PGE 338 Data Analytics and Geostatistics

Lecture 10: Spatial Continuity

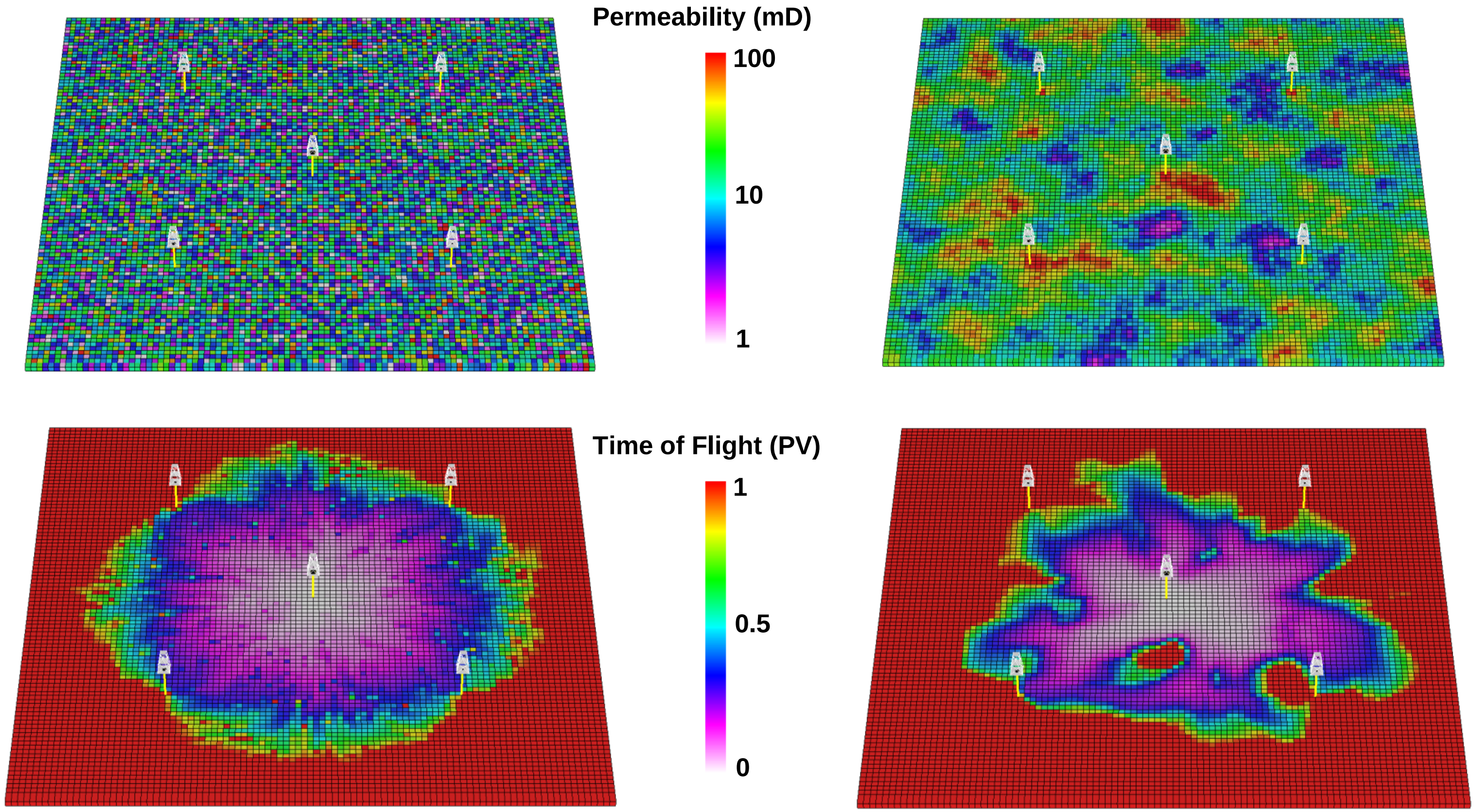
Lecture outline . . .

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

Motivation

Spatial continuity has a significant impact on reservoir flow rates and overall recovery.

- We need to quantify spatial continuity and predict / forecast with it.



Two models with very different spatial continuity permeability (above), and time of flight from fast marching (below).

Recorded Lectures

PGE 338 Data Analytics and Geostatistics

Lecture 10: Spatial Continuity

Lecture outline . . .

- Random Variables and Functions

Introduction

General Concepts

Univariate

Bivariate

Spatial

Calculation

Variogram Modeling

Kriging

Simulation

Time Series

Machine Learning

Uncertainty Analysis

Now We Begin Spatial / Geostatistics

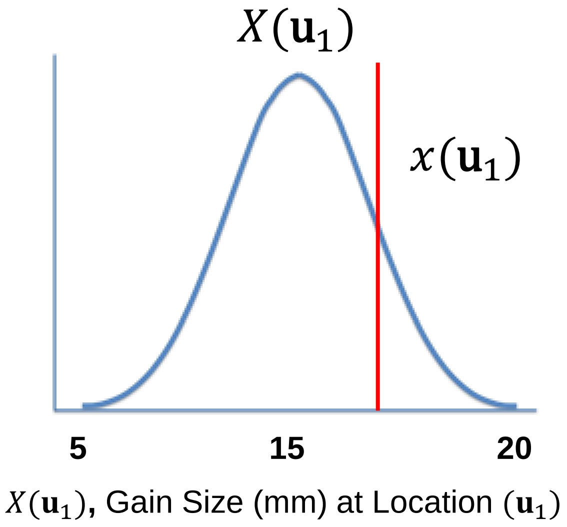
- We Build on the past lectures:
  - probability theory
  - univariate and bivariate statistics
  - sampling and bootstrap

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

Recall: 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., $x(\mathbf{u})$, $X(\mathbf{u})$

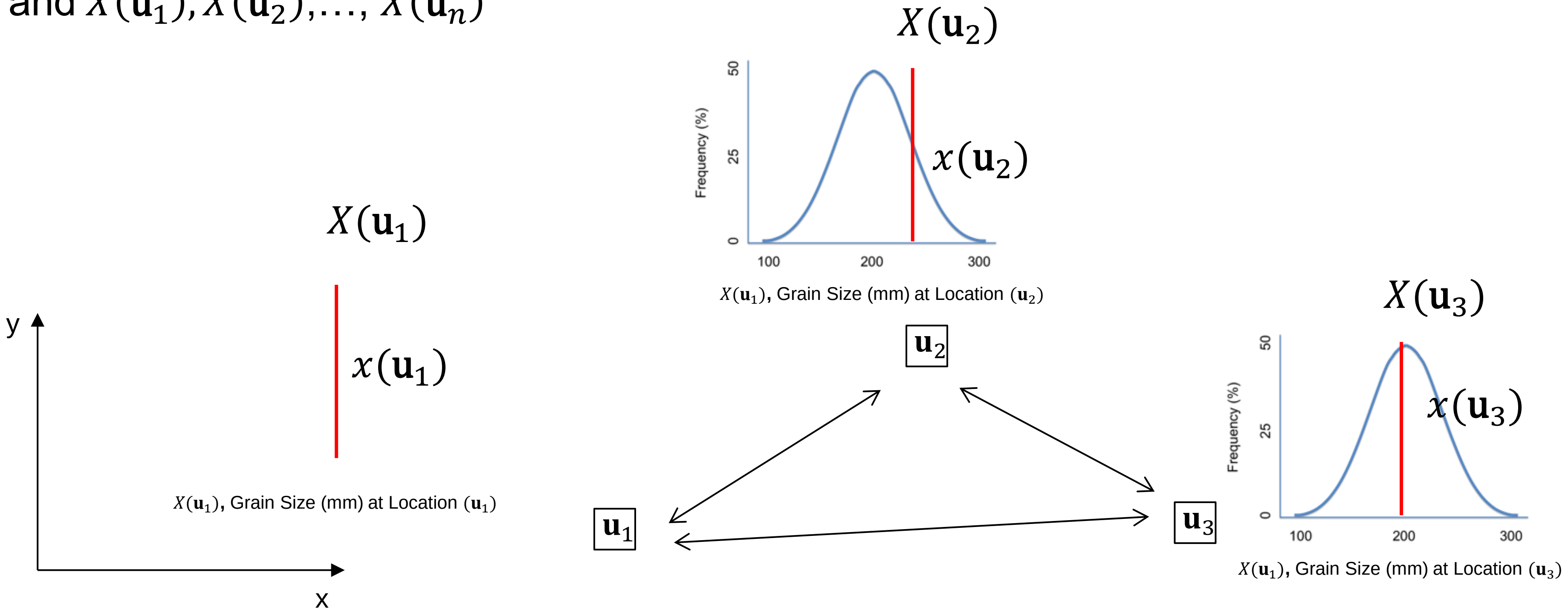


Example of a random variable at location $\mathbf{u}_1$.

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 $X(\mathbf{u}_1), X(\mathbf{u}_2), \ldots, X(\mathbf{u}_n)$

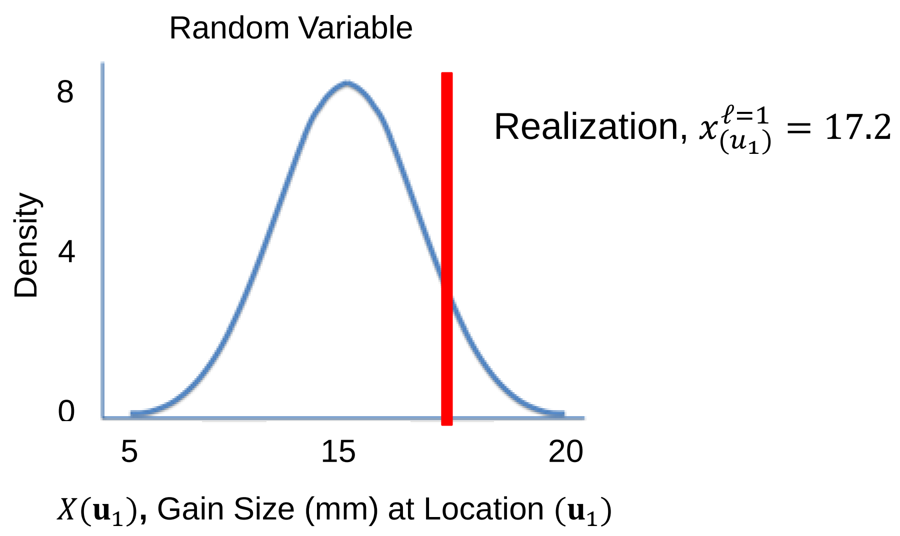


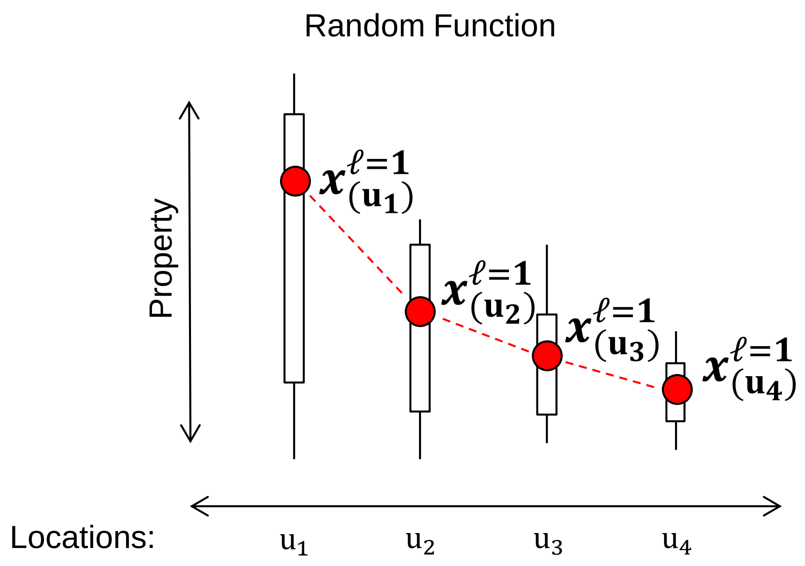
Example of a random function over locations $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$.

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, methods to sample from the RVs or RFs.
- each realization is considered equiprobable



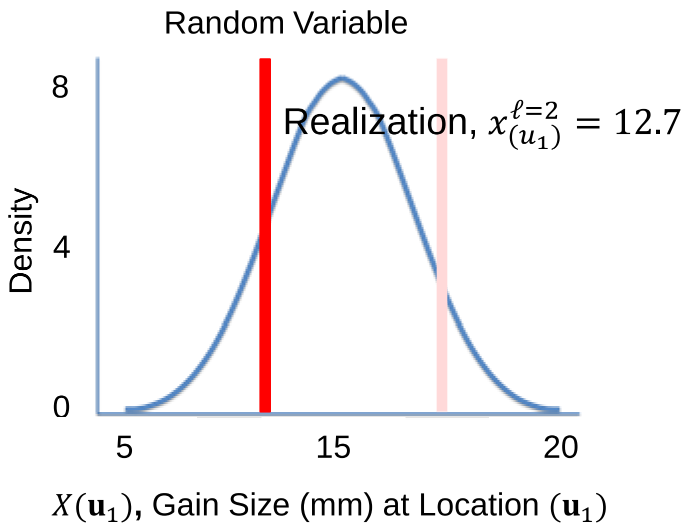


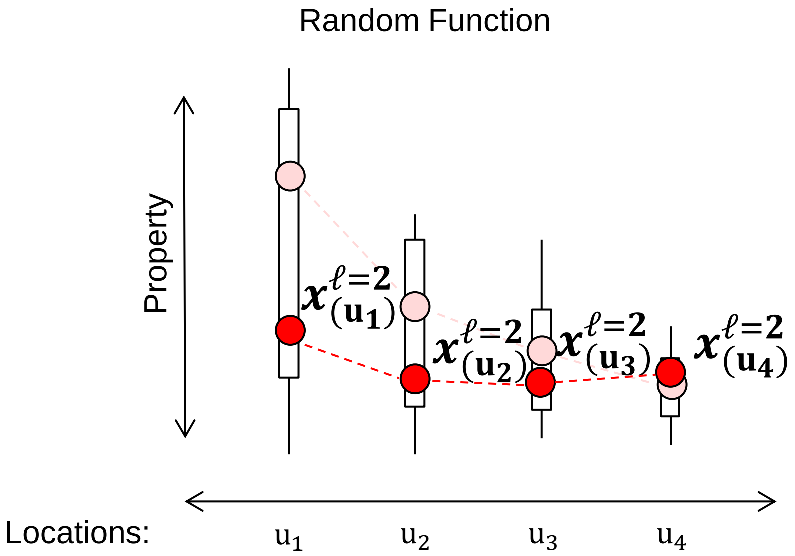
Example realizations from a random variable (left) and a random function (right).

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, methods to sample from the RVs or RFs.
- each realization is considered equiprobable



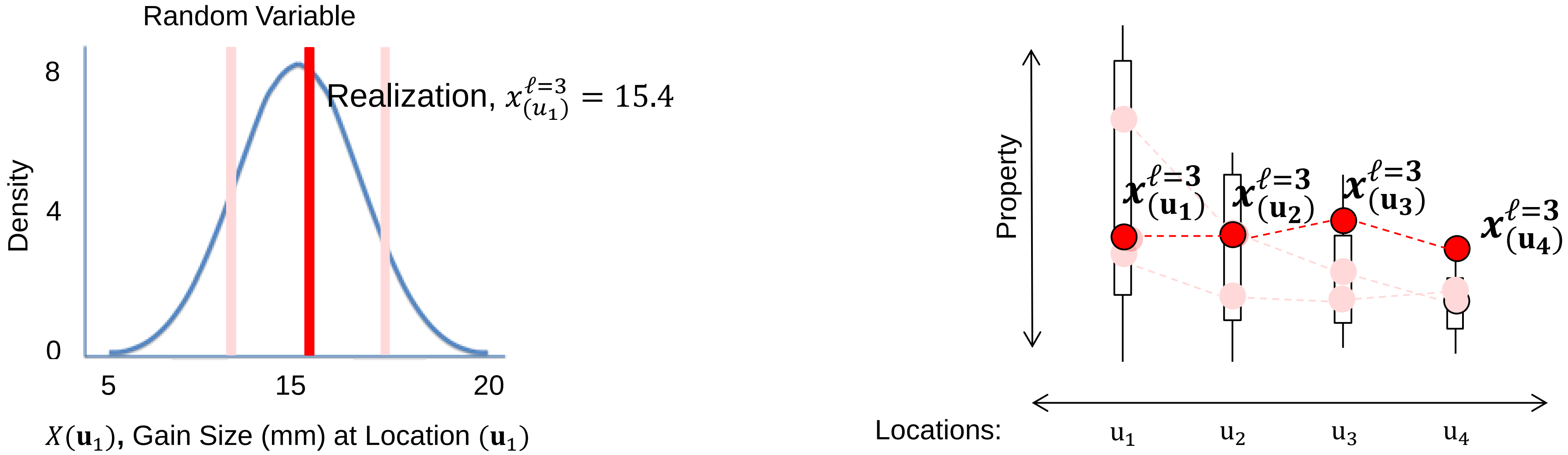


Example realizations from a random variable (left) and a random function (right).

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, methods to sample from the RVs or RFs.
- each realization is considered equiprobable



Example realizations from a random variable (left) and a random function (right).

PGE 338 Data Analytics and Geostatistics

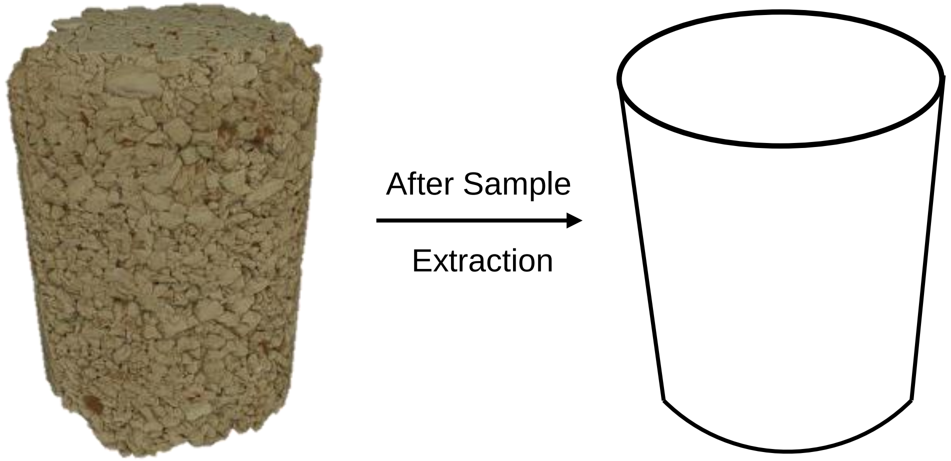
Lecture 10: Spatial Continuity

Lecture outline . . .

- Stationarity

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.



