{C}^{-1} \cdot \mathbf{D} \\ \mathbf{w} &= \mathbf{C}^{-1} \cdot \mathbf{D} \end{aligned}$$

Kriging in ArcGIS Pro Geostatistical Wizard



Geostatistical Wizard - Kriging - Semivariogram/Covariance Modeling

Model : 0*Nugget+0,092671*Stable(53,589,0,577

General Properties

Optimize model

Function Type: Semivariogram

Lag Size: 7,45755207724854

Number of Lags: 12

Model Nugget: Enable

Nugget: 0

Model #1: Stable

Parameter: 0,5779296875

Major Range: 53,5892444025322

Anisotropy: False

Partial Sill: 0,0926713828724028

Model #2: <none>

Model #3: <none>

Source

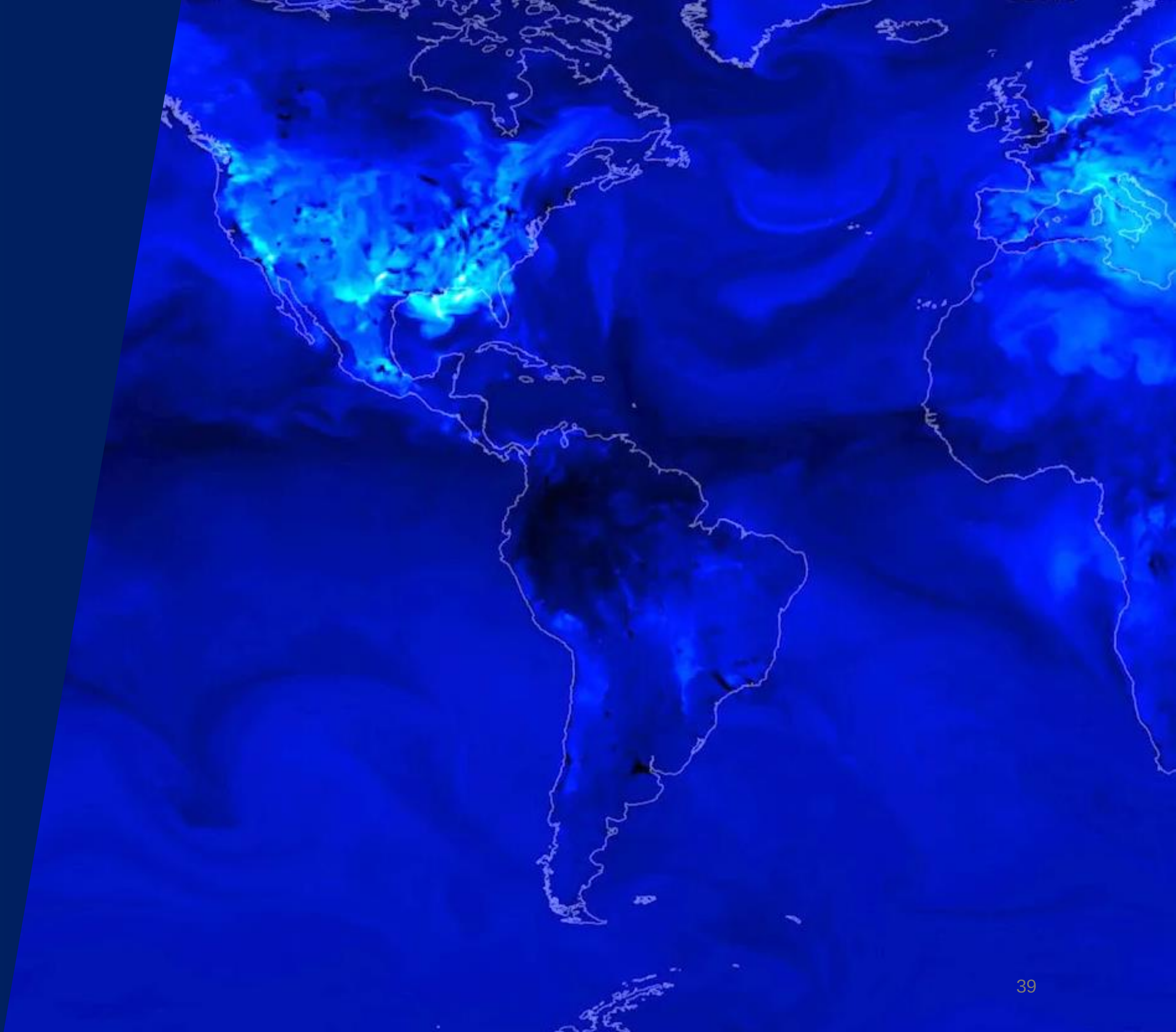
Some key points about kriging

- Different types of kriging (ordinary, simple, universal, block, indicator, co-kriging)
- Co-kriging includes the use of multiple variables together
- Anisotropy should be considered (Spatial variation is not the same in all directions)
- Leave-one-out cross-validation is important to understand the quality of interpolation
- Scatterplot of errors from observed samples and predicted values can be useful

In summary

- Two types of interpolation
  - Deterministic and stochastic
- When you have dense sample, deterministic methods are good enough
- IDW is the most used deterministic interpolation method
- Kriging is useful when the samples are sparse
- Variogram cloud is very important to figure out kriging raster

SAINT LOUIS
UNIVERSITY.



Thank You