2-Day Course – Spatial Modeling with Geostatistics

Prof. Michael J. Pyrcz, Ph.D., P.Eng.
Associate Professor

Hildebrand Department of Petroleum & Geosystems Engineering
University of Texas at Austin

Bureau of Economic Geology, Jackson School of Geosciences
University of Texas at Austin

“In two days, what a geoscientists needs to know about geostatistics, and workflows to get you started with applying geostatistics to impact your work.”

Spatial Modeling with Geostatistics

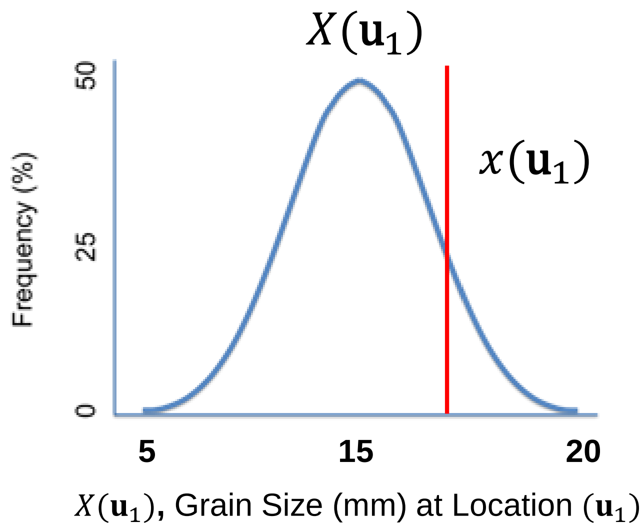
Spatial Data Analysis - Calculation

Lecture outline . . .

- **Stationarity**
- **Quantifying Spatial Continuity**

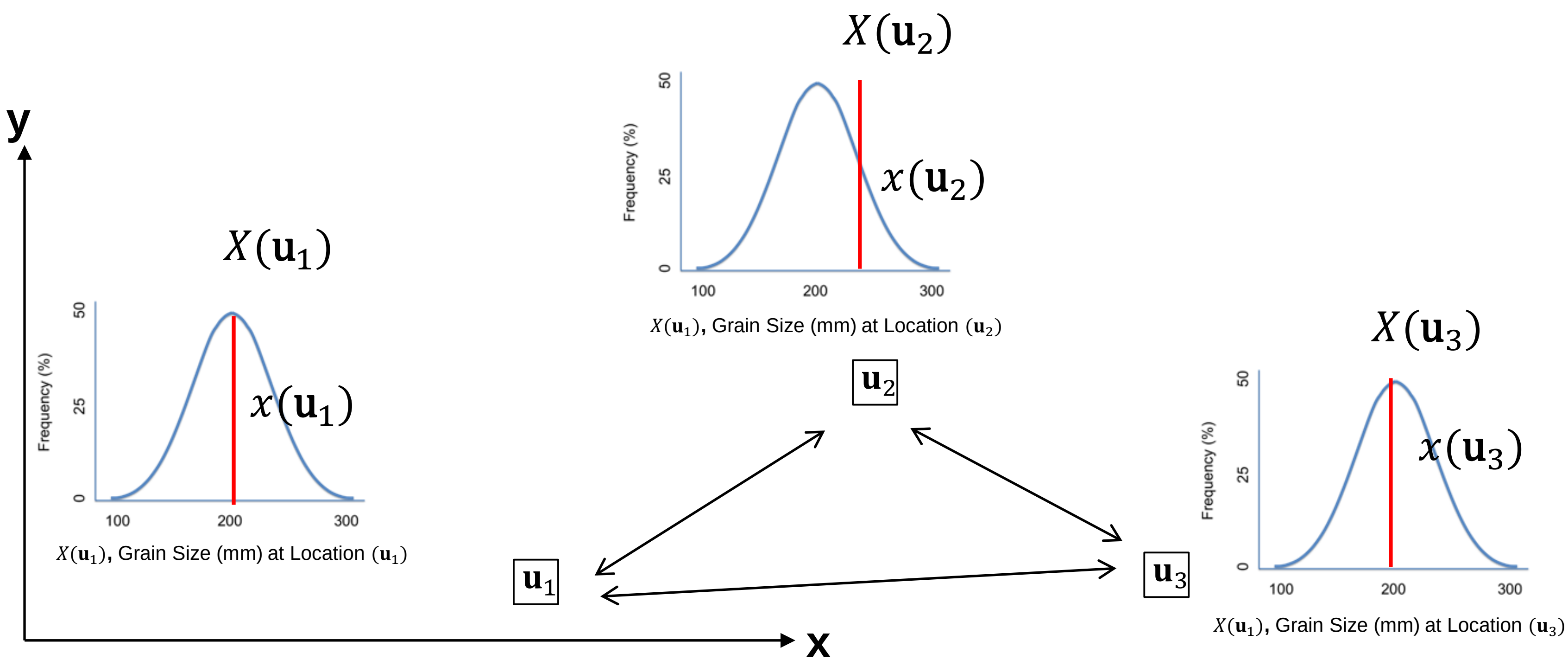
Random Variable (RV) Definition

- **Random Variable**
  - we do not know the value at a location / time, it can take on a range of possible values, fully described with a PDF.
  - represented as an upper case variable, e.g. X, while possible outcomes or data measures are represented with lower case, e.g. x.
  - in spatial context common to use a location vector, $\mathbf{u}$, to describe a location, e.g. $x(\mathbf{u})$, $X(\mathbf{u})$



Random Function (RF) Definition

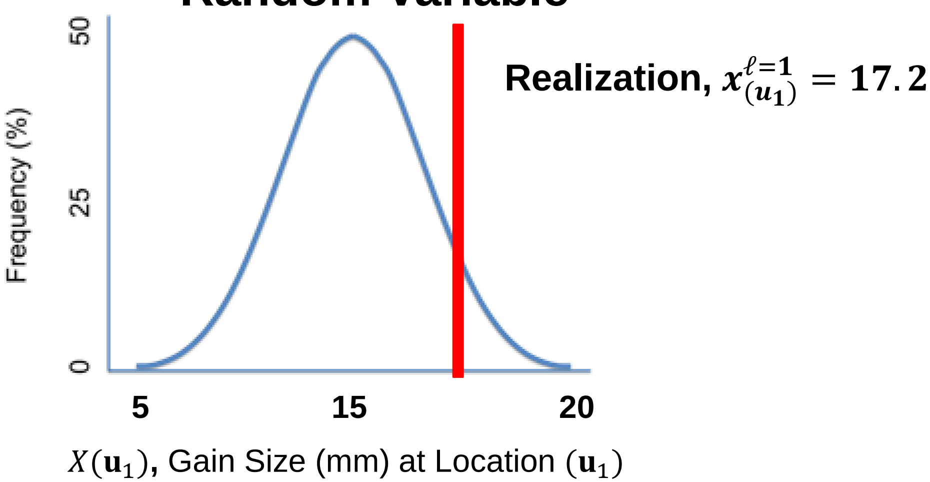
- **Random Function**
  - set of random variables correlated over space and / or time
  - represented as an upper case variable, e.g. $X_1, X_2, \ldots, X_n$, while possible joint outcomes or data measures are represented with lower case, e.g. $x_1, x_2, \ldots, x_n$
  - in spatial context common to use a location vector, $\mathbf{u}_\alpha$, to describe a location, e.g. $x(\mathbf{u}_1), x(\mathbf{u}_2), \ldots, x(\mathbf{u}_n)$, and $X(\mathbf{u}_1), X(\mathbf{u}_2), \ldots, X(\mathbf{u}_n)$



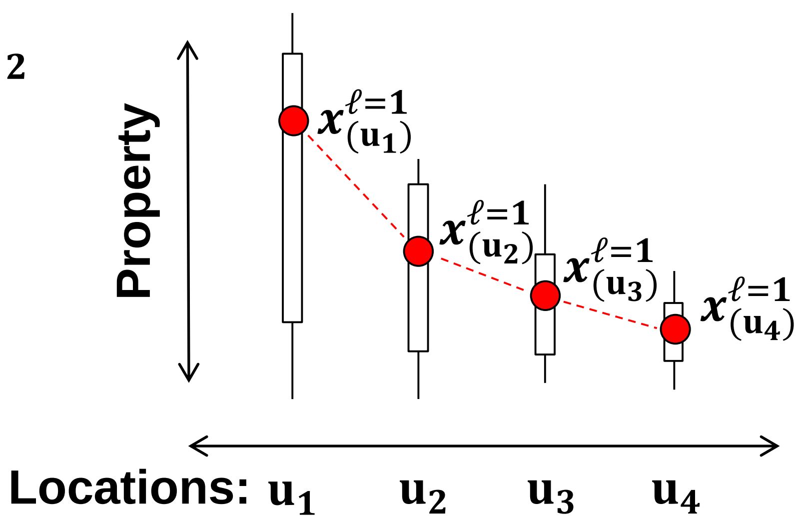
Realization Definition

- **Realization**
- an outcome from a random variable or joint set of outcomes from a random function.
- represented with lower case, e.g. x.
- in spatial context common to use a location vector, $\mathbf{u}$, to describe a location, e.g. $x(\mathbf{u})$, $X(\mathbf{u})$
- resulting from simulation, e.g. Monte Carlo simulation, sequential Gaussian simulation ← a method to sample (jointly) from the RV (RF)
- each realization considered equiprobable

Random Variable



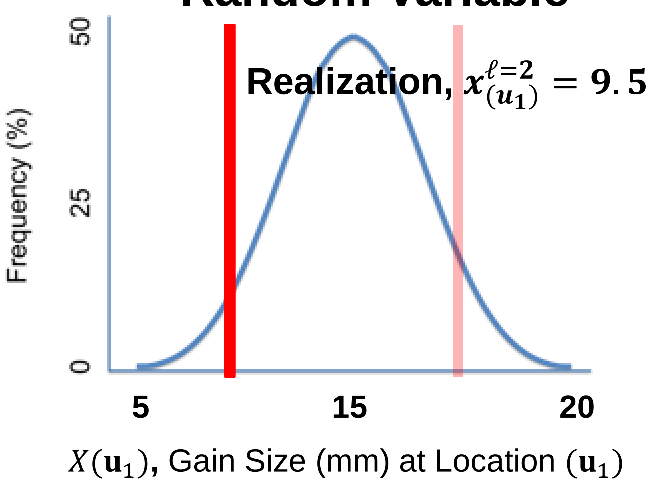
Random Function



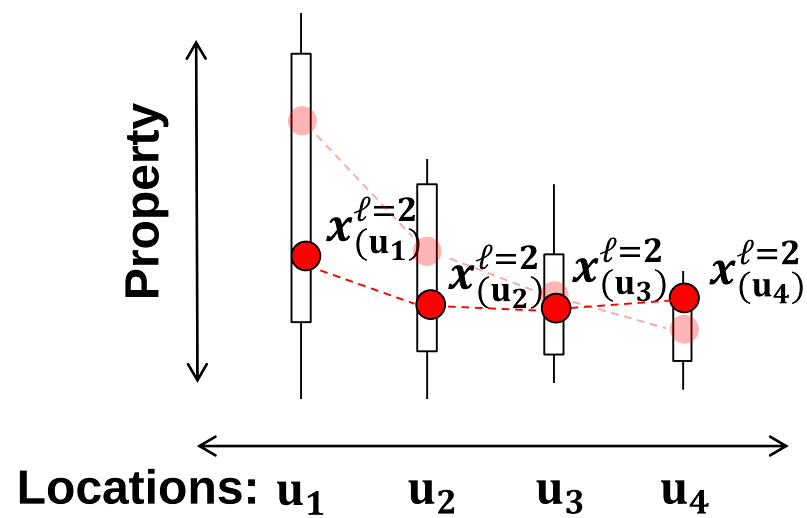
Realization Definition

- **Realization**
- an outcome from a random variable or joint set of outcomes from a random function.
- represented with lower case, e.g. x.
- in spatial context common to use a location vector, $\mathbf{u}$, to describe a location, e.g. $x(\mathbf{u})$, $X(\mathbf{u})$
- resulting from simulation, e.g. Monte Carlo simulation, sequential Gaussian simulation ← a method to sample (jointly) from the RV (RF)
- each realization is considered equiprobable

Random Variable



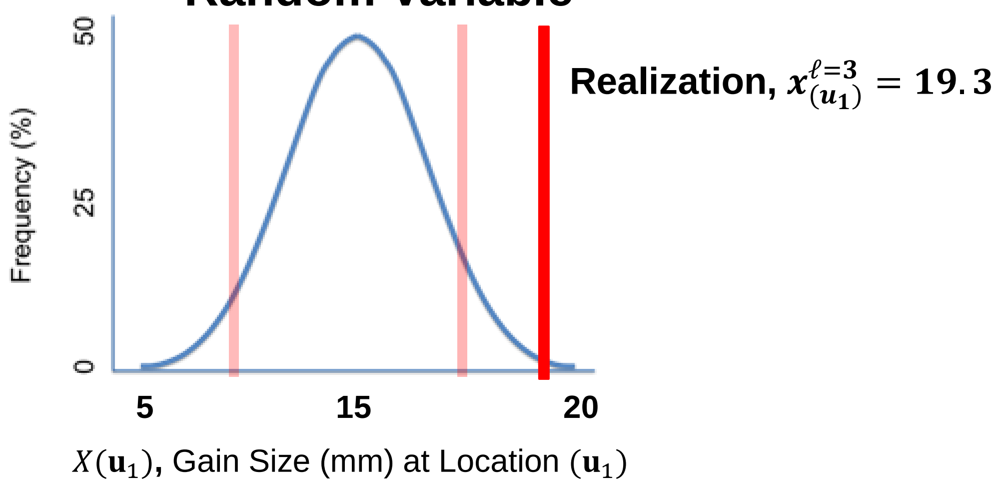
Random Function



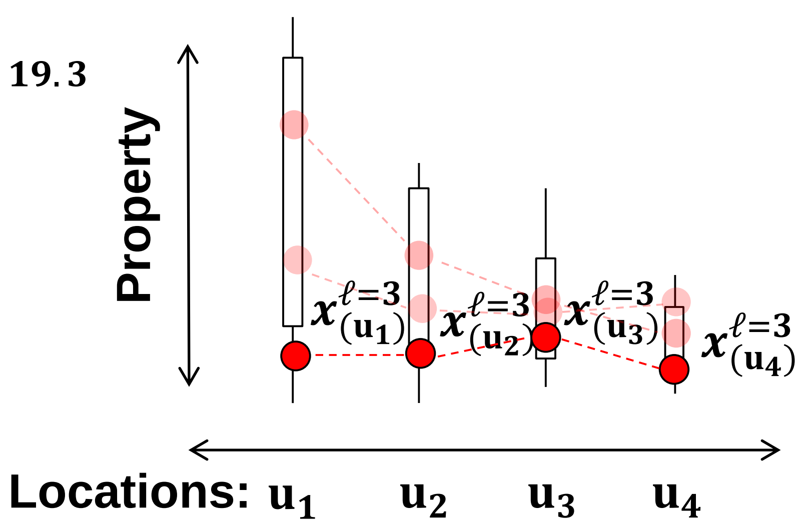
Realization Definition

- **Realization**
- an outcome from a random variable or joint set of outcomes from a random function.
- represented with lower case, e.g. x.
- in spatial context common to use a location vector, $\mathbf{u}$, to describe a location, e.g. $x(\mathbf{u})$, $X(\mathbf{u})$
- resulting from simulation, e.g. Monte Carlo simulation, sequential Gaussian simulation ← a method to sample (jointly) from the RV (RF)
- each realization is considered equiprobable

Random Variable



Random Function



Decision of Stationarity

- Consider a random variable $X(\mathbf{u}_\alpha) \rightarrow F_x(x;\mathbf{u}_\alpha) = Prob(X \leq x)$
- What is the practical meaning of $F_x(x;\mathbf{u}_\alpha)$? There can only be one sample at any specific time/location.
- There is a need to pool samples coming from different times and/or locations to come up with $F_x(x;\mathbf{u}_\alpha)$ (or any statistic).