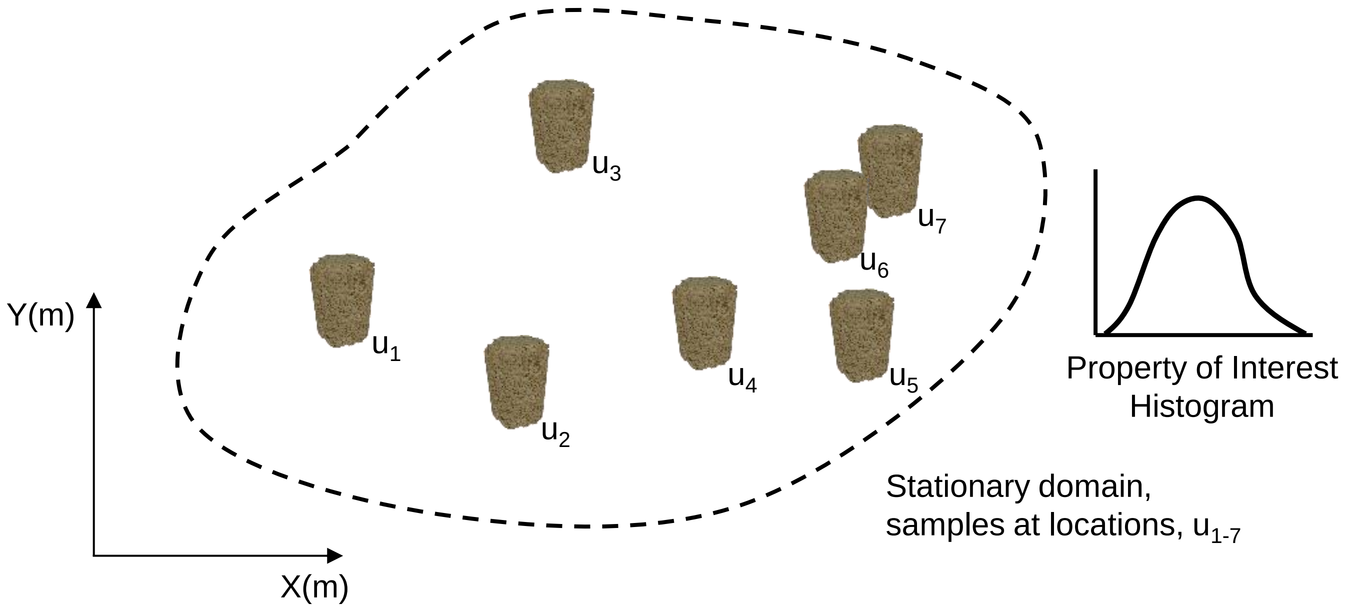
Extracted Sample at Location, u_1 Hole Remains at Location, u_1

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**.



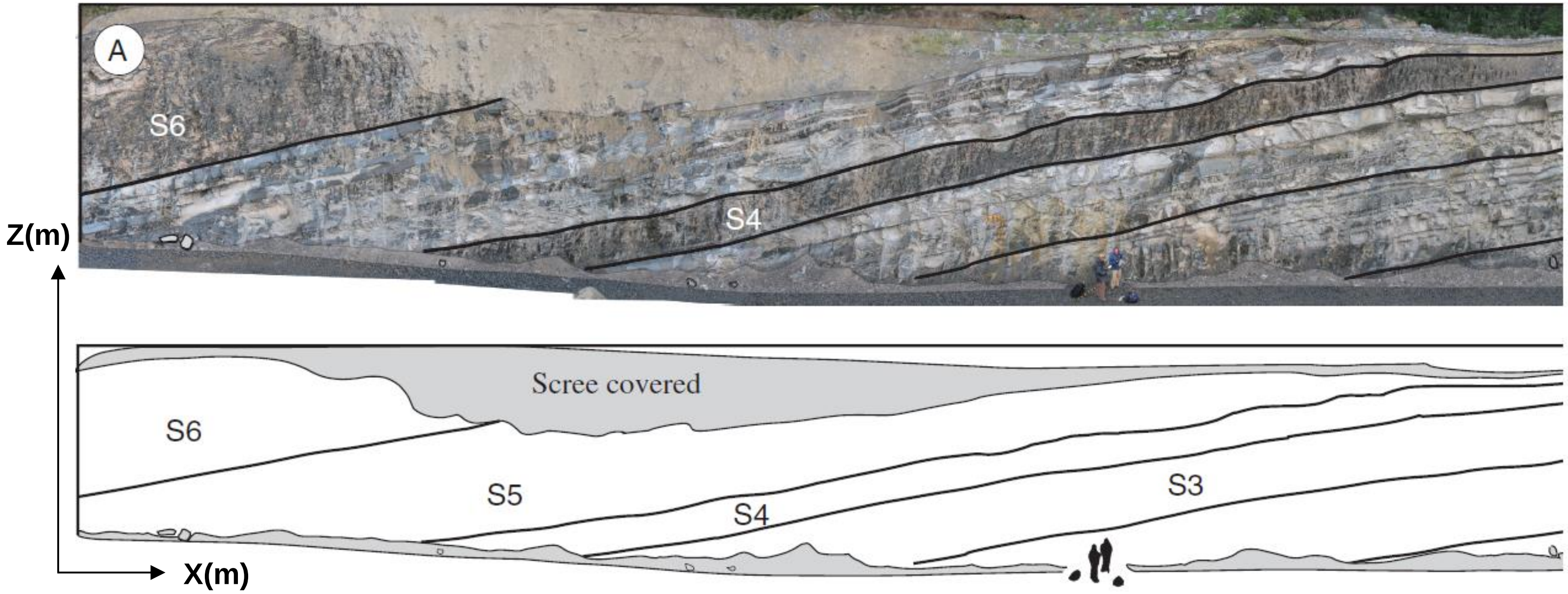
Core samples pooled over space, statistic of interest and boundary to apply the statistic.

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

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

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

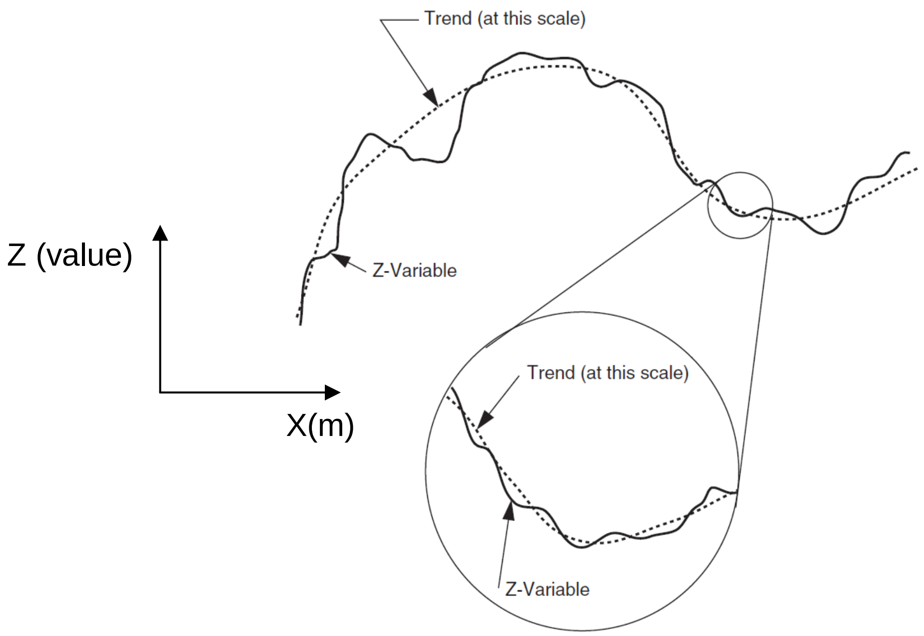
Stationarity: *What metric / statistic? Over what volume?*

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 a 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.



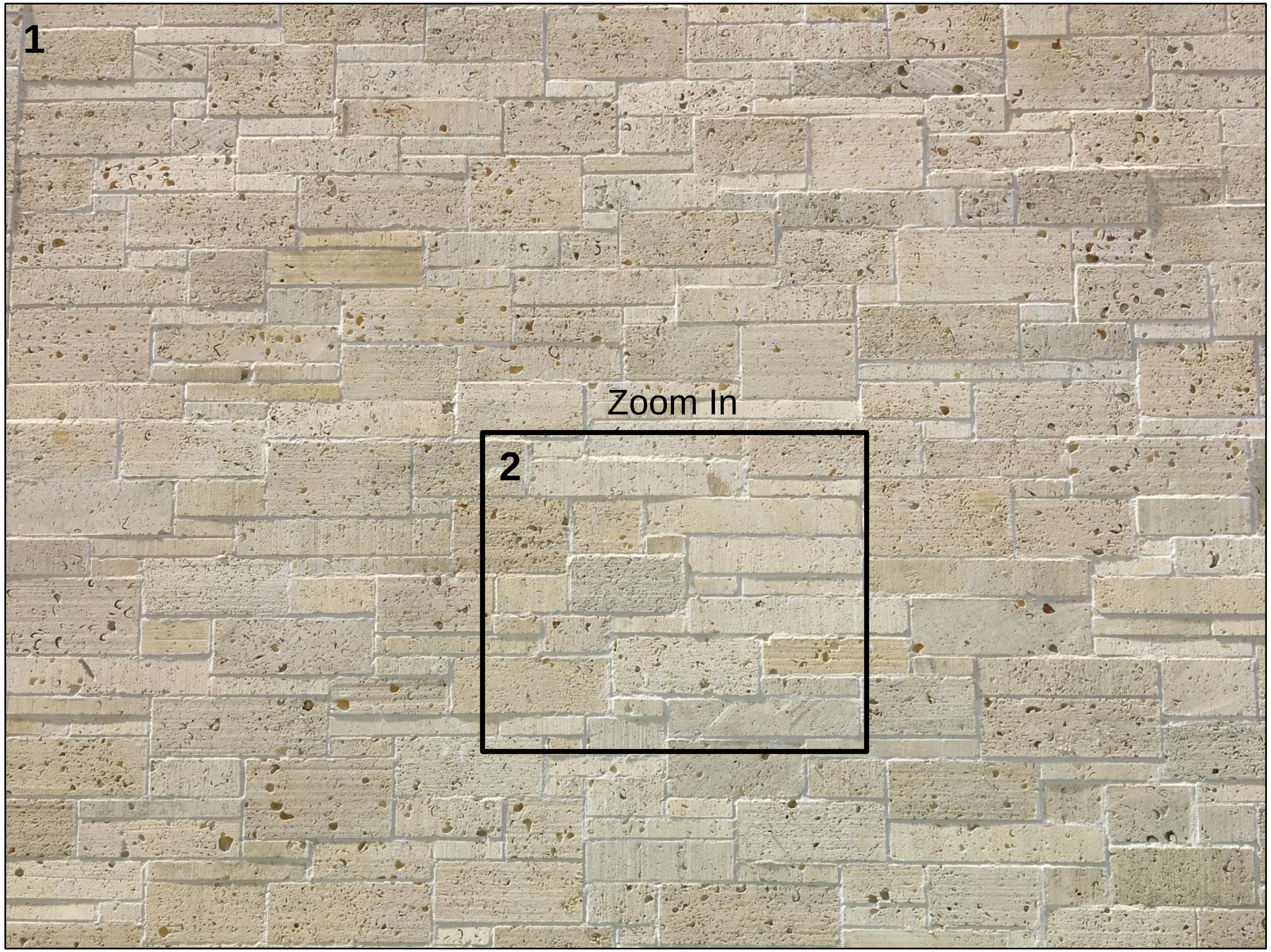
Scales of stationarity, modified from Pyrcz and Deutsch (2014).

Stationarity Example

Is this image stationary?

Stationarity Example

Let’s zoom in.



Stationarity Example

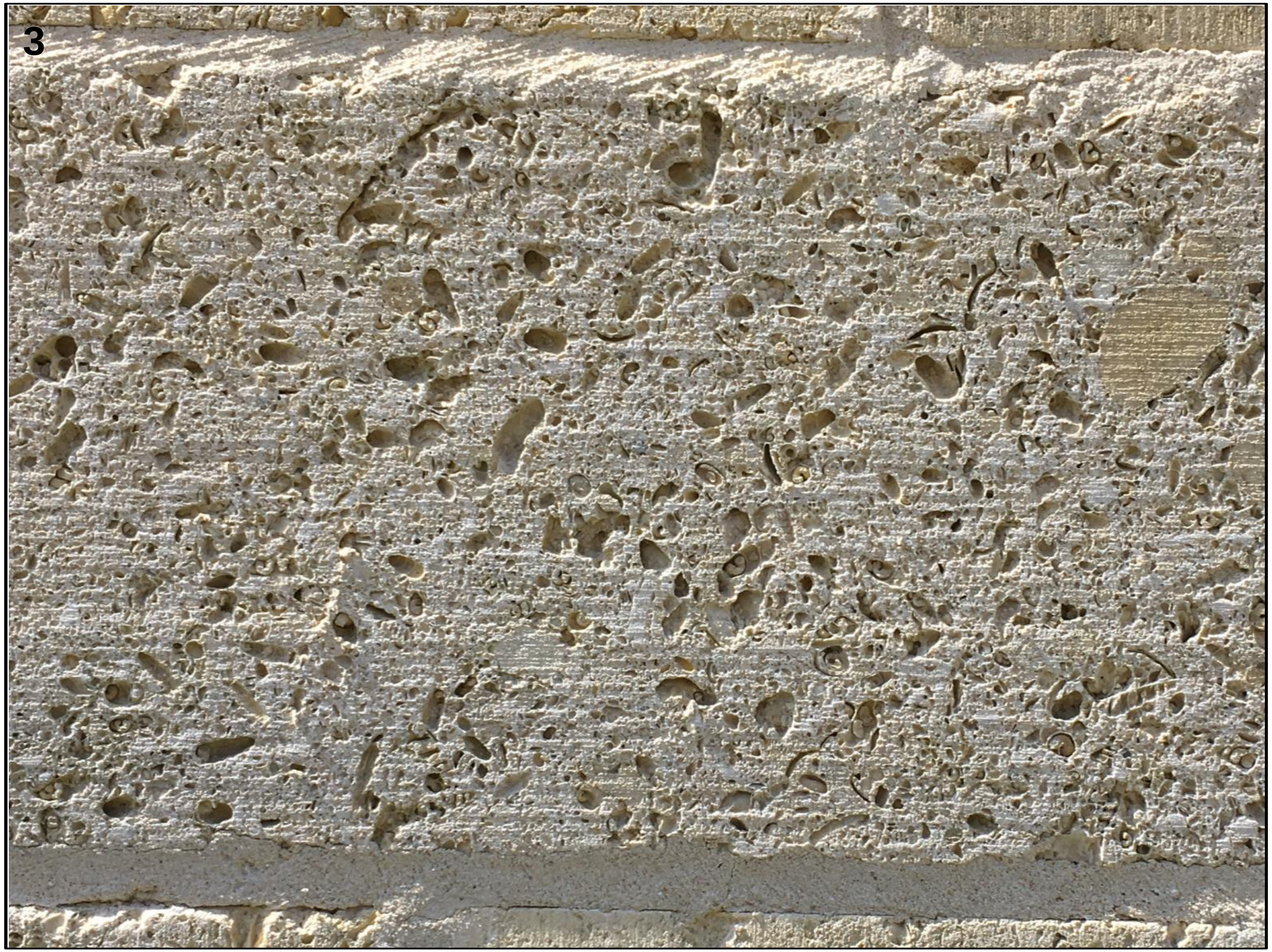
Is this image stationary?

Stationarity Example

Let's zoom in.

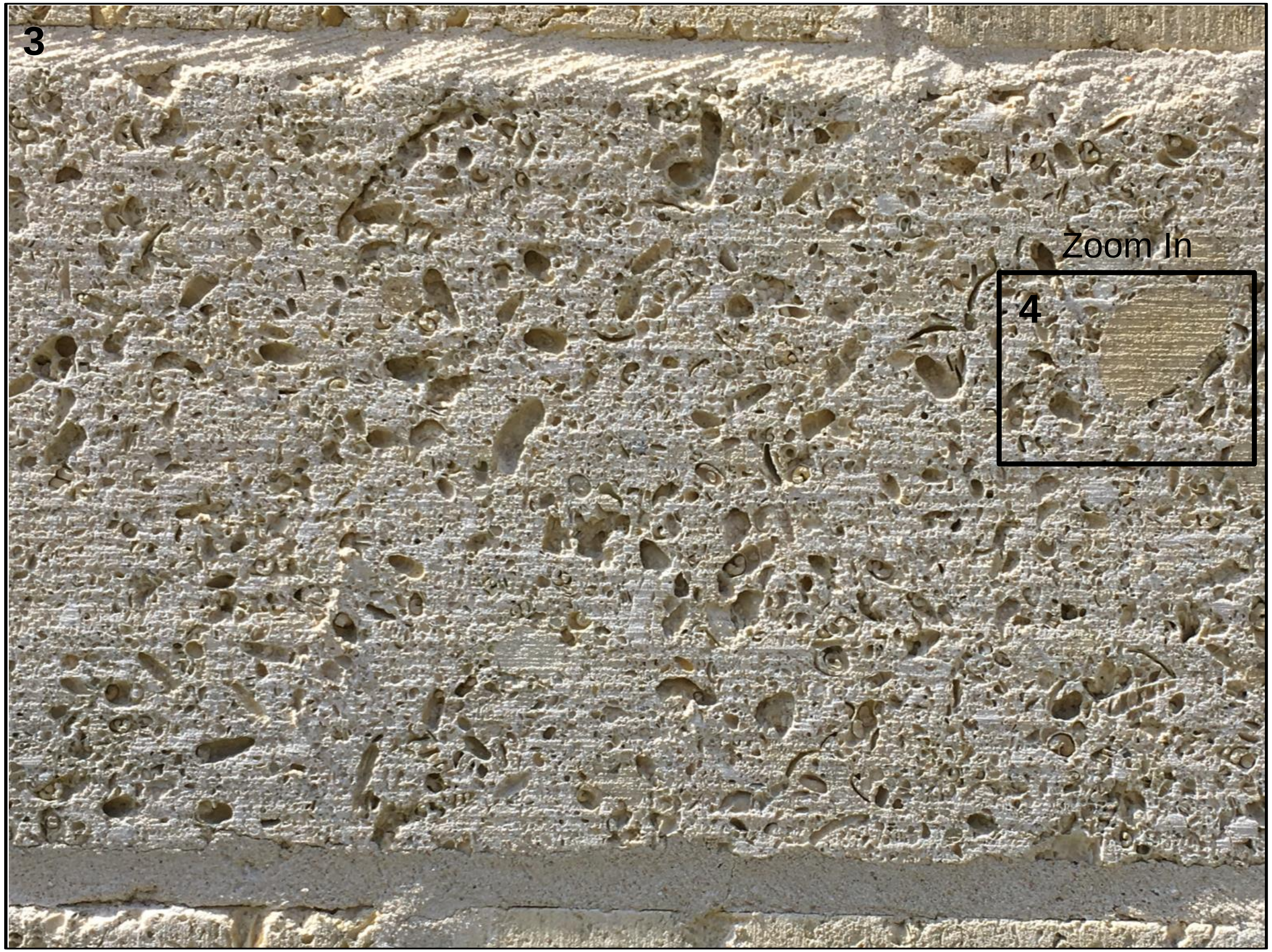
Stationarity Example

Is this image stationary?



Stationarity Example

Let's zoom in.



Stationarity Example

Is this image stationary?

