Multivariate Modeling: Spatial Continuity

Lecture outline . . .

- Random Variables and Functions
- Stationarity
- Spatial Continuity
- Variogram Calculation

Introduction

Prerequisites

Probability

Multivariate Analysis

Spatial Estimation

Statistical Learning

Feature Selection

Multivariate Modeling

Conclusions

Other Resources

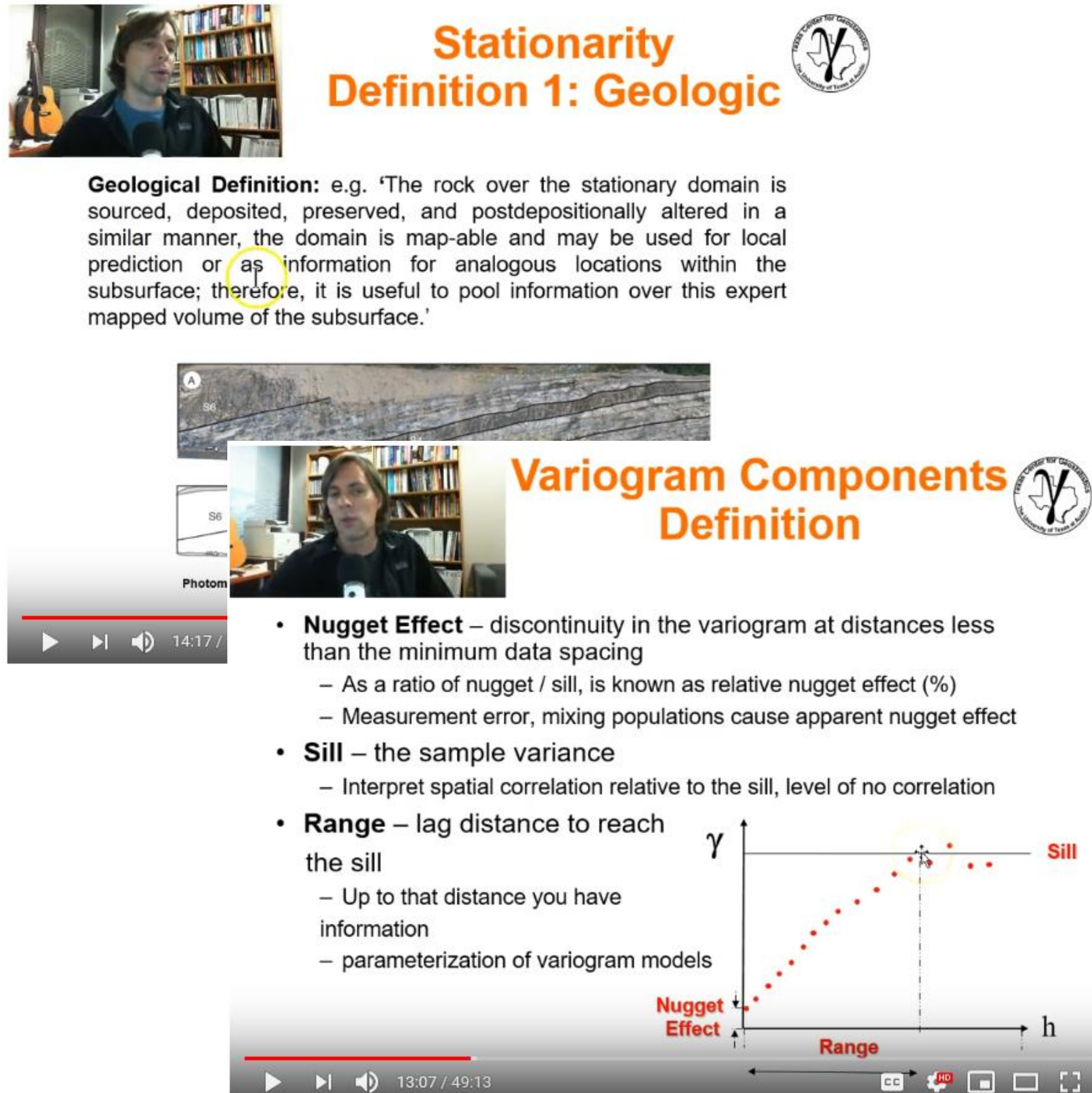
YouTube Lectures on:

Stationarity:
https://youtu.be/QwxQ9xuUHlU

Variogram Calculation:
https://youtu.be/mzPLicovE7Q

Variogram Interpretation:
https://youtu.be/Li-Xzlu7hvs

Variogram Modeling:
https://youtu.be/-Bi63Y3u6TU



Multivariate Modeling: Spatial Continuity

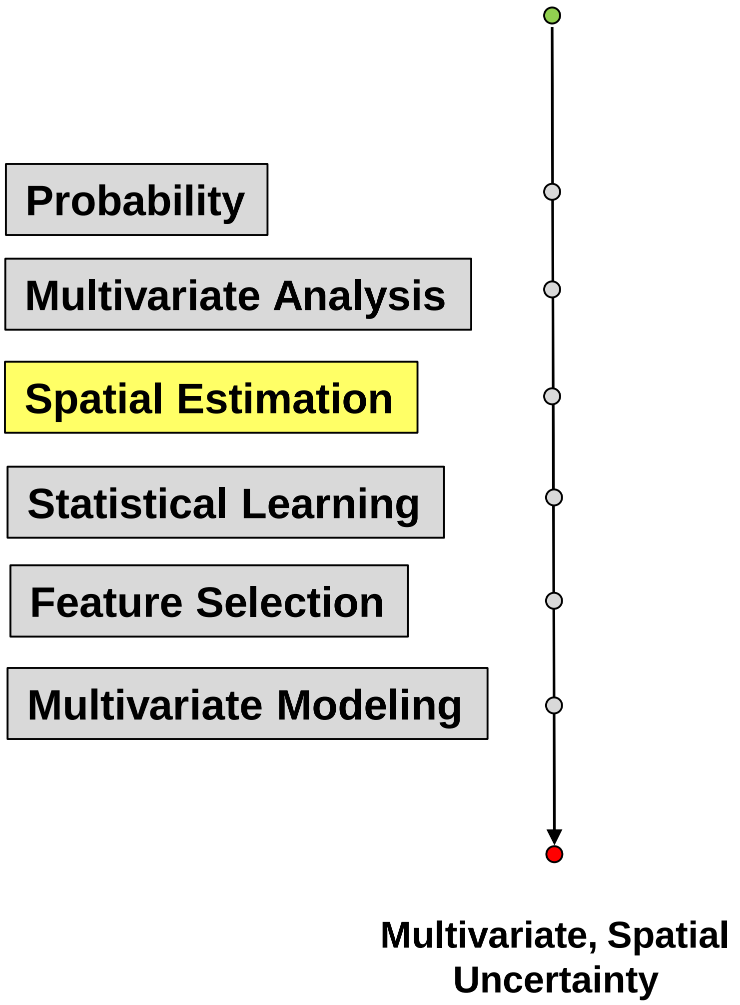
Lecture outline . . .

- **Random Variables and Functions**
- **Stationarity**

What Will You Learn?

Why Cover Spatial Continuity?

- We will use spatial continuity to calculate the spatial prior probabilities from well data.
- The heart of any spatial model.
- Note we skip the topic of variogram modeling and accounting for scale.



Now We Begin Spatial / Geostatistics

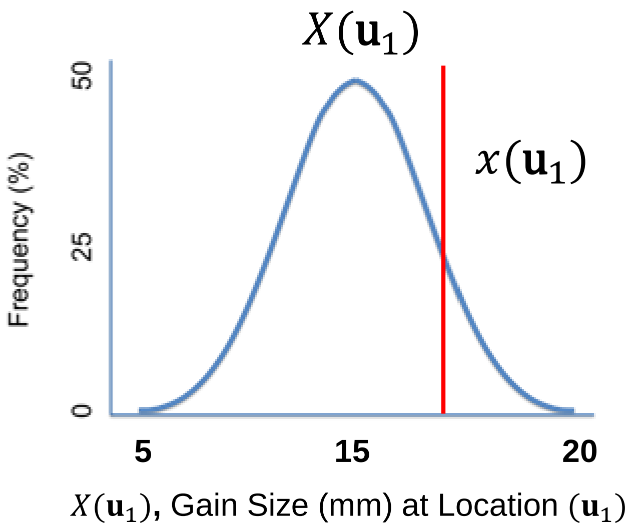
- **We Build on the past lectures:**
  - probability theory
  - univariate and bivariate statistics

- To begin let's provide a concise and practical definition for random variable.

Random Variable (RV) Definition

Random Variable

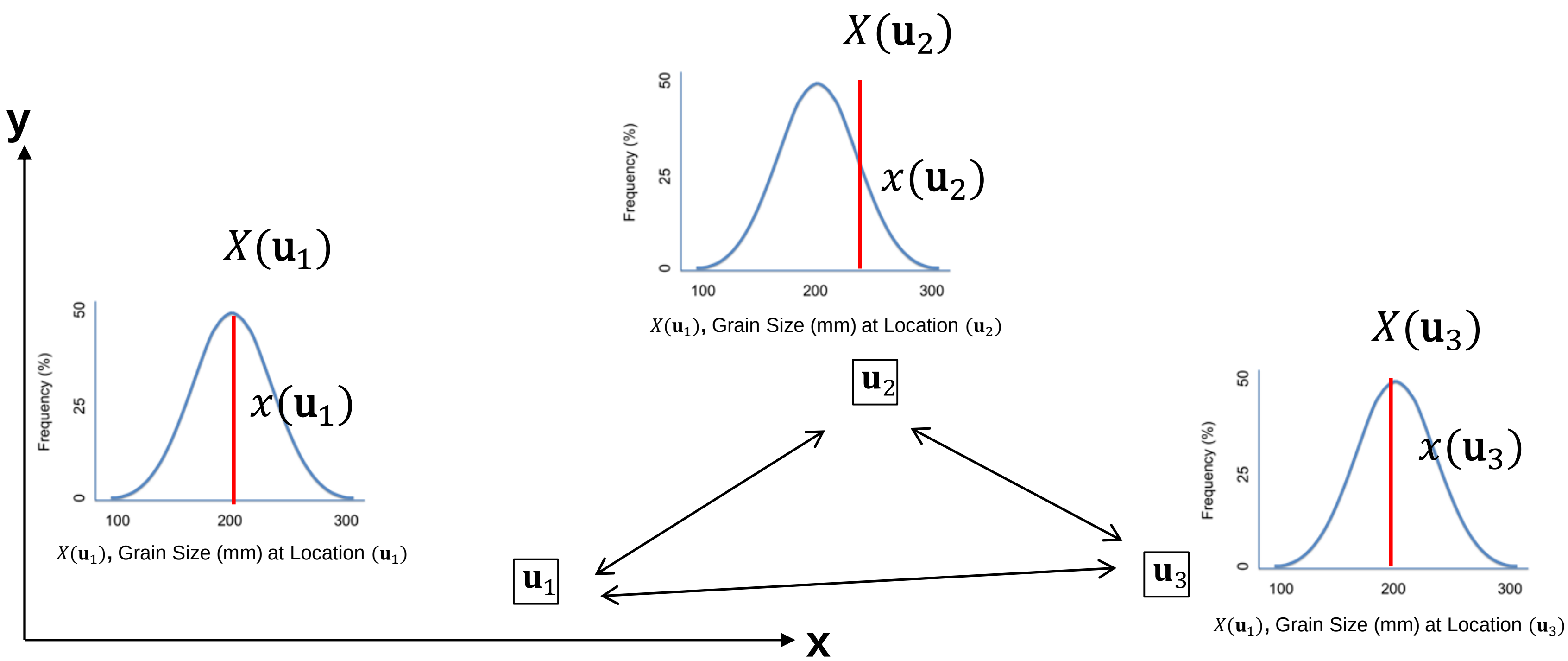
- we do not know the value at a location / time, it can take on a range of possible values, fully described with a PDF.
- represented as an upper case variable, e.g. X, while possible outcomes or data measures are represented with lower case, e.g. x.
- in spatial context common to use a location vector, $\mathbf{u}$, 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 $X(\mathbf{u}_1), X(\mathbf{u}_2), \ldots, X(\mathbf{u}_n)$

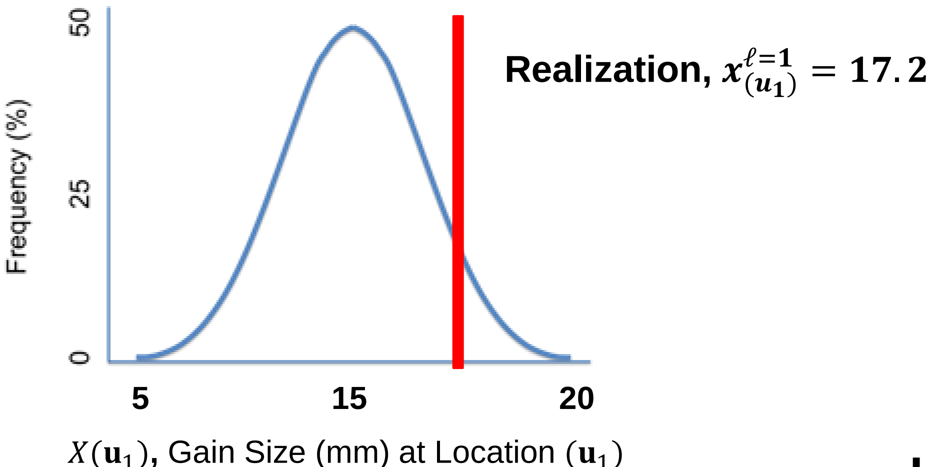


Realization Definition

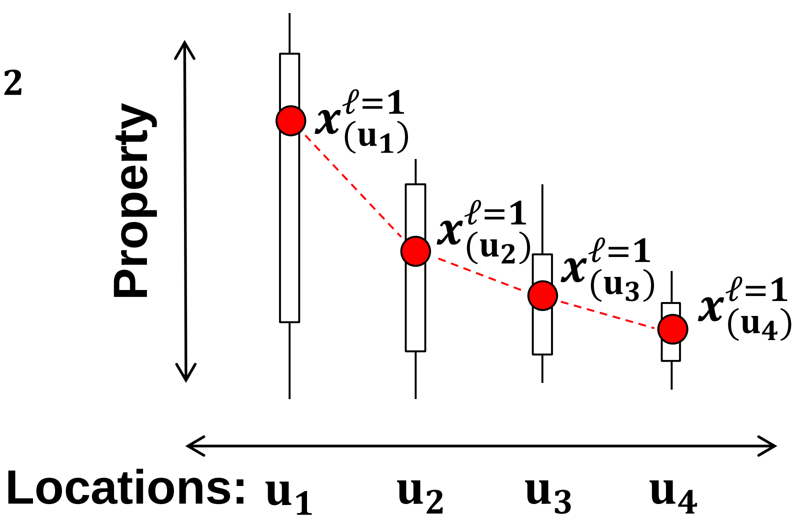
Realization

- 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 is considered equiprobable