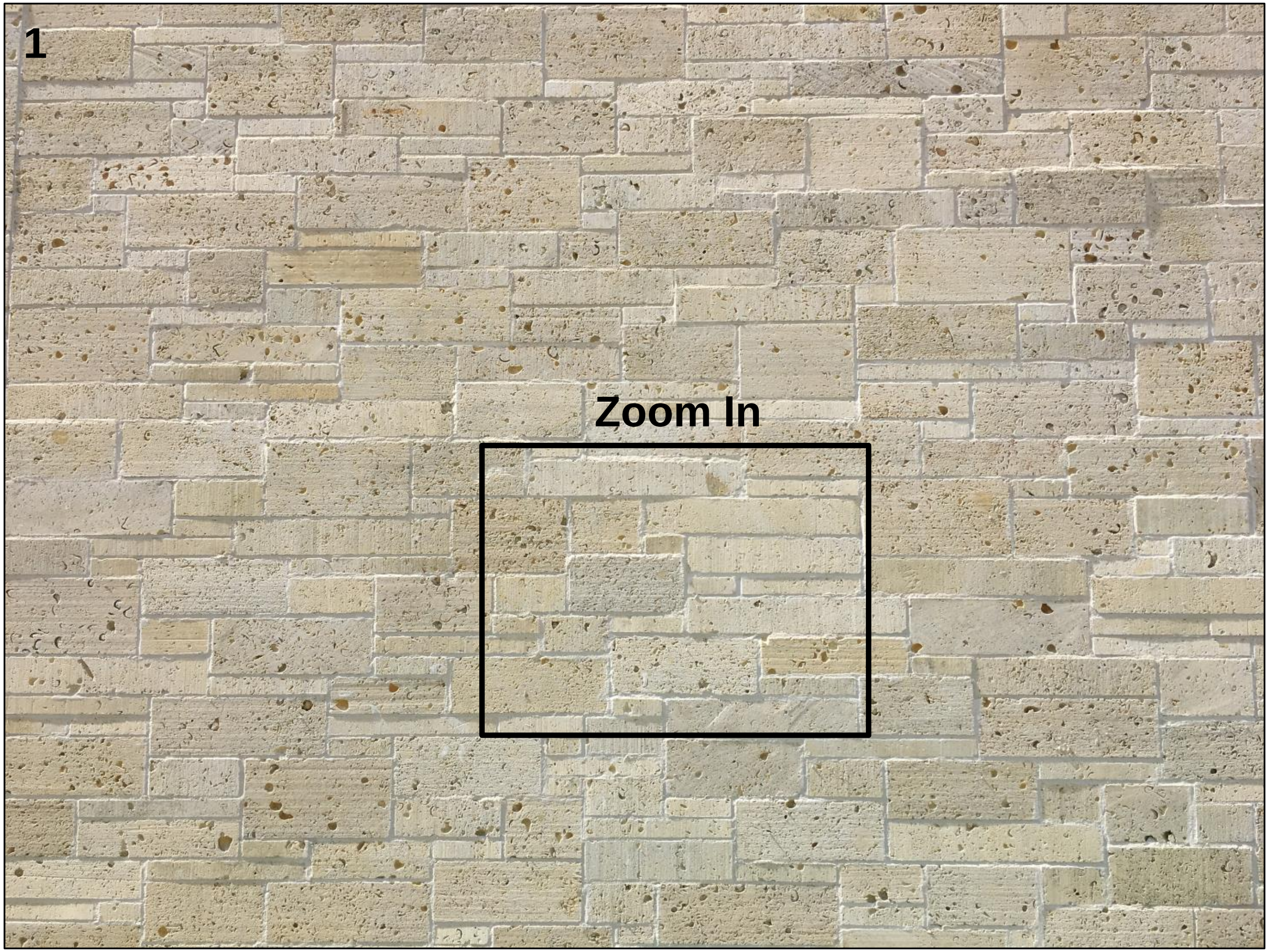
Choice of the Pool = Decision of Stationarity

Import License to pool samples over an area / volume.

Export License to use these statistics over an area / volume.

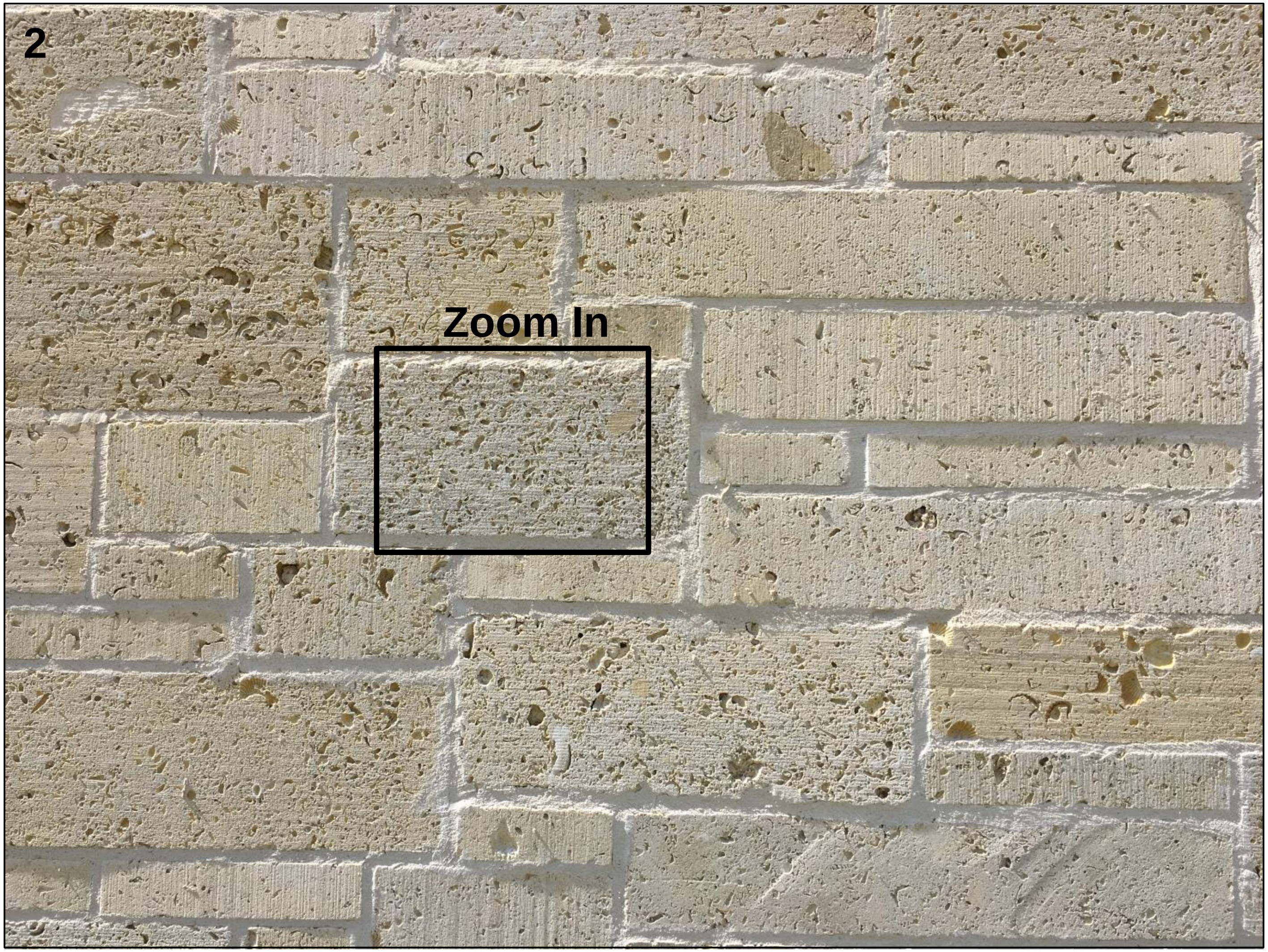
- Stationarity in the mean, variance and entire CDF.
  - stationary mean, $m_x(\mathbf{u}) = m_x$
  - stationary variance, $\sigma_x^2(\mathbf{u}) = \sigma_x^2$
  - stationary CDF, $F_x(x;\mathbf{u}) = F_x(x)$
  - etc.
- Depends on scale of observation

Stationarity



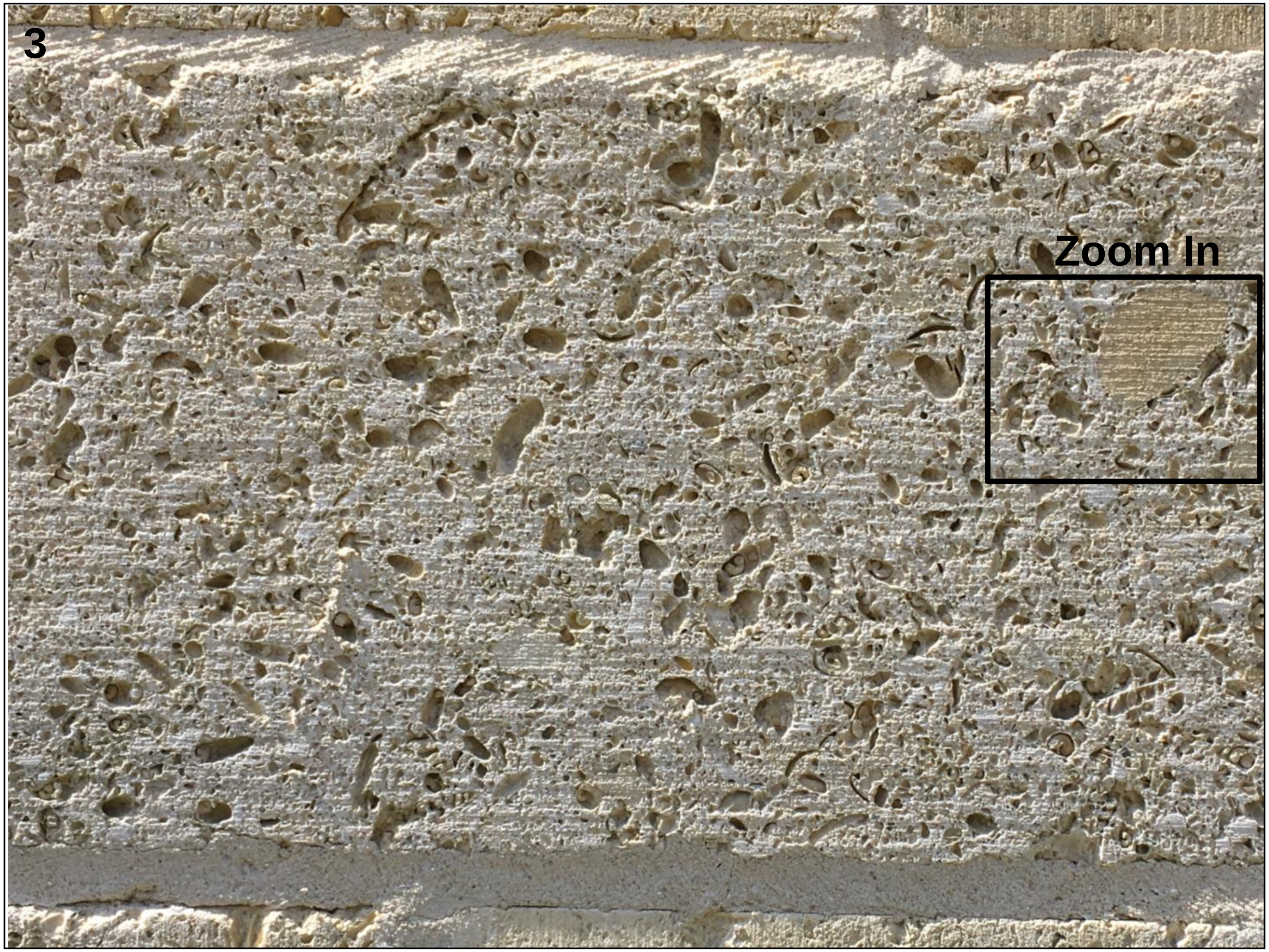
- Is this image stationary? What metric do you consider?

Stationarity



- A smaller group of bricks?

Stationarity

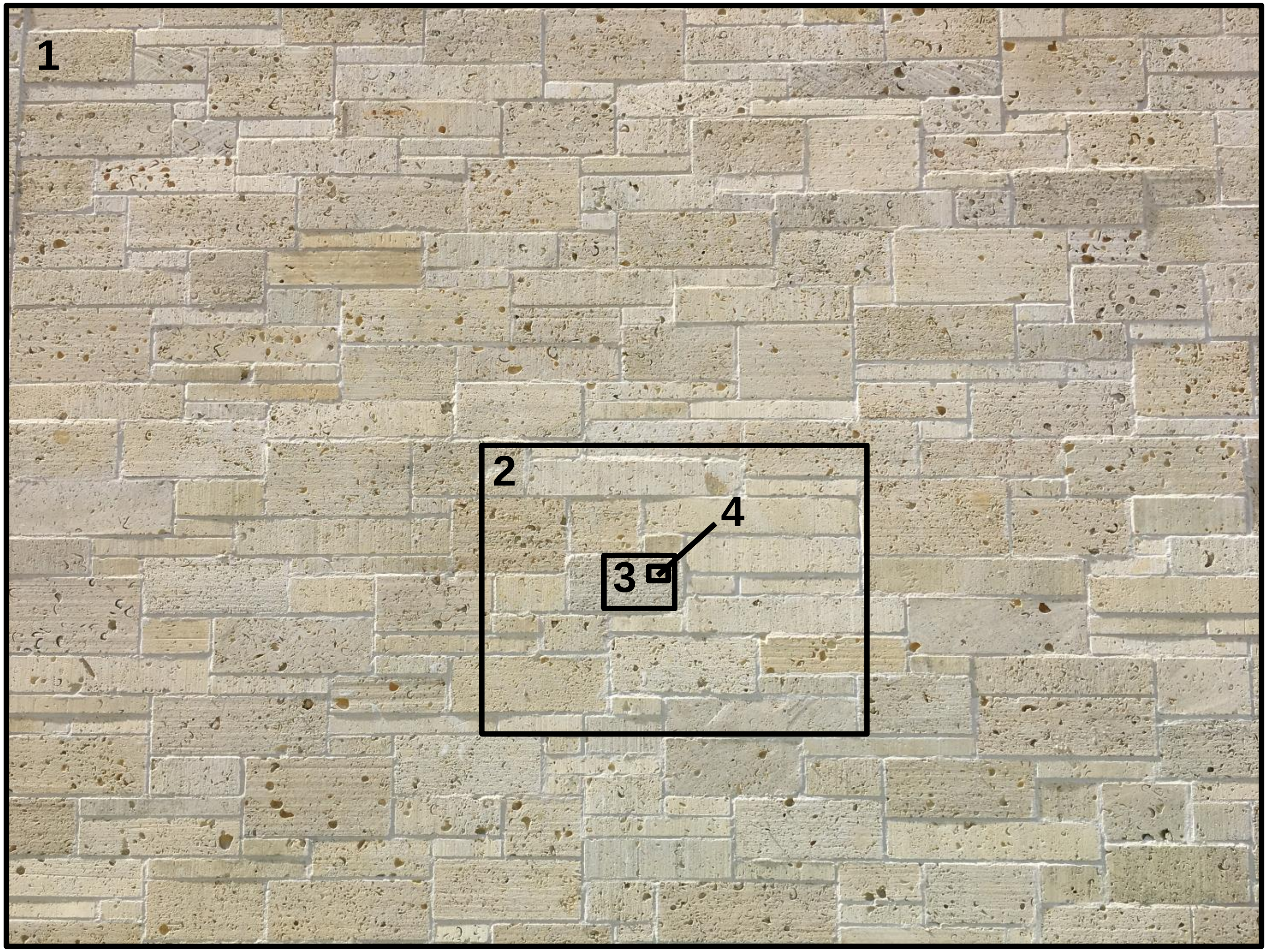


- A single brick?

Stationarity

- Small part of a brick?

Stationarity



- Is this image stationary? What metric do you consider?

Stationarity Definition

- **Stationarity** – "*statistic/metric*" is invariant under translation over an *interval* e.g. region, time period etc.
  - **What is stationary?** Need a metric.
  - **Over what interval?** Need a time or volume of interest
  - Depends on the model purpose
  - Depends on the scale of observation
  - Decision not an hypothesis; therefore, if cannot be tested

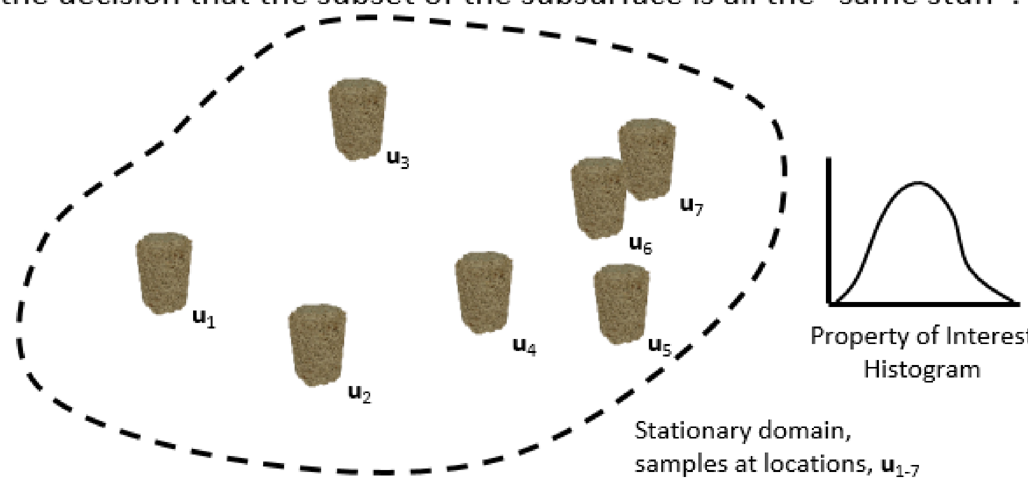
Stationarity Summary

1. Substituting time for space.

Any statistic requires replicates, repeated sampling (e.g. air or water samples from a monitoring station). In our geospatial problems repeated samples are not available at a location in the subsurface.