Stationarity Example

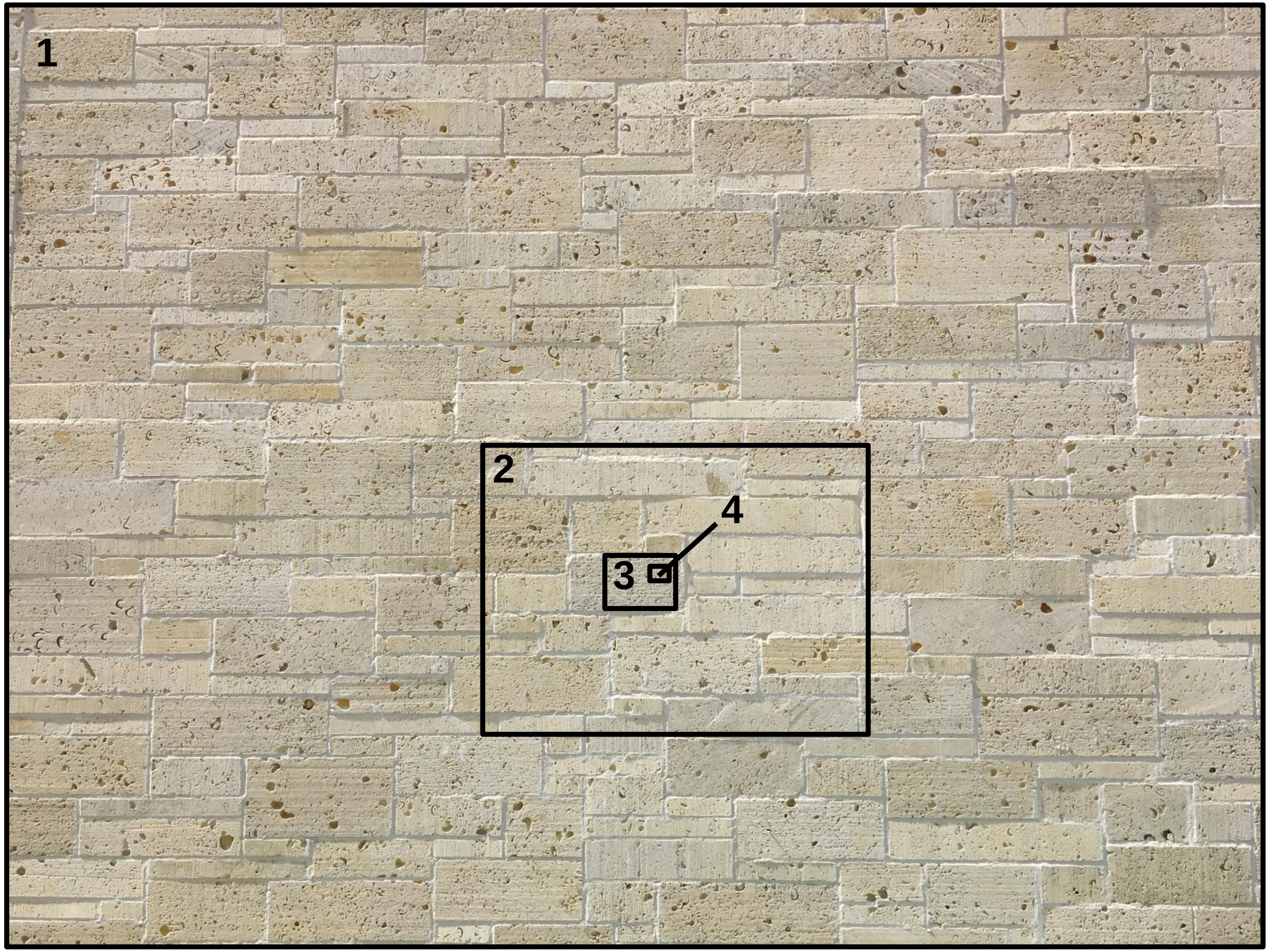
What metric?

- e.g., brick size, brick color, fossils, brick size anisotropy

What scale?

1. stationary
2. nonstationary
3. stationary
4. nonstationary

Stationarity depends on metric and scale.



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., geological unit, mapped region, early production 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 a 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 infer $F_x(x;\mathbf{u}_\alpha)$ (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 the metric and the scale, area of investigation.

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

Michael Pyrcz, University of Texas at Austin, @GeostatsGuy

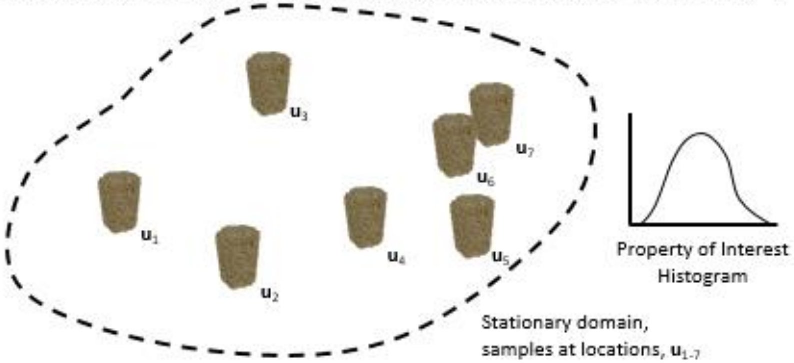
A description of the concepts of stationarity that are central to collecting geoscience information and applying it in subsurface modeling.

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".

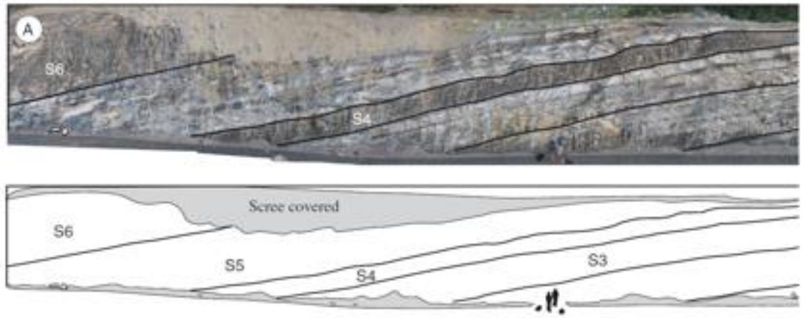


Stationary domain, samples at locations, $\mathbf{u}_{1-7}$

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

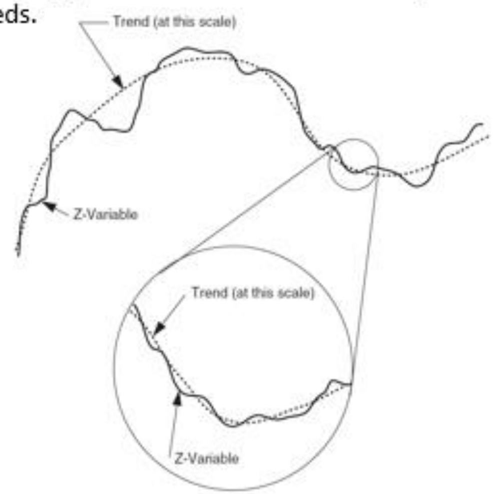
Stationary Semivariogram: $\gamma_z(\mathbf{u}; \mathbf{h}) = \gamma_z(\mathbf{h}), \ \forall \, \mathbf{u}$

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.

PGE 338 Data Analytics and Geostatistics

Lecture 10: Spatial Continuity

Lecture outline . . .

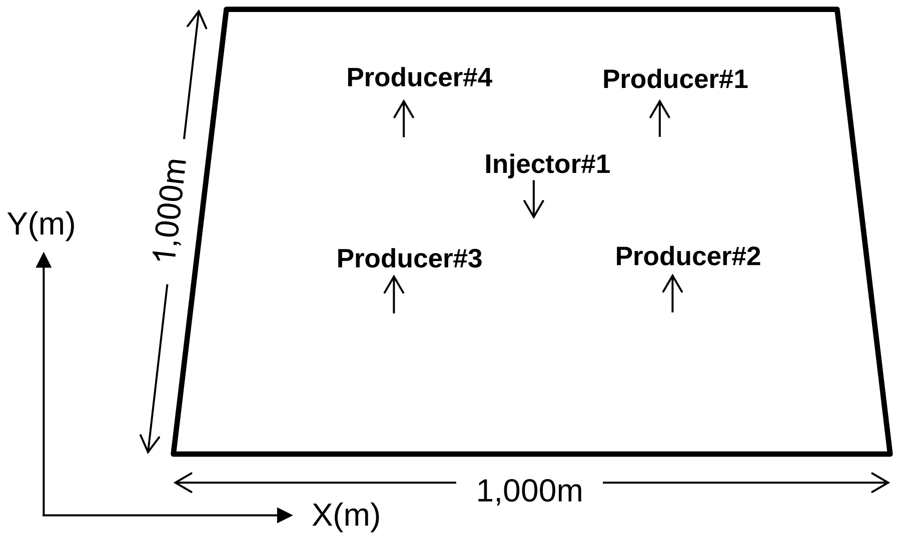
- Spatial Continuity

Motivation for Measuring Spatial Continuity

Simple Example

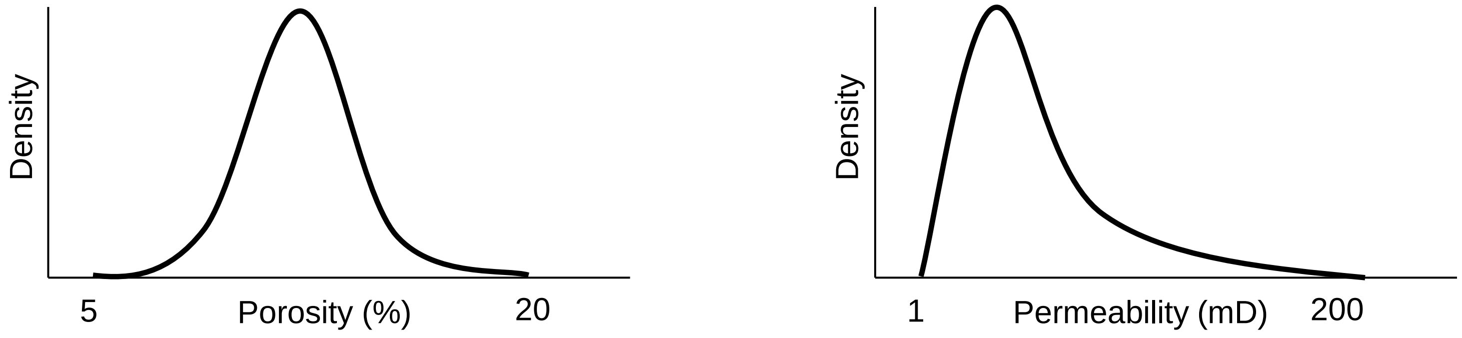
Area of interest

- 2D 1,000 m x 1,000 m
- 1 Injector and 4 producers



Area of interest, 2D model with 5 wells, 1 injector, and 4 producers.

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

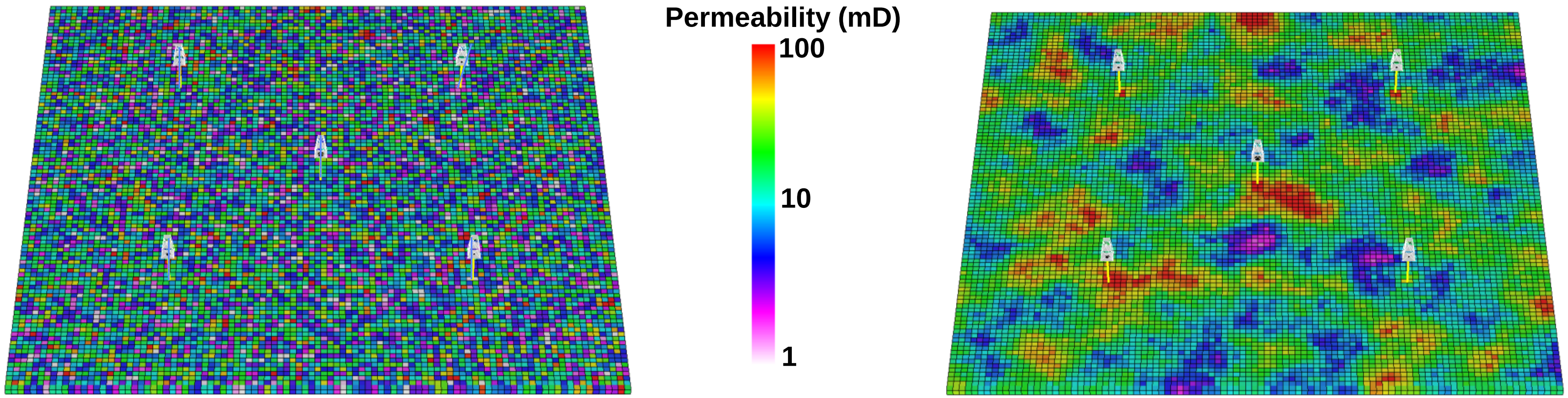


Schematics of the porosity (left) and permeability (right) probability density functions.

Motivation for Measuring Spatial Continuity

Does spatial continuity of reservoir properties matter?

Consider these models of permeability



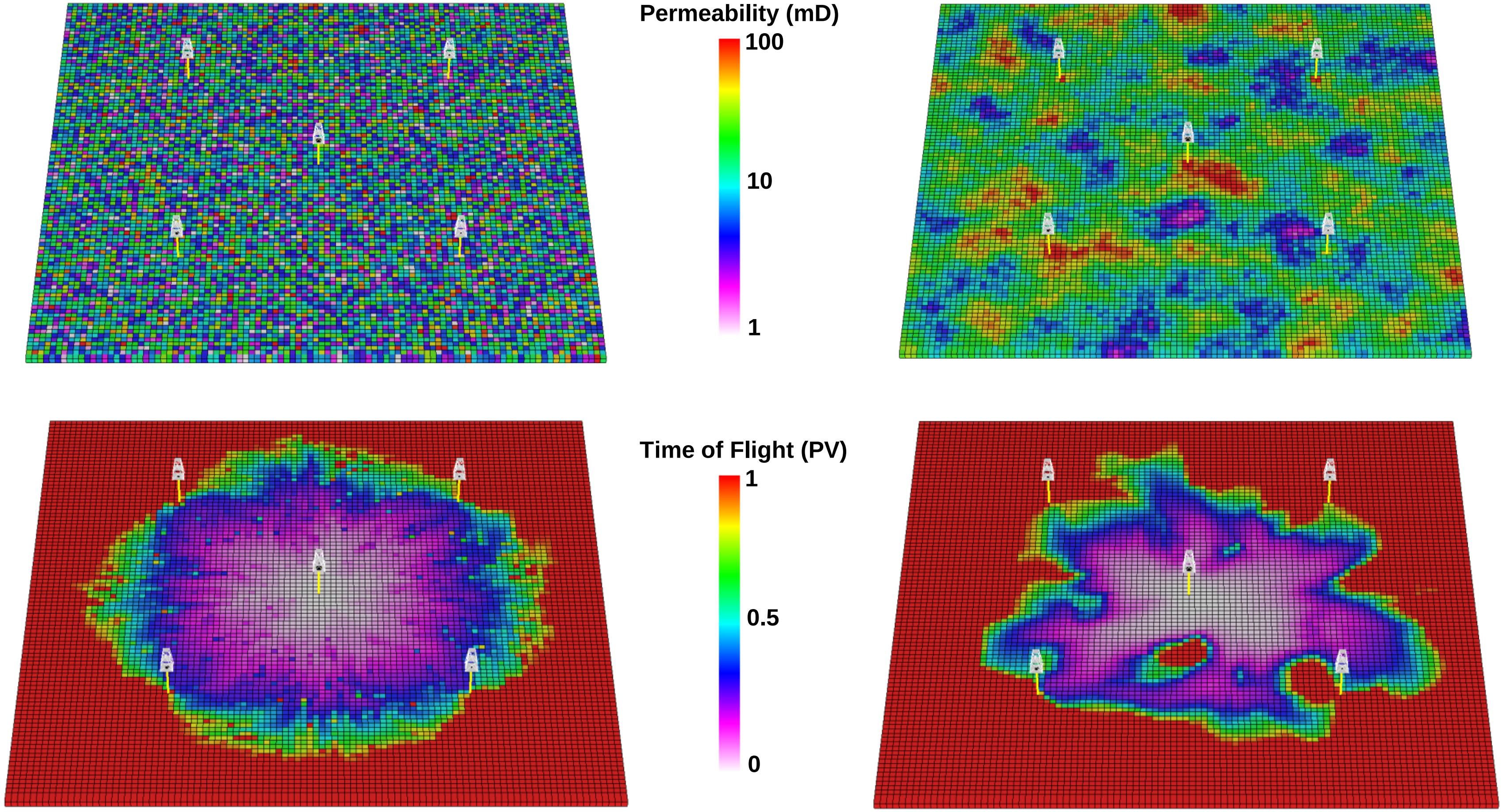
Two models with very different spatial continuity permeability.

Recall – all models have the same porosity and permeability distributions

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

Motivation for Measuring Spatial Continuity

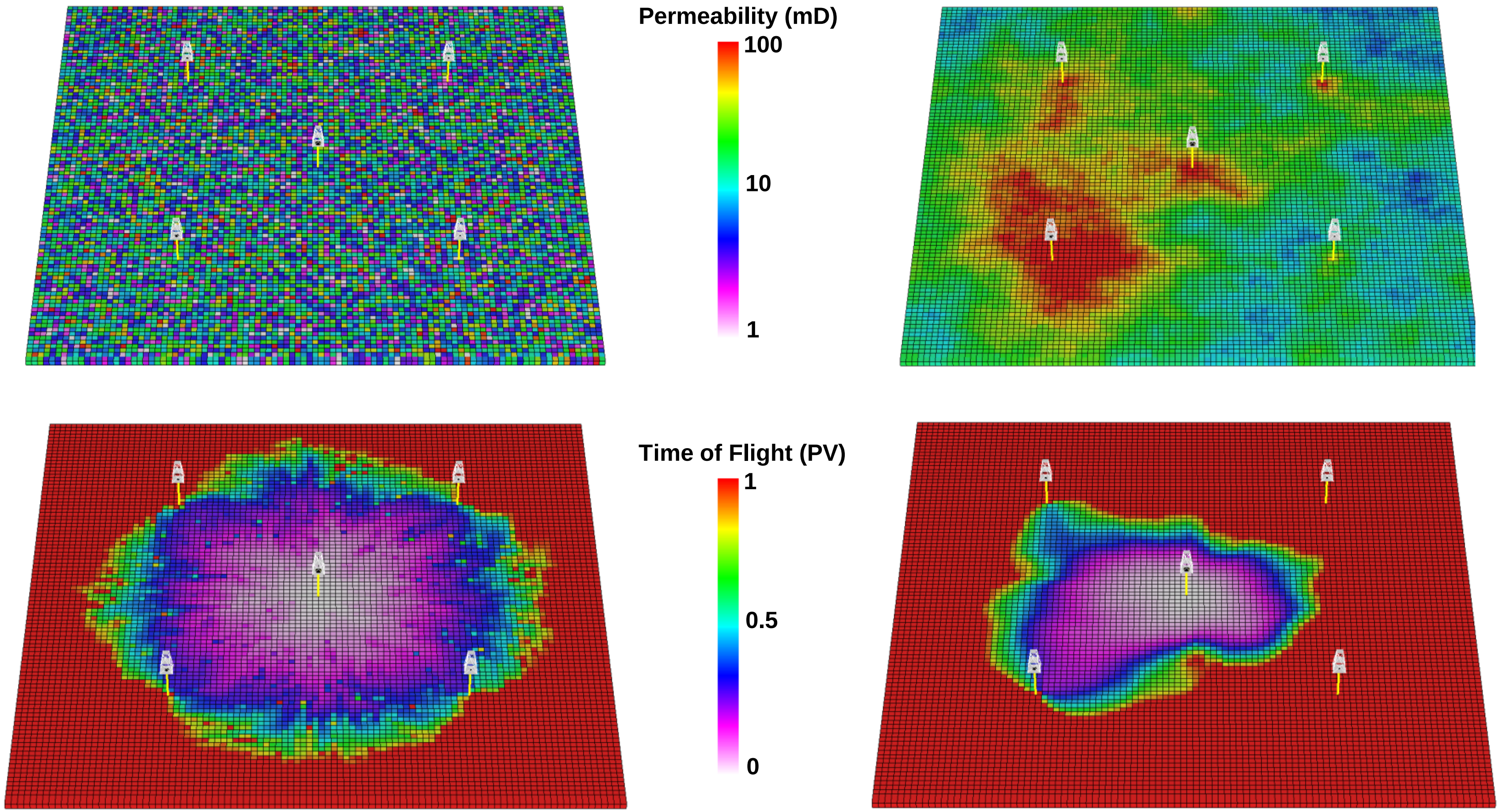
Does spatial continuity of reservoir properties matter?



Two models with very different spatial continuity permeability (above), and time of flight from fast marching (below).