Random Variable



Random Function

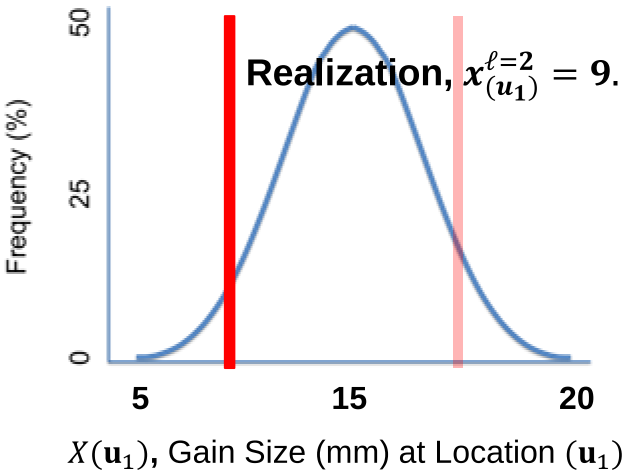


Realization Definition

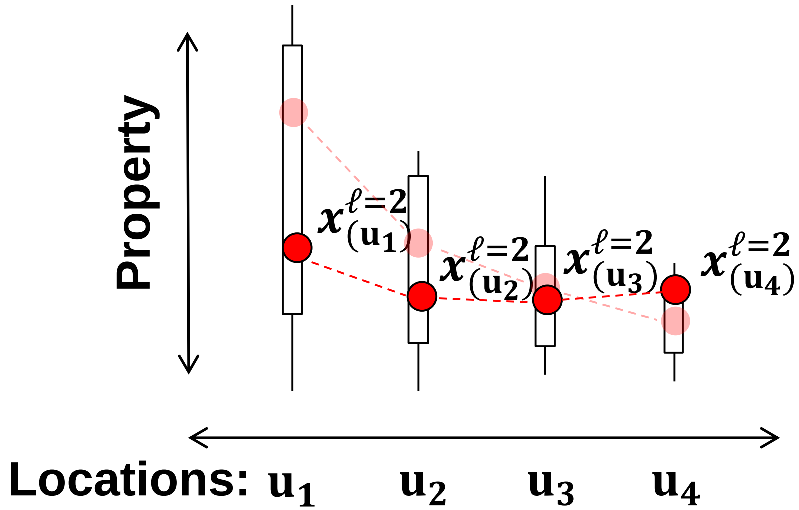
Realization

- 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

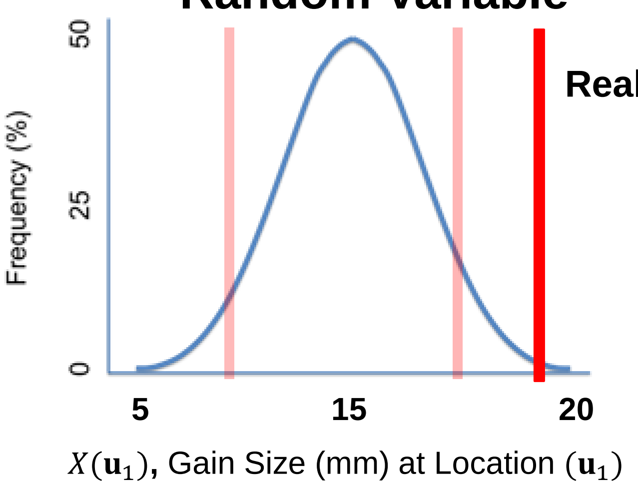


Realization Definition

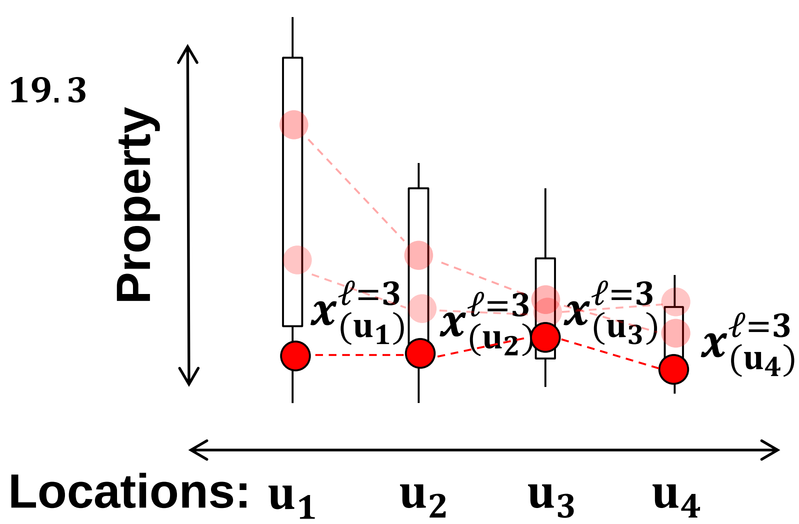
Realization

- 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



Stationarity
Substituting Time for Space

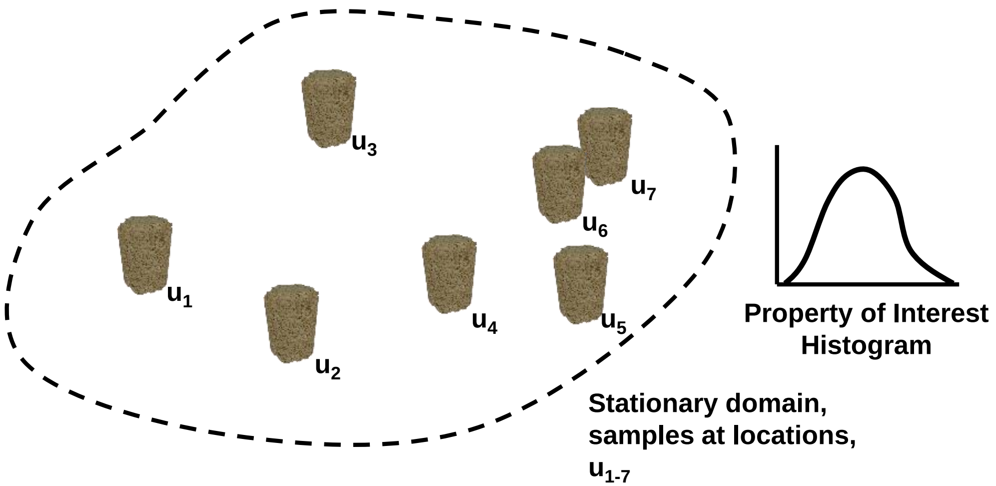
Any statistic requires replicates, repeated sampling (e.g. air or water samples from a monitoring station). In our geospatial problems repeated samples are not available at a location in the subsurface.

Extracted Sample at Location, u_1 **Hole Remains at Location, 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”.

Stationarity Substituting Time for Space

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**.

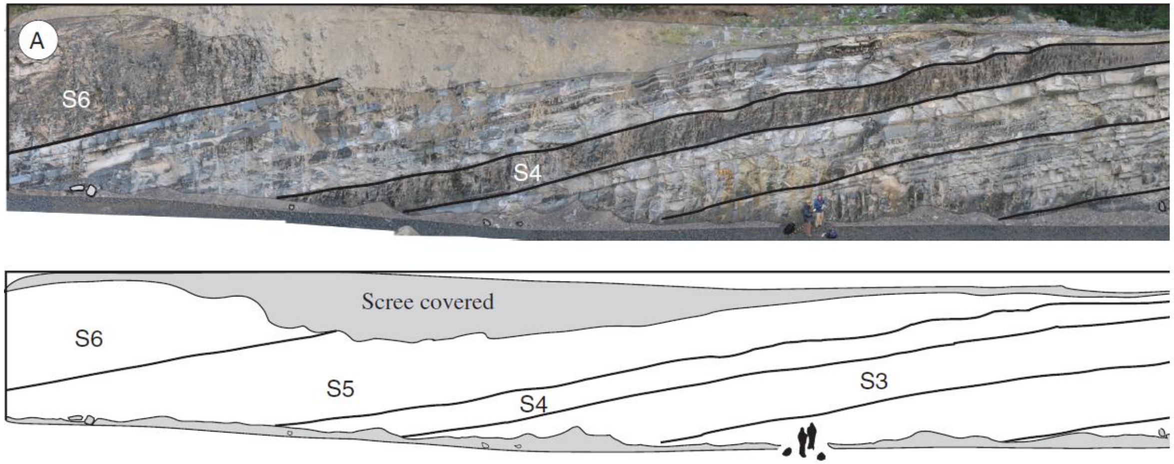


Import License: choice to pool specific samples to evaluate a statistic.

Export License: choice of where in the subsurface this statistic is applicable.

Stationarity
Definition 1: Geologic

Geological Definition: e.g. **'**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.'



Photomosaic, line drawing Punta Barrosa Formation sheet complex (Fildani et al. (2009).

Stationarity
Definition 2: Statistical

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

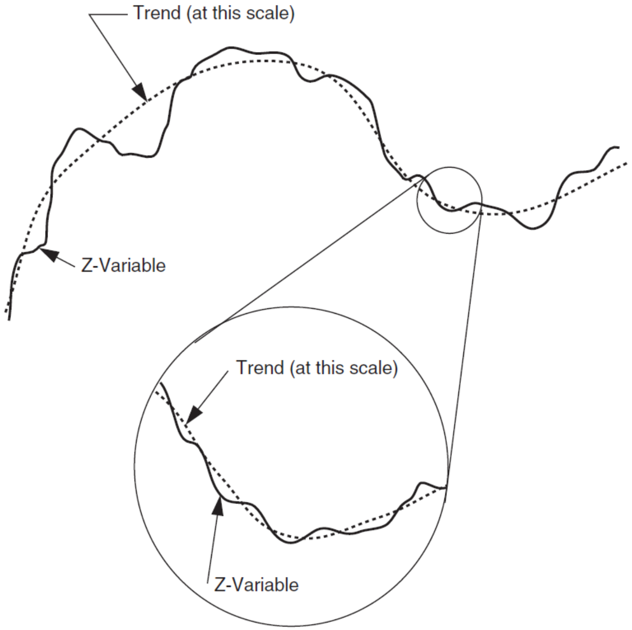
Stationarity: *What metric / statistic? Over what volume?*

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

Stationarity Comments on Stationarity

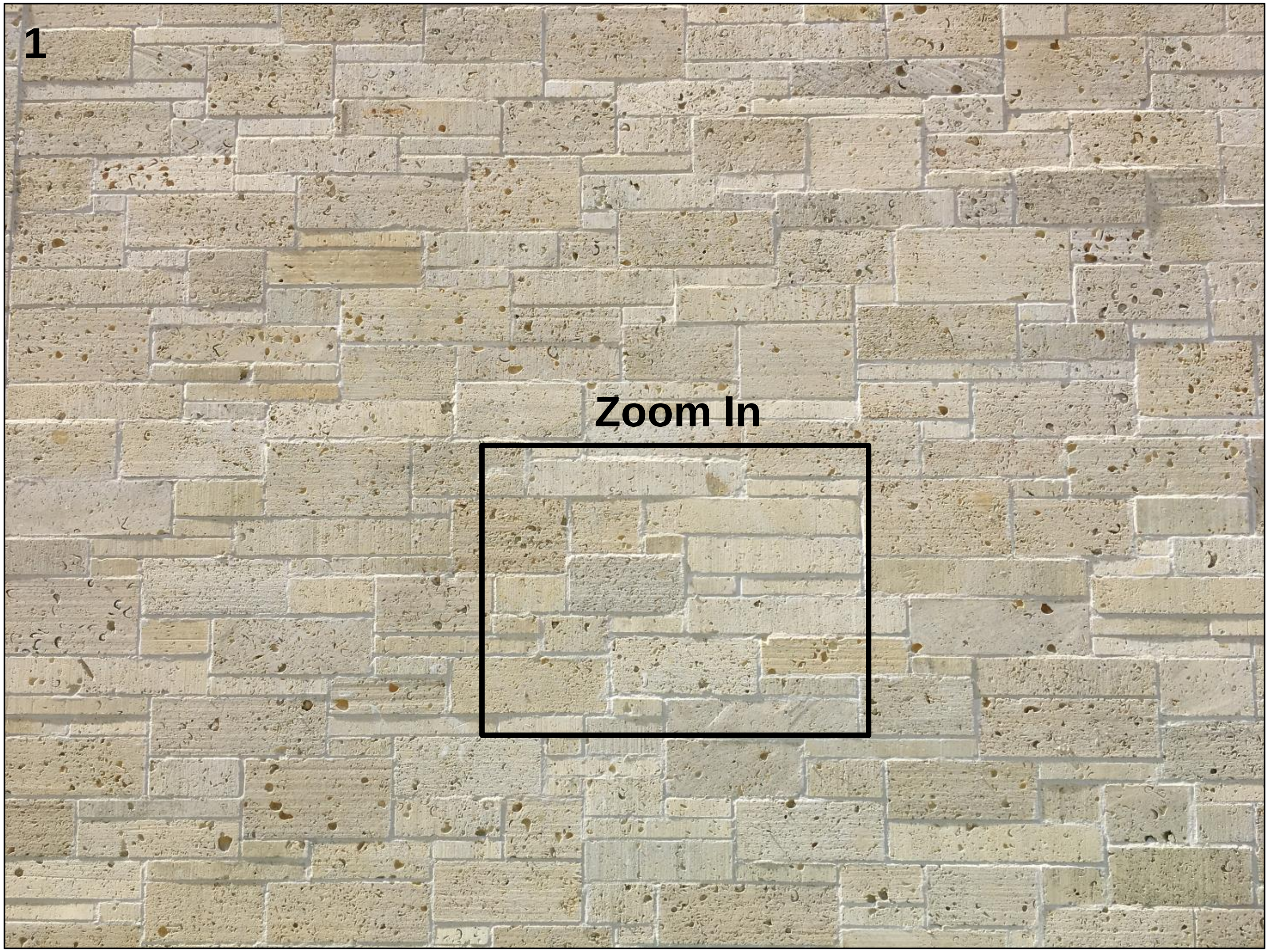
Stationarity is a decision, not an hypothesis; therefore it cannot be tested. Data may demonstrate that it is inappropriate.