Extracted Sample at Location, $\mathbf{u}_1$ Hole Remains at Location, $\mathbf{u}_1$

Instead of time, we must pool samples over space to calculate our statistics. This decision to pool is the decision of stationarity. It is the decision that the subset of the subsurface is all the “same stuff”.

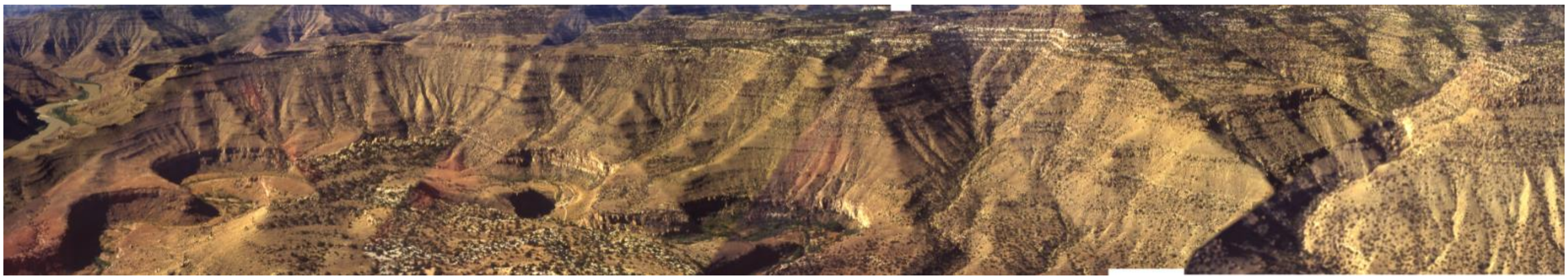


The decision of the stationary domain for sampling is an expert choice. Without it we are stuck in the “hole” and cannot calculate any statistics nor say anything about the behavior of the subsurface between the sample data. Core image from https://www.fei.com/oil-gas/

Stationarity Summary

2. Definitions of Stationarity.

Geological Definition: The rock over the stationary domain is sourced, deposited, preserved, and postdepositionally altered in a similar manner, the domain is map-able and may be used for local prediction or as information for analogous locations within the subsurface; therefore, it is useful to pool information over this expert mapped volume of the subsurface.



Statistical Definition: The metrics of interest are invariant under translation over the domain. For example, one point stationarity indicates the that histogram and associated statistics do not rely on location, $\mathbf{u}$. Statistical stationarity for some common statistics:

Stationary Mean: $E\{Z(\mathbf{u})\} = \text{m}, \ \forall\, \mathbf{u}$

Stationary Distribution: $F(\mathbf{u}; z) = F(z), \ \forall\, \mathbf{u}$

Stationary Semivariogram: $\gamma_z(\mathbf{u}; \mathbf{h}) = \gamma_z(\mathbf{h}), \ \forall\, \mathbf{u}$

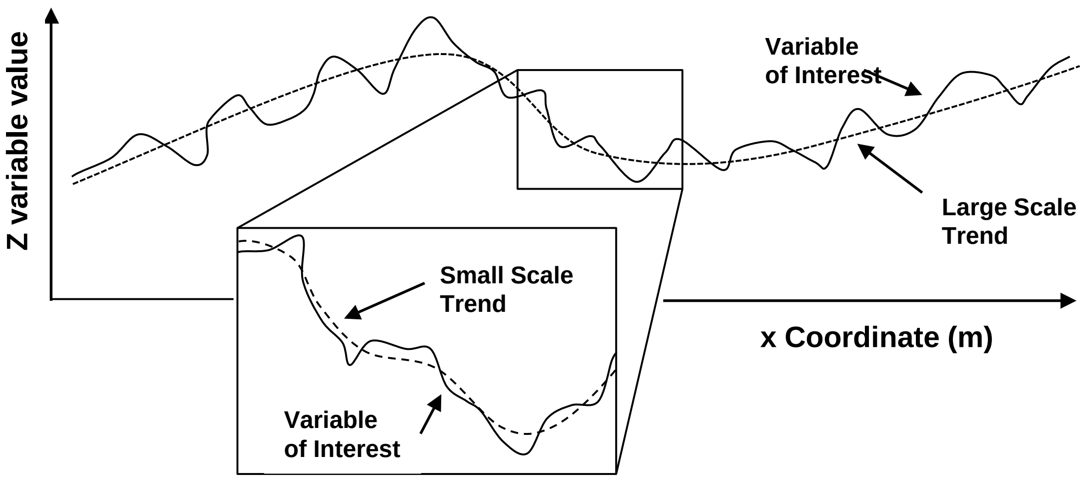
May be extended to any statistic of interest including, facies proportions, bivariate distributions and multiple point statistics.

Stationarity Summary

3. Comments on Stationarity.

Stationarity is a decision, not an hypothesis; therefore it cannot be tested. Data may demonstrate that it is inappropriate.

The **stationarity assessment depends on scale**. This choice of modeling scale(s) should be based on the specific problem and project needs.



We cannot avoid a decision of stationarity. No stationarity decision and we cannot move beyond the data. Conversely, assuming broad stationarity over all the data and over large volumes of the earth is naïve. Good geological mapping is essential.

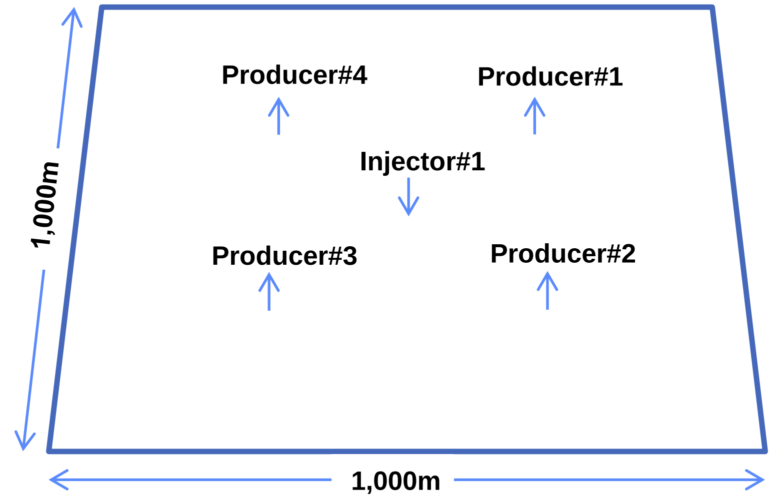
Geomodeling stationarity is the decision (1) over what region to pool data and (2) over what region to use the resulting statistics.

Nonstationary trends may be mapped and the remaining stationary residual modelled stochastically, trends may be treated uncertain.

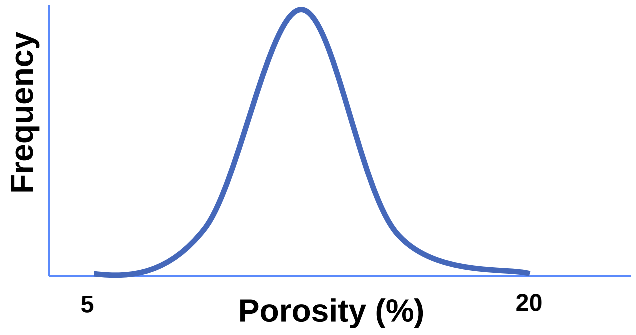
Motivation for Measuring Spatial Continuity

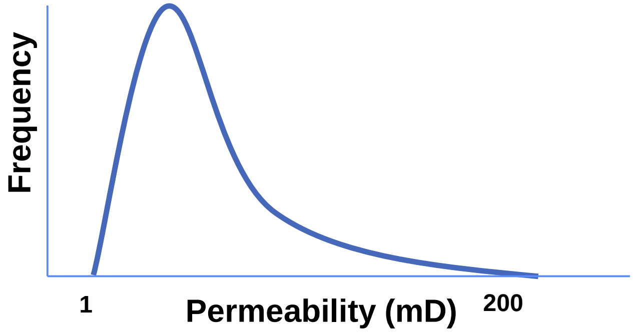
Simple Example

- Area of interest
- 1 Injector and 4 producers



- Porosity and permeability distributions (held constant for all cases)

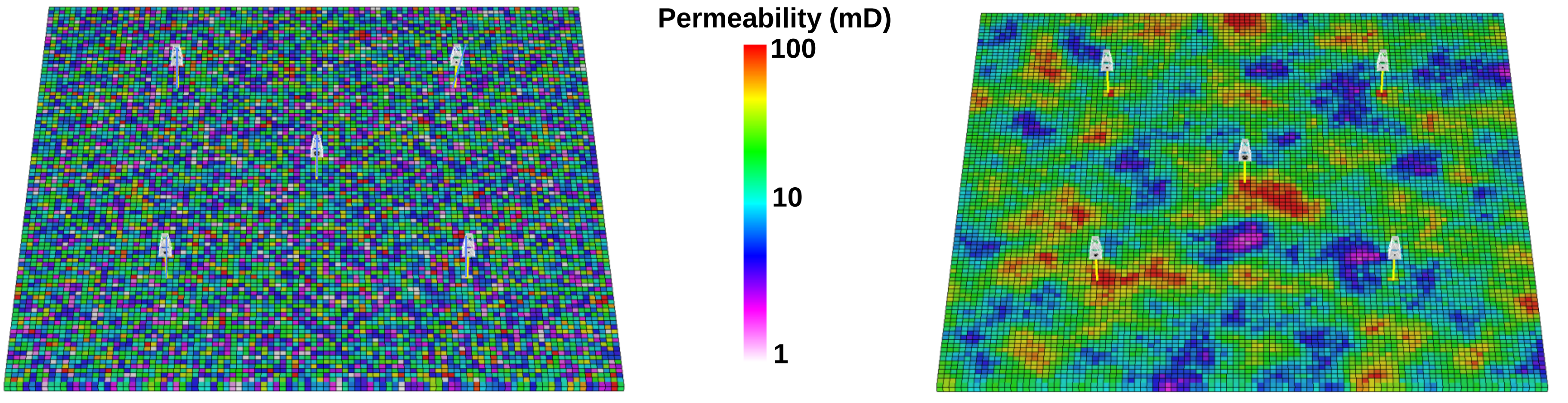




Motivation for Measuring Spatial Continuity

Does spatial continuity of reservoir properties matter?

Consider these models of permeability

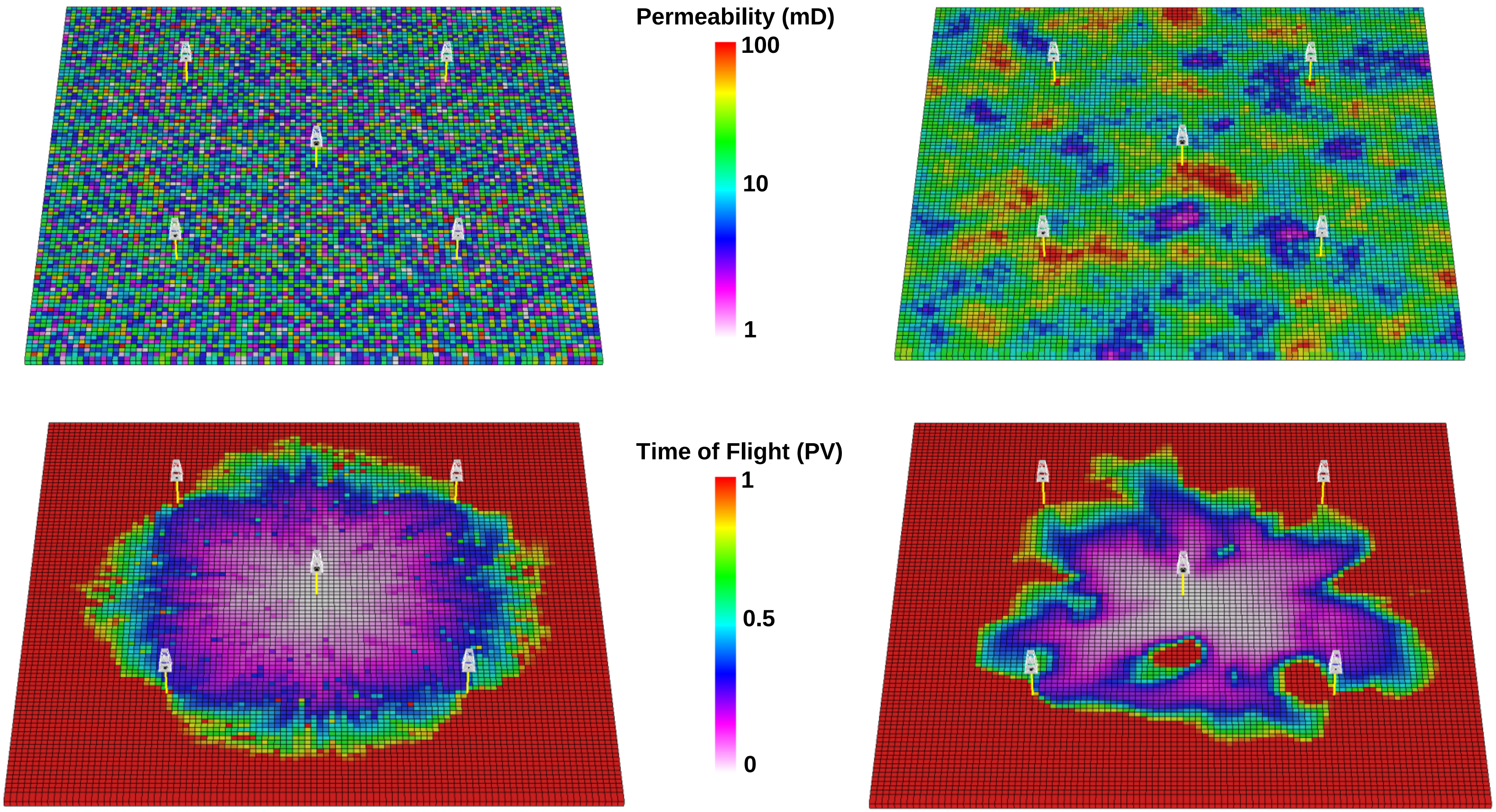


Recall – all models have the same porosity and permeability distributions

- Mean, variance, P10, P90 …
- Same static oil in place!

Motivation for Measuring Spatial Continuity

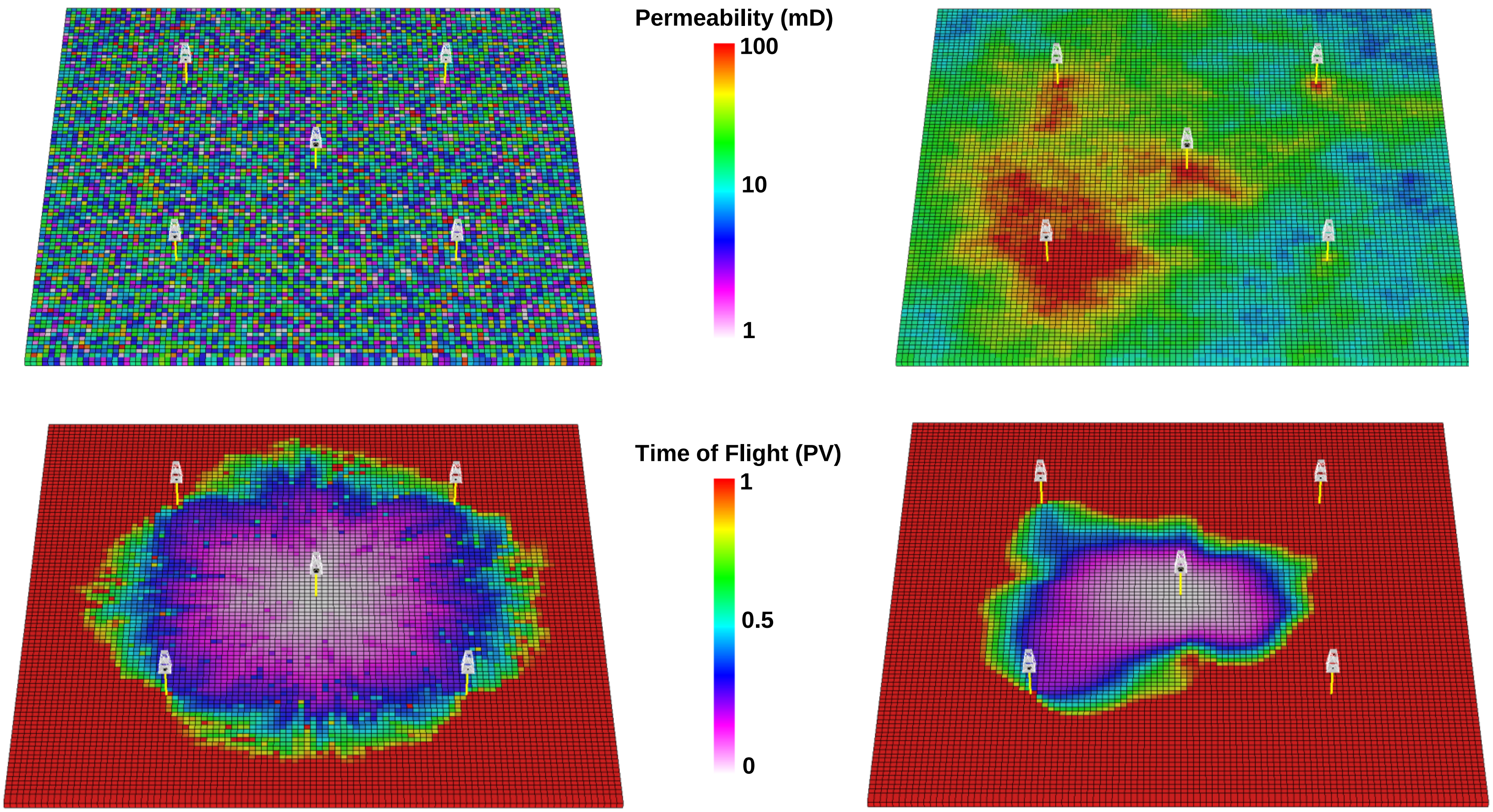
Does spatial continuity of reservoir properties matter?