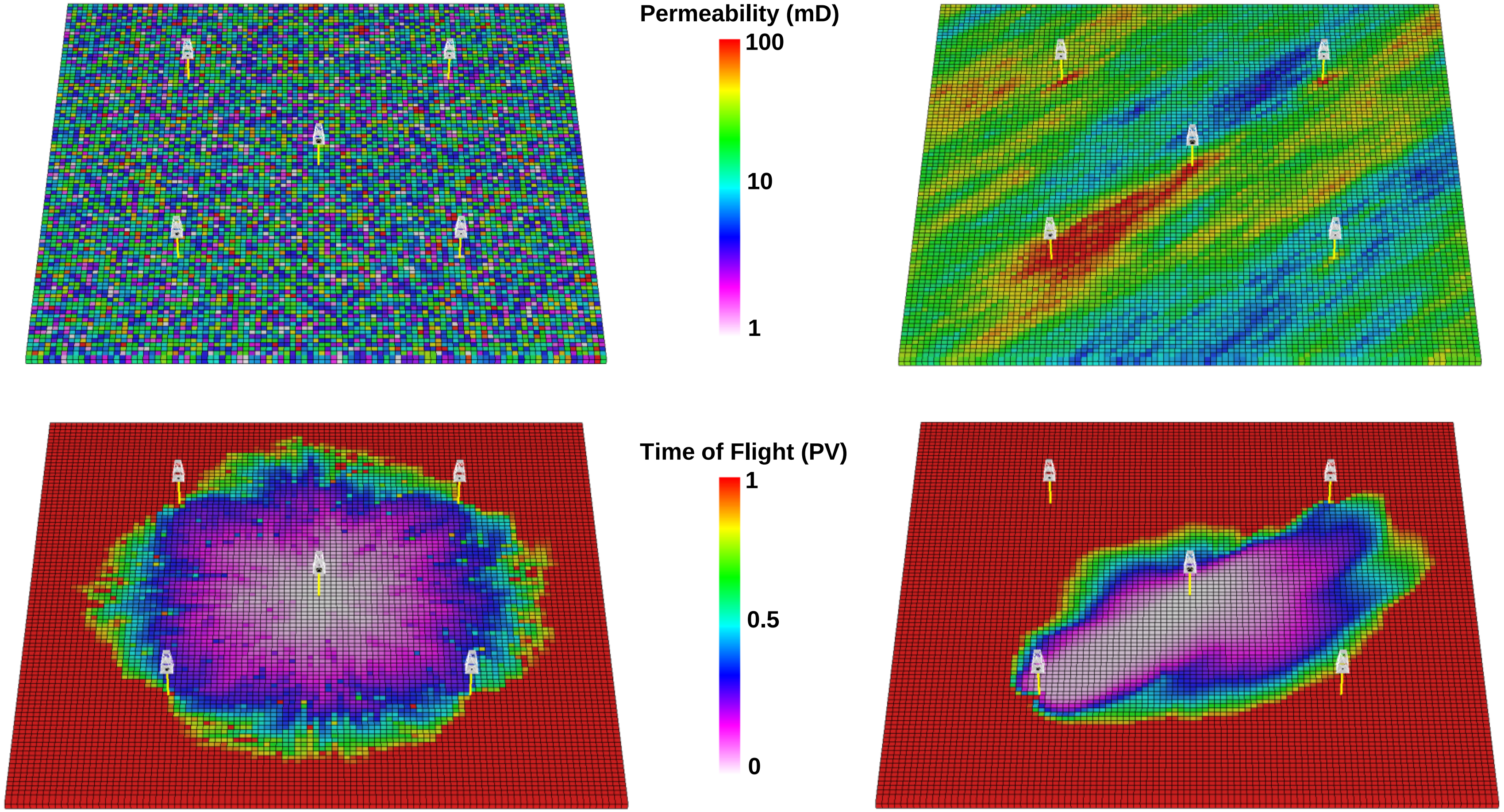
Motivation for Measuring Spatial Continuity

Does spatial continuity of reservoir properties matter?



Motivation for Measuring Spatial Continuity

Does spatial continuity of reservoir properties matter?



Motivation for Measuring Spatial Continuity

For the same reservoir feature 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

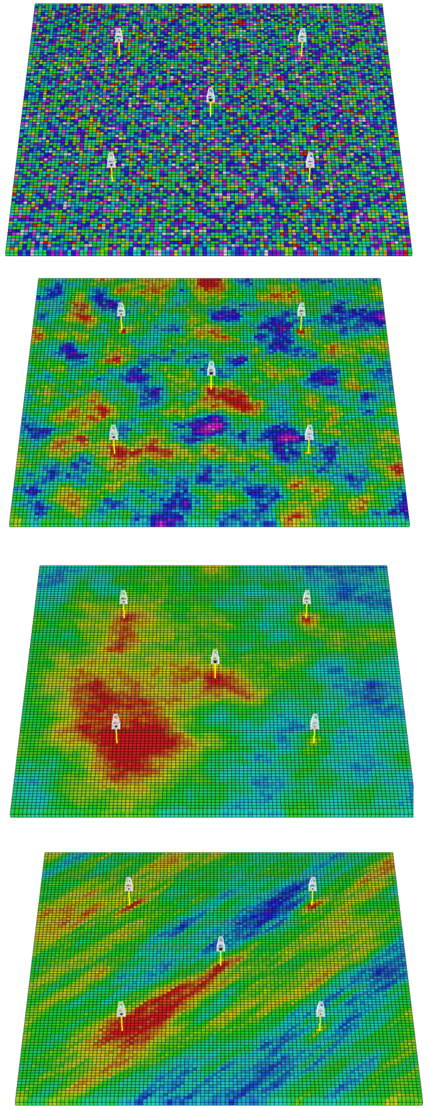
Spatial Continuity

“Very Short”

“Medium”

“Long”

“Anisotropic “



Spatial Continuity Definition

Spatial Continuity – correlation between spatial samples of a feature over distance.

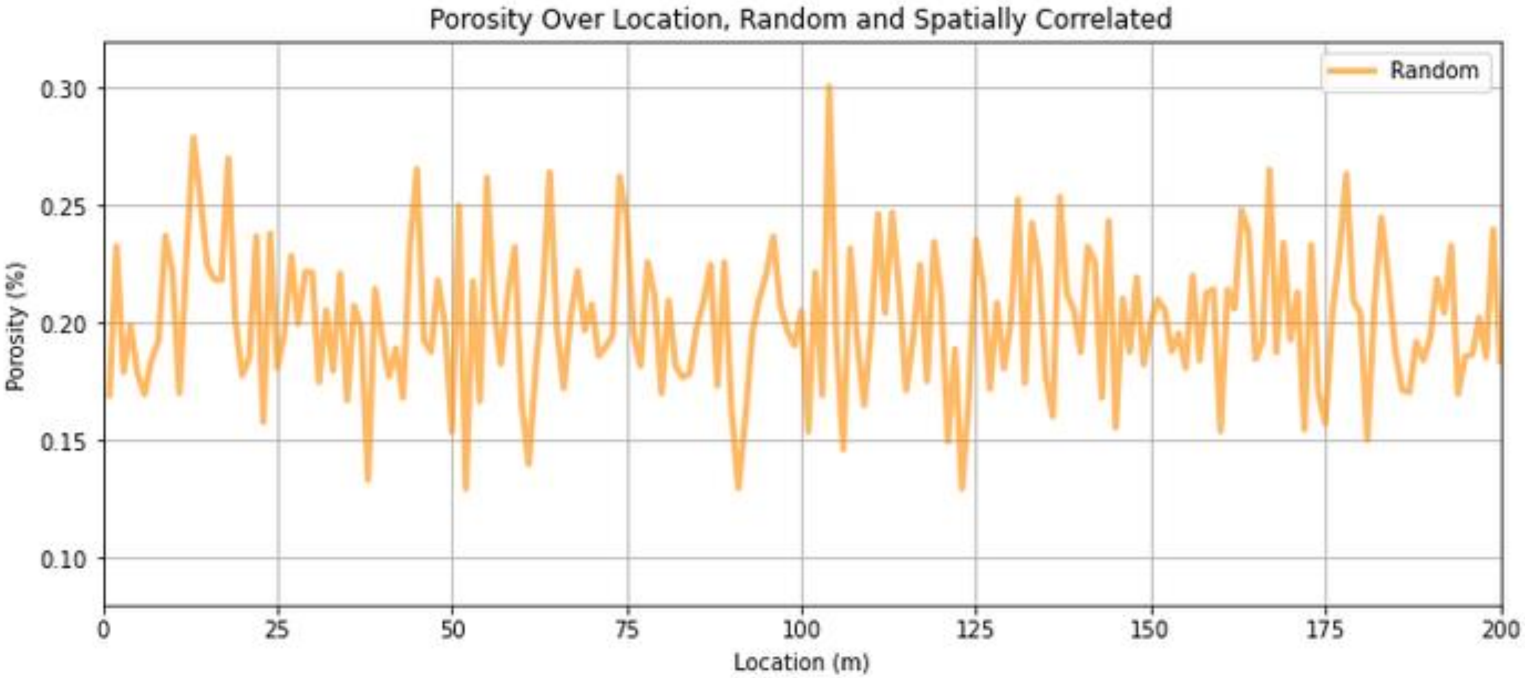
- **No spatial continuity** – no correlation between spatial samples over distance, random values at each location in space regardless of separation distance.

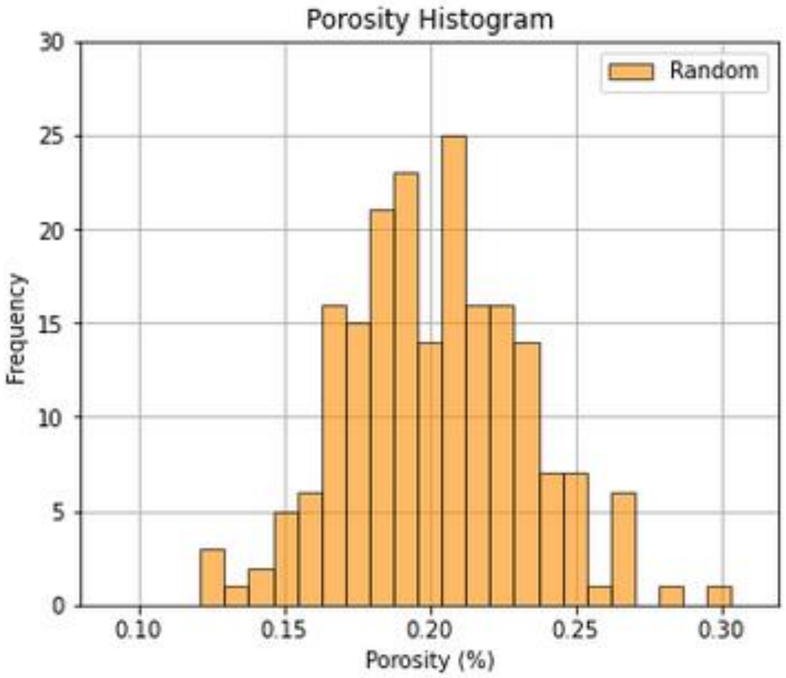
- **Homogenous phenomena** have perfect spatial continuity since all values are the same (or very similar) they are correlated.

Spatial Continuity Definition

Let's Tune Our Eyes to Spatial Continuity

- **No spatial continuity** – random values at each location in space regardless of separation distance.



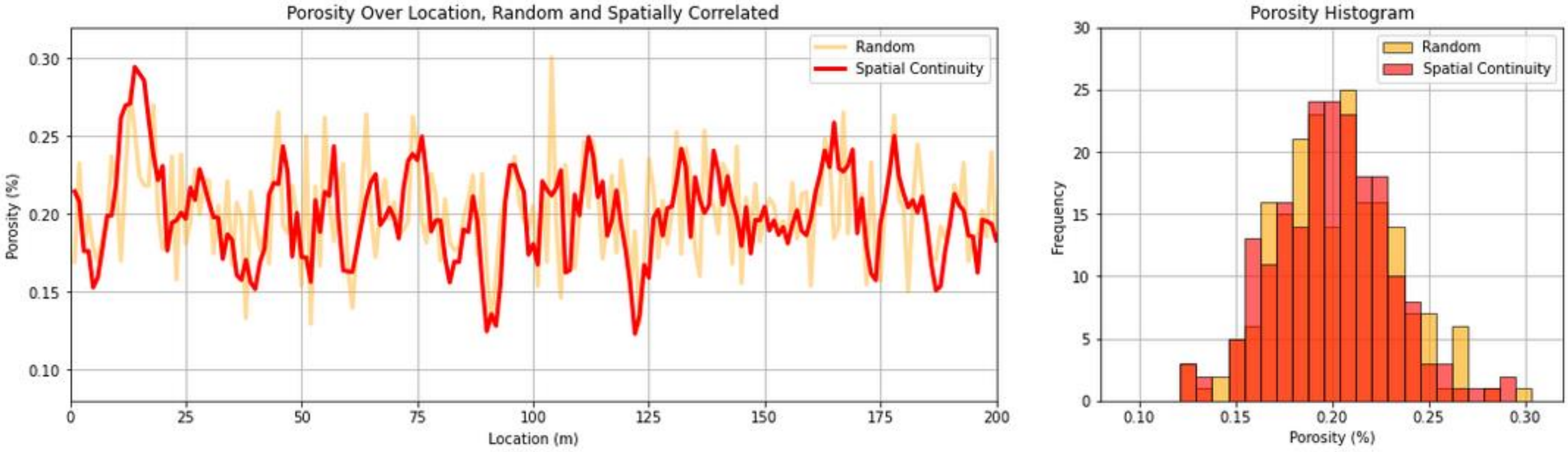


1D spatial porosity model without correlation, plot (left) and histogram (right), the file is PythonDataBasics_Visualize_Spatial_Continuity.ipynb.

Spatial Continuity Definition

Let’s Tune Our Eyes to Spatial Continuity

- **Spatial correlation over 5 meters** – used a convolution kernel to impose spatial correlation in the random model.

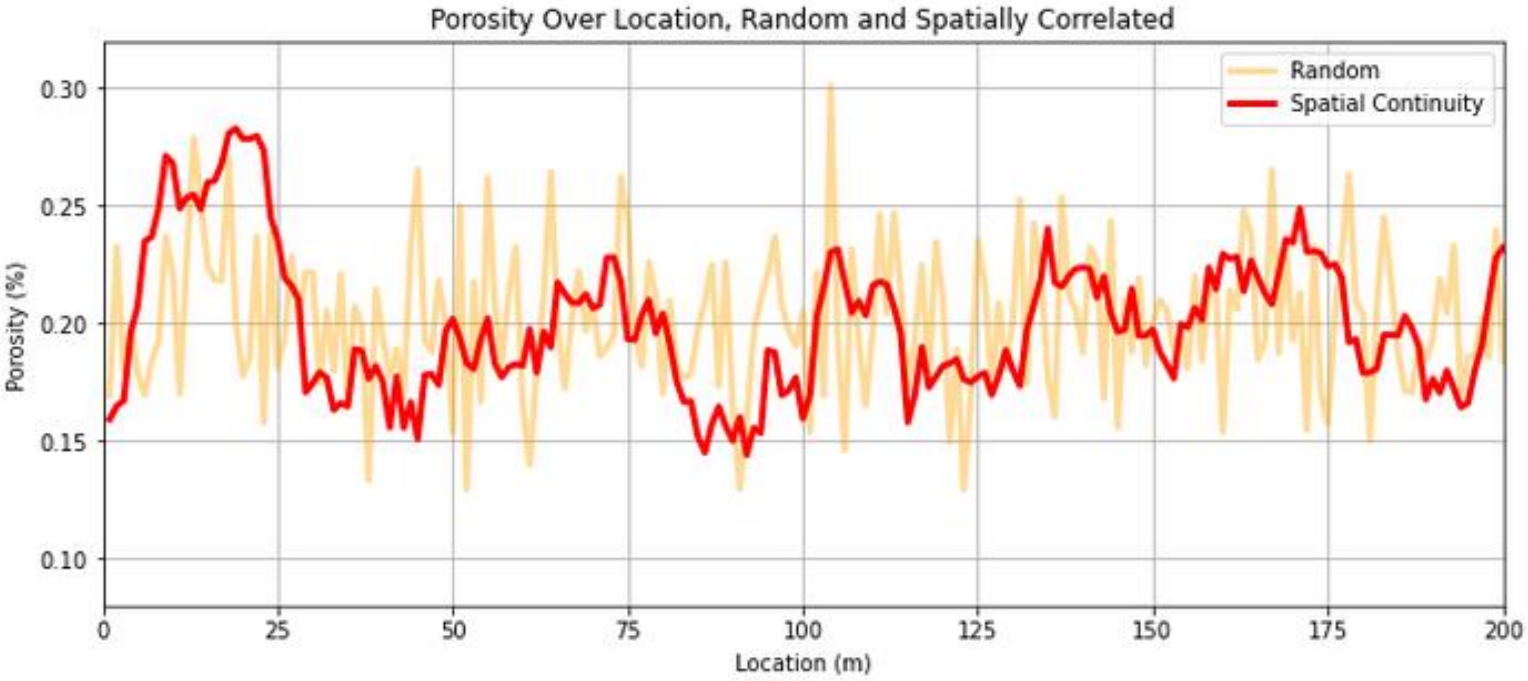


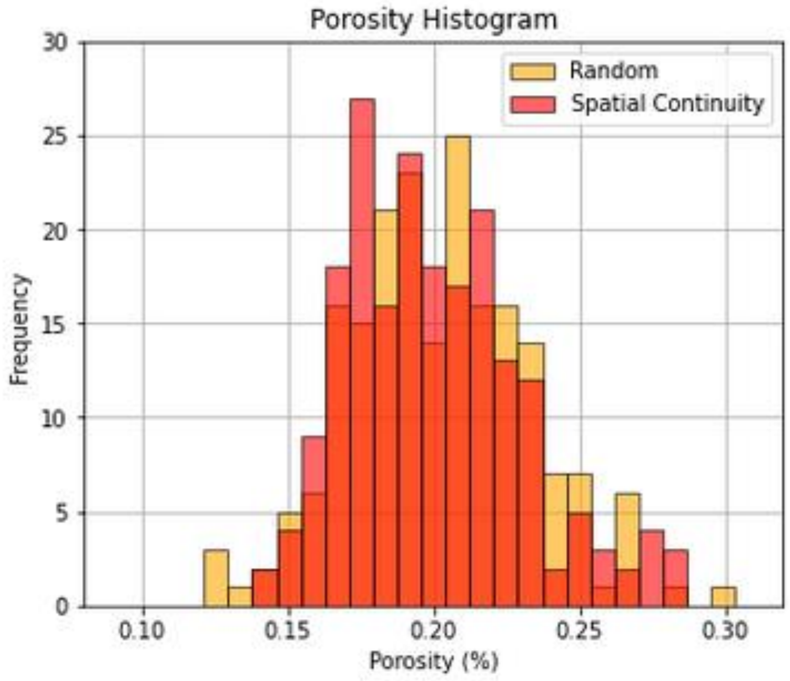
1D spatial porosity model without correlation, plot (left) and histogram (right), the file is PythonDataBasics_Visualize_Spatial_Continuity.ipynb.

Spatial Continuity Definition

Let's Tune Our Eyes to Spatial Continuity

- **Spatial correlation over 20 meters** – used a convolution kernel to impose spatial correlation in the random model.