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



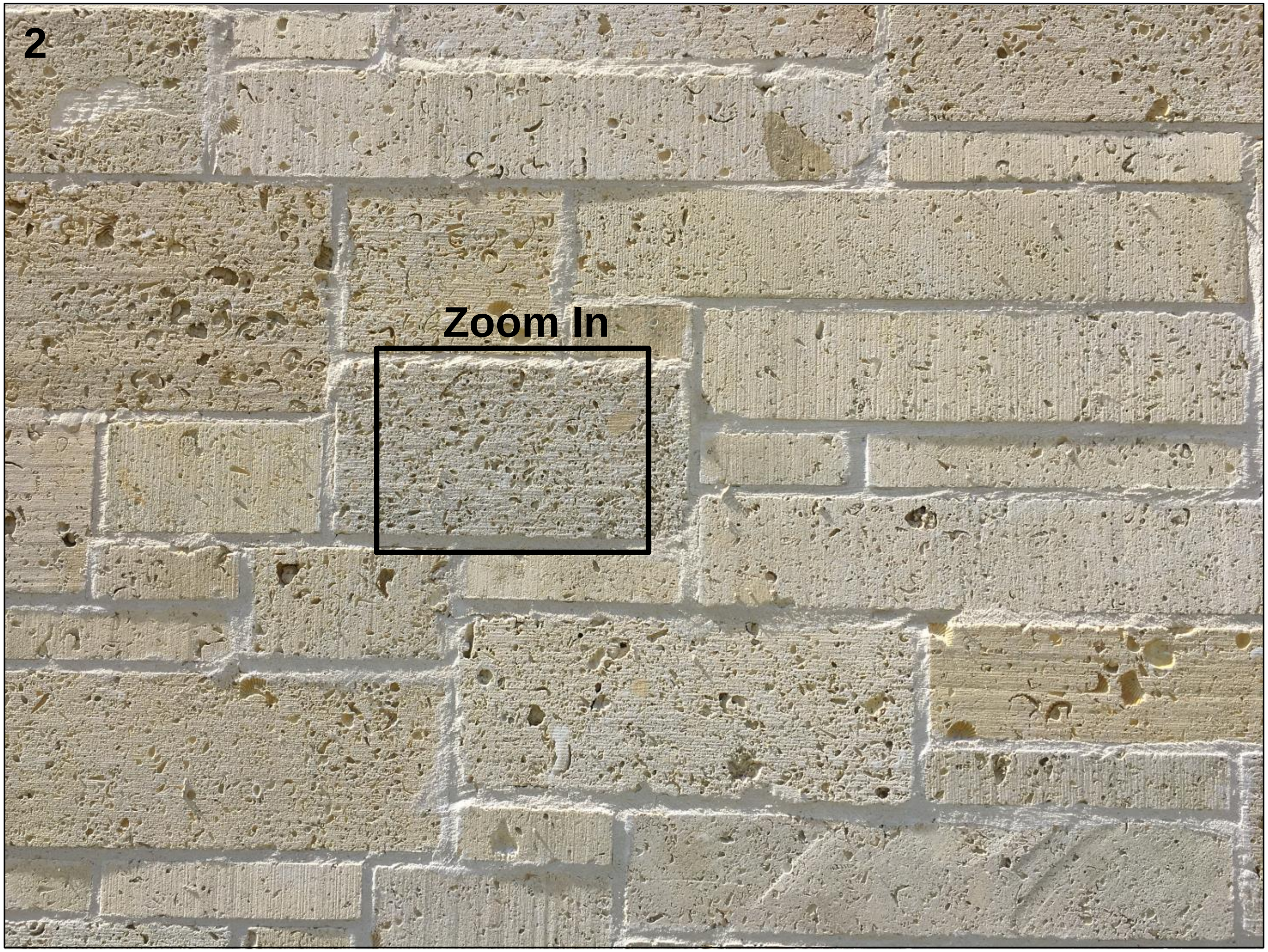
Scales of stationarity from Pyrcz and Deutsch (2014).

Stationarity Example



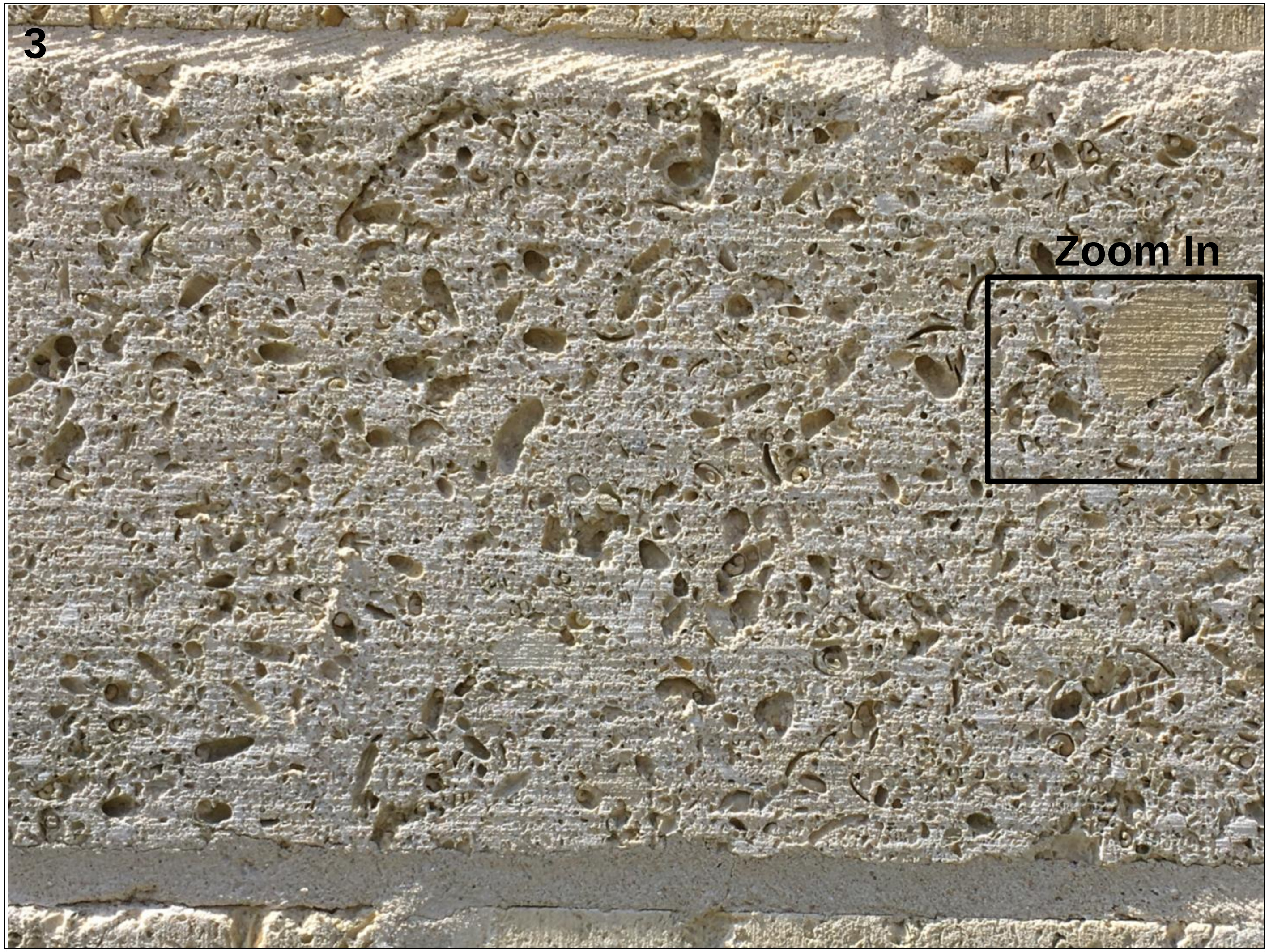
- Is this image stationary? What metric do you consider?

Stationarity Example



- A smaller group of bricks?

Stationarity Example

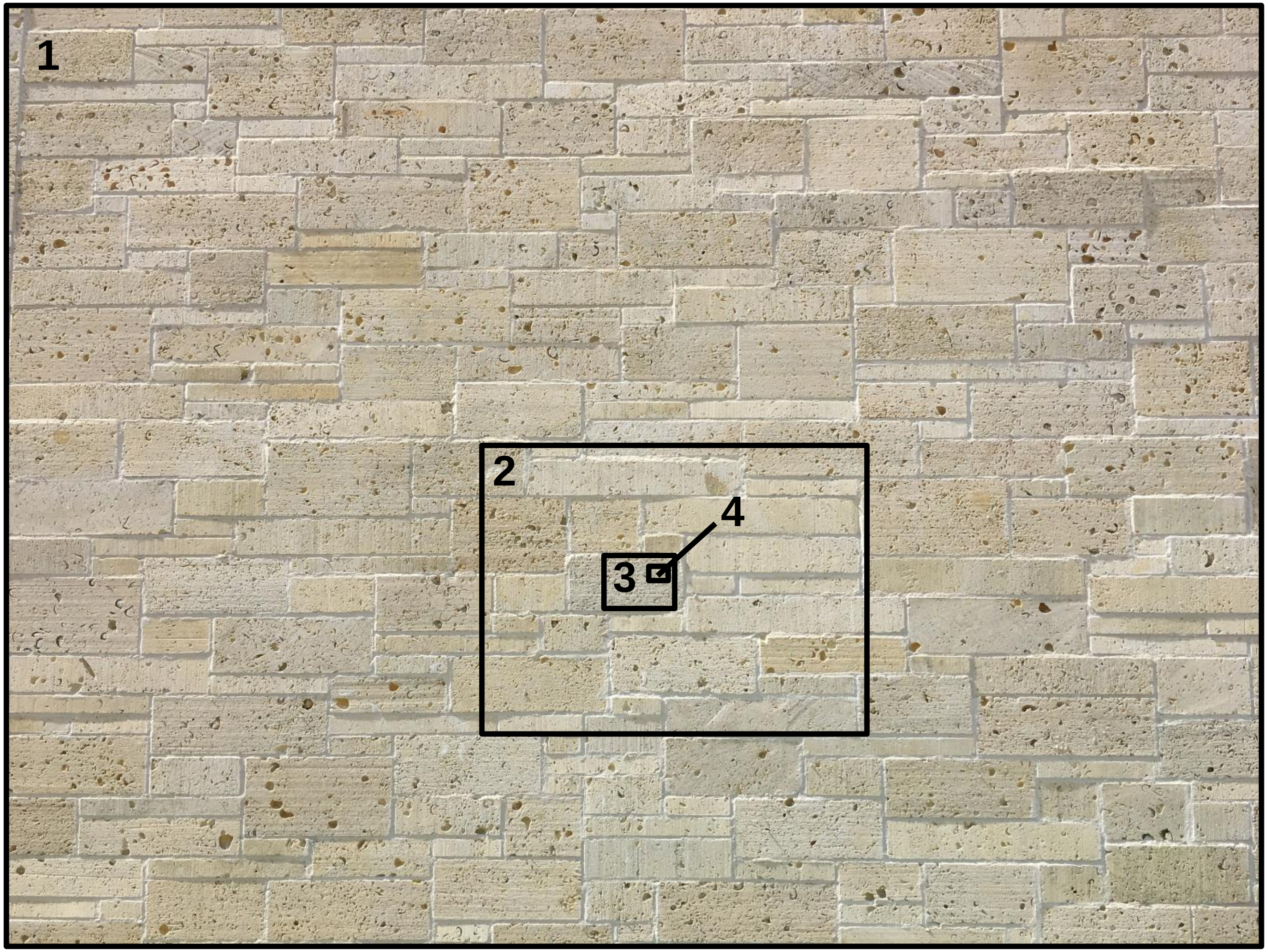


- A single brick?

Stationarity Example

- Small part of a brick?

Stationarity Example



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

Comments on Stationarity

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.

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

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

Good geological mapping and data integration is essential!

it is the framework of any subsurface model.

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

- 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 Summary on GitHub / Twitter

An explanation of **STATIONARITY** for geoscientists and geo-engineers.

Michael Pyrcz, University of Texas at Austin, @GeostatsGuy

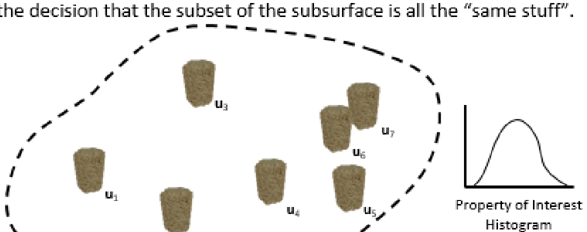
A description of the concepts of stationarity that are central to collecting geoscience information and applying it in subsurface modeling.

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".

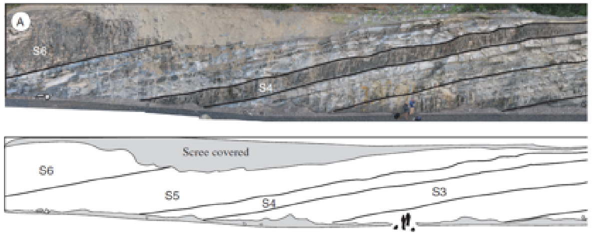


Stationary domain, samples at locations, $\mathbf{u}_{1-7}$

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/

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.



Photomosaic, line drawing Punta Barrosa Formation sheet complex (Fildani et al. (2009).

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

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

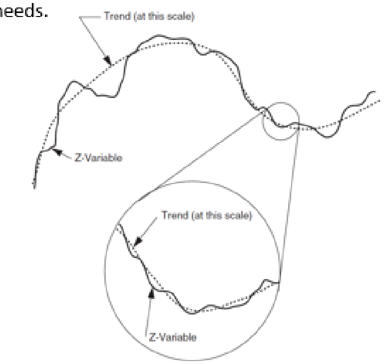
Stationary Semivariogram: $\gamma_z(\mathbf{u}; \mathbf{h}) = \gamma_z(\mathbf{h}), \ \forall \mathbf{u}$

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

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.

For more information check out Pyrcz, M.J., and Deutsch, C.V., 2014, Geostatistical Reservoir Modeling, 2nd edition, Oxford University Press.

Multivariate Modeling:
Spatial Continuity

Lecture outline . . .