How does heterogeneity impact recovery factor? Well Estimated Ultimate Recovery?

Motivation for Measuring Spatial Continuity

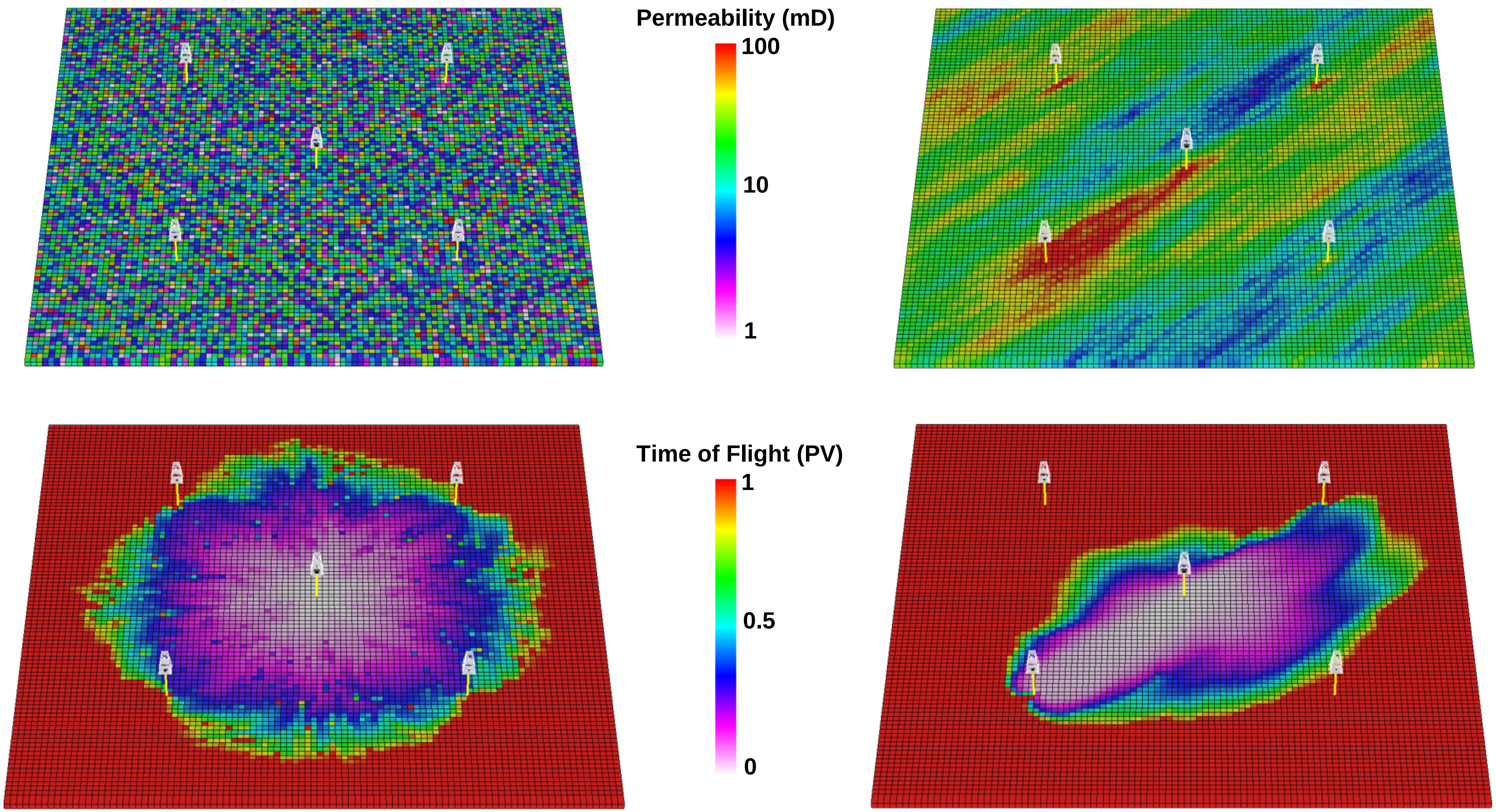
Does spatial continuity of reservoir properties matter?



How does heterogeneity impact recovery factor? Well Estimated Ultimate Recovery?

Motivation for Measuring Spatial Continuity

Does spatial continuity of reservoir properties matter?



How does heterogeneity impact recovery factor? Well Estimated Ultimate Recovery?

Motivation for Measuring Spatial Continuity

- For the same reservoir property distributions a wide range of spatial continuities are possible.
- Spatial continuity often impacts reservoir forecasts.
- Need to be able to:

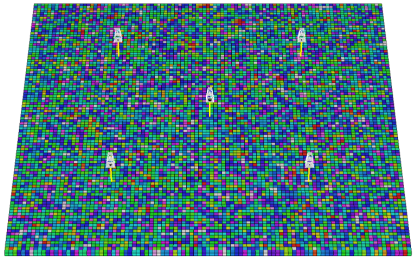
Characterize / Quantify

Spatial Continuity Measures

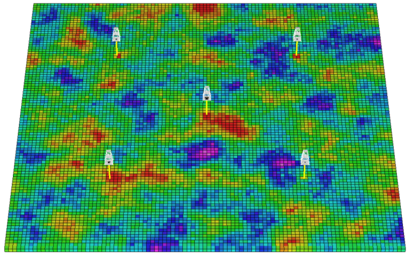
Impose in Reservoir Models

Spatial Continuity

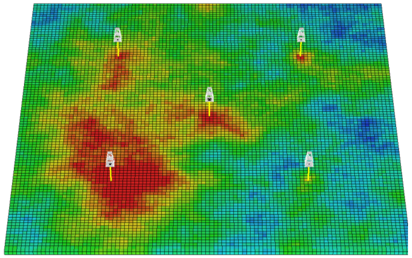
"Very Short"



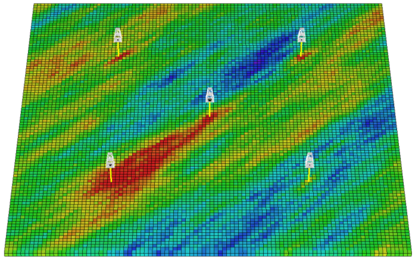
"Medium"



"Long"



"Anisotropic "

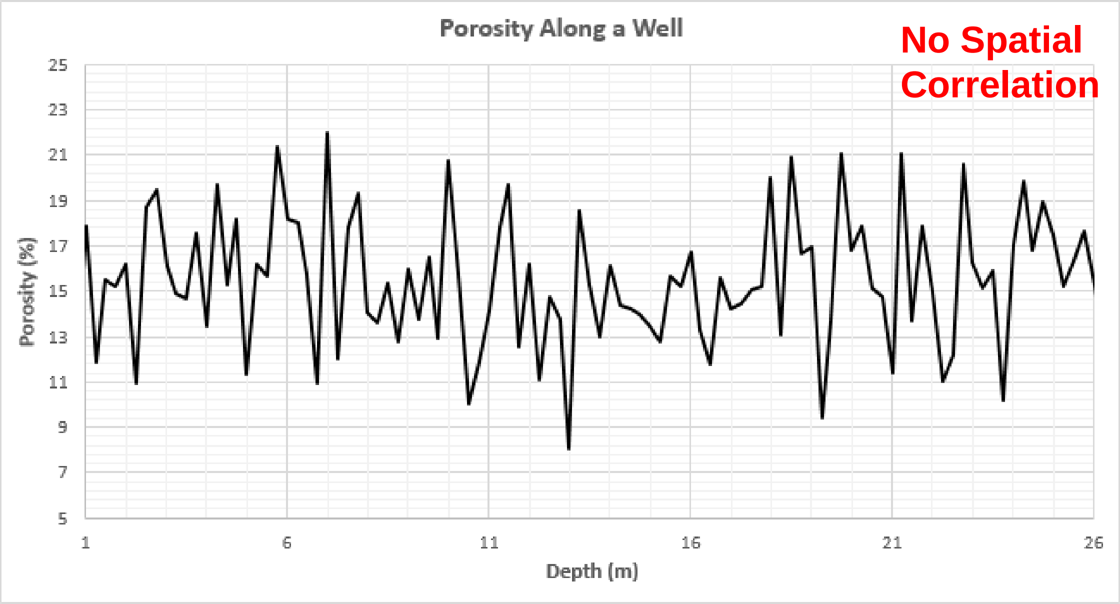


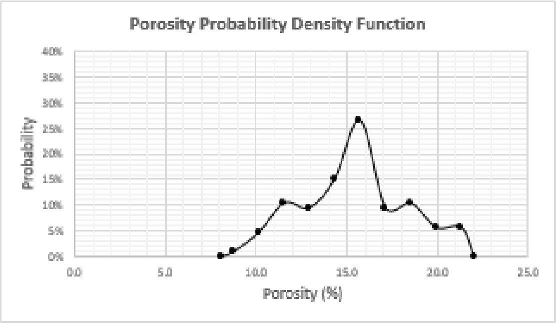
Spatial Continuity Definition

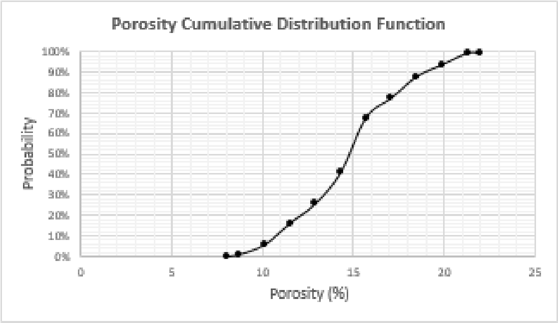
- **Spatial Continuity** – correlation between values over distance.
  - No spatial continuity – no correlation between values over distance, random values at each location in space regardless of separation distance.
  - Homogenous phenomenon have perfect spatial continuity, since all values as the same (or very similar) they are correlated.

Spatial Continuity Definition

- No spatial continuity – random values at each location in space regardless of separation distance. Example from RAND().

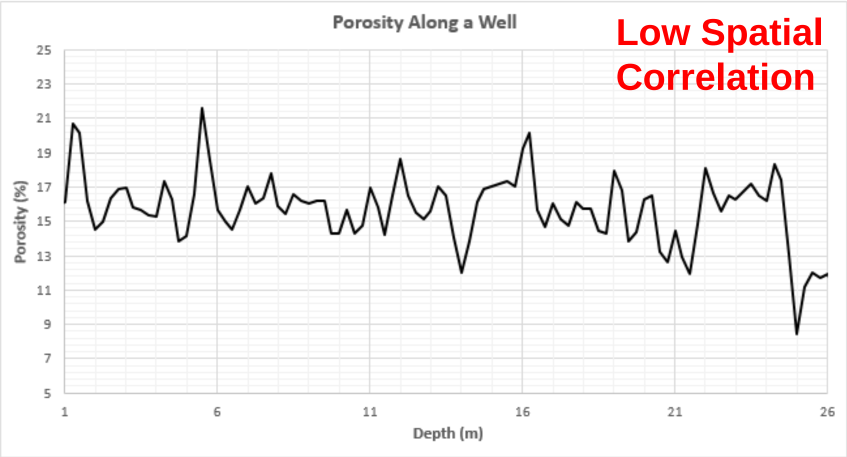


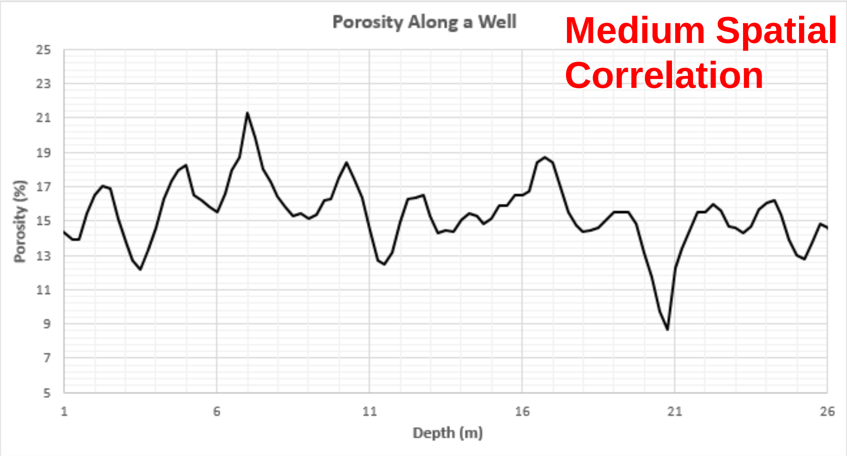


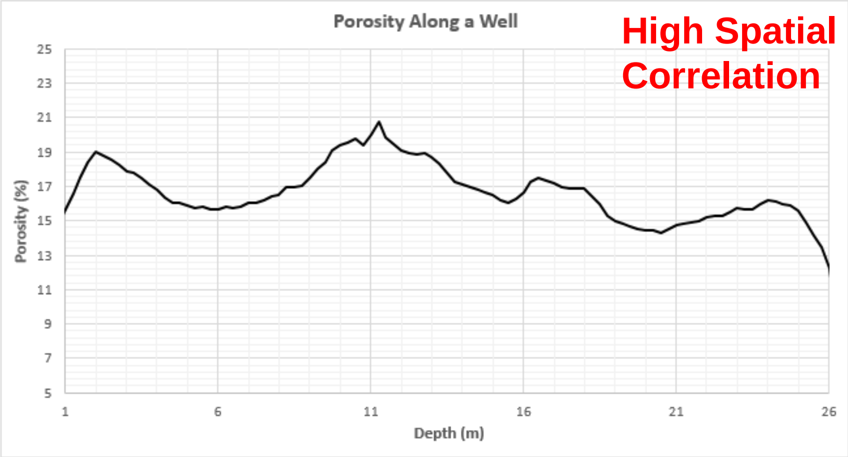


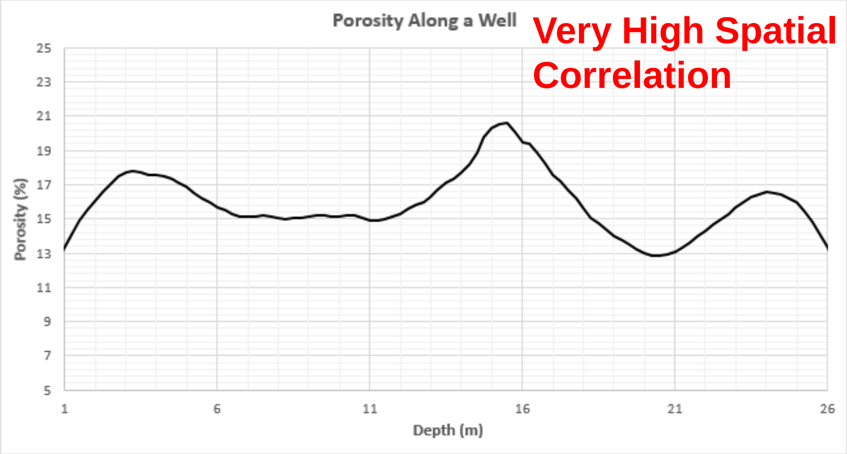
Spatial Continuity Definition

- No spatial continuity – random values at each location in space regardless of separation distance. Imposed variable level of correlation.