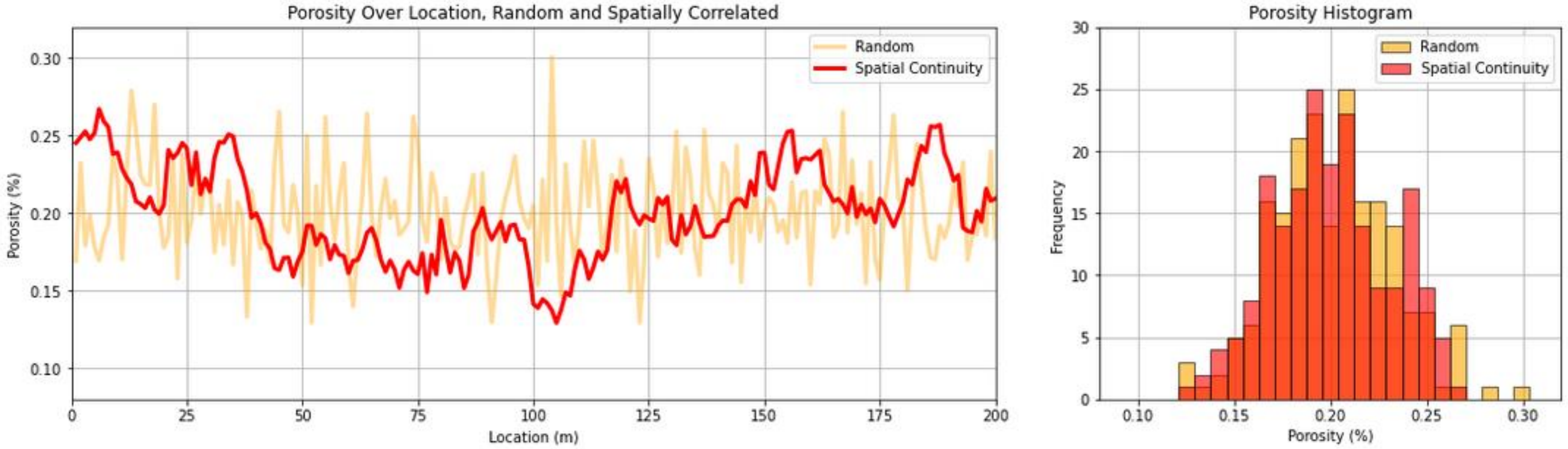


1D spatial porosity model without correlation, plot (left) and histogram (right), the file is PythonDataBasics_Visualize_Spatial_Continuity.ipynb.

Spatial Continuity Definition

Let's Tune Our Eyes to Spatial Continuity

- **Spatial correlation over 50 meters** – used a convolution kernel to impose spatial correlation in the random model.



1D spatial porosity model without correlation, plot (left) and histogram (right), the file is PythonDataBasics_Visualize_Spatial_Continuity.ipynb.

PGE 338 Data Analytics and Geostatistics

Lecture 10: Spatial Continuity

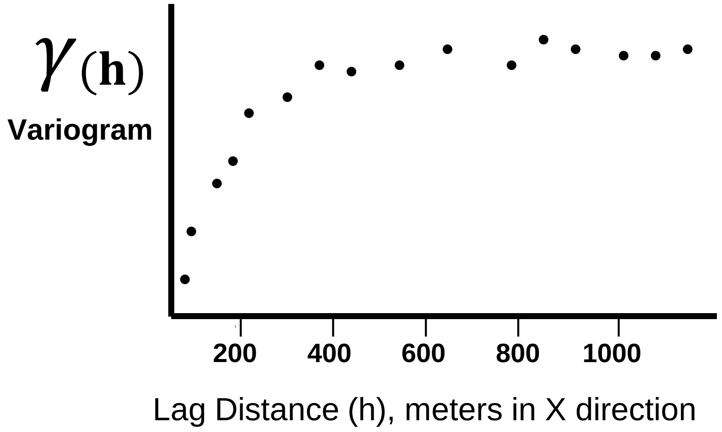
Lecture outline . . .

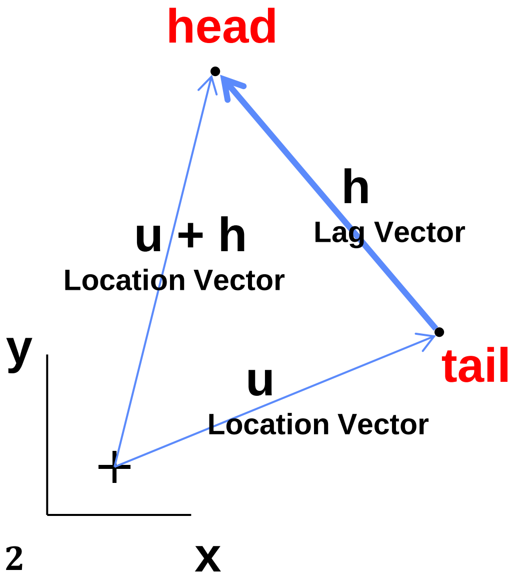
- **Variogram Calculation**

Measuring Spatial Continuity

The Semivariogram:

Function of difference over distance. Experimental variogram calculated at specific distances and plotted as points.



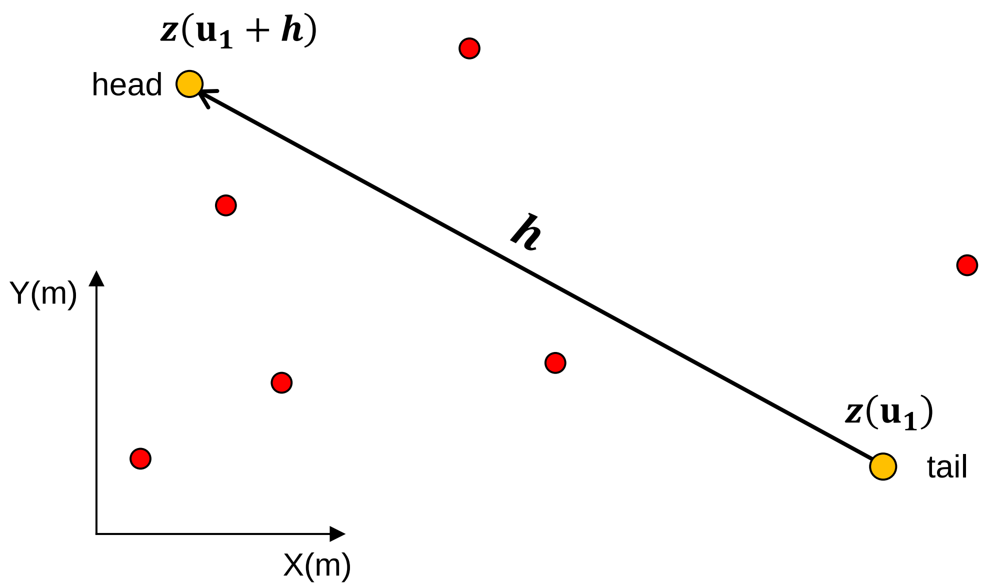


- The equation:

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

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

h-Scatterplot



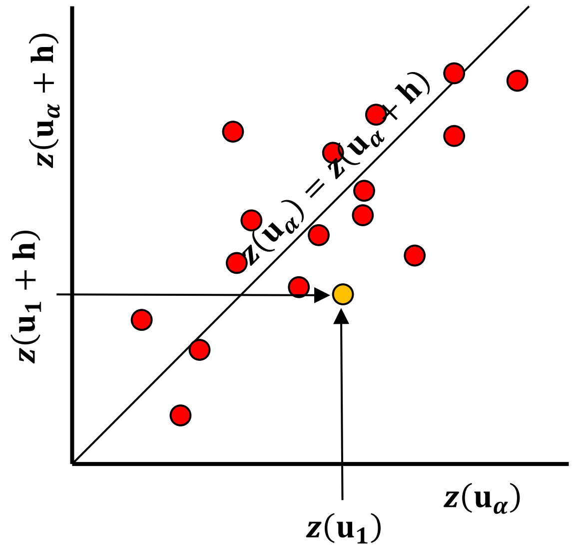


Illustration of h scatter plot, example data pair separated by lag distance h (left) and h scatter plot with example data pair highlighted.

Take all data pairs separated by lag vector h and put them on a scatter plot.

- x axis is the tail data values and y axis is the head data values for each data pair separated by lag vector h.

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 (or variogram if you remove the ½), but in practice, the term variogram is used.

- The ½ is used so that the covariance function and variogram may be related directly:

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

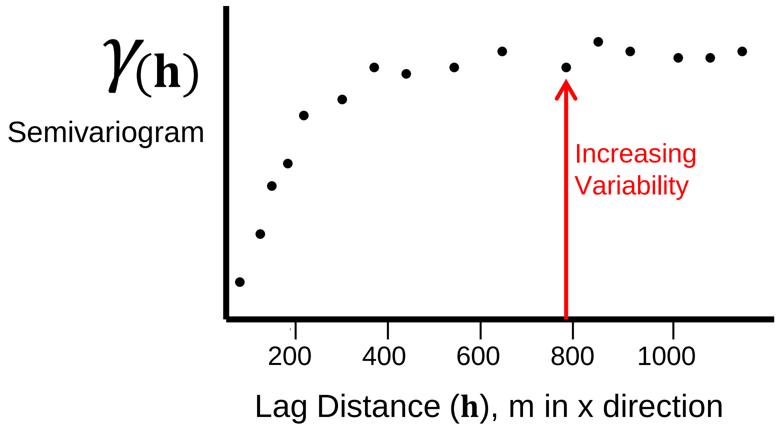
- Note the correlogram, $\rho_x(h)$, is related to the covariance function, $C_x(h)$, as:

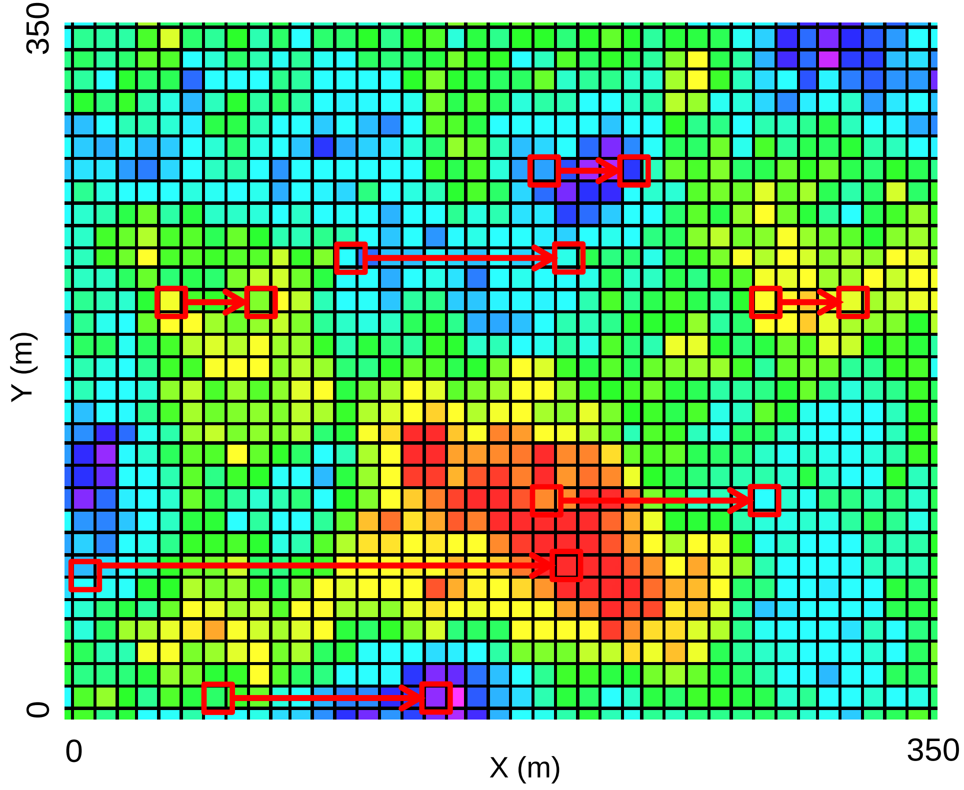
$$\rho_x(h) = \frac{C_x(h)}{\sigma_x^2}$$, h-scatterplot correlation vs. lag distance

Variogram Observations

Observation #1

As distance increases, variability increase (in general).



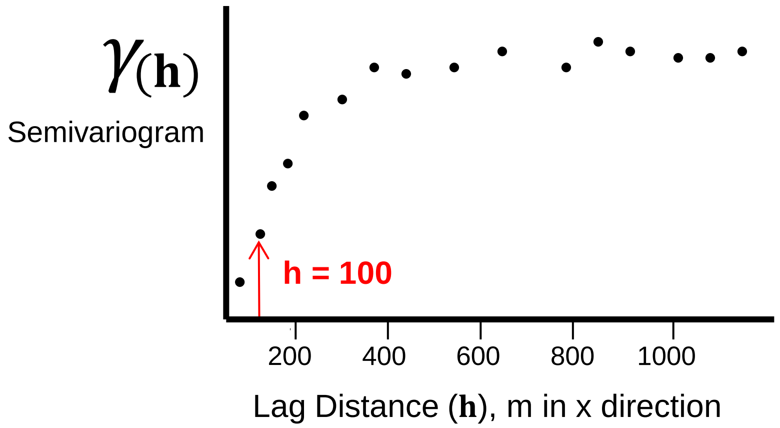


Increase in variability with increase in lag distance.

Variogram Observations

Observation #2

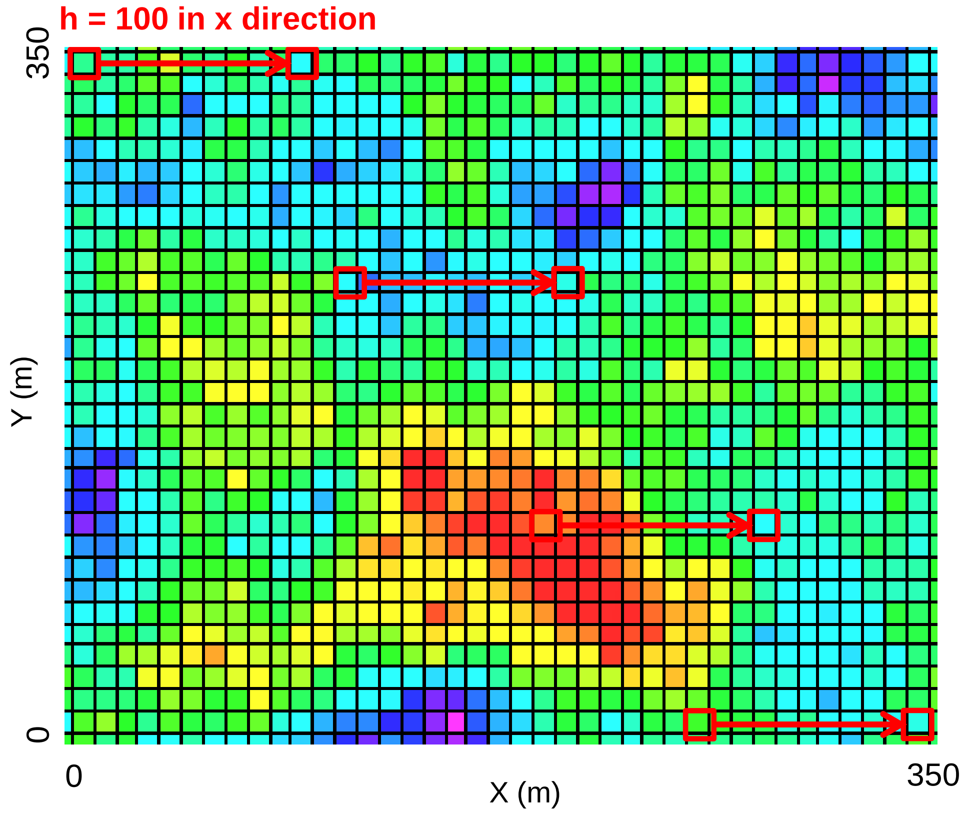
Calculated with over all possible pairs separated by lag vector, $\mathbf{h}$.



- Recall the variogram equation:

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

given the number of pairs available $N(h)$.



Use all possible data pairs separated by lag vector h.

Variogram Observations

Observation #3

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

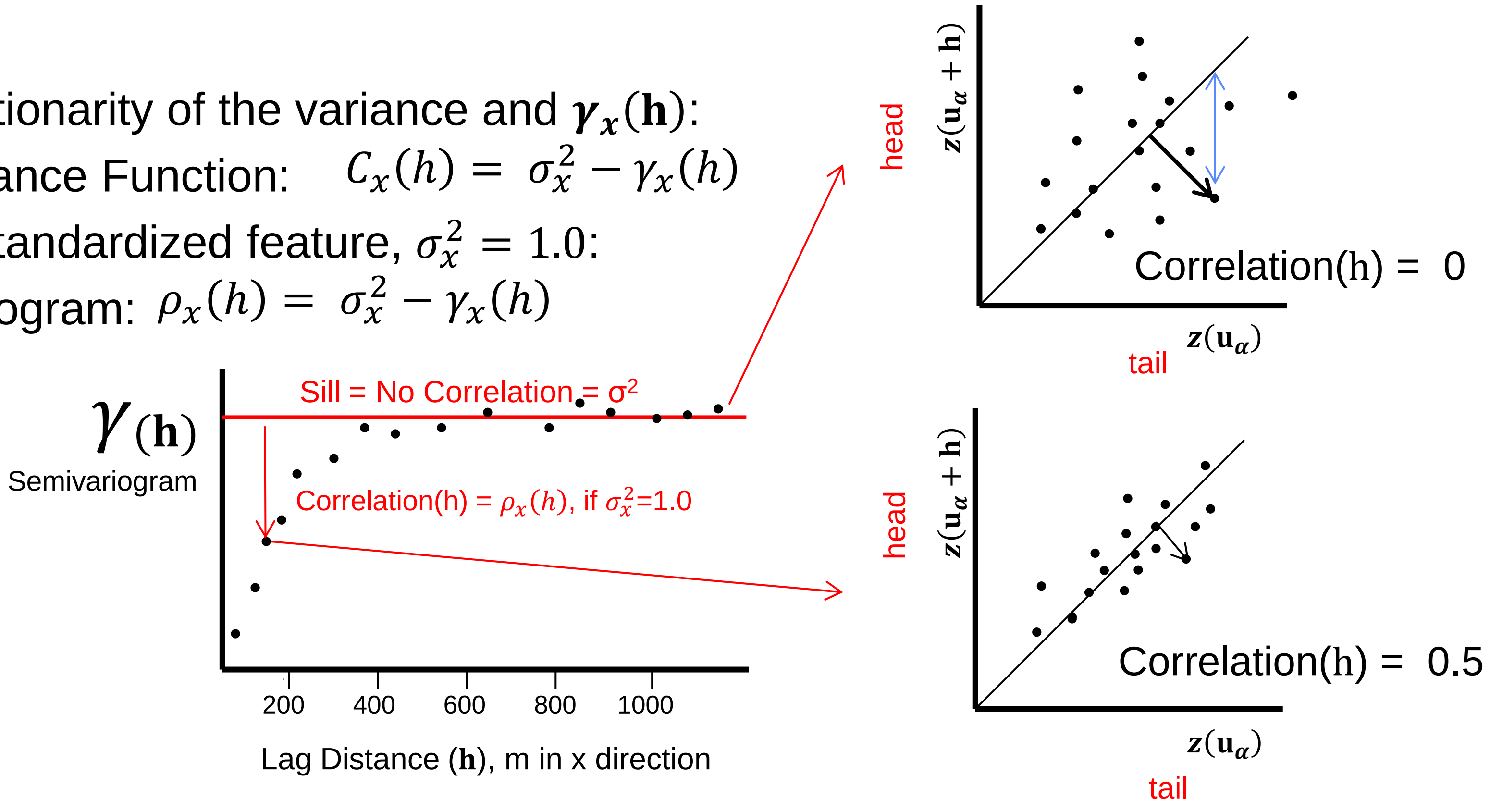
Sill is the Variance, σ^2

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

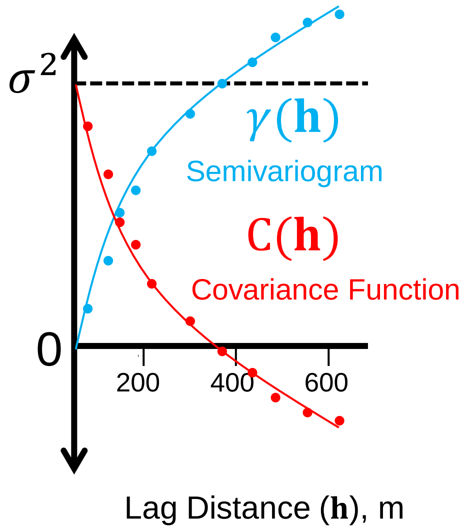
  Covariance Function: $C_x(h) = \sigma_x^2 - \gamma_x(h)$

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

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



Experimental variogram (left) and h scatter plots for 2 points (right).