- **Importance of Spatial Continuity**

Introduction

Prerequisites

Probability

Multivariate Analysis

Spatial Estimation

Statistical Learning

Feature Selection

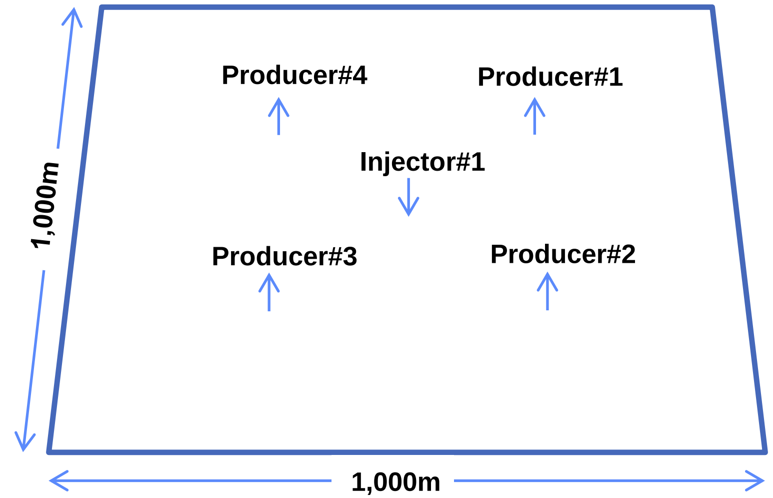
Multivariate Modeling

Conclusions

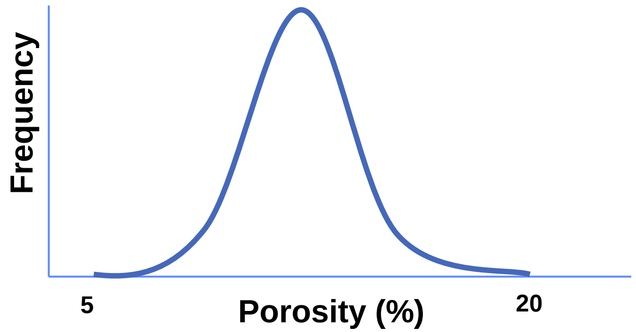
Motivation for Measuring Spatial Continuity

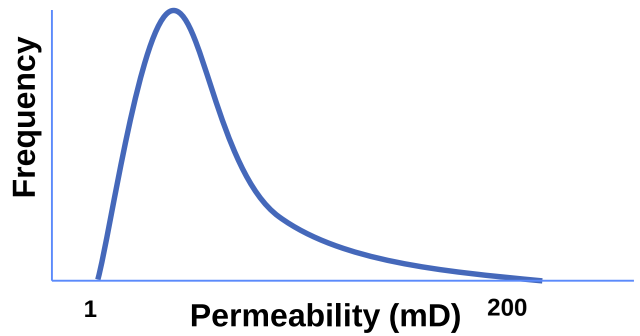
Simple Example

- Area of interest
- 1 Injector and 4 producers



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

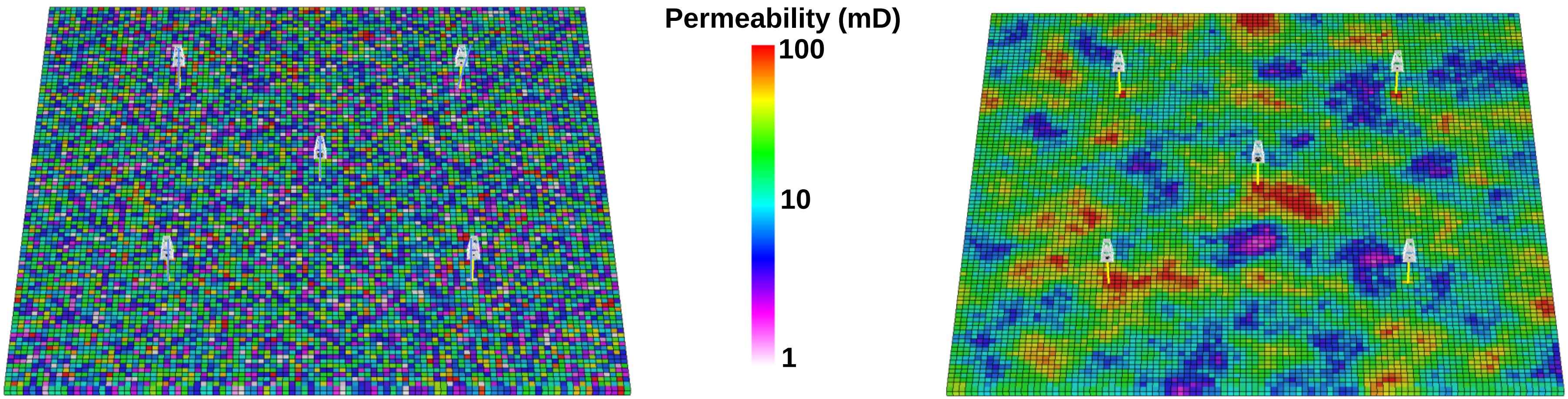




Motivation for Measuring Spatial Continuity

Does spatial continuity of reservoir properties matter?

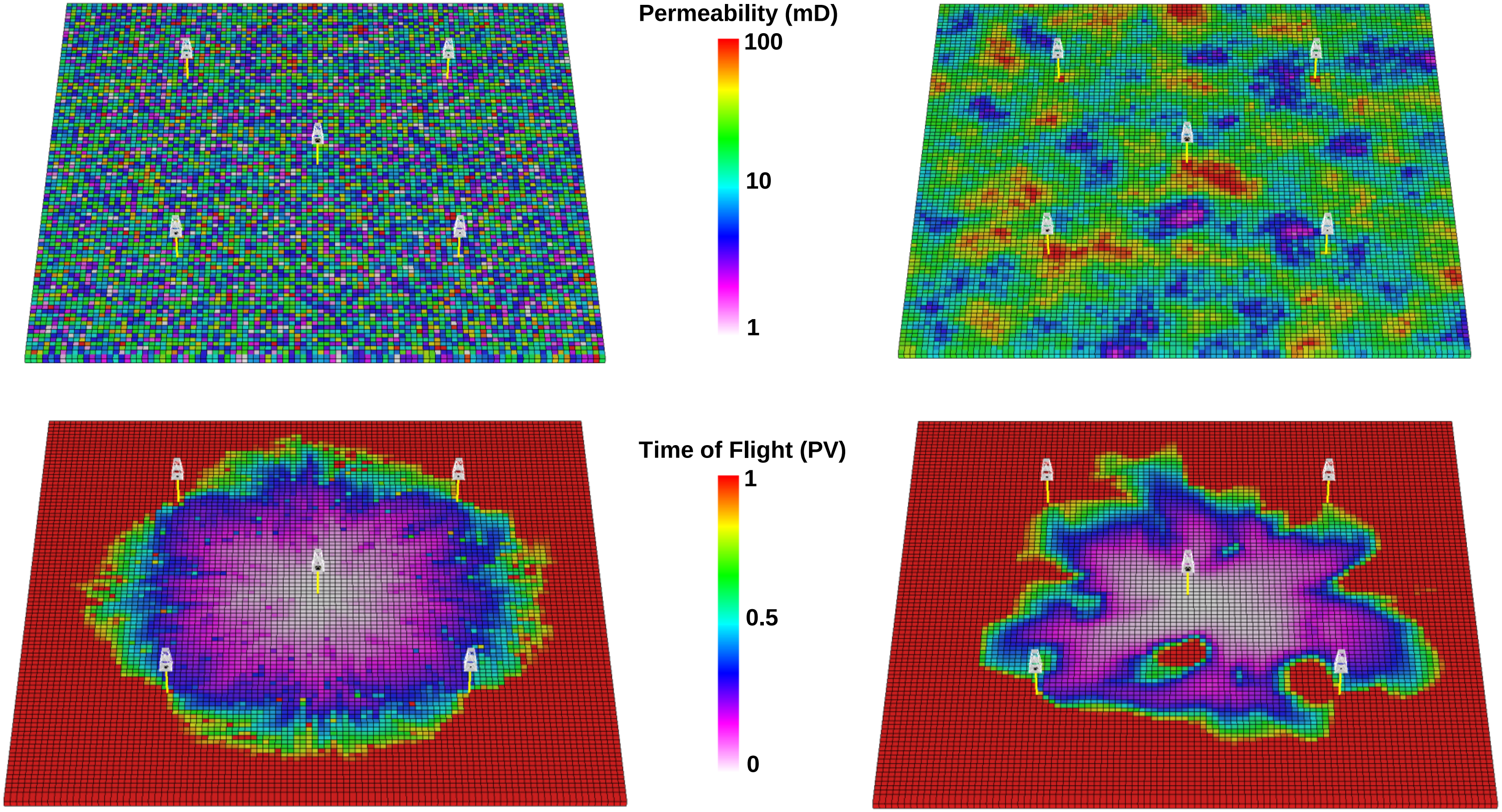
Consider these models of permeability



Recall – all models have the same porosity and permeability distributions

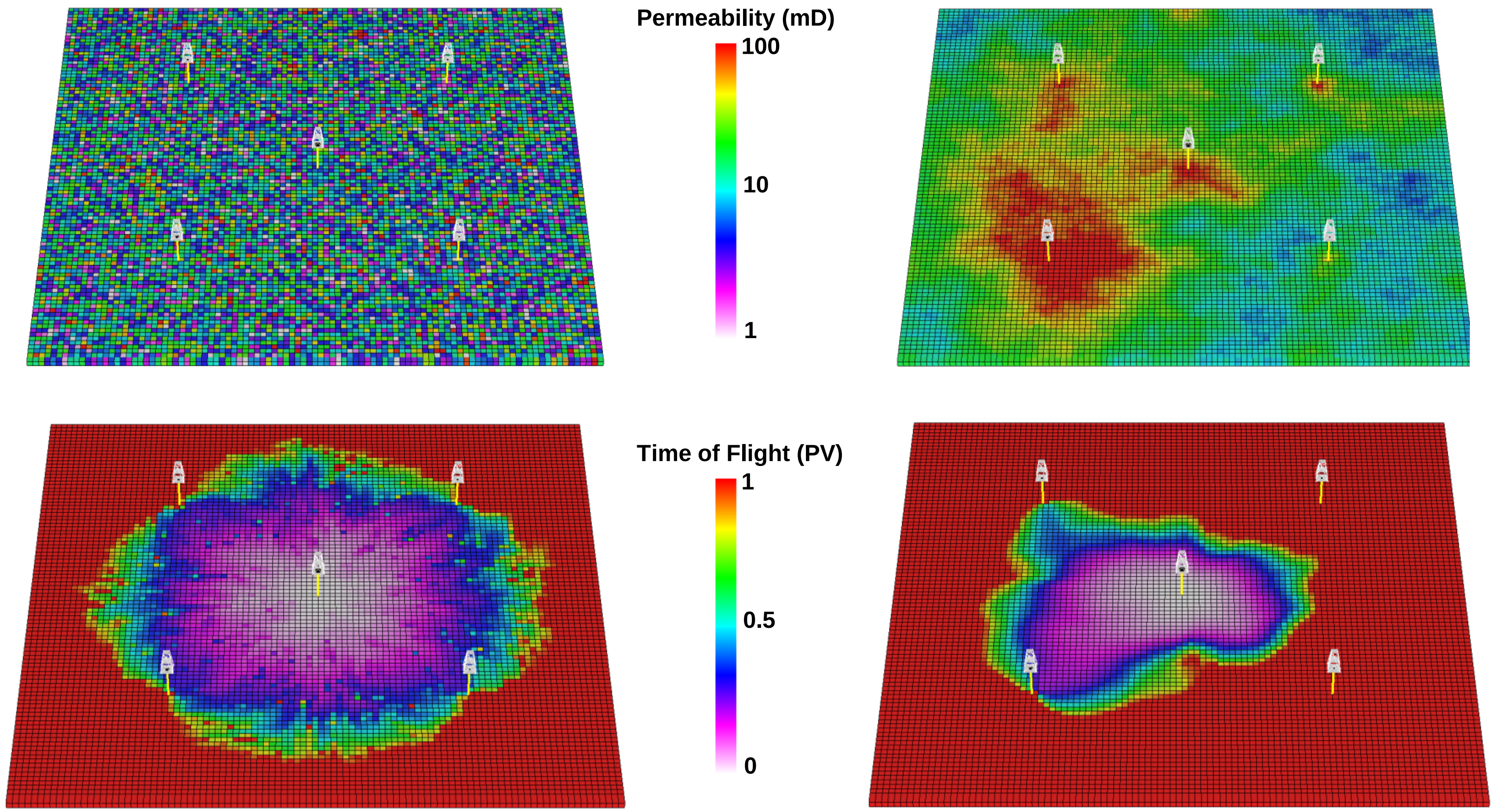
- Mean, variance, P10, P90 …
- Same static oil in place!

Motivation for Measuring Spatial Continuity



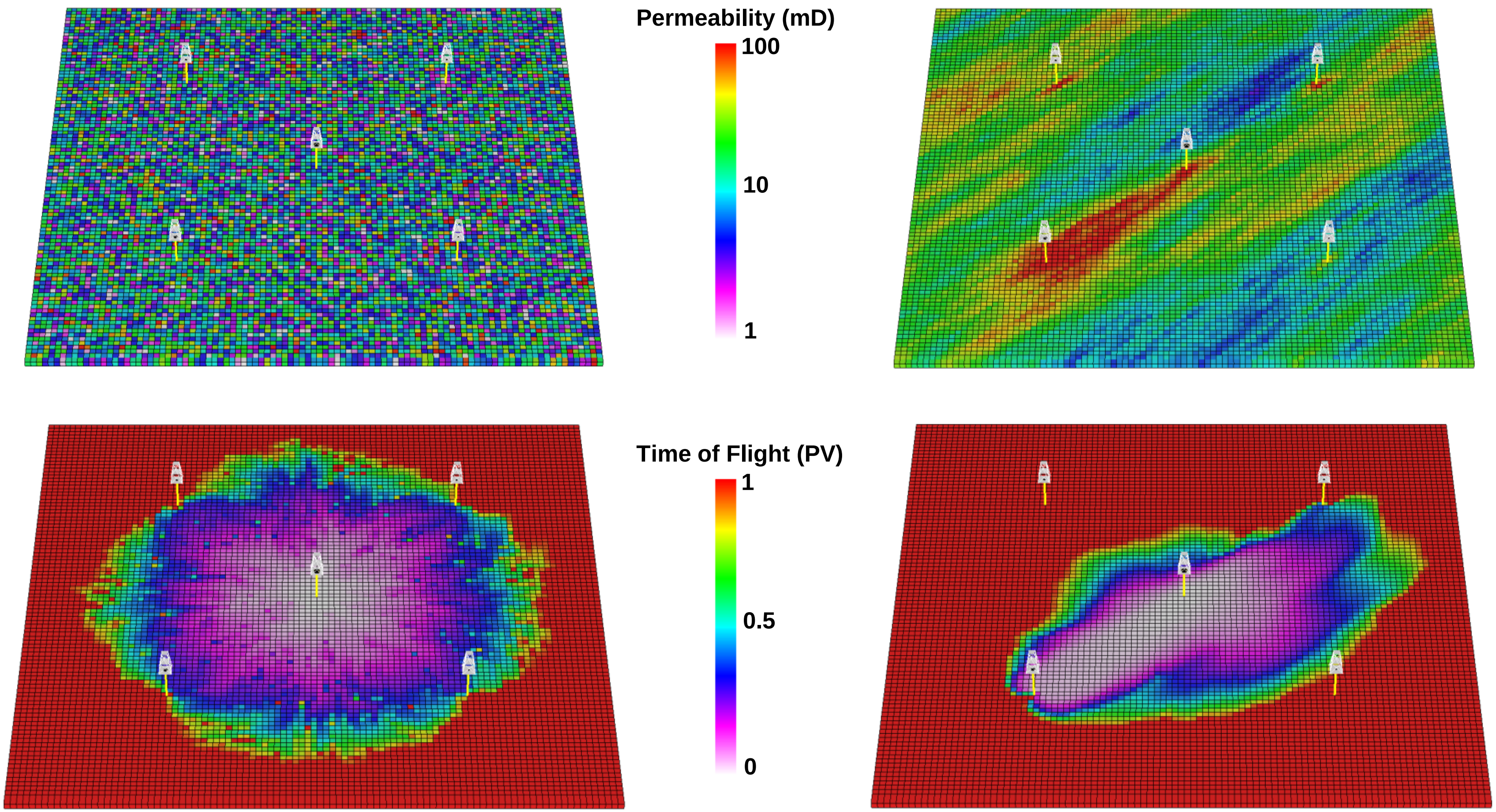
How does heterogeneity impact recovery factor?

Motivation for Measuring Spatial Continuity



How does heterogeneity impact recovery factor?

Motivation for Measuring Spatial Continuity



How does heterogeneity impact recovery factor?