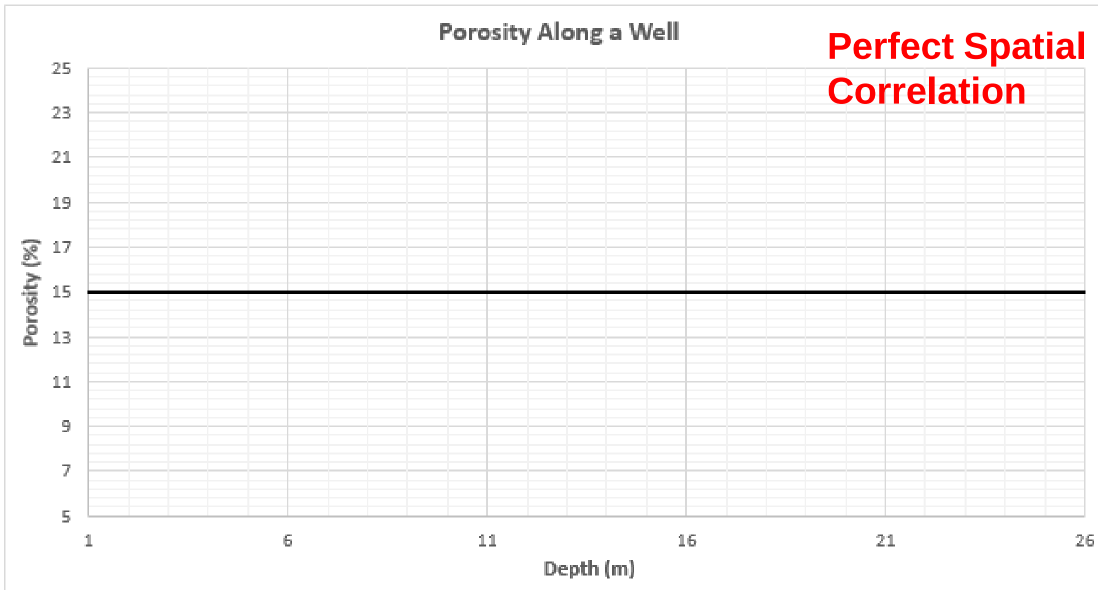


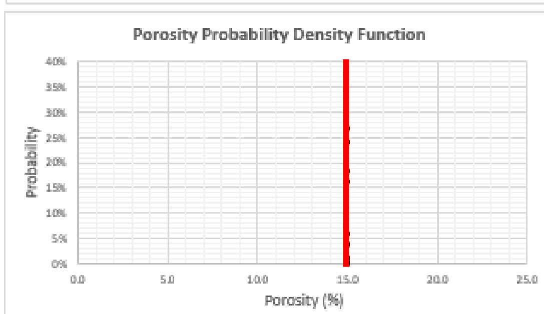


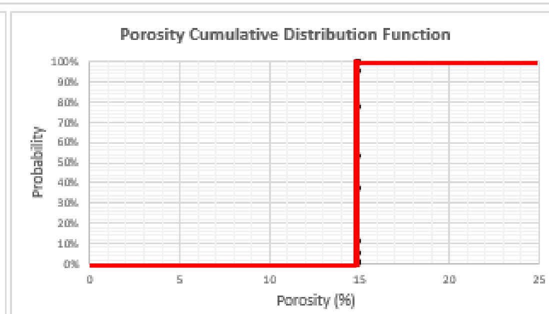


Spatial Continuity Definition

- Homogenous phenomenon have perfect spatial continuity, since all values as the same (or very similar).





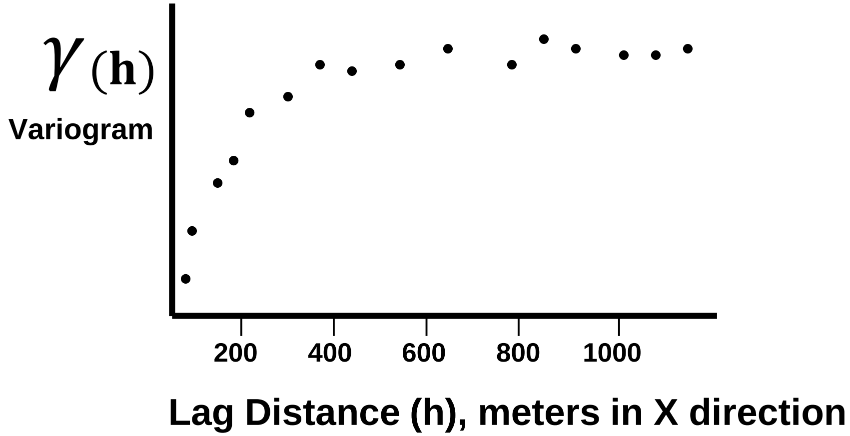


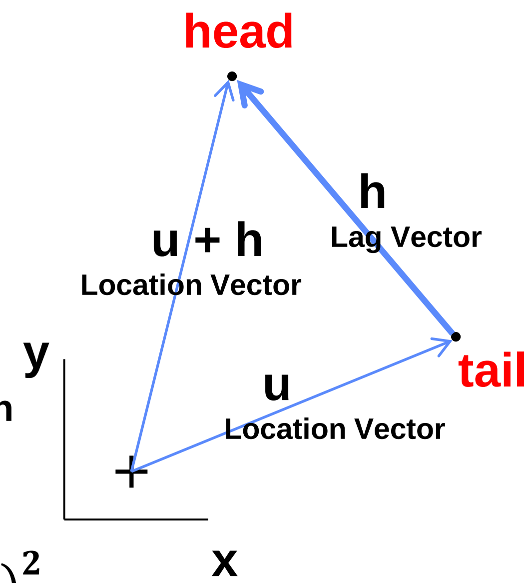
Measuring Spatial Continuity

We need a statistic to quantify spatial continuity!

The Semivariogram:

- Function of difference over distance.



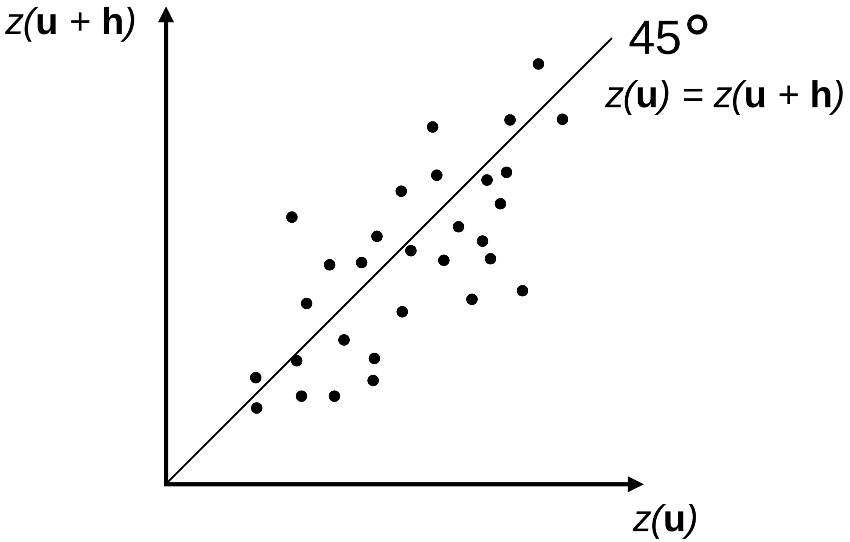


- The equation:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{\alpha=1}^{N(\mathbf{h})}\left(\overset{\text{tail}}{z(\mathbf{u}_\alpha)} - \overset{\text{head}}{z(\mathbf{u}_\alpha+\mathbf{h})}\right)^2$$

One half the average squared difference over lag distance, $\mathbf{h}$, over all possible pairs of data, $N(\mathbf{h})$.

“h” Scatterplot



- The variogram calculated for lag distance, $\mathbf{h}$, corresponds to the expected value of squared difference:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left(z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})\right)^2$$

- Calculate for a suite of lag distances to obtain a continuous function.

Variogram Definition

- **Variogram** – a measure of dissimilarity vs. distance. Calculated as ½ the average squared difference of values separated by a lag vector.

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \left(z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h}) \right)^2$$

  - The precise term is semivariogram (variogram if you remove the 1/2), but in practice the term variogram is used.
  - The ½ is used so that the covariance function and variogram may be related directly:

$$C_x(h) = \sigma_x^2 - \gamma_x(\mathbf{h})$$

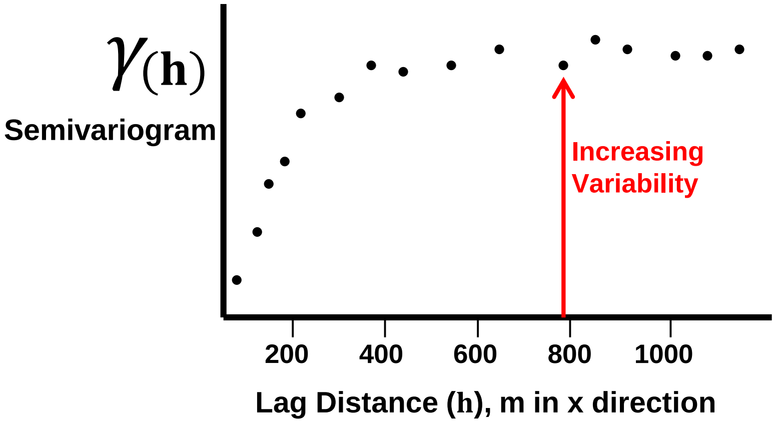
  - Note the correlogram is related to the covariance function as:

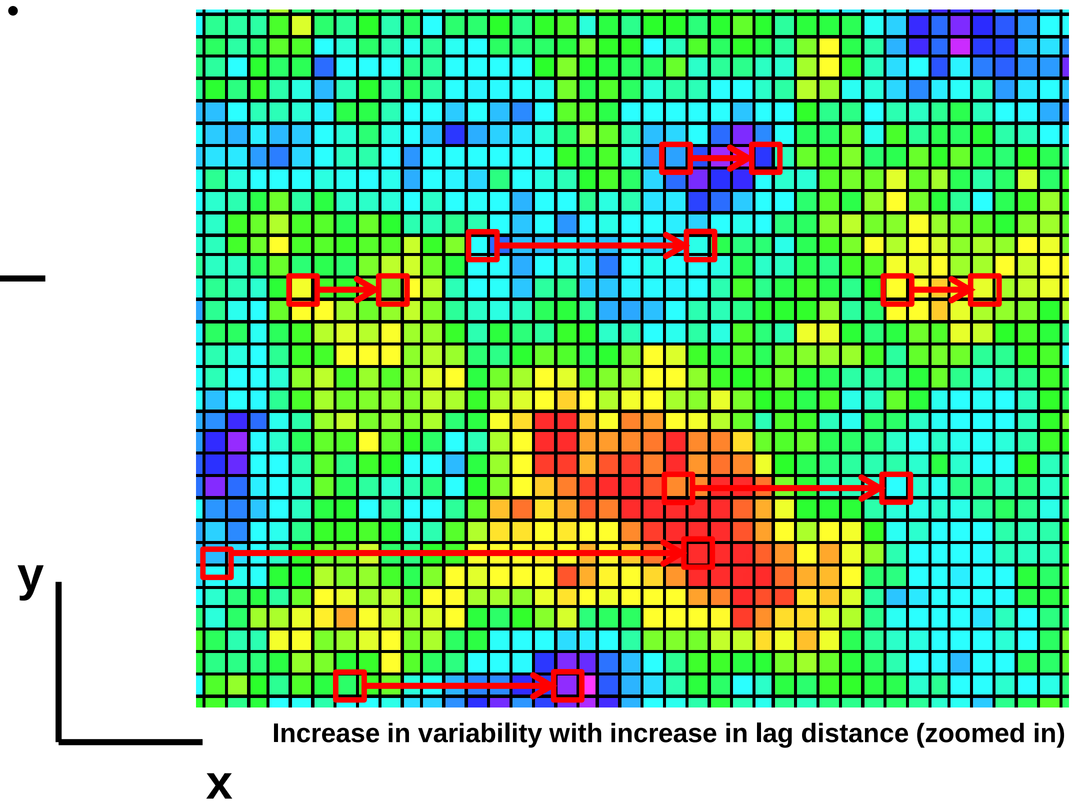
$$\rho_x(h) = \frac{C_x(h)}{\sigma_x^2}$$, h-scatter plot correlation vs. lag distance

Variogram Interpretation

Observation #1

- As distance increases, variability increase (in general).



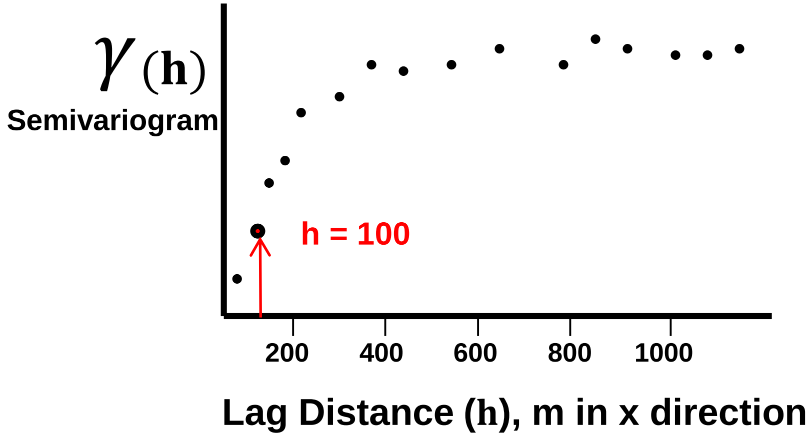


Increase in variability with increase in lag distance (zoomed in)

Variogram Interpretation

Observation #2

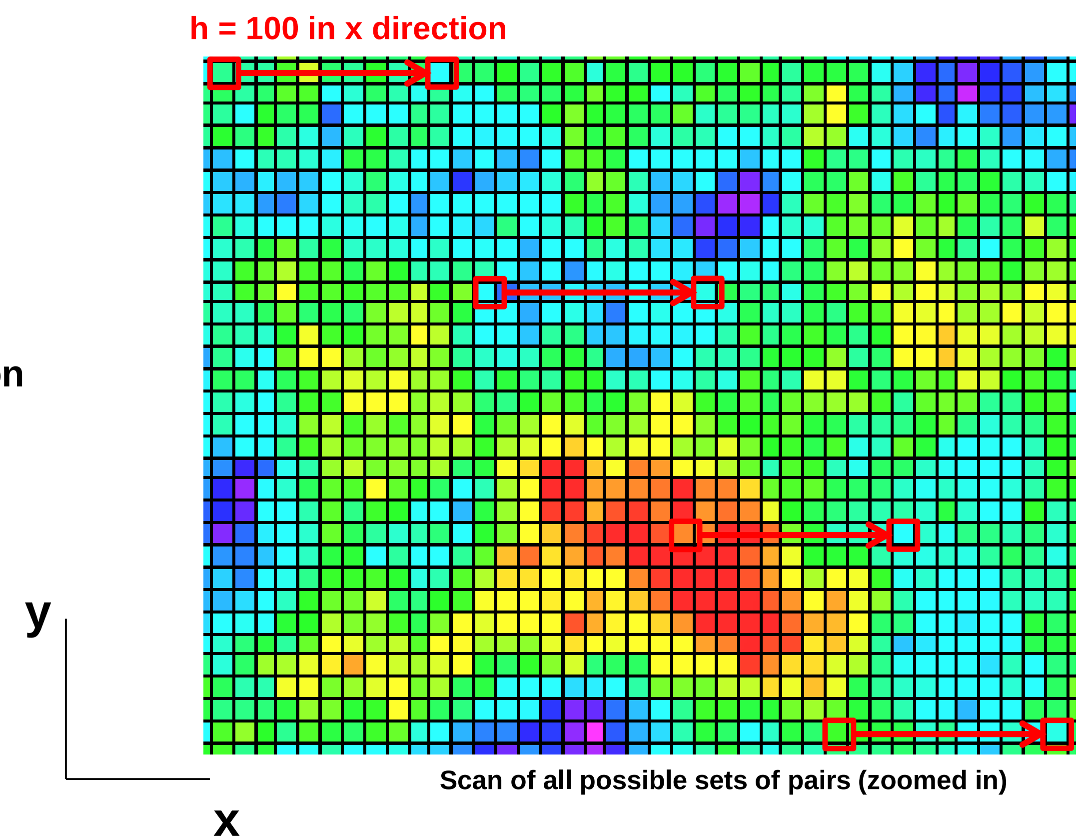
- Calculated with over all possible pairs separated by lag vector, $\mathbf{h}$.