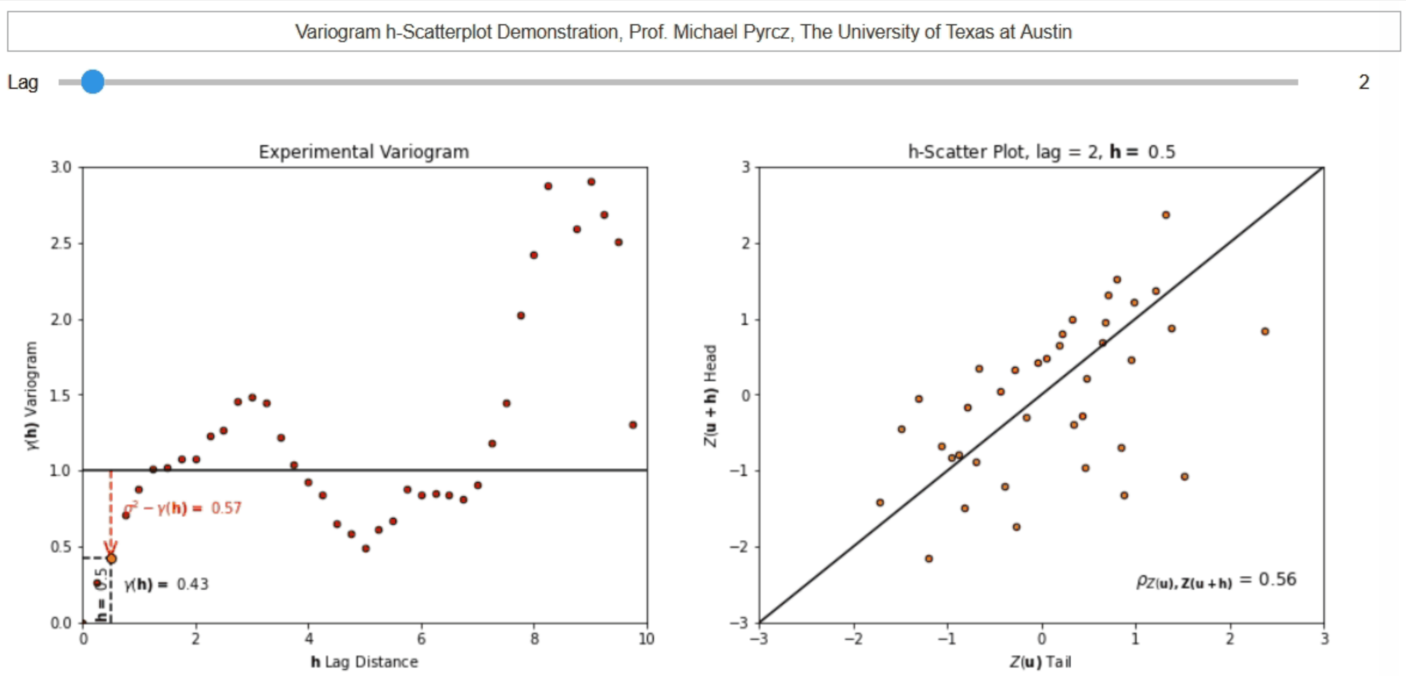
Semivariogram and covariance function compared.

Variogram Observations

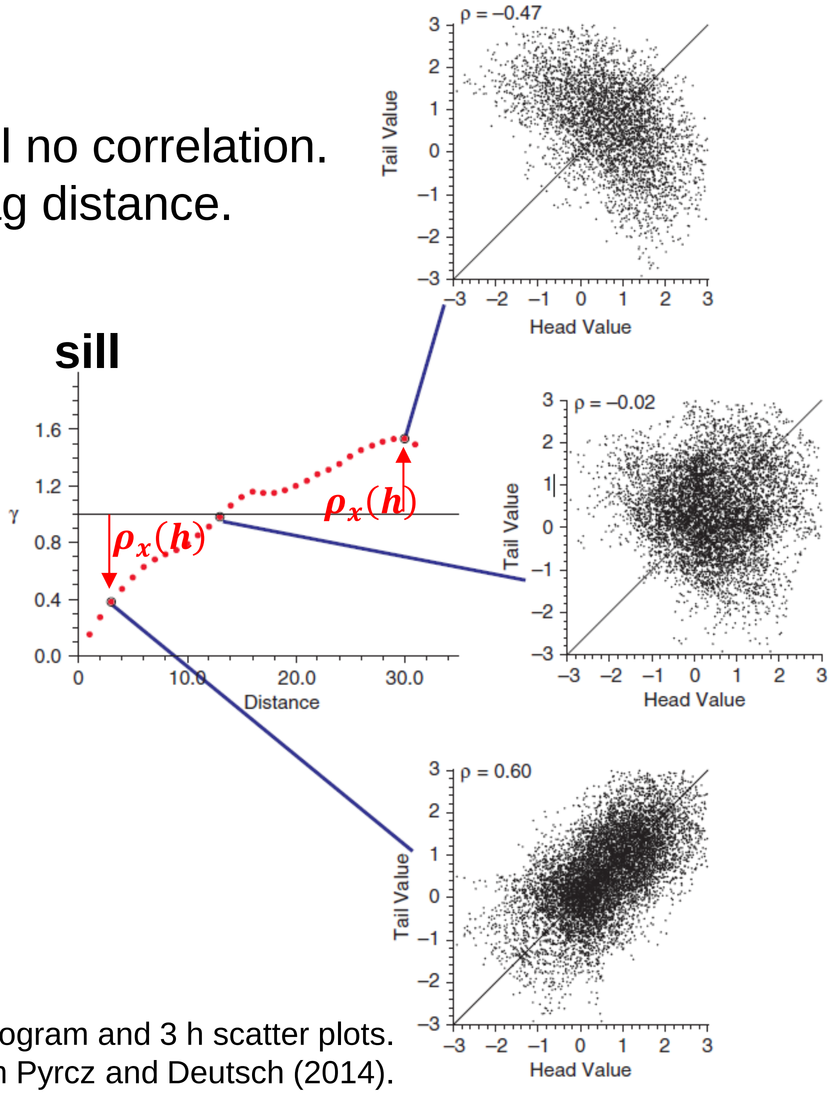
Observation #3

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

- Below sill positive correlation, above sill negative correlation, and on sill no correlation.
- Demonstration and another illustration of h-scatterplot correlation vs. lag distance.



Demonstration of h-scatter plots, file is Interactive_Variogram_h-scatter.ipynb.

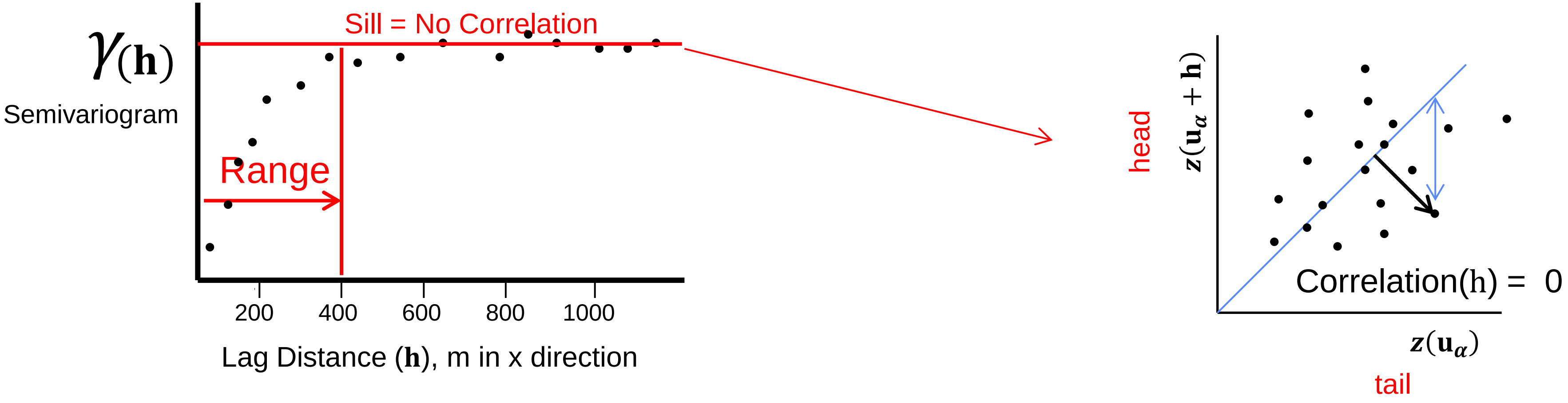


Experimental variogram and 3 h scatter plots.
Image modified from Pyrcz and Deutsch (2014).

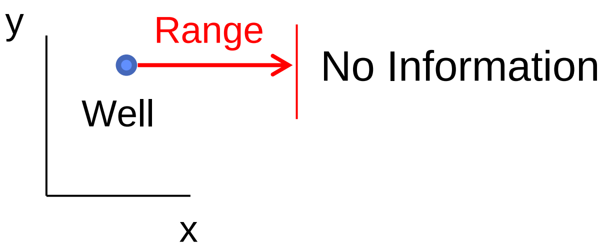
Variogram Observations

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.

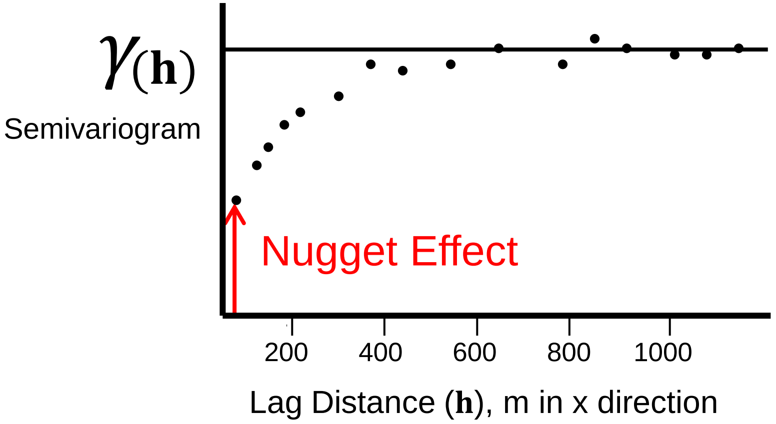


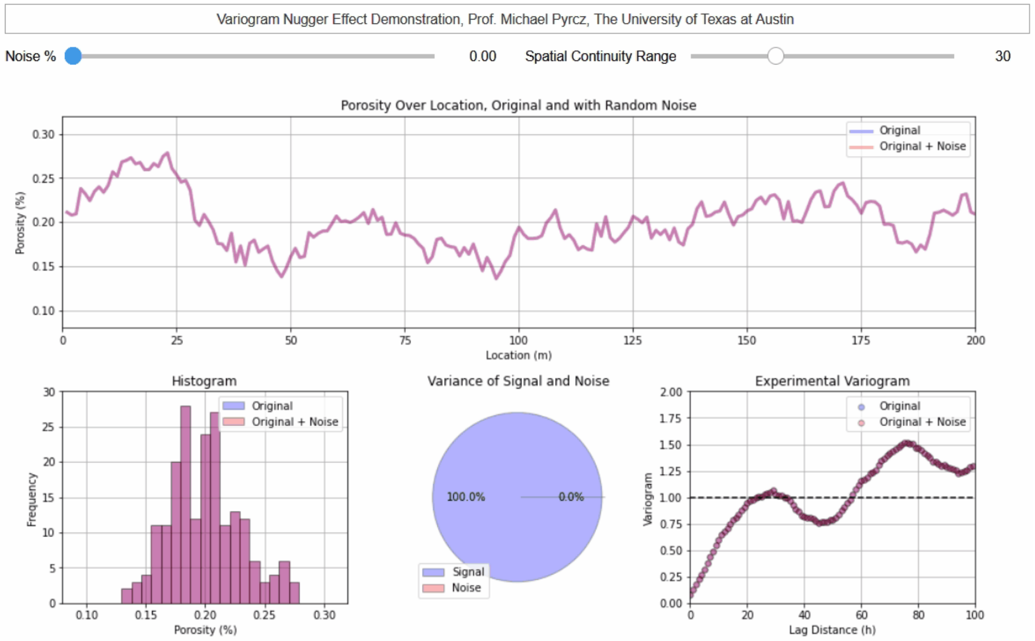
The h-scatterplot (above) and interpretation (below) at the range.

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.





Demonstration of variogram nugget effect, file is Interactive_Variogram_Nugget_Effect.ipynb.

- As a ratio of nugget / sill, is known as relative nugget effect (%)
- Modeled as a no correlation structure over all lags, $h > \varepsilon$, an infinitesimal distance
- Measurement error, mixing populations cause apparent nugget effect

Spatial Variability

You can interpret multiscale spatial continuity with the variogram.

- The three maps are remarkably similar: all three have the same 140 data, the same histograms, and the same variogram, and yet their spatial variability/continuity is quite different over short distances, small scales.

- Our map-making efforts should consider the spatial variability/continuity of the variable we are mapping:
  - Variability
  - Uncertainty

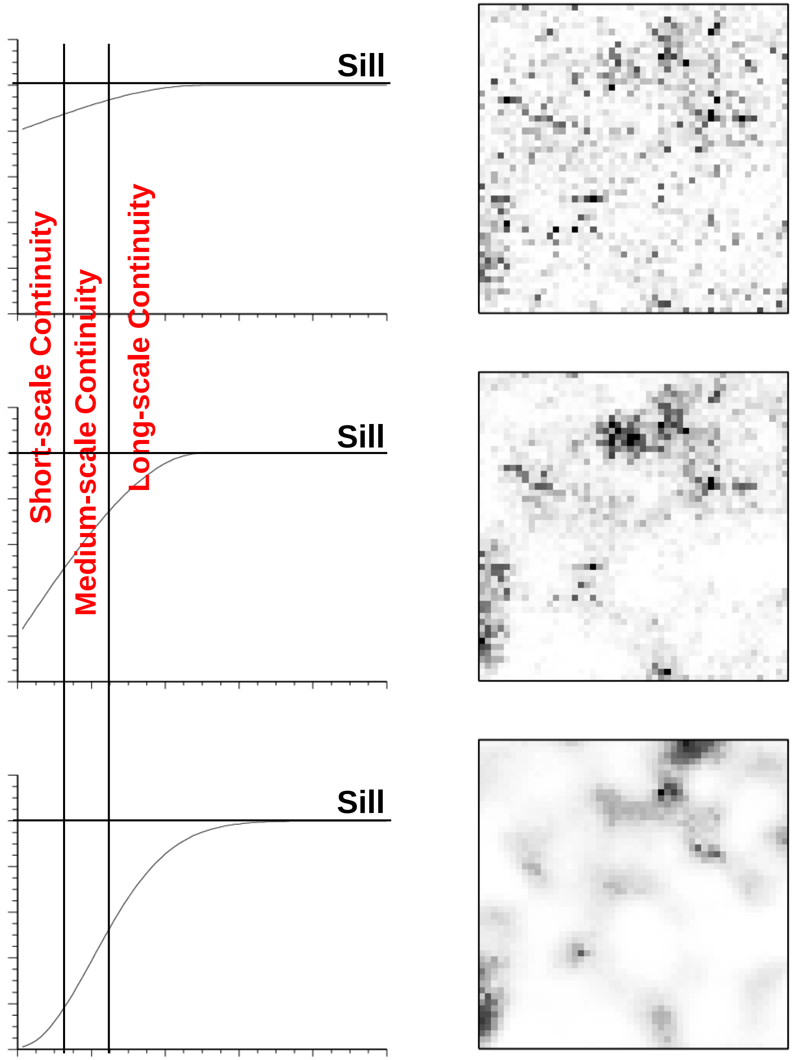
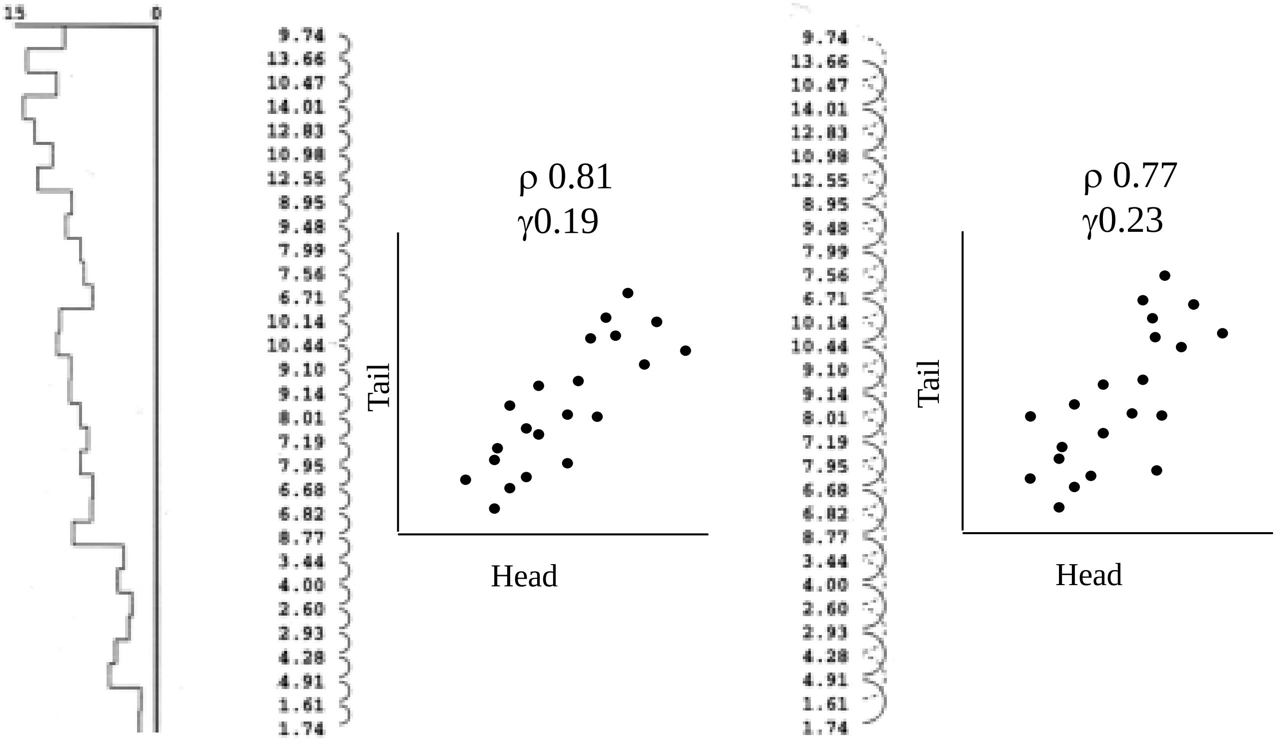


Image modified from Pyrcz and Deutsch (2014)

Variogram Calculation
Hands On

Variogram Calculation by Hand

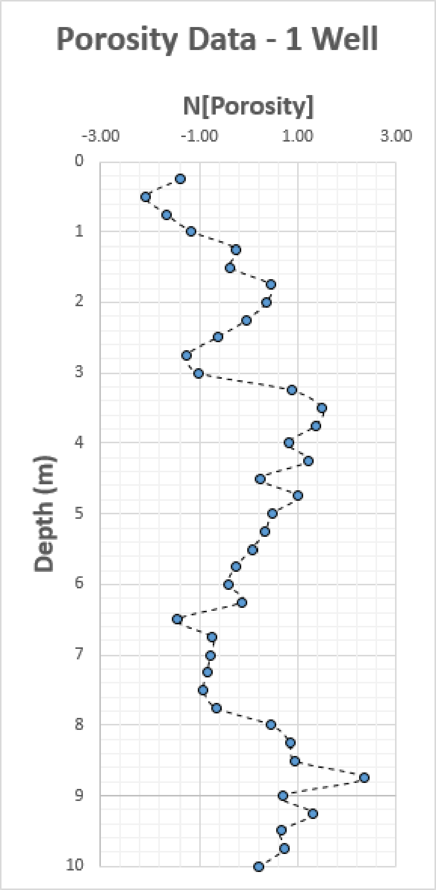
- Consider data values separated by *lag* vectors (the h values)
- Here are two examples of a lag vector equal to the data spacing and then twice the data spacing:



Variogram Calculation Example

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

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



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.