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:

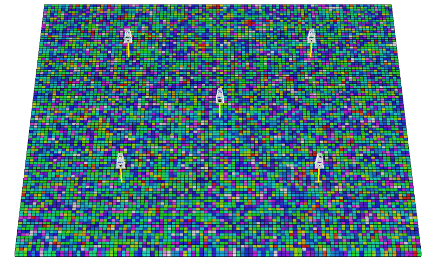
Characterize / Quantify

↓ **Spatial Continuity Measures**

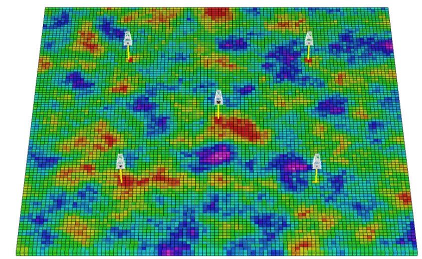
Impose in Reservoir Models

Spatial Continuity

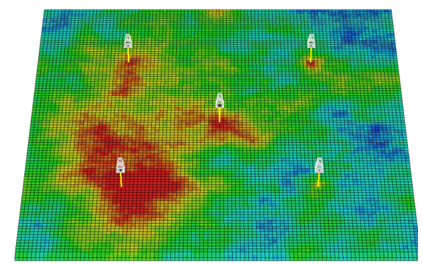
"Very Short"



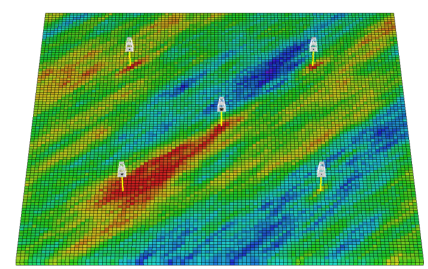
"Medium"



"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.

Multivariate Modeling: Spatial Continuity

Lecture outline . . .

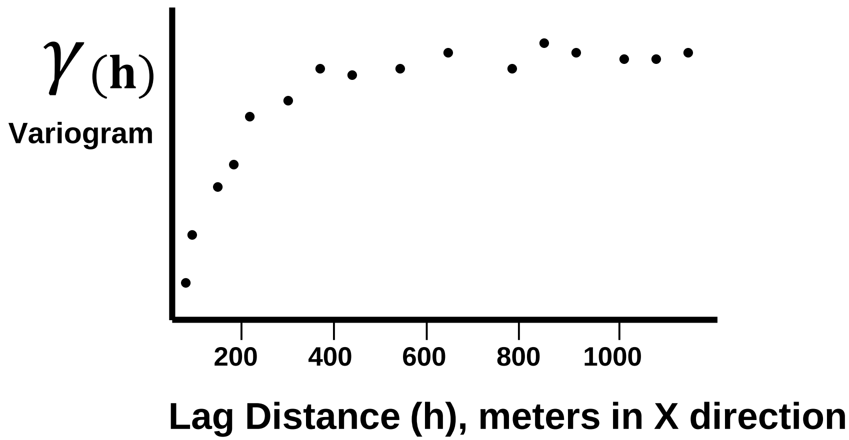
- Spatial Continuity

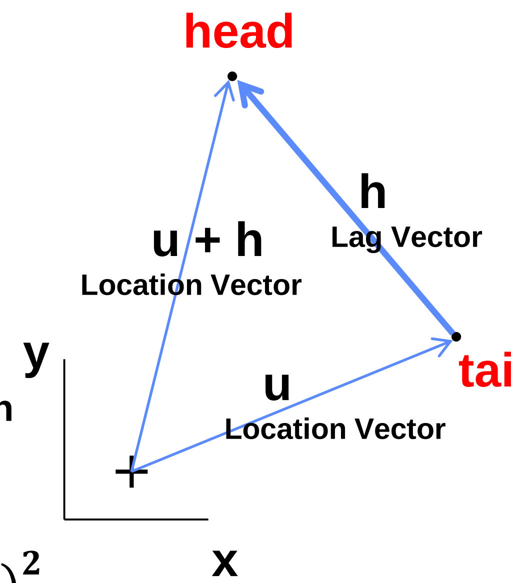
Measuring Spatial Continuity

We need a statistic to quantify spatial continuity!

The Semivariogram:

- Function of difference over distance.





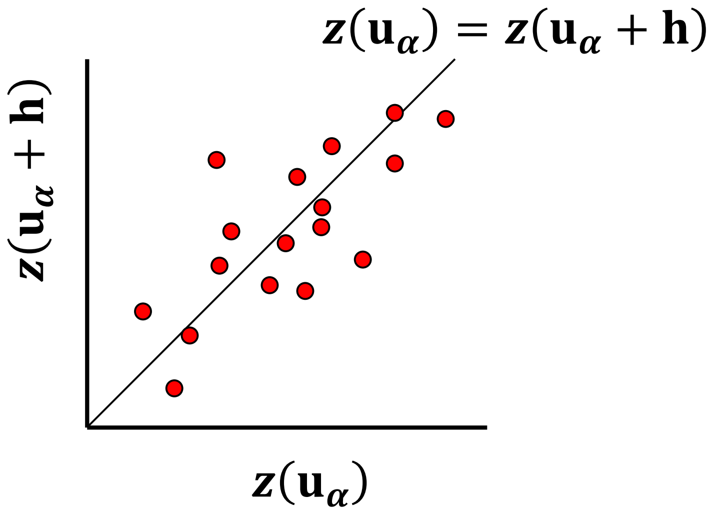
- The equation:

tail head

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

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:

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

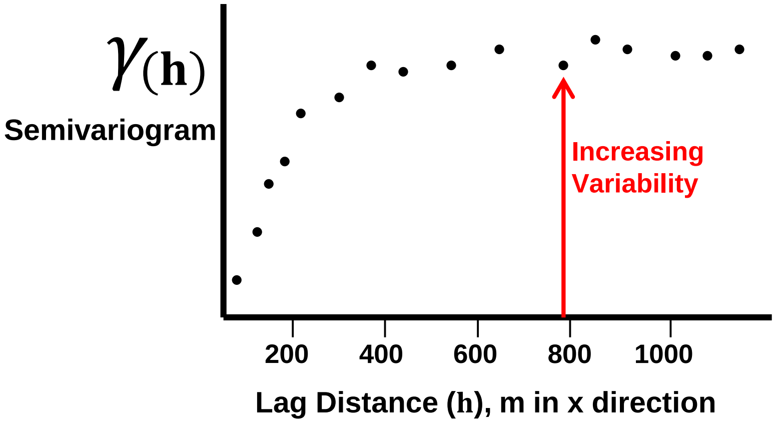
  - Note the correlogram is related to the covariance function as:

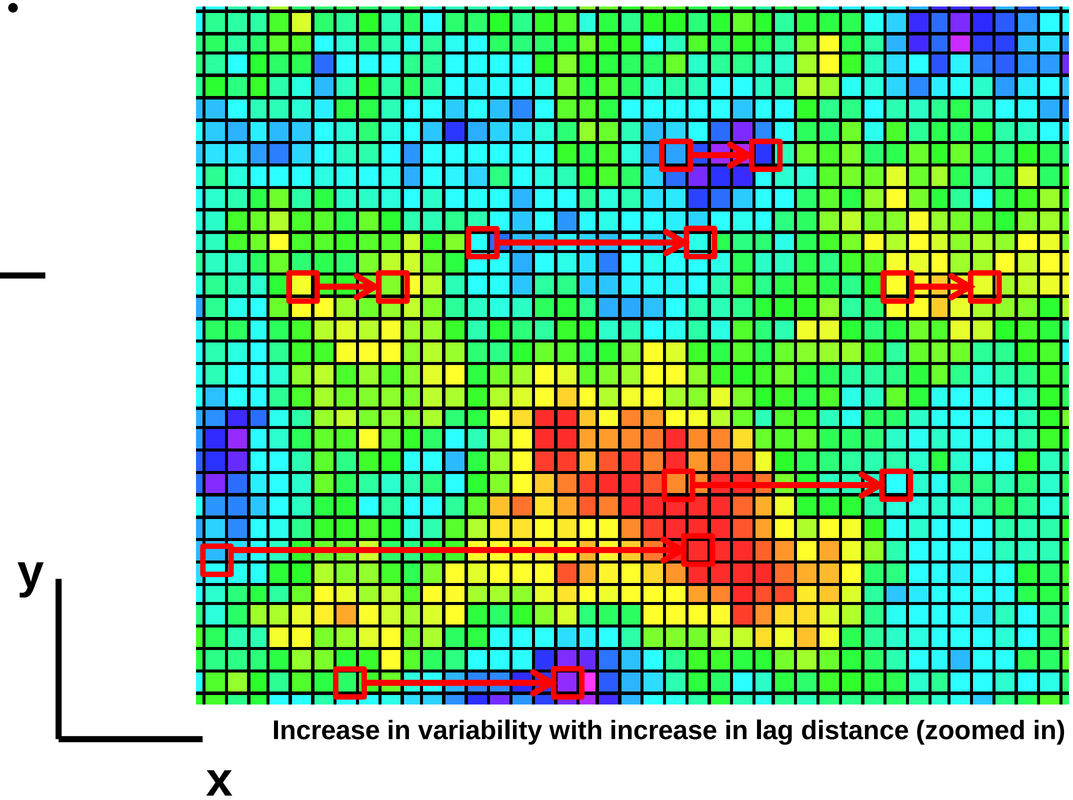
$$\rho_x(h) = \frac{C_x(h)}{\sigma_x^2}$$, h-scatter plot correlation vs. lag distance

Variogram Observations

Observation #1

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



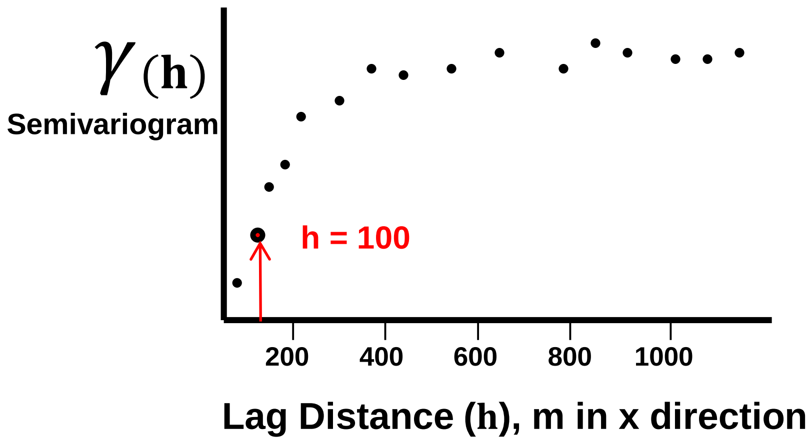


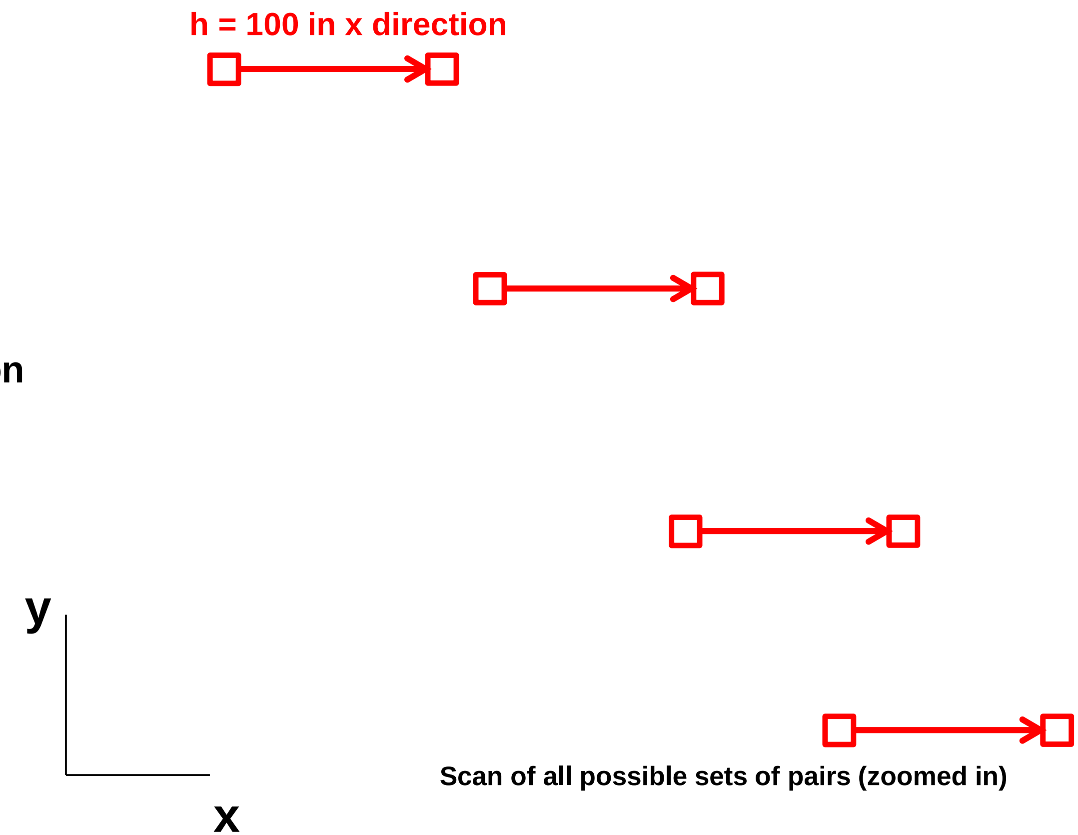
Increase in variability with increase in lag distance (zoomed in)

Variogram Observations

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).