- **The variogram:**

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{\alpha=1}^{N(\mathbf{h})}\left(z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha+\mathbf{h})\right)^2$$

Given the number of pairs available N(h).



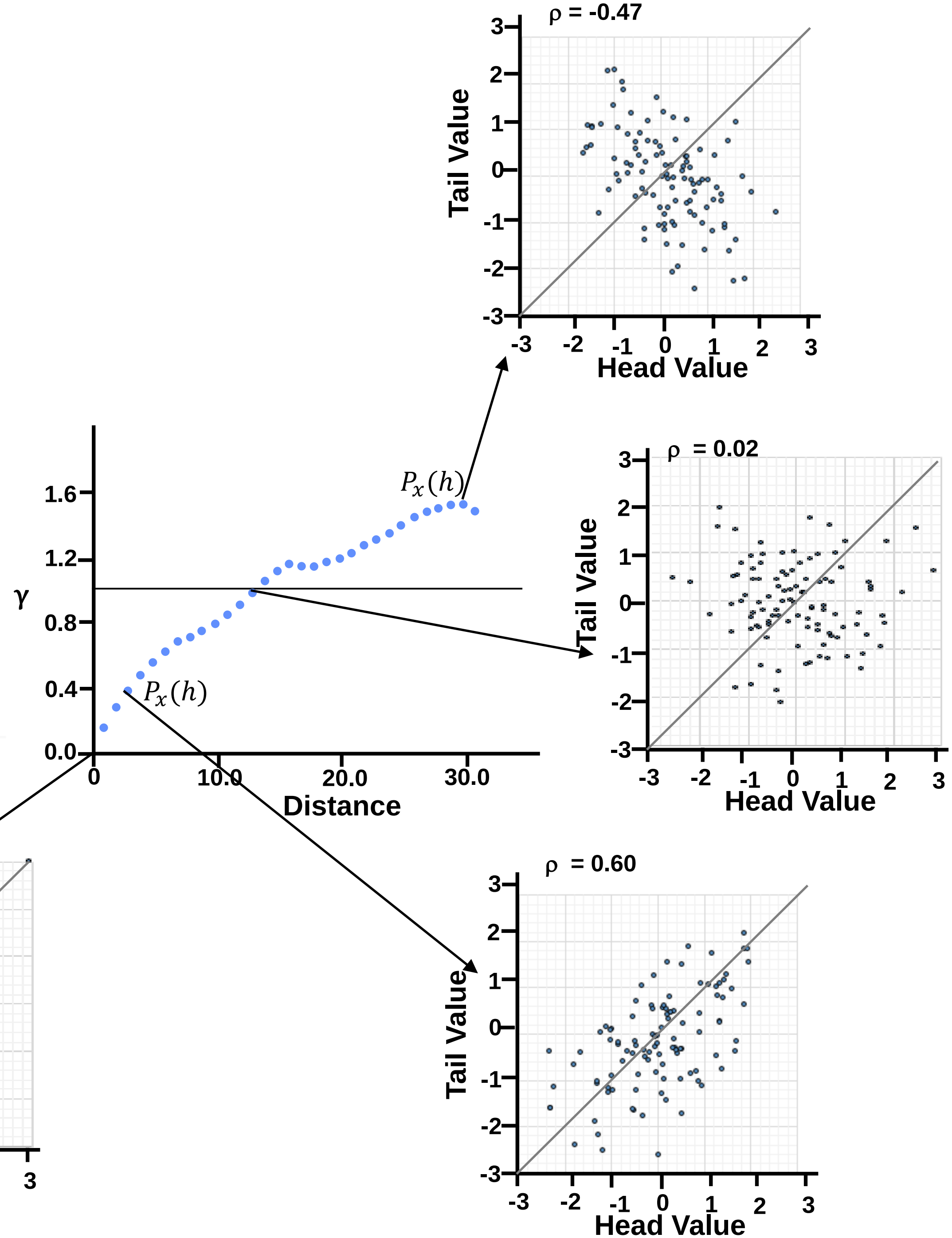
Scan of all possible sets of pairs (zoomed in)

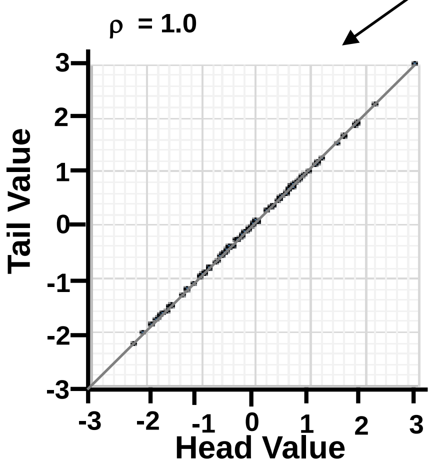
Variogram Interpretation

Observation #3

Need to plot the sill to know the degree of correlation.

Another illustration of h-scatter plot correlation vs. lag distance.

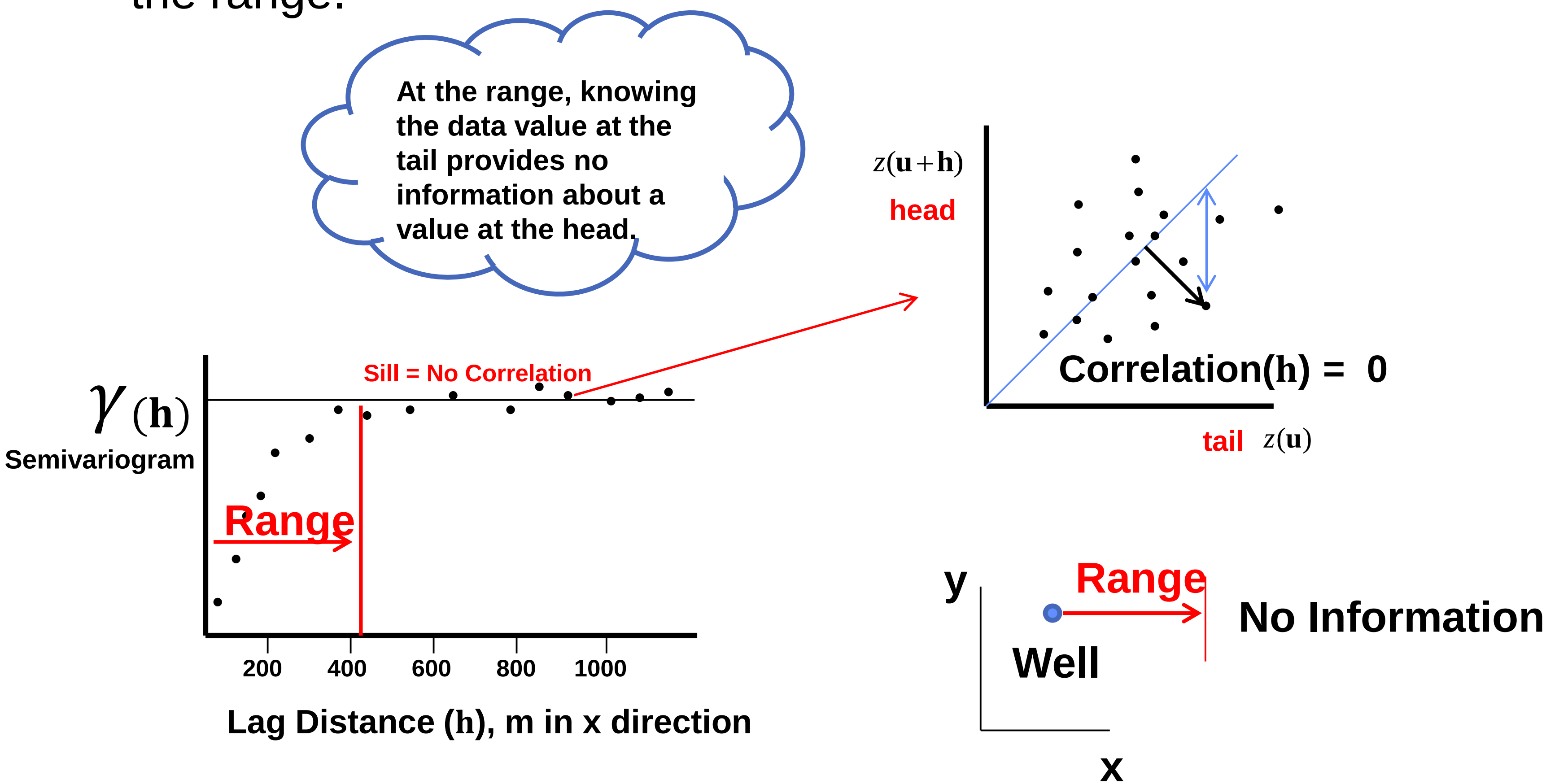




Variogram Interpretation

Observation #4

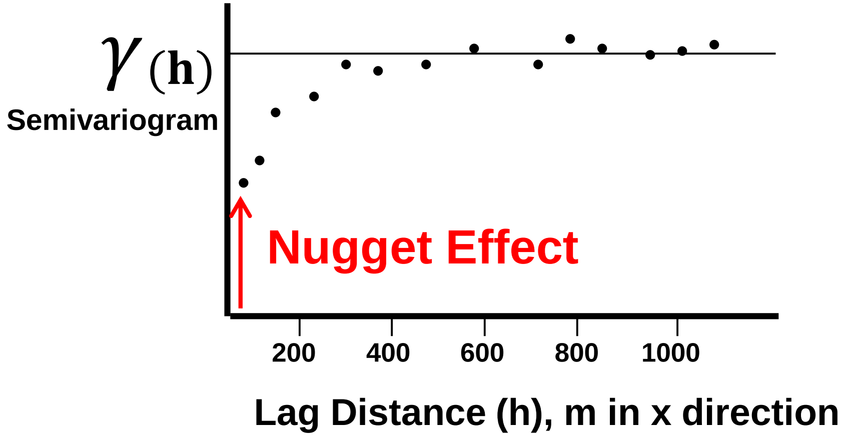
- The lag distance at which the variogram reaches the sill is know as the range.



Variogram Interpretation

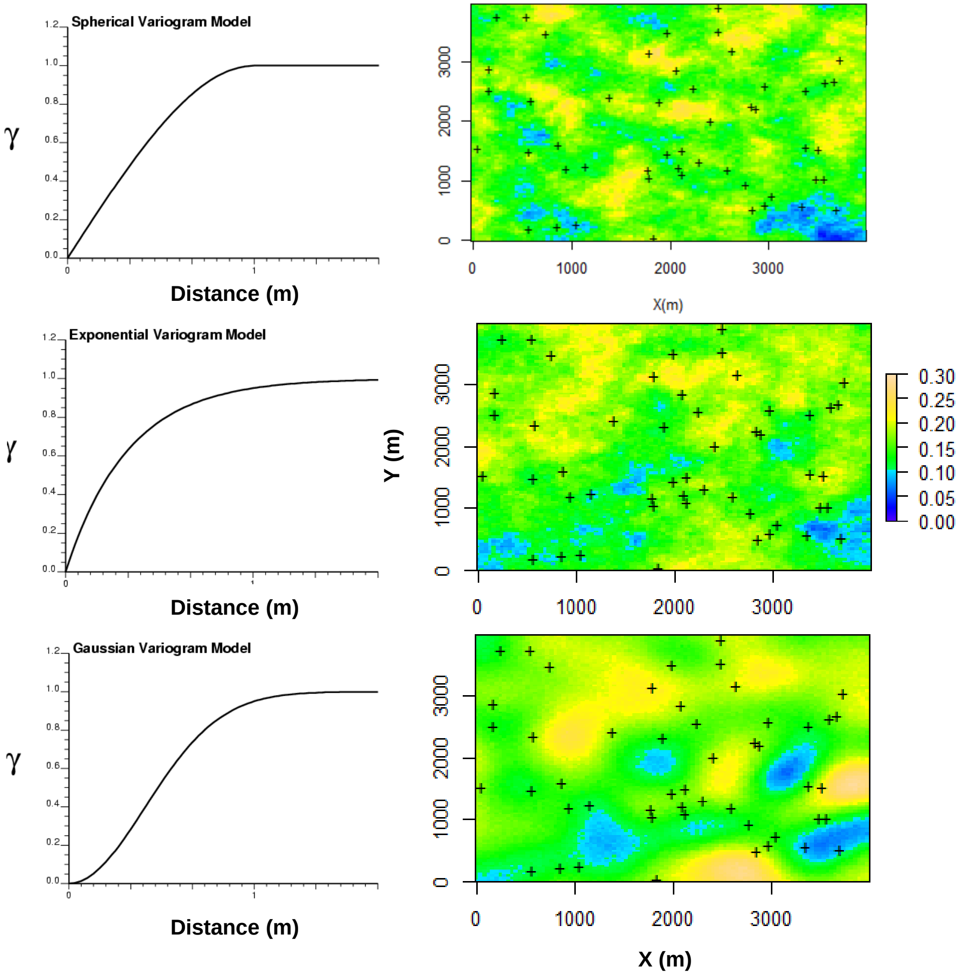
Observation #5

- Sometimes there is a discontinuity in the variogram at distances less than the minimum data spacing. This is known as nugget effect.
  - As a ratio of nugget / sill, is known as relative nugget effect (%)
  - Modeled as a no correlation structure that at lags, $h > \varepsilon$, an infinitesimal distance
  - Measurement error, mixing populations cause apparent nugget effect



Spatial Variability

- The three maps are remarkably similar: all three have the same 50 data, same histograms and same range of correlation, and yet their spatial variability/continuity is quite different
- The spatial variability/continuity depends on the detailed distribution of the petrophysical attribute (ϕ,K)
- The charts on the left are "variograms"
- Our map-making efforts should consider the spatial variability/continuity of the variable we are mapping:
  - Variability
  - Uncertainty

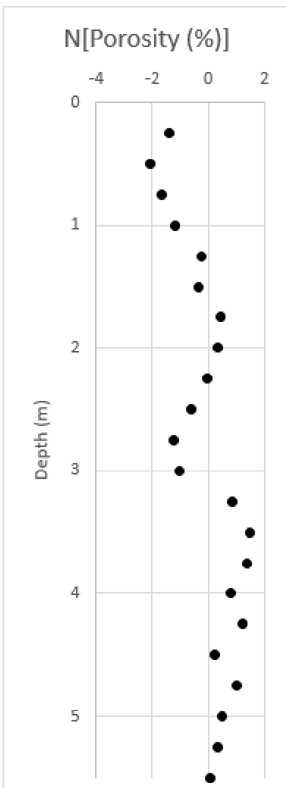


Porosity Realizations with 3 isotropic variograms.