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

Excel Solution (and all lags) is in:

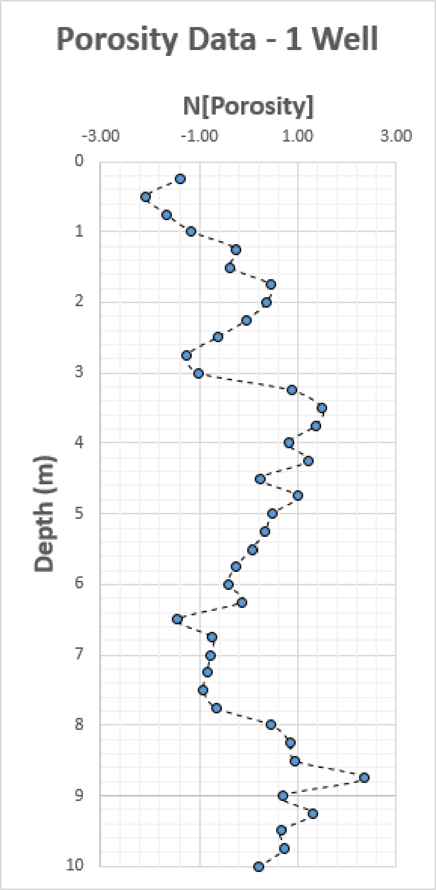
- GeoDataSets/1D_Porosity.xls

Excel Variogram for 1D Regularly Spaced Data:

1. use relative references and add a new column with cell – cell shifted down by lag
2. take the average of the new column and divide by 2

Python Variogram for 1D Regularly Spaced Data:

```
npor = df['Porosity']                    # this is a Pandas series
np.average(np.square((npor - npor.shift(lag)).dropna()))*0.5
```



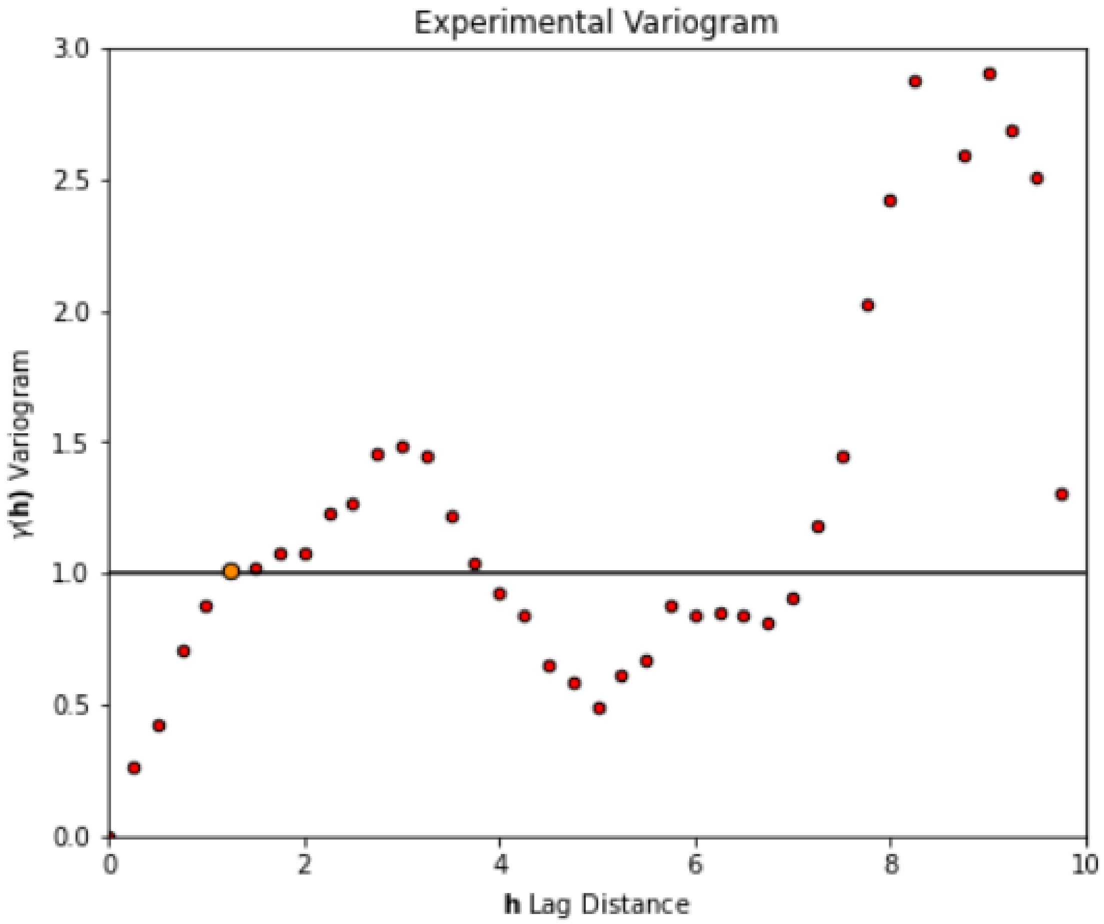
Depth	N[Porosity]
0.25	-1.37
0.5	-2.08
0.75	-1.67
1	-1.16
1.25	-0.24
1.5	-0.36
1.75	0.44
2	0.36
2.25	-0.02
2.5	-0.63
2.75	-1.26
3	-1.03
3.25	0.88
3.5	1.51
3.75	1.37
4	0.81
4.25	1.21
4.5	0.24
4.75	0.99
5	0.49
5.25	0.34
5.5	0.07
5.75	-0.26
6	-0.41
6.25	-0.14
6.5	-1.44
6.75	-0.75
7	-0.78
7.25	-0.85
7.5	-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 are all of the experimental variogram values over all the possible lag distances
- Calculated in Python by looping this code over all the possible lag distances for the dataset 1D_Porosity

```
np.average(np.square((npor - npor.shift(lag)).dropna()))*0.5
```



Experimental variogram for 1D_Porosity.csv.

The Variogram and Covariance Function

The variogram, $\gamma_x(h)$, covariance function, $C_x(h)$, and correlogram, $\rho_x(h)$, are all tools for characterizing spatial two-point correlation (assuming stationarity):

$$\gamma_x(h) = \sigma_x^2 - C_x(h)$$
$$= \sigma_x^2(1 - \rho_x(h))$$

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

where:

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

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

$$C_x(h) = \frac{\sum_{\alpha=1}^{n} x(u_\alpha) \cdot x(u_\alpha + h)}{n} - (\bar{x})^2$$, if stationary mean

Stationarity entails that:

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

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

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.

$$C_x(h) = \frac{\sum_{\alpha=1}^{n} x(u_\alpha)\cdot x(u_\alpha+h)}{n} - (\bar{x})^2, \text{ if stationary mean}$$

- The covariance function is the variogram upside down. Recall: $\gamma_x(\mathbf{h}) = \sigma_x^2 - C_x(\mathbf{h})$

- We model variograms, but inside the kriging and simulation methods they are converted to covariance values for numerical convenience.

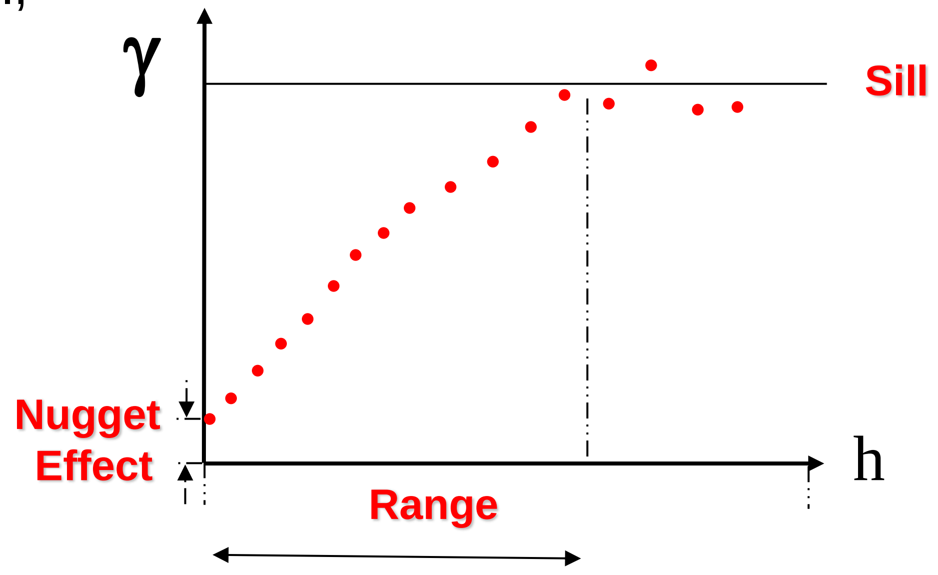
Variogram Components Definition

Primary variogram definitions up to now.

- **Nugget Effect** – discontinuity in the variogram at distances less than the minimum data spacing
  - Often communicated as the ratio, nugget effect / sill, known as the 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

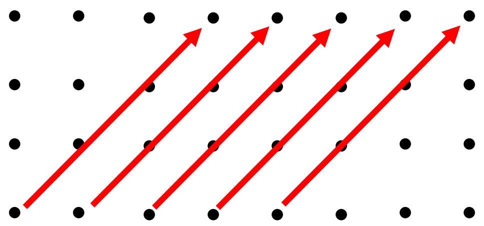
3. Full 3D spatial continuity model
   – Model primary directions, i.e. major horizontal, minor horizontal and vertical and combine together with assumption of geometric anisotropy

Calculating Experimental Variograms

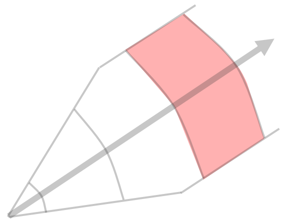
How do we get pairs separated by lag vector?

- Regularly spaced data:
  - Specify as offsets in x and y
  - e.g. 045, 4^{th} lag is offset [3,3]

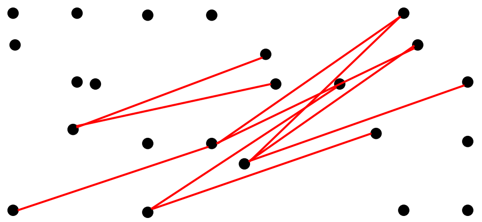
- Irregularly spaced data in 2D:
  - Unit lag distance and distance tolerance
  - Azimuth direction and azimuth tolerance
  - Bandwidth (maximum deviation) from azimuth



All pairs for 4^{th} lag 045 azimuth.



4^{th} lag, 4 x lag distance in 045 azimuth

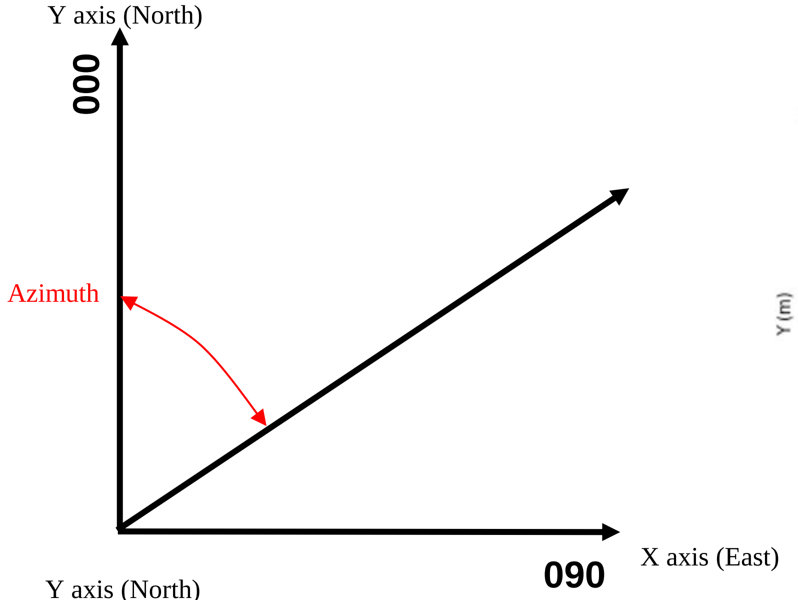


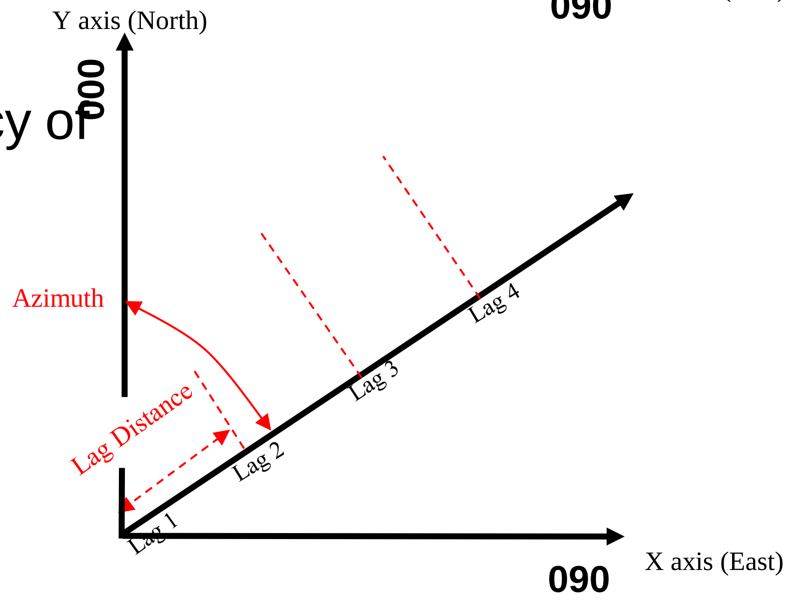
All pairs for 4^{th} lag 045 azimuth.

Calculating Experimental Variograms Sparse Spatial Data

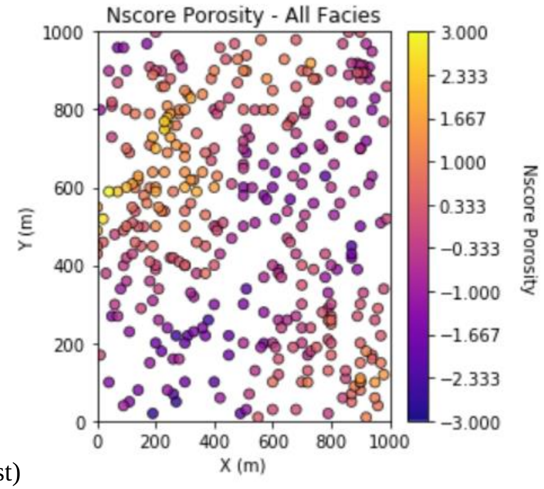
Search Template Explained

- **Select the direction as an azimuth.**
- Set **lag distance** to at least the 'common' minimum data spacing → no data pairs available at shorter lag distances
- **Lag distance** controls the number/frequency of experimental variogram points.





Variogram search template.



Porosity 2D dataset.

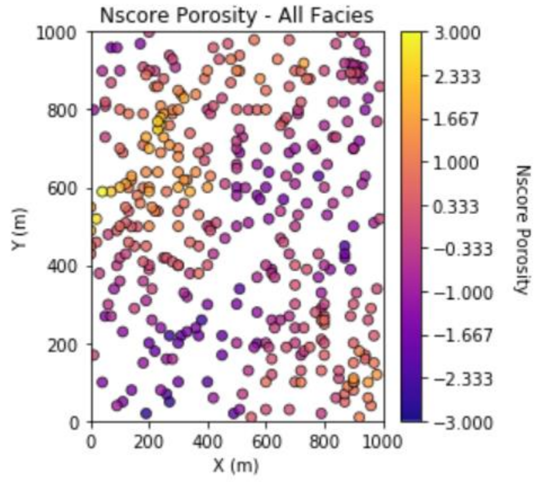
Calculating Experimental Variograms Sparse Spatial Data

Search Template Explained

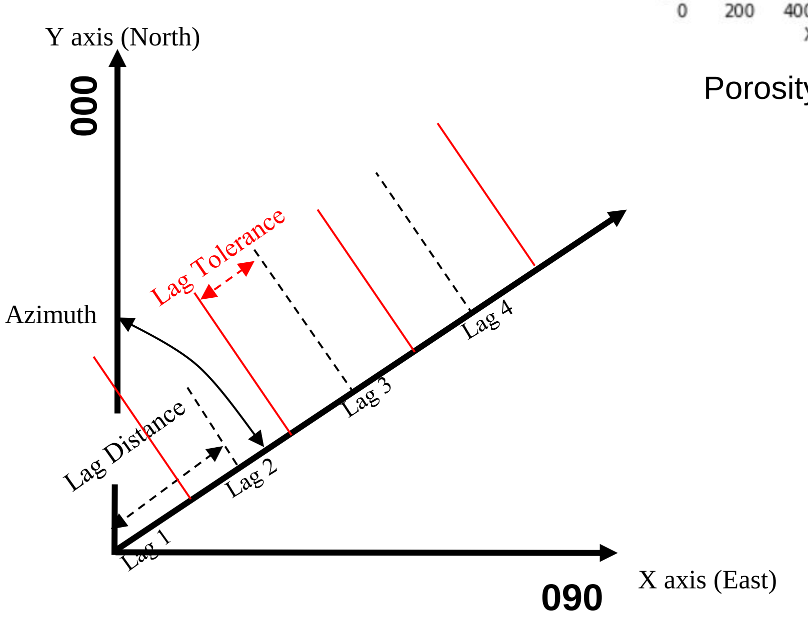
- Set the **lag tolerance** to ½ the unit lag distance.

- Lower **lag tolerance** will omit possible pairs

- Larger **lag tolerance** will overlap the lag bins, smooth the results.



Porosity 2D dataset.

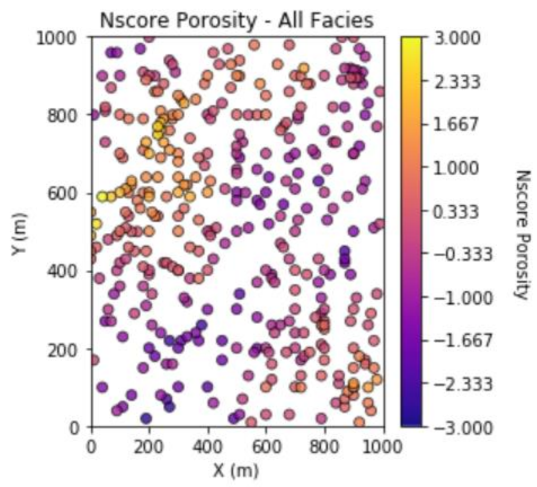


Variogram search template.

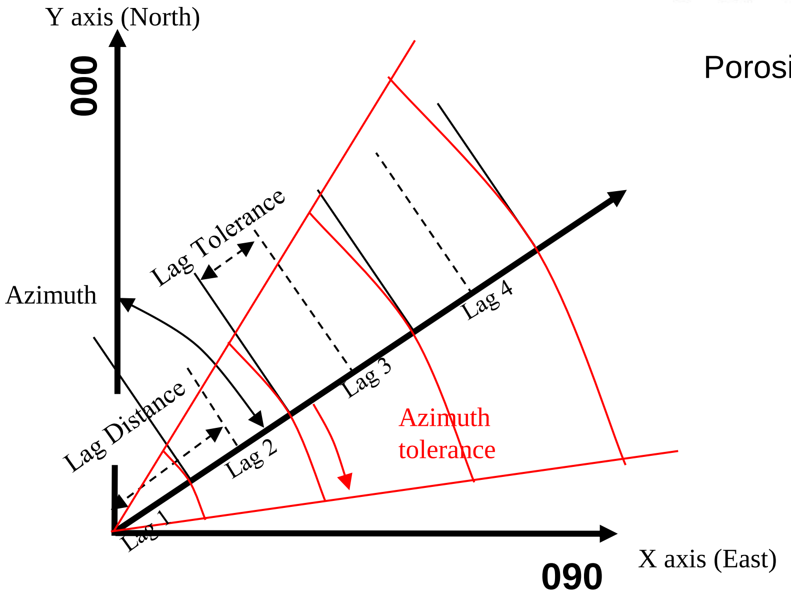
Calculating Experimental Variograms Sparse Spatial Data

Search Template Explained

- Set the **azimuth tolerance** to 90.0 for isotropic variograms.
- Set **azimuth tolerance** to about 22.5 degrees for directional variograms.
- Lower **azimuth tolerance** is more direction specific and noisier
- Larger **azimuth tolerance** is more isotropic and smoother



Porosity 2D dataset.



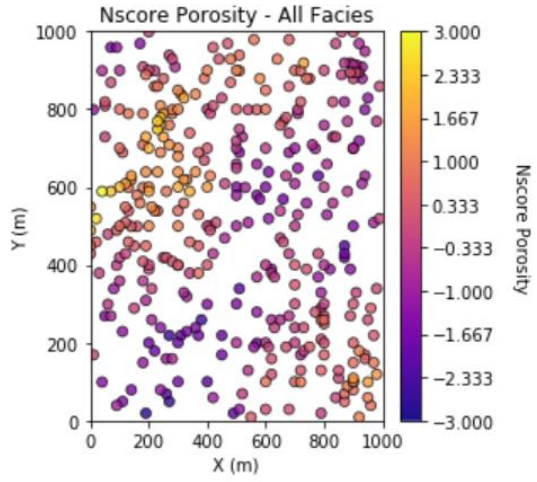
Variogram search template.

Calculating Experimental Variograms Sparse Spatial Data

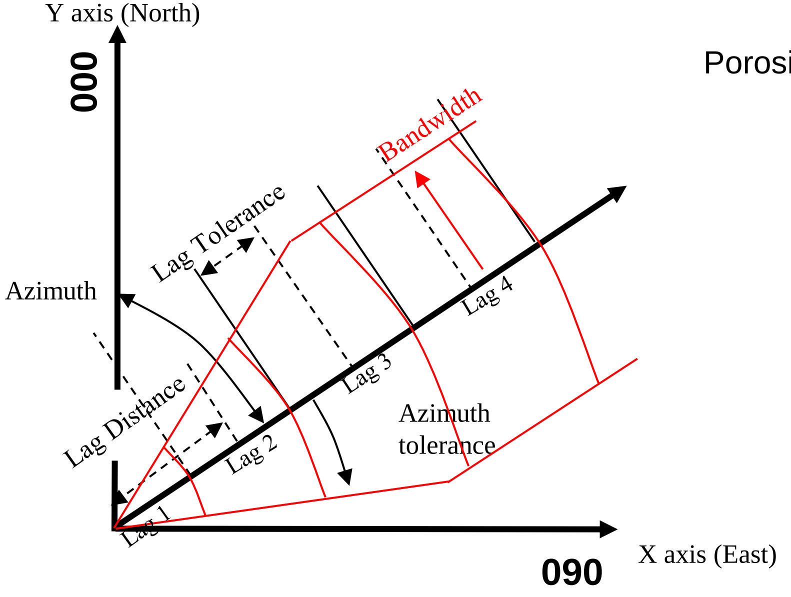
Search Template Explained

- Set a **bandwidth** limit is needed to limit the distance of investigation away from the lag vector.

- Set **bandwidth** very large if isotropic to avoid artifacts.



Porosity 2D dataset.



Variogram search template.

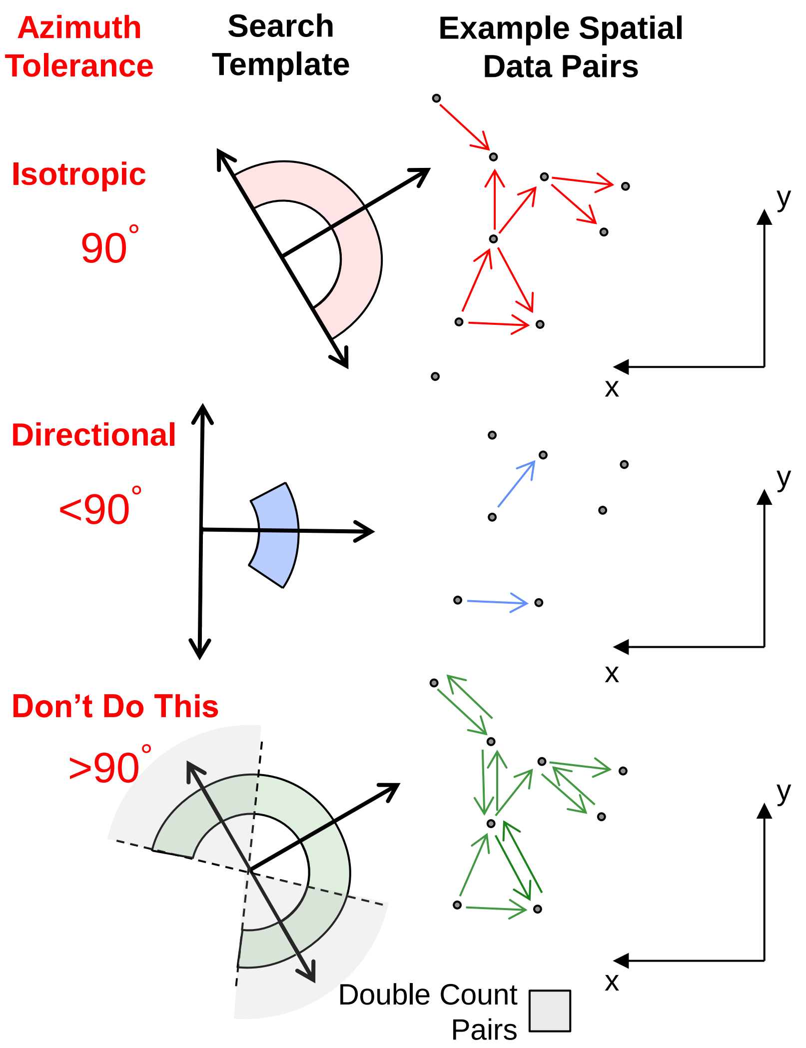
Calculating Experimental Variograms Sparse Spatial Data

Isotropic and Anisotropic Variogram Calculation

- **Isotropic / Omnidirectional** – Azimuth Tolerance of 90°, all pairs lag distance $\pm$ lag [distance] tolerance.

- **Anisotropic / Directional** – Azimuth Tolerance of $<90^\circ$. Only considering pairs in a specified direction, larger azimuth tolerance will approach isotropic. Calculate in the **major and minor directions** to support interpretation and variogram modeling.

- Do not extend the azimuth tolerance beyond 90°
  – Results in a bias as some of the data pairs are counted twice in a lag due to arbitrary orientation.



Azimuth tolerance and identified data pairs.

Calculating Experimental Variograms Sparse Spatial Data

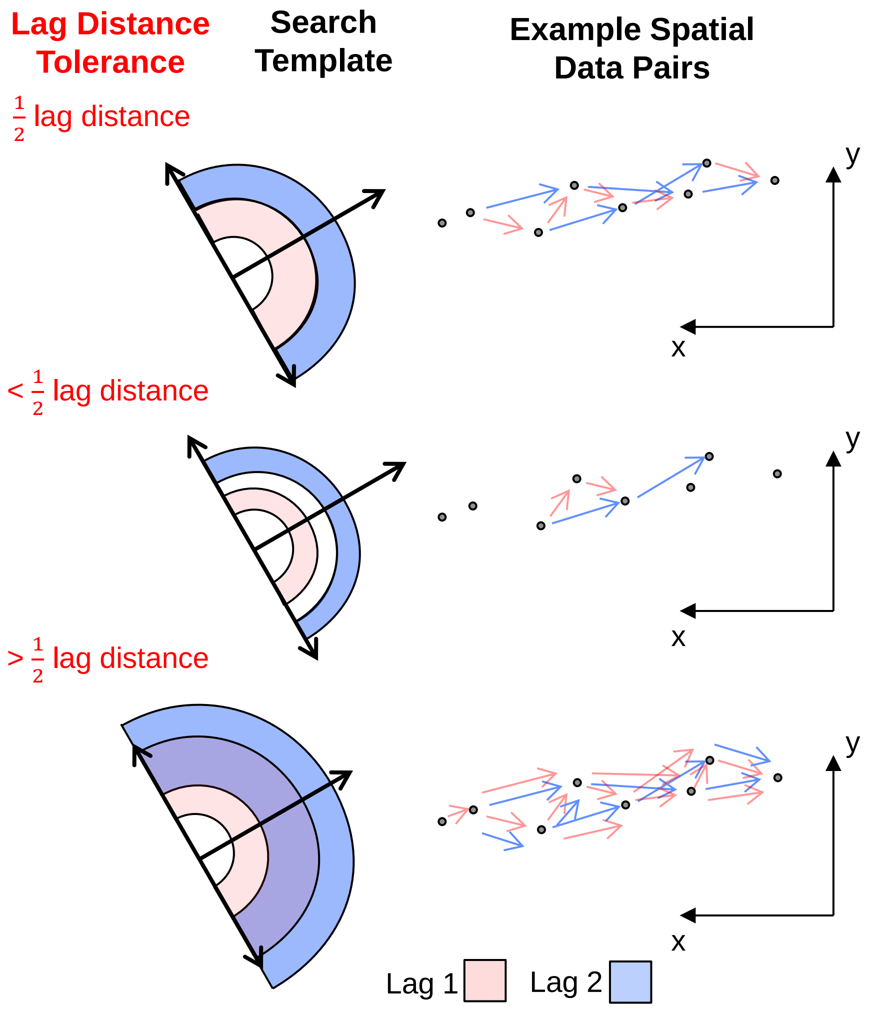
Lag Distance Tolerance

Commonly called “lag tolerance”

- $\frac{1}{2}$ **lag distance** – all possible pairs are included in a lag distance bin

- $< \frac{1}{2}$ **lag distance** – some pairs fall between the cracks.
  - This is not typically done.

- $> \frac{1}{2}$ **lag distance** – some pairs are counted twice, causing smoothing / correlation between the experimental variogram points / lags.
  - This is done with to smooth a noisy experimental variogram



Lag distance tolerance and identified data pairs for an example isotropic variogram.

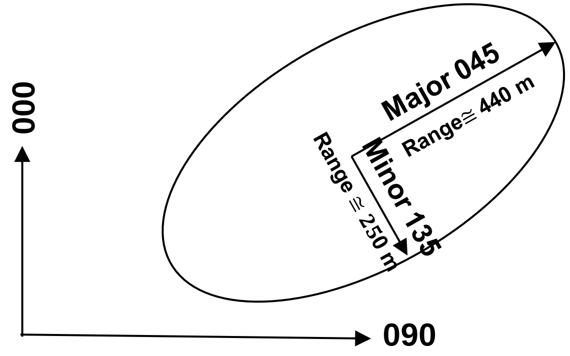
Calculating Directional Variograms

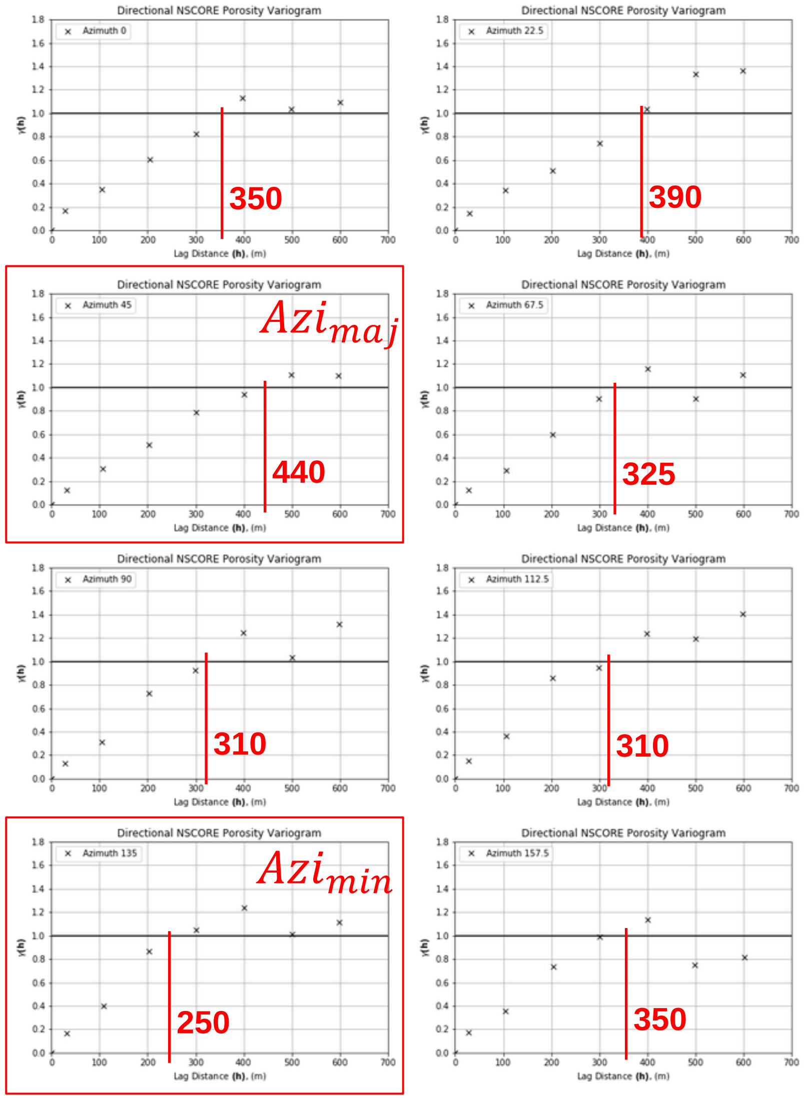
Identify and Calculate the Directional 2D Variograms in the:

Major Direction (Azi_{maj}) – direction with the largest variogram range

Minor Direction (Azi_{min}) – direction with the smallest variogram range, orthogonal to the major.

$$Azi_{maj} = Azi_{min} + 90$$





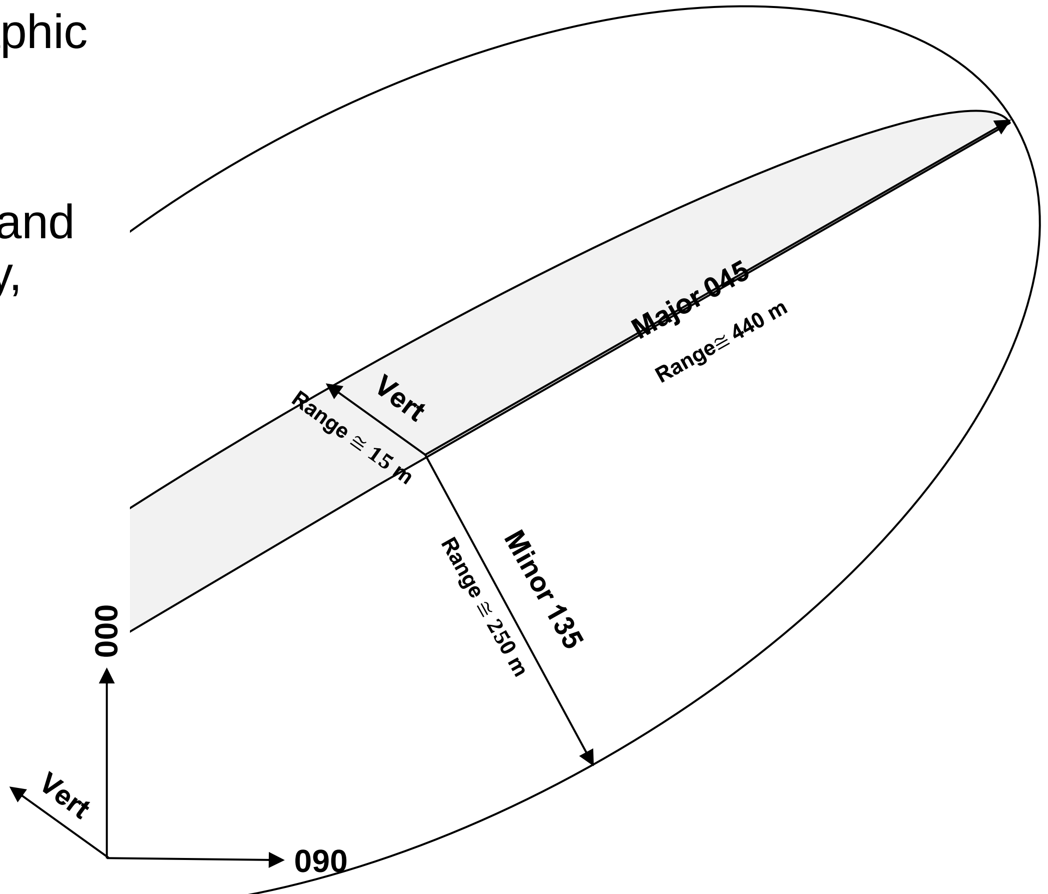
Directional experimental variograms with ranges and major and minor directions indicated.

Calculating Directional Variograms

Identify and Calculate the Directional 3D Variograms, it is typical to work with:

Horizontal Major and Minor Directions – the maximum and minimum range aligned with stratigraphic layering.

Vertical Direction – orthogonal to horizontal major and minor, orthogonal to stratigraphic layering. Generally, the lowest range of all directions.

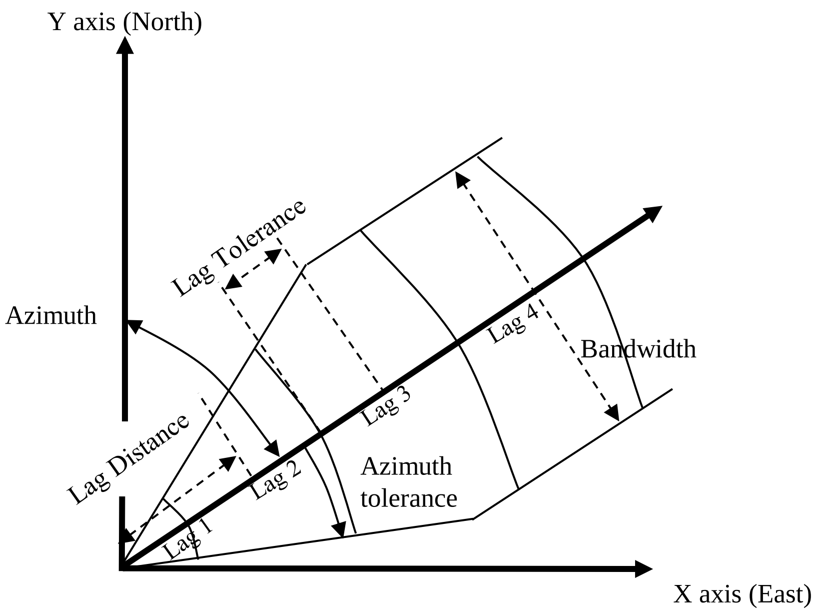


Variogram Maps