Variogram Observations

Observation #3

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

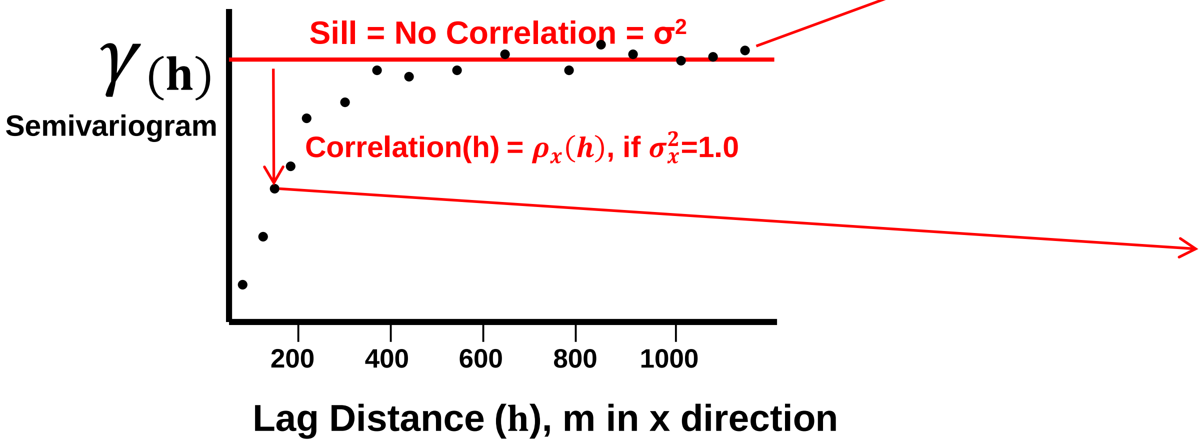
  Sill is the Variance, σ^2

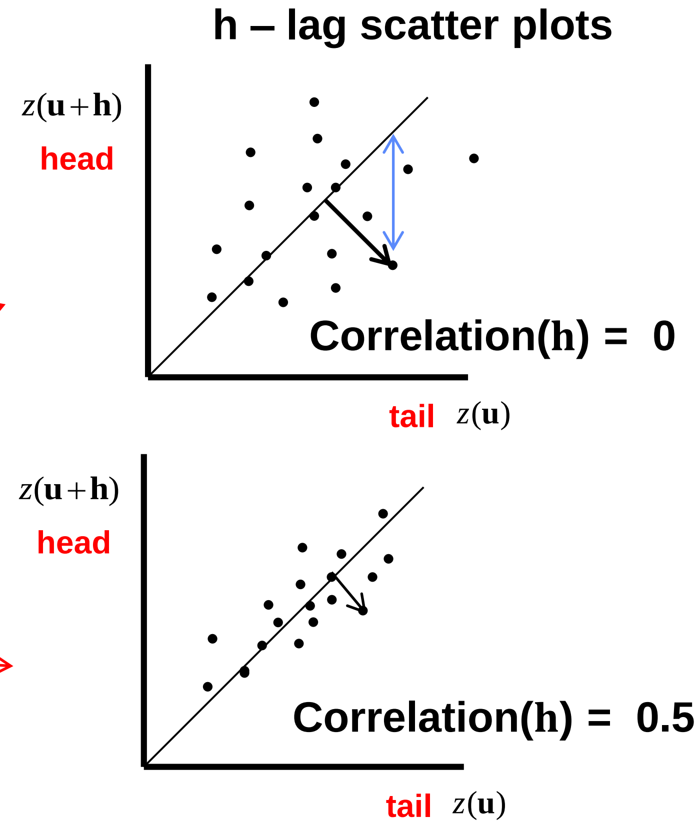
- Given stationarity of the variance and $\gamma_x(\mathbf{h})$:

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

- Given a standardized distribution $\sigma_x^2 = 1.0$:

Correlogram: $\rho_x(h) = \sigma_x^2 - \gamma_x(\mathbf{h})$





Variogram Observations

Observation #3

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

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

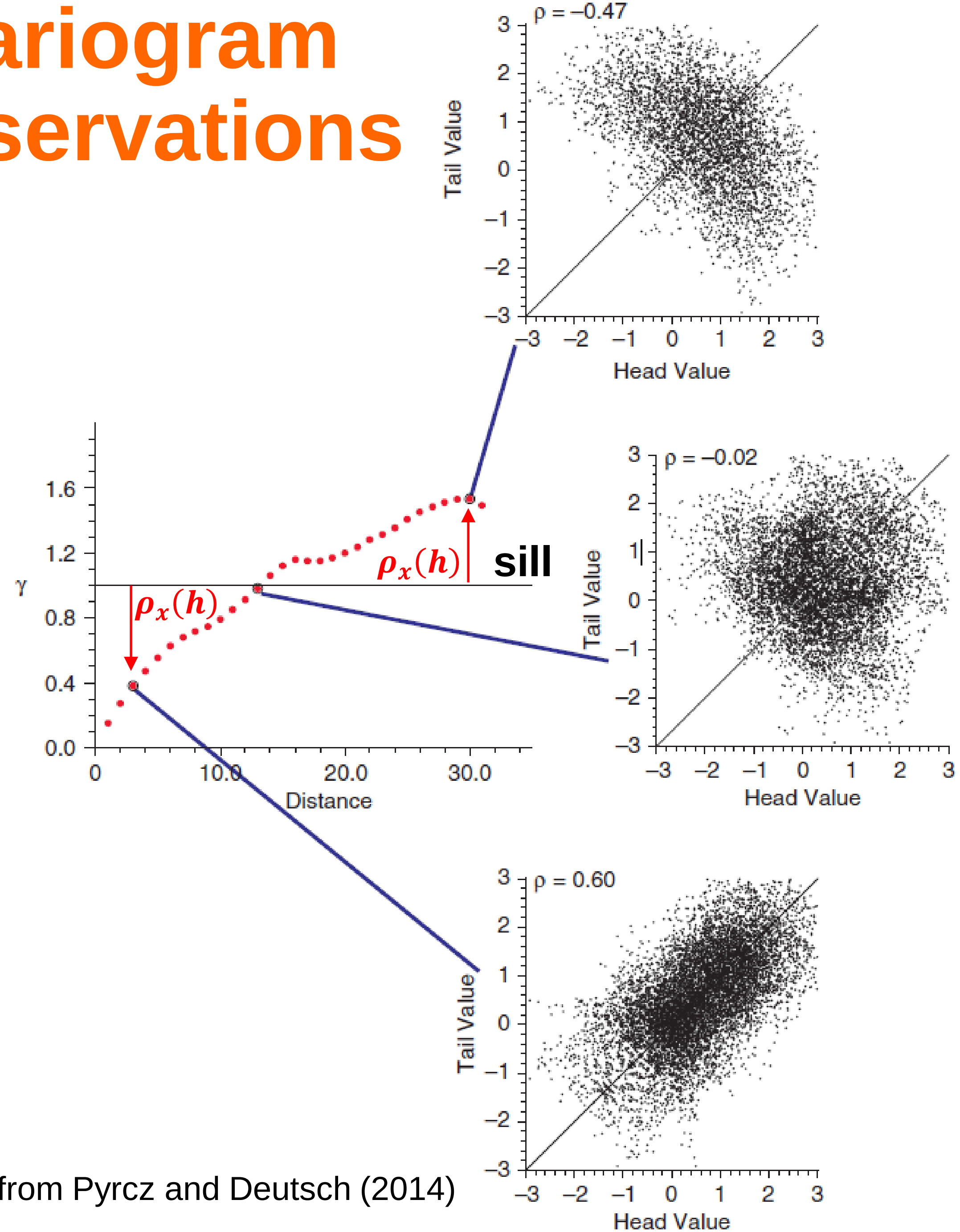


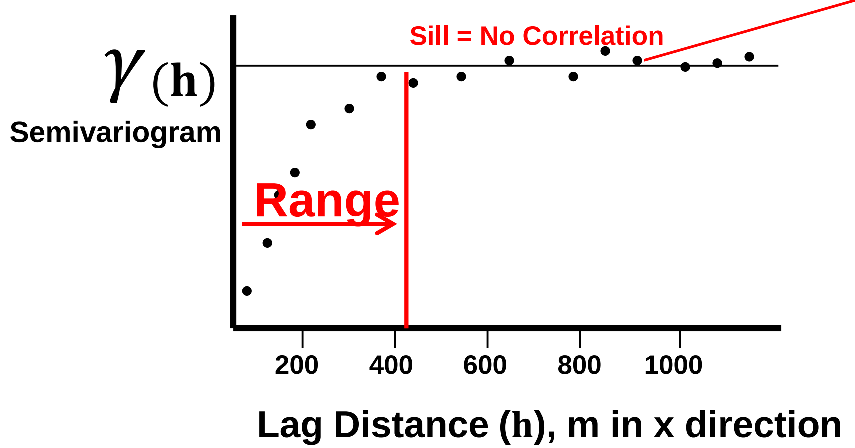
Image modified from Pyrcz and Deutsch (2014)

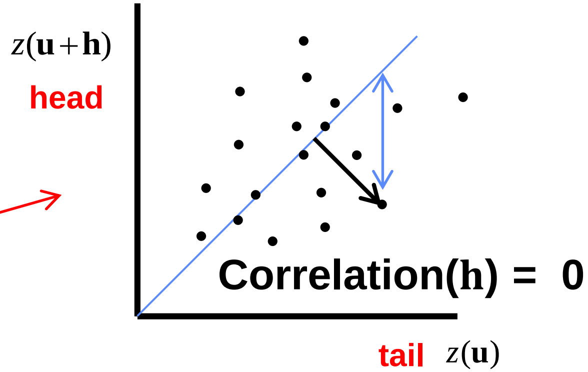
Variogram Observations

Observation #4

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

At the range, knowing the data value at the tail provides no information about a value at the head.

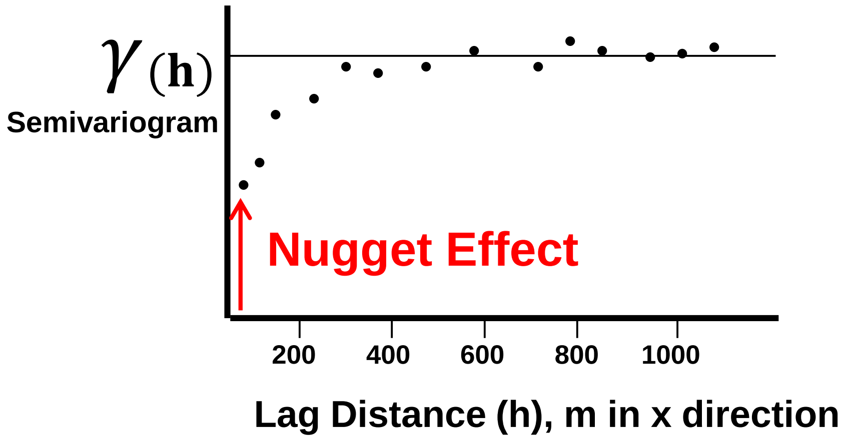




Variogram Observations

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 140 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

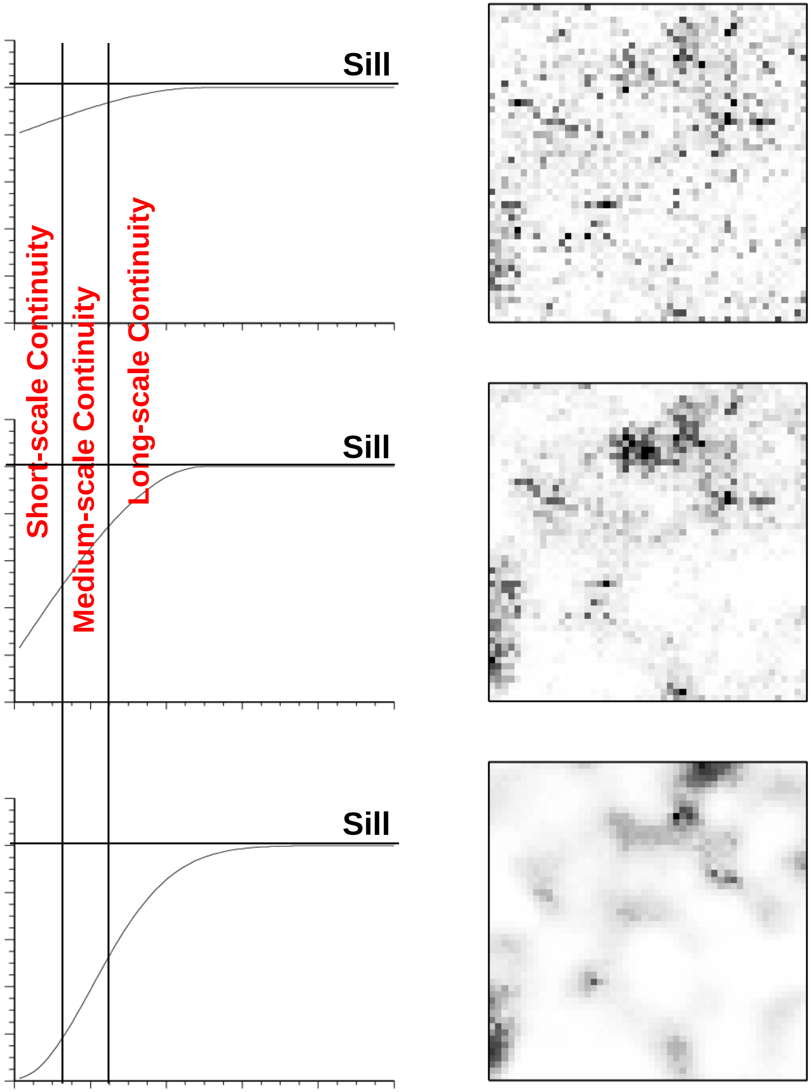


Image modified from Pyrcz and Deutsch (2014)

Multivariate Modeling: Spatial Continuity

Lecture outline . . .

- **Variogram Calculation**

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:

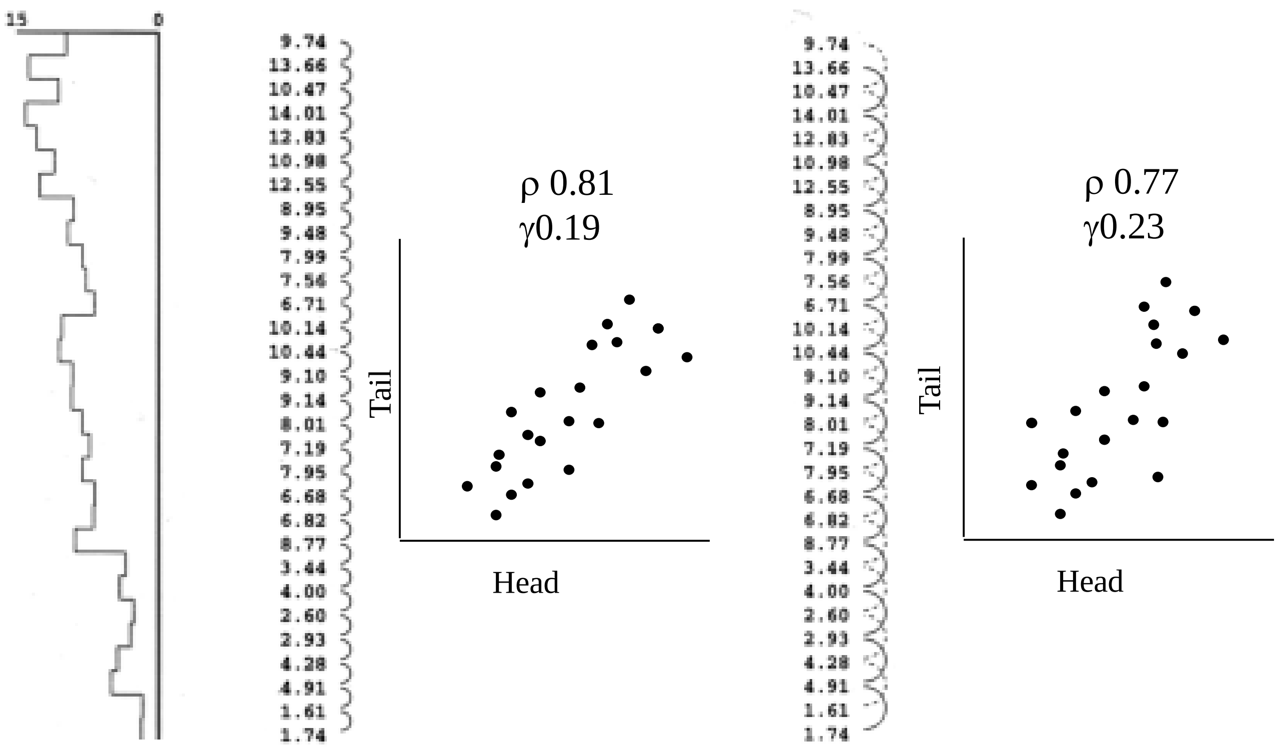
$$C_x(h) = \sigma_x^2 - \gamma_x(\mathbf{h})$$