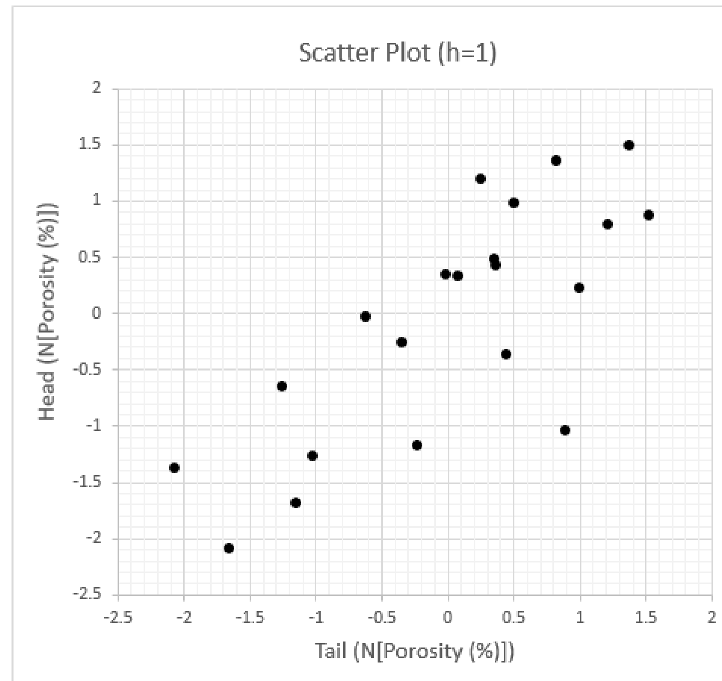
Variogram Calculation

- Consider data values separated by *lag* vectors (the h values)
- Here are two examples of a lag vector equal to the data spacing and then twice the data spacing:

Depth	N[Porosity]
0.25	-1.37
0.5	-2.08
0.75	-1.67
1	-1.16
1.25	-0.24
1.5	-0.36
1.75	0.44
2	0.36
2.25	-0.02
2.5	-0.63
2.75	-1.26
3	-1.03
3.25	0.88
3.5	1.51
3.75	1.37
4	0.81
4.25	1.21
4.5	0.24
4.75	0.99
5	0.49
5.25	0.34
5.5	0.07



Squared Difference	-1.37	
0.50	-2.08	-1.37
0.17	-1.67	-2.08
0.26	-1.16	-1.67
0.85	-0.24	-1.16
0.01	-0.36	-0.24
0.64	0.44	-0.36
0.01	0.36	0.44
0.14	-0.02	0.36
0.37	-0.63	-0.02
0.40	-1.26	-0.63
0.05	-1.03	-1.26
3.65	0.88	-1.03
0.40	1.51	0.88
0.02	1.37	1.51
0.31	0.81	1.37
0.16	1.21	0.81
0.94	0.24	1.21
0.56	0.99	0.24
0.25	0.49	0.99
0.02	0.34	0.49
0.07	0.07	0.34
Average / 2		0.07
0.23		

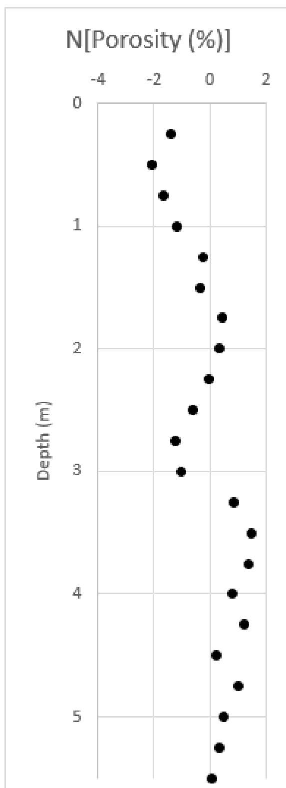


Correlation Coefficient (h=1)	0.77
Variogram (h=1)	0.23

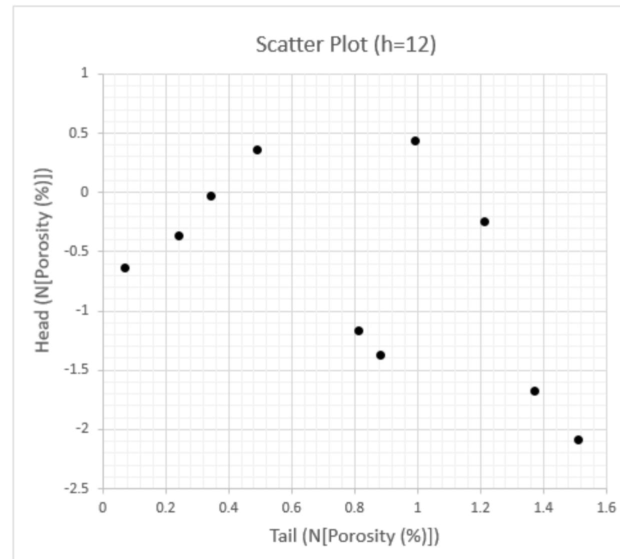
Variogram Calculation

- Consider data values separated by *lag* vectors (the h values)
- Here are two examples of a lag vector equal to the data spacing and then twice the data spacing:

Depth	N[Porosity]
0.25	-1.37
0.5	-2.08
0.75	-1.67
1	-1.16
1.25	-0.24
1.5	-0.36
1.75	0.44
2	0.36
2.25	-0.02
2.5	-0.63
2.75	-1.26
3	-1.03
3.25	0.88
3.5	1.51
3.75	1.37
4	0.81
4.25	1.21
4.5	0.24
4.75	0.99
5	0.49
5.25	0.34
5.5	0.07



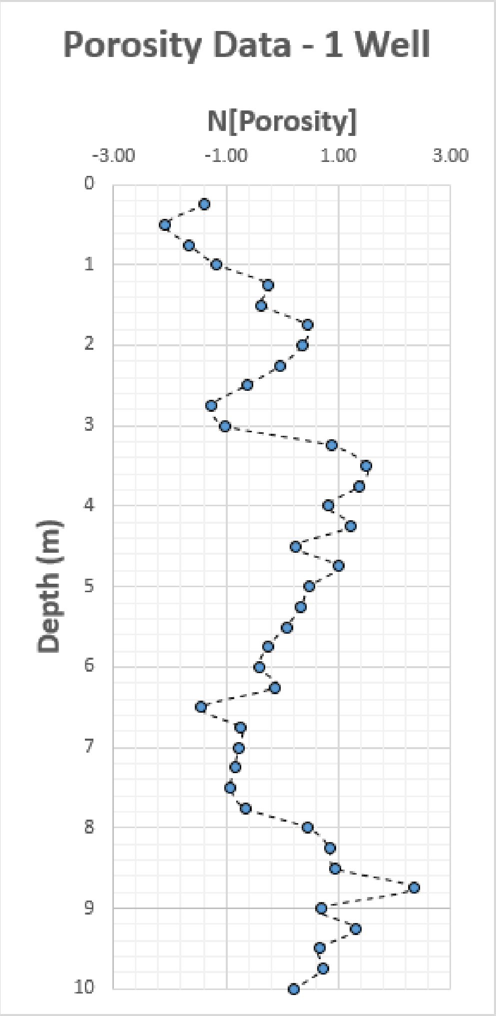
Squared Difference		
	-1.37	
	-2.08	
	-1.67	
	-1.16	
	-0.24	
	-0.36	
	0.44	
	0.36	
	-0.02	
	-0.63	
	-1.26	
	-1.03	
5.06	0.88	-1.37
12.89	1.51	-2.08
9.24	1.37	-1.67
3.88	0.81	-1.16
2.10	1.21	-0.24
0.36	0.24	-0.36
0.30	0.99	0.44
0.02	0.49	0.36
0.13	0.34	-0.02
0.49	0.07	-0.63
Average / 2		-1.26
1.72		-1.03
		0.88
		1.51
		1.37
		0.81
		1.21
		0.24
		0.99
		0.49
		0.34
		0.07



Correlation Coefficient (h=12)	-0.54
Variogram (h=12)	1.72

Variogram Calculation

- Pick a lag distance and calculate the variogram for that one lag distance.

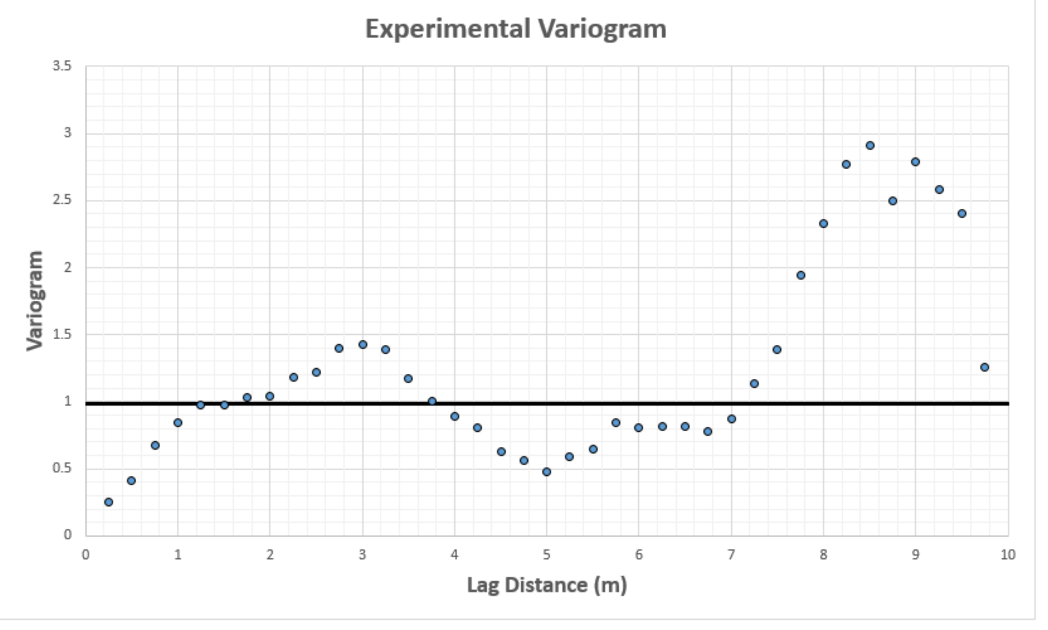


Depth	N[Porosity]
0.25	-1.37
0.5	-2.08
0.75	-1.67
1	-1.16
1.25	-0.24
1.5	-0.36
1.75	0.44
2	0.36
2.25	-0.02
2.5	-0.63
2.75	-1.26
3	-1.03
3.25	0.88
3.5	1.51
3.75	1.37
4	0.81
4.25	1.21
4.5	0.24
4.75	0.99
5	0.49
5.25	0.34
5.5	0.07
5.75	-0.26
6	-0.41
6.25	-0.14
6.5	-1.44
6.75	-0.75
7	-0.78
7.25	-0.85
7.5	-0.92
7.75	-0.66
8	0.47
8.25	0.85
8.5	0.95
8.75	2.35
9	0.69
9.25	1.31
9.5	0.66
9.75	0.72
10	0.21

Variogram Calculation

- Pick a lag distance and calculate the variogram for that one lag distance.

Here's all of them:

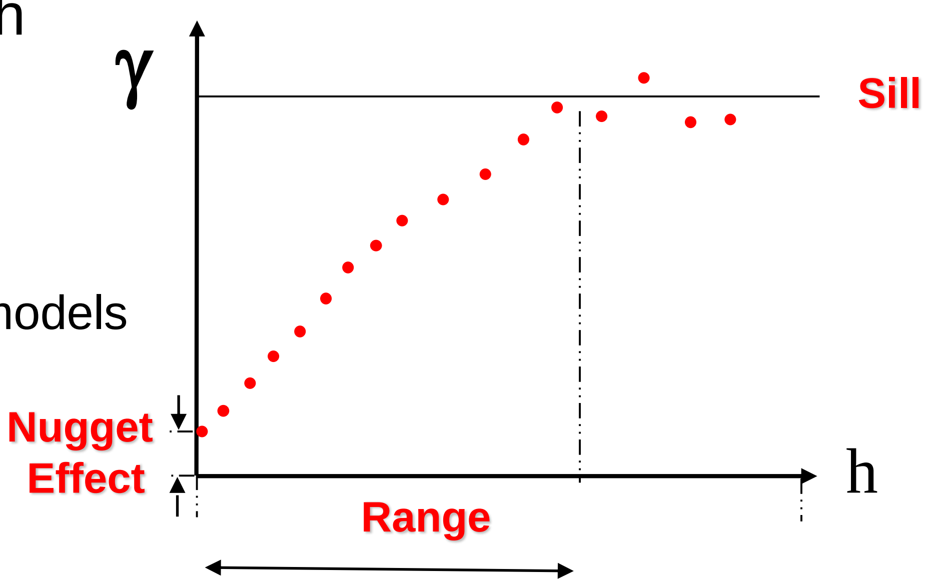


- Interpretation:
  - Cyclicity in the variogram (more later)
  - Range of about 1.25 m, no nugget effect
  - Unreliable as we go beyond ½ data extent

Depth	N[Porosity]
0.25	-1.37
0.5	-2.08
0.75	-1.67
1	-1.16
1.25	-0.24
1.5	-0.36
1.75	0.44
2	0.36
2.25	-0.02
2.5	-0.63
2.75	-1.26
3	-1.03
3.25	0.88
3.5	1.51
3.75	1.37
4	0.81
4.25	1.21
4.5	0.24
4.75	0.99
5	0.49
5.25	0.34
5.5	0.07
5.75	-0.26
6	-0.41
6.25	-0.14
6.5	-1.44
6.75	-0.75
7	-0.78
7.25	-0.85
7.5	-0.92
7.75	-0.66
8	0.47
8.25	0.85
8.5	0.95
8.75	2.35
9	0.69
9.25	1.31
9.5	0.66
9.75	0.72
10	0.21

Variogram Components Definition

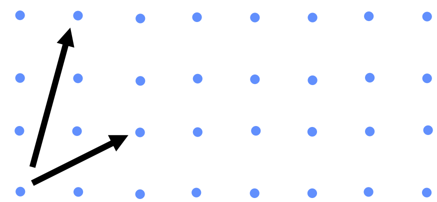
- **Nugget Effect** – discontinuity in the variogram at distances less than the minimum data spacing
  - As a ratio of nugget / sill, is known as relative nugget effect (%)
  - Measurement error, mixing populations cause apparent nugget effect
- **Sill** – the sample variance
  - Interpret spatial correlation relative to the sill, level of no correlation
- **Range** – lag distance to reach the sill
  - Up to that distance you have information
  - parameterization of variogram models



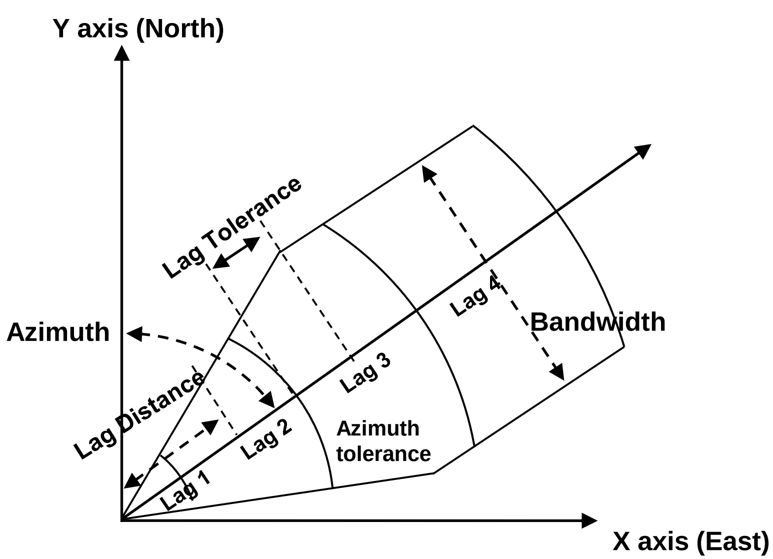
Calculating Experimental Variograms

How do we get pairs separated by lag vector?

- Regular spaced data:
  - Specify as offsets of grid units
  - Fast calculation
  - Diagonal directions are awkward

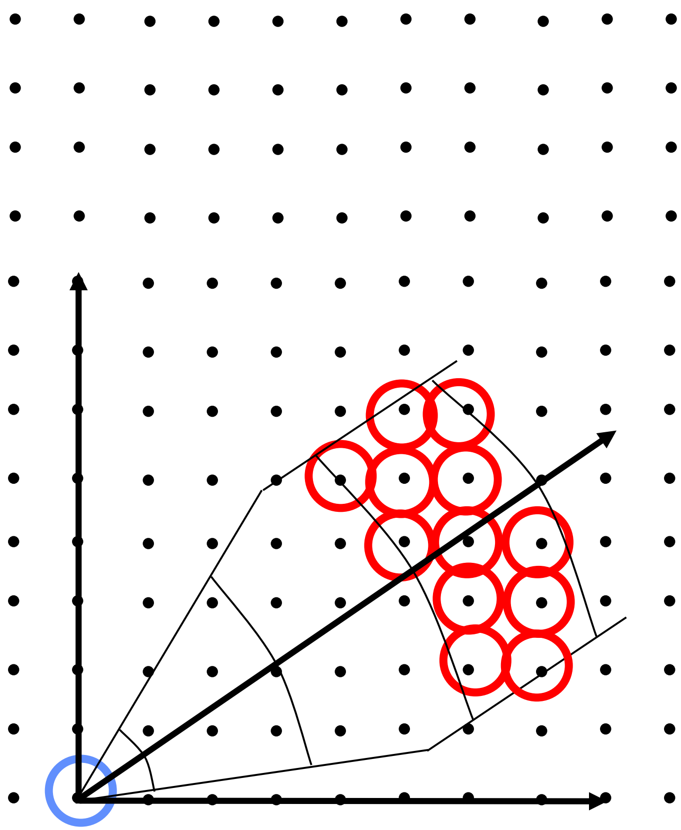


- Irregular spaced data:
  - Nominal distance for each lag
  - Distance tolerance
  - Azimuth direction
  - Azimuth tolerance
  - Dip direction
  - Dip tolerance
  - Bandwidth (maximum deviation) in originally horizontal plane
  - Bandwidth in originally vertical plane



Calculating Experimental Variograms

Example: Starting With One Lag (i.e. #4)

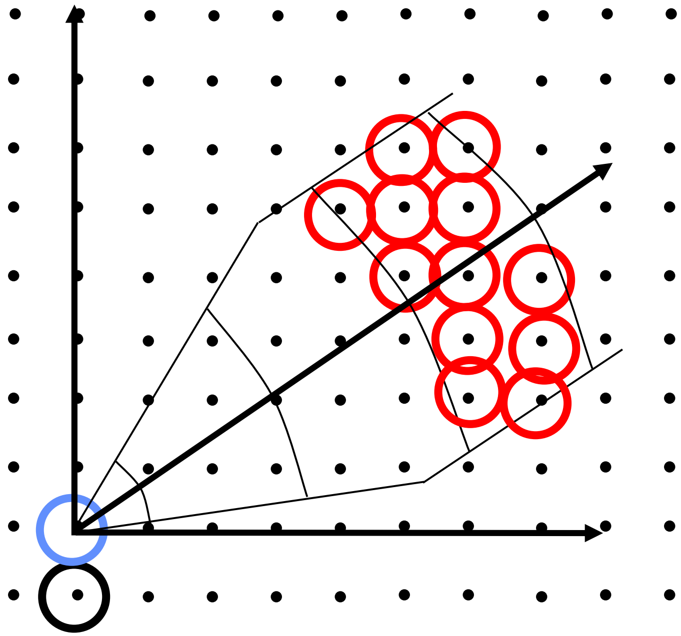


$$2\gamma(\boldsymbol{h}) = \frac{1}{N(\boldsymbol{h})} \sum_{N(\boldsymbol{h})} [z(\boldsymbol{u}) - z(\boldsymbol{u}+\boldsymbol{h})]^2$$

Start at a node, and compare value to all nodes which fall in the lag and angle tolerance.

…

Calculating Experimental Variograms



$$2\gamma(h) = \frac{1}{N(h)} \sum_{N(h)} [z(u) - z(u+h)]^2$$

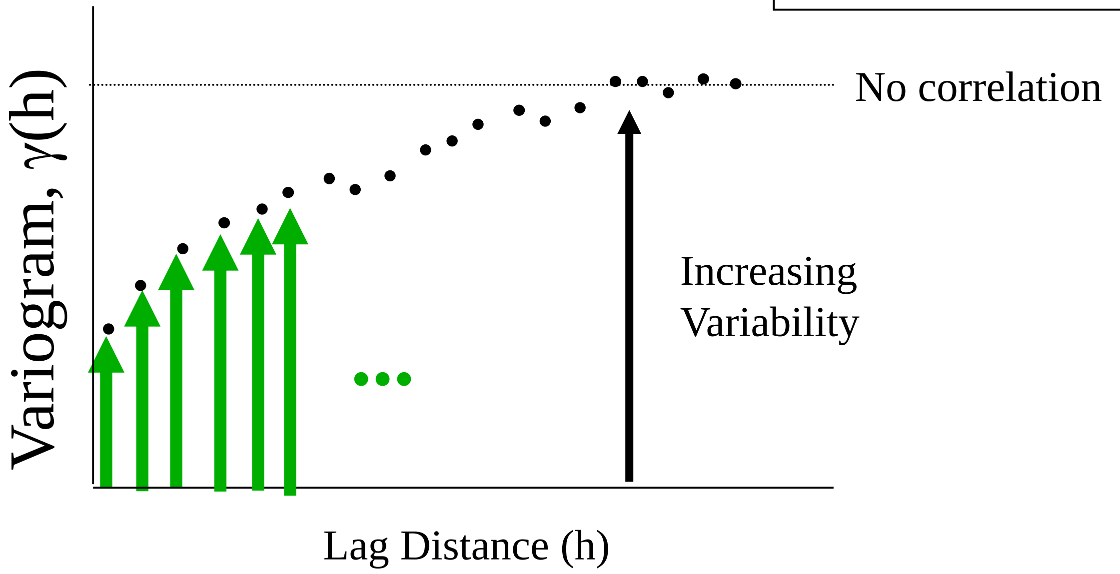
Move to next node.

...

Calculating Experimental Variograms

Now Repeat for All Nodes

And Repeat for All Lags

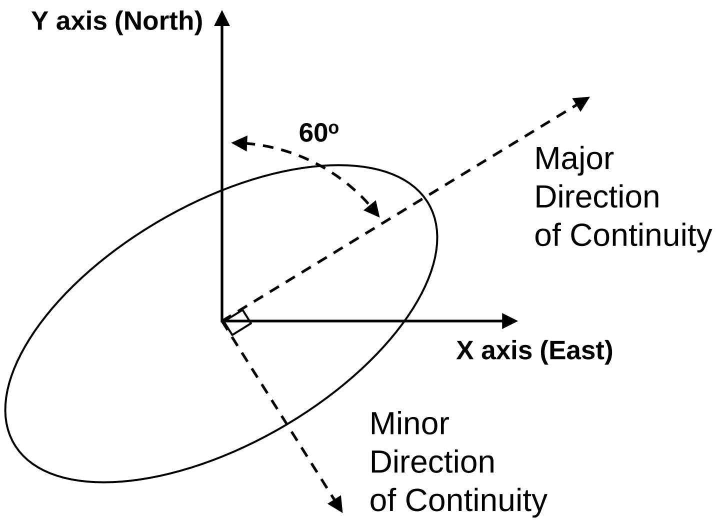


Some Options

- Data transformation:
  - Commonly transform a continuous variable to a Gaussian or normal distribution because that is the most common simulation algorithm
  - Commonly transform a categorical variable to a series of indicator variables
- Coordinate transformation:
  - Vertical variograms are not usually sensitive to coordinate transformation, but
  - Horizontal variograms are often very sensitive to stratigraphic deformation or geologic variability
- Should calculate the variogram on the variable being modeled with transforms (data and coordinates)
- Calculate the traditional variogram (not alternatives such as correlograms and relative variograms) since that is what is going to be used later

Choosing the Directions

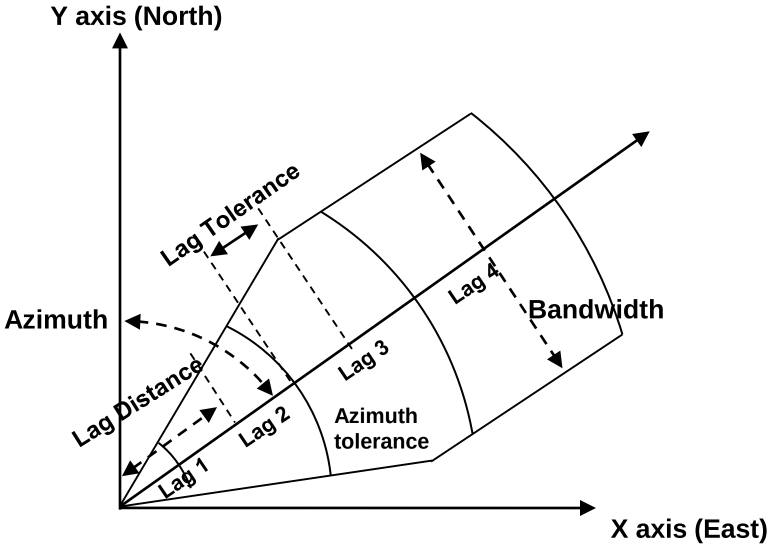
- Look at maps – sections, plan views, …
- Choose directions based on orientation of reservoir
- Azimuth angles are entered in degrees clockwise from north
- Review multiple directions before choosing a set of 3 perpendicular directions
  - Omnidirectional: all directions taken together → often yields the most well-behaved variograms.
  - Major horizontal direction & two perpendicular to major direction
  - All anisotropy in geostatistics is geometric – three mutually orthogonal directions



Choosing the Lag Distances and Tolerances

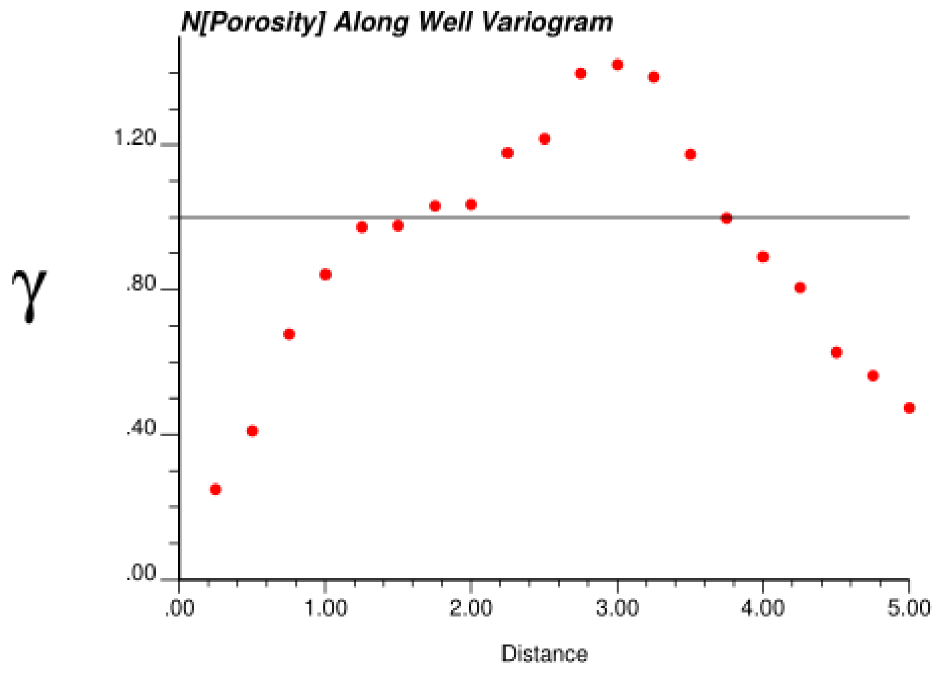
Rules of Thumb on Variogram Calculation Parameters:

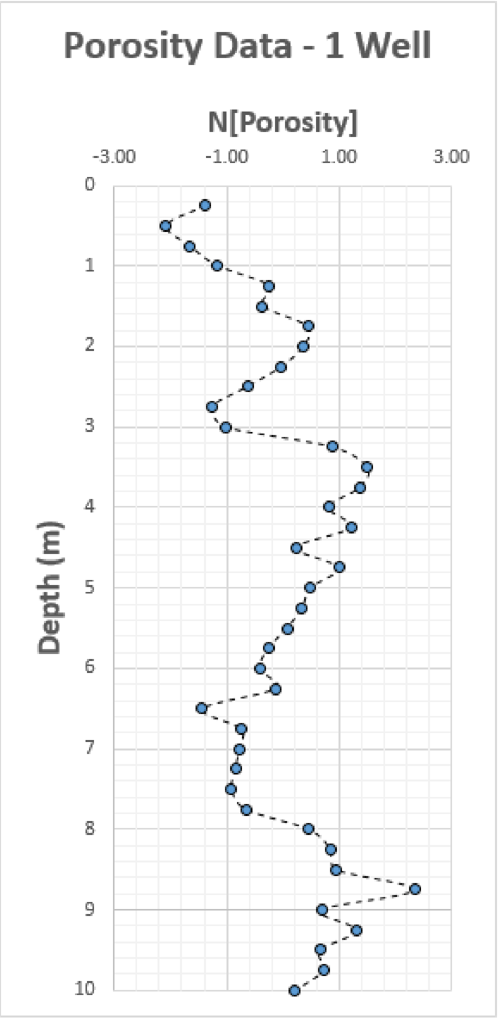
- Lag separation distance should coincide with data spacing
- Lag tolerance typically chosen to be ½ lag separation distance
  - in cases of erratic variograms, may choose to overlap calculations so lag tolerance > ½ lag separation.
- The variogram is only valid for a distance one half of the field size: start leaving data out of calculations with larger distances
- The lag distances, tolerances and number of lags should be changed for each lag distance



Variogram Calculation

First Walkthrough Together with GSLIB



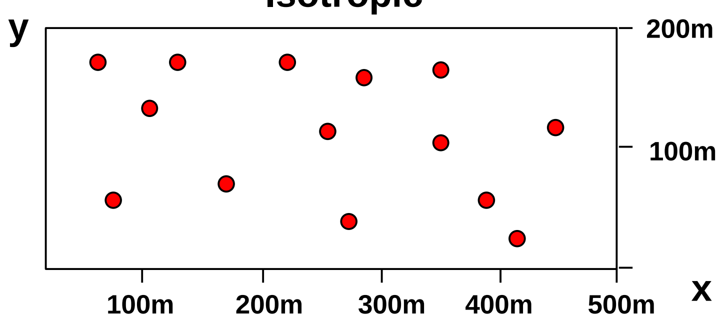


How would you interpret this result?

Choosing the Lag Distances and Tolerances

Let's pick some variogram calculation parameters:

Isotropic

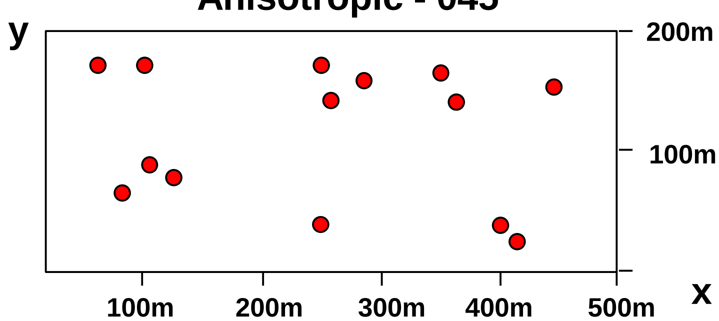


lag size ______, lag tolerance ______, number of lags ___

horizontal bandwidth ___

azimuth ____, azimuth tolerance ______

Anisotropic - 045

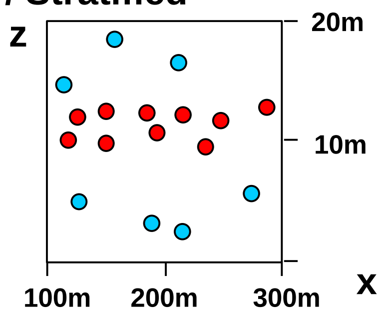


lag size ______, lag tolerance ______, number of lags ___

horizontal bandwidth ___

azimuth ____, azimuth tolerance ______

Isotropic / Stratified



lag size ______, lag tolerance ______, number of lags ___

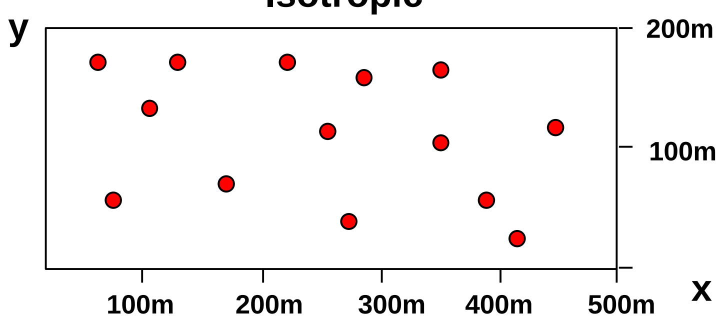
vertical bandwidth ___

azimuth ____, azimuth tolerance ______

Choosing the Lag Distances and Tolerances

Let’s pick some variogram calculation parameters in groups (estimate):

Isotropic



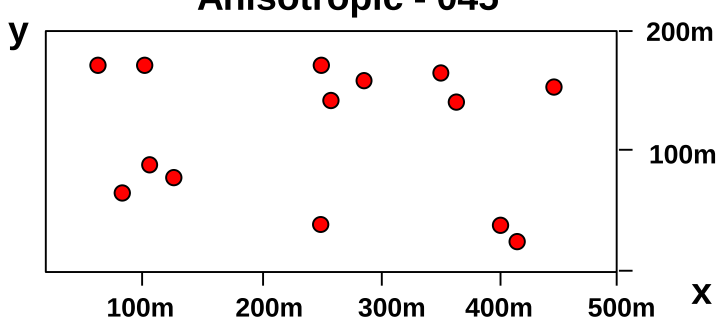
lag size **50**, lag tolerance **25**, number of lags **5**

horizontal bandwidth **∞**

azimuth **0**, azimuth tolerance **90**

Any

Anisotropic - 045

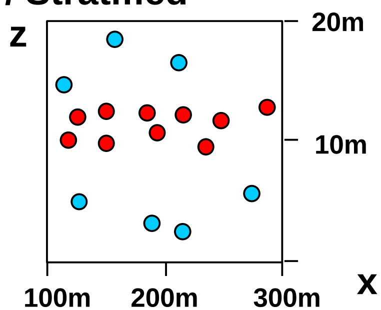


lag size **20**, lag tolerance **10**, number of lags **12**

horizontal bandwidth **100**

azimuth **45**, azimuth tolerance **22.5**

Isotropic / Stratified



lag size **15**, lag tolerance **7.5**, number of lags **10**

vertical bandwidth **5**

azimuth **90**, azimuth tolerance **22.5**

Along x

Spatial Modeling with Geostatistics

Spatial Data Analysis - Calculation

Lecture outline . . .

- **Stationarity**
- **Quantifying Spatial Continuity**

Prerequisites

Introduction

Probability Theory

Representative Sampling

Spatial Data Analysis

Spatial Estimation

Stochastic Simulation

Uncertainty Management

Machine Learning