  - 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 Calculation

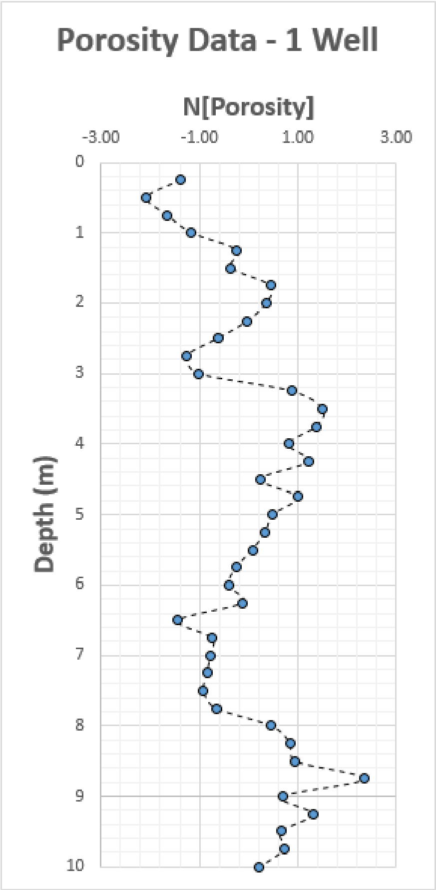
- Consider data values separated by *lag* vectors (the h values)
- Here are two examples of a lag vector equal to the data spacing and then twice the data spacing:



Variogram Calculation Example

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

- Dataset is at GitHub/GeostatsGuy
- GeoDataSets/1D_Porosity.xlsx

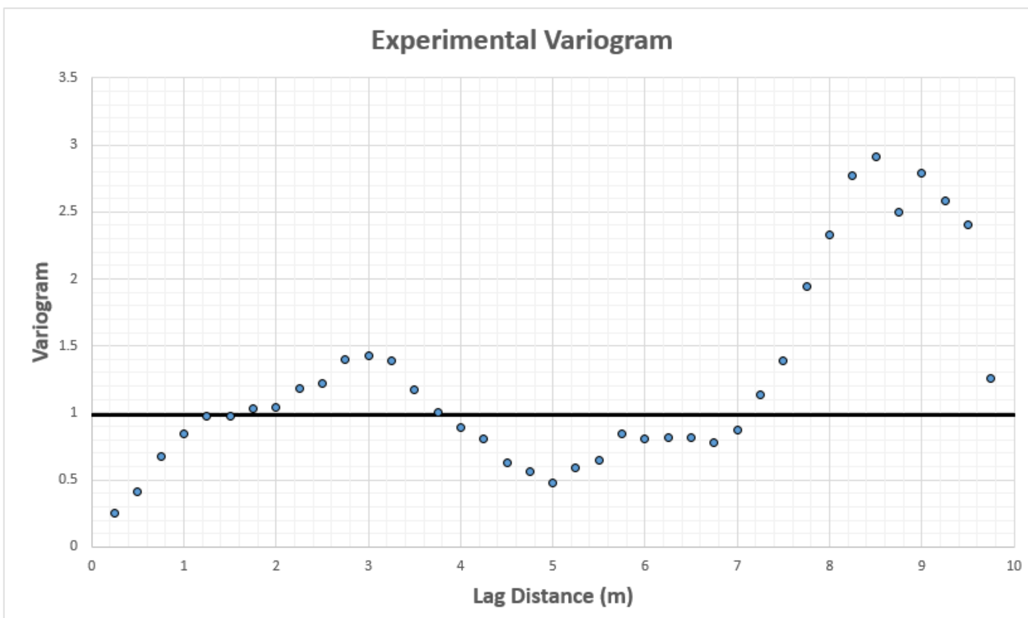


Depth	N[Porosity]
0.25	-1.37
0.5	-2.08
0.75	-1.67
1	-1.16
1.25	-0.24
1.5	-0.36
1.75	0.44
2	0.36
2.25	-0.02
2.5	-0.63
2.75	-1.26
3	-1.03
3.25	0.88
3.5	1.51
3.75	1.37
4	0.81
4.25	1.21
4.5	0.24
4.75	0.99
5	0.49
5.25	0.34
5.5	0.07
5.75	-0.26
6	-0.41
6.25	-0.14
6.5	-1.44
6.75	-0.75
7	-0.78
7.25	-0.85
7.5	-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 Example

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

- Here's all of them:



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

The Variogram and Covariance Function

- **The variogram, covariance function and correlation coefficient are equivalent tools for characterizing spatial two-point correlation** (assuming stationarity):

$$\boldsymbol{\gamma_x(\mathbf{h}) = \sigma_x^2 - C_x(h)}$$
$$\boldsymbol{= \sigma_x^2(1 - \rho_x(h))}$$

$$\boldsymbol{\rho_x(h) = \frac{C_x(h)}{\sigma_x^2}}$$

where:

$$\boldsymbol{C_x(\mathbf{h}) = E\{X(\mathbf{u}) \cdot X(\mathbf{u}+\mathbf{h})\} - [E\{X(\mathbf{u})\} \cdot E\{X(\mathbf{u}+\mathbf{h})\}], \forall\, \mathbf{u}, \mathbf{u}+\mathbf{h} \in A}$$

$$C_x(0) = \sigma_x^2$$

$\boldsymbol{C_x(\mathbf{h}) = \frac{\sum_{\alpha=1}^{n} x(\mathbf{u}_\alpha)\cdot x(\mathbf{u}_\alpha+\mathbf{h})}{n} - (\overline{x})^2}$, if stationary mean

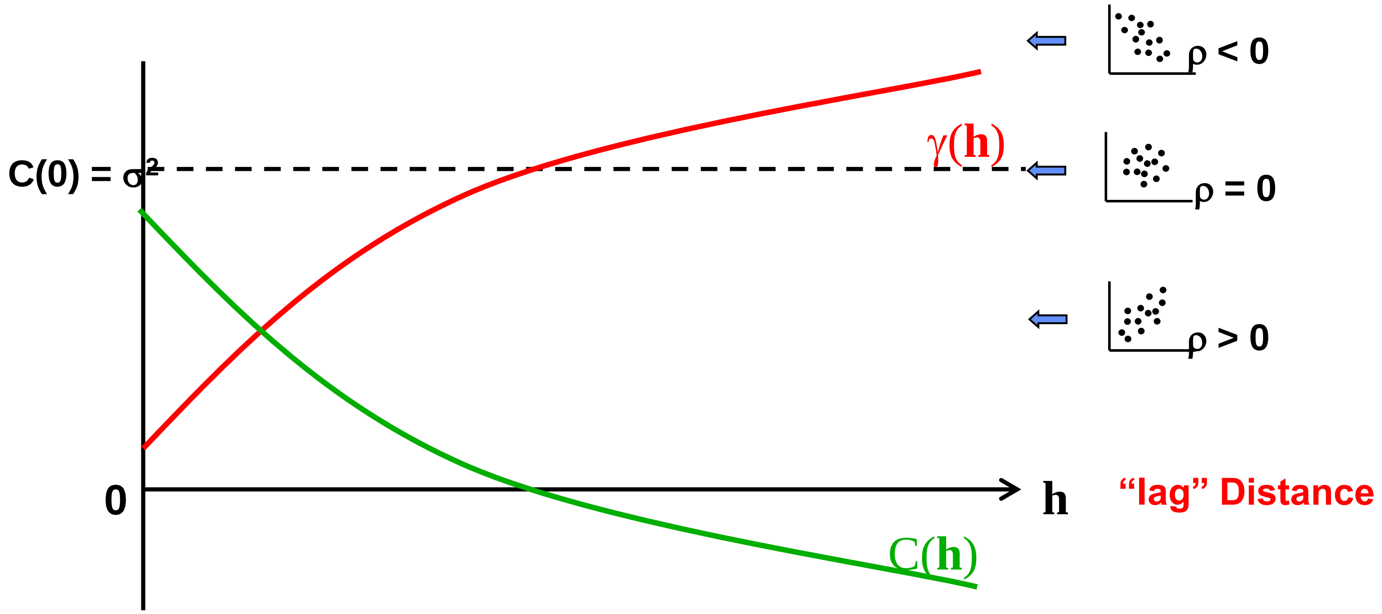
Stationarity entails that:

$$m(\mathbf{u}) = m(\mathbf{u}+\mathbf{h}) = m = E\{Z\}, \forall \mathbf{u} \in A$$

$$Var(\mathbf{u}) = Var(\mathbf{u}+\mathbf{h}) = \sigma^2 = Var\{Z\}, \forall \mathbf{u} \in A$$

The Variogram and Covariance Function

- **Must plot variance to interpret variogram:**
  - Positive correlation when semivariogram less than variance
  - No correlation when the semivariogram is equal to the variance
  - Negative correlation when the semivariogram points above variance



Covariance Function Definition

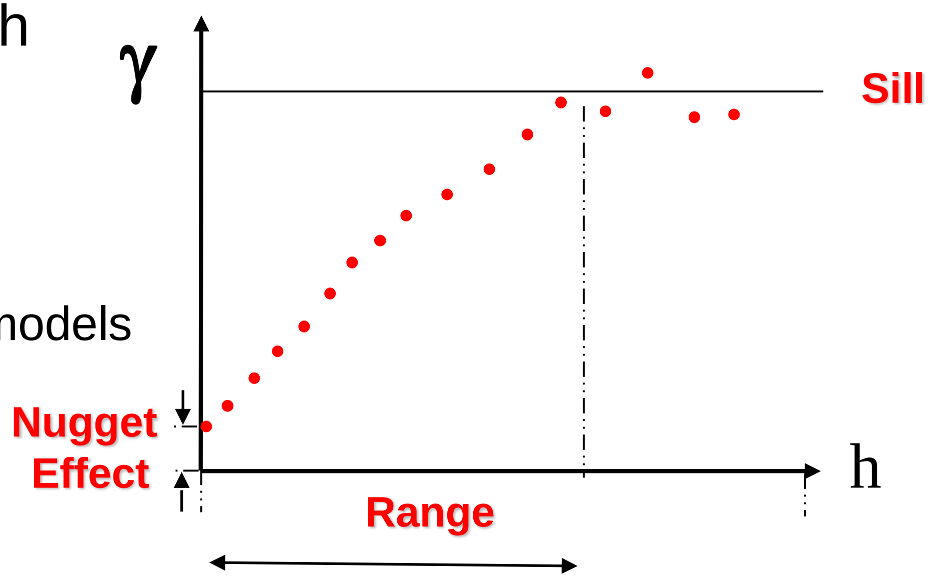
- **Covariance Function** – a measure of similarity vs. distance. Calculated as the average product of values separated by a lag vector centered by the square of the mean.

$$\boldsymbol{C_x}(\mathbf{h}) = \frac{\sum_{\alpha=1}^{n} x(\mathbf{u}_\alpha) \cdot x(\mathbf{u}_\alpha + \mathbf{h})}{n} - (\overline{x})^2,$$ if stationary mean

  - The covariance function is the variogram upside down. $\boldsymbol{\gamma_x}(\mathbf{h}) = \boldsymbol{\sigma_x^2} - \boldsymbol{C_x}(\boldsymbol{h})$
  - We model variograms, but inside the kriging and simulation methods they are converted to covariance values for numerical convenience.

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



Spoiler Alert

We need to practically calculate and model spatial continuity. From the available and often sparse subsurface data.

1. Calculate variogram with irregularly spaced data
   – Search templates with parameters

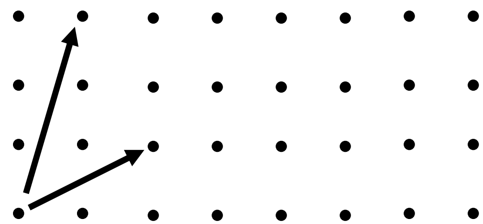
2. Valid spatial model
   – Fit with a couple different, nest (additive) spatial continuity models e.g. nugget, spherical, exponential and Gaussian

3. Full 3D spatial continuity model
   – Model primary directions, i.e. major horizontal, minor horizontal and vertical and combine together with assumption of geometric anisotropy

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



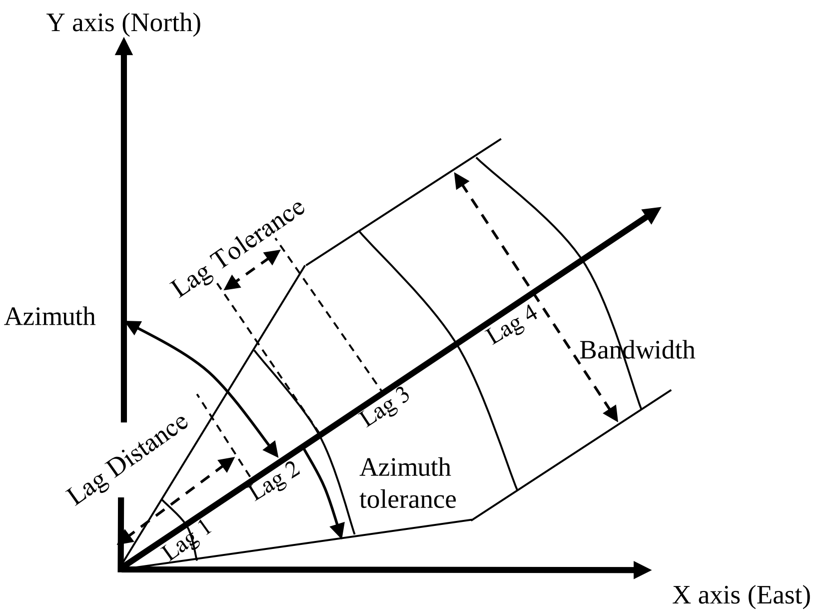
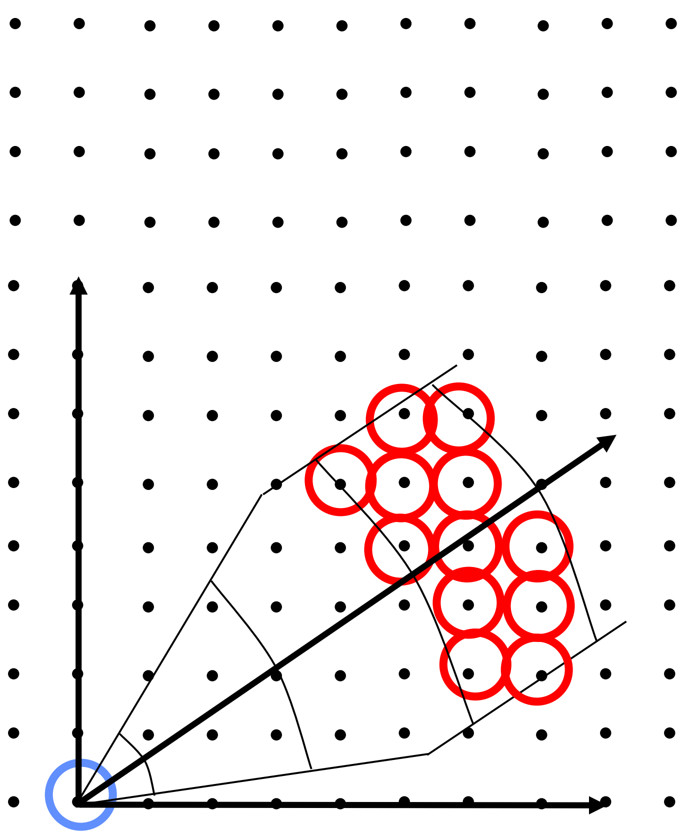


Image from Pyrcz and Deutsch, 2014

Calculating Experimental Variograms

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

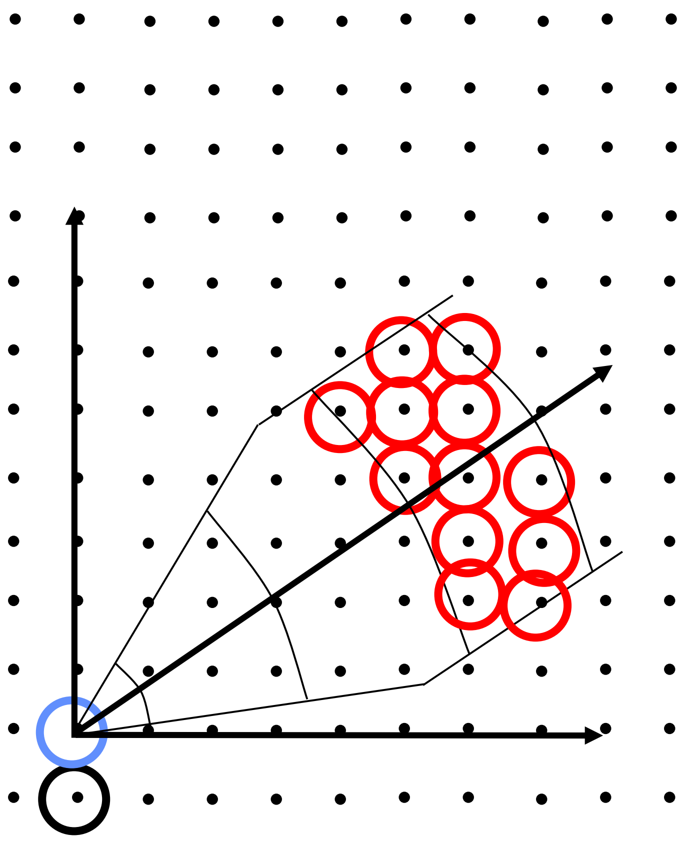


$$2\gamma(\mathbf{h}) = \frac{1}{N(\mathbf{h})}\sum_{N(\mathbf{h})}[z(\mathbf{u}) - z(\mathbf{u}+\mathbf{h})]^2$$

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

...

Calculating Experimental Variograms



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

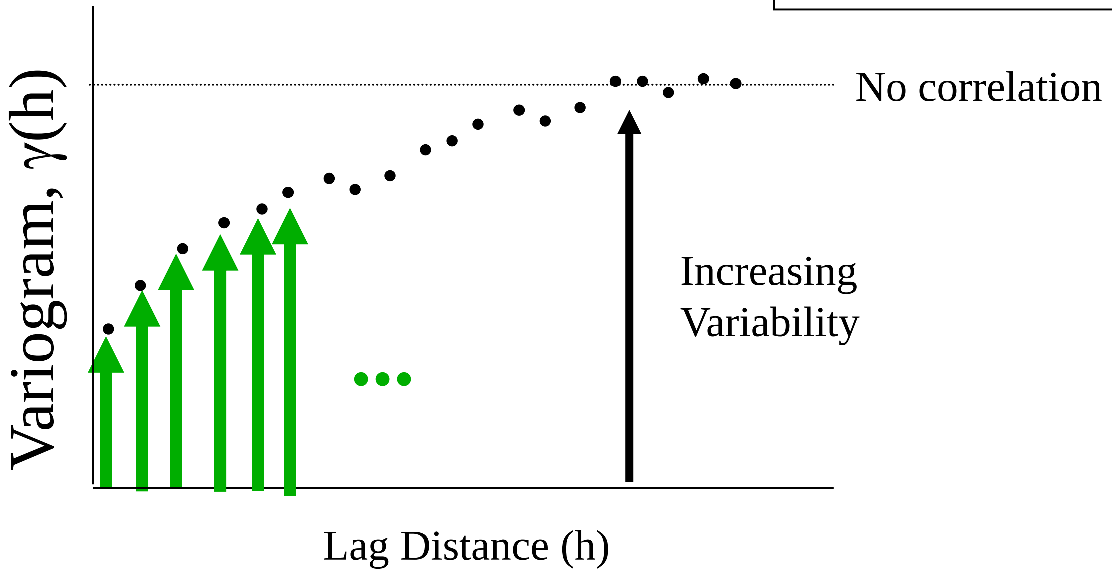
Move to next node.

...

Calculating Experimental Variograms

Now Repeat for All Nodes

And Repeat for All Lags



Some Options

- Data transformation:
  - Transform a continuous variable to a Gaussian or normal distribution, for use with / consistency with Gaussian simulation methods
  - Transform a categorical variable to a series of indicator variables for indicator methods and categorical to continuous for truncated Gaussian methods
- Coordinate transformation:
  - Variograms are calculated aligned with the stratigraphic framework
  - Otherwise the spatial continuity will be underestimated
- Should calculate the variogram on the variable being modeled with transforms (data and coordinates)
- Calculate the variogram, as this is what we model and apply in estimation and simulation (more later).

Choosing the Directions

- Inspect the data and interpretations, sections, plan views, …
- Azimuth angles 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 with ellipsoidal change in the other directions:

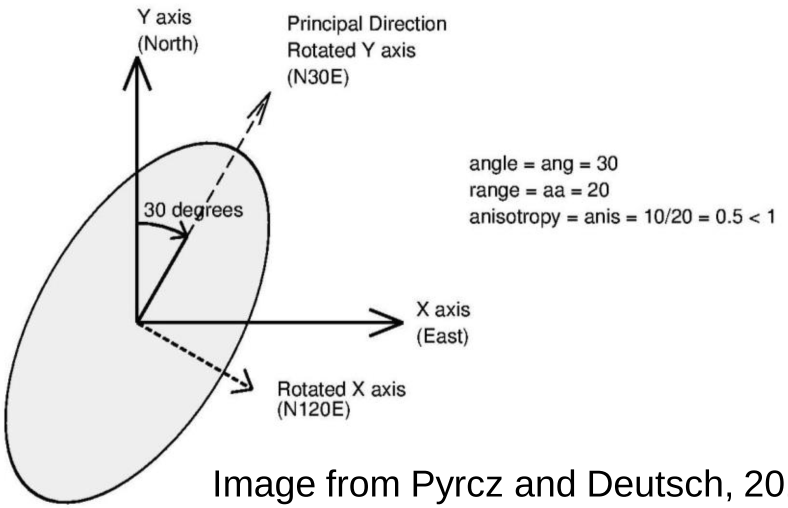


Image from Pyrcz and Deutsch, 2014

Choosing the Lag Distances and Tolerances

Guidance for 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, results in more pairs.
- The variogram is only valid for a distance one half of the field size: start leaving data out of calculations with larger distances

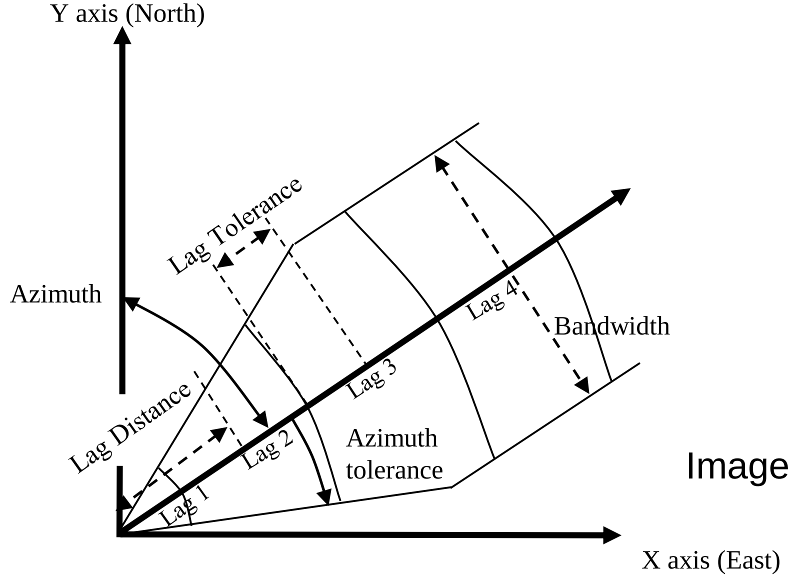
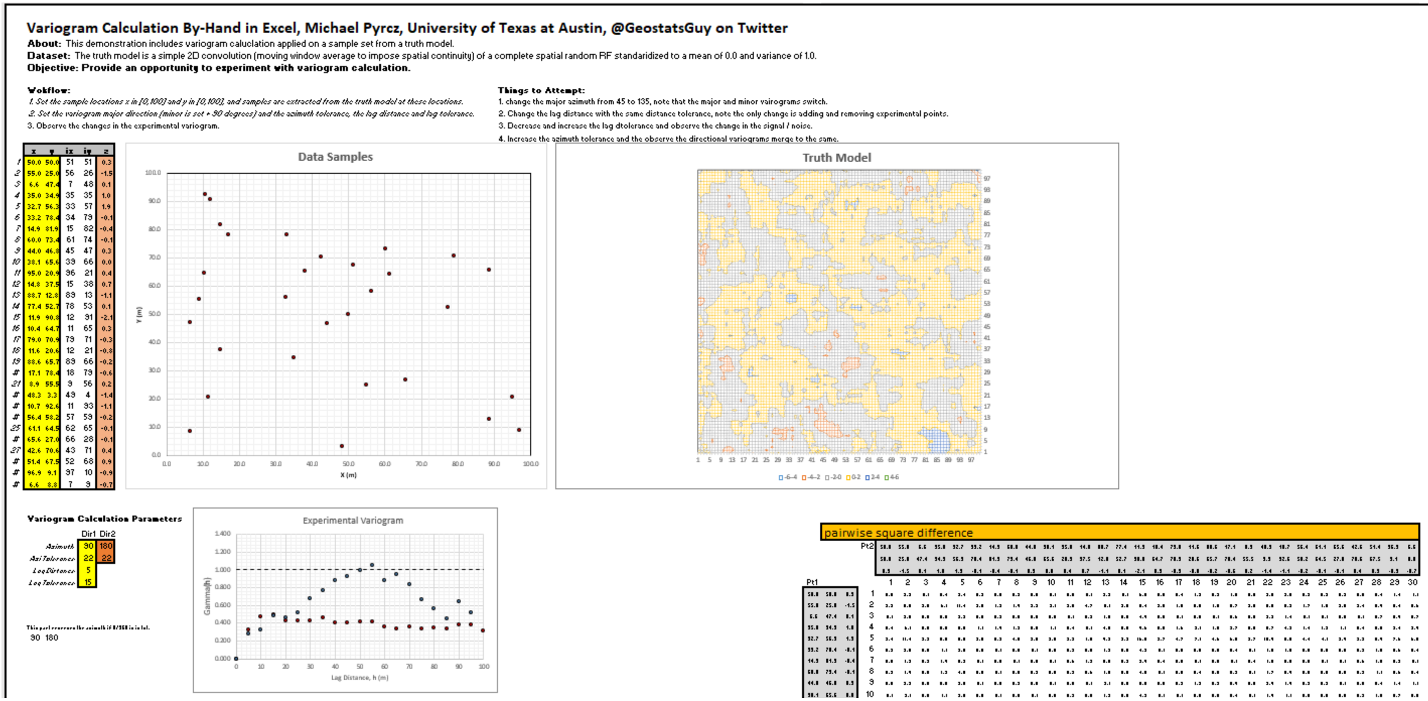


Image from Pyrcz and Deutsch, 2014

Spatial Calculation in Hands-on in Excel

Experiment with Variogram Calculation:



Things to try:

1. Set azimuth / lag tolerance small and large, 2. change directions.

Demo available in Variogram Calc_Model_Demo_v2.0.xlsx.

Spatial Calculation in Hands-on in Python

Experiment with Variogram Calculation:

Things to Try:

Variogram maps

- Relate to the data location maps

Directional variograms

- Change lag tolerance
- Change lag distance

GeostatsPy: Spatial Continuity Directions for Subsurface Data Analytics in Python

Michael Pyrcz, Associate Professor, University of Texas at Austin

Twitter | *GitHub* | *Website* | *GoogleScholar* | *Book* | *YouTube* | *LinkedIn*

PGE 383 Exercise: Methods to Detect Directions of Continuity with GeostatsPy

Here's a simple workflow on detecting the major spatial continuity directions in a spatial dataset with variogram analysis. This information is essential to optimum well placement and prectiction away from wells. First let's explain the concept of spatial continuity and the variogram.

Spatial Continuity

Spatial Continuity is the 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.

We need a statistic to quantify spatial continuity! A convenient method is the Semivariogram.

The Semivariogram

Function of difference over distance.

- The expected (average) squared difference between values separated by a lag distance vector (distance and direction), h:

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

where $z(\mathbf{u}_\alpha)$ and $z(\mathbf{u}_\alpha + \mathbf{h})$ are the spatial sample values at tail and head locations of the lag vector respectively.

- Calculated over a suite of lag distances to obtain a continuous function.
- the $\frac{1}{2}$ term converts a variogram into a semivariogram, but in practice the term variogram is used instead of semivariogram.
- We prefer the semivariogram because it relates directly to the covariance function, $C_x(\mathbf{h})$ and univariate variance, σ_x^2:

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

Note the correlogram is related to the covariance function as:

$$\rho_x(\mathbf{h}) = \frac{C_x(\mathbf{h})}{\sigma_x^2}$$

The correlogram provides of function of the $\mathbf{h} - \mathbf{h}$ scatter plot correlation vs. lag offset $\mathbf{h}$.

$$-1.0 \le \rho_x(\mathbf{h}) \le 1.0$$

Variogram Observations

The following are common observations for variograms that should assist with their practical use.

Observation #1 - As distance increases, variability increase (in general).

Use Python notebook file: GeostatsPy_spatial_continuity_directions.ipynb

Spatial Continuity New Tools

Topic	Application to Subsurface Modeling
Stationarity	In the presence of multivariate relationships, must jointly model variables. *Summarize with bivariate statistics, and visualize and use conditional statistics to go beyond linear measures.*
Random Variables and Functions	Random variables and random functions are used to represent spatial uncertainty. *Porosity at a pre-drill location has the uncertainty model based on a random variable with Gaussian mean of 15% and standard deviation of 3%.*
Variogram Calculation	*Calculate spatial continuity from spatial data to use for spatial prediction.* *From the available wells the porosity spatial continuity range is 300 m in the 030 azimuth, we have no information beyond this spacing from existing wells.*

Multivariate Modeling: Spatial Continuity

Lecture outline . . .

- **Variogram Calculation**

Multivariate Modeling: Spatial Continuity

Lecture outline . . .

- **Random Variables and Functions**
- **Stationarity**
- **Spatial Continuity**
- **Variogram Calculation**
- **Basic Variogram Interpretation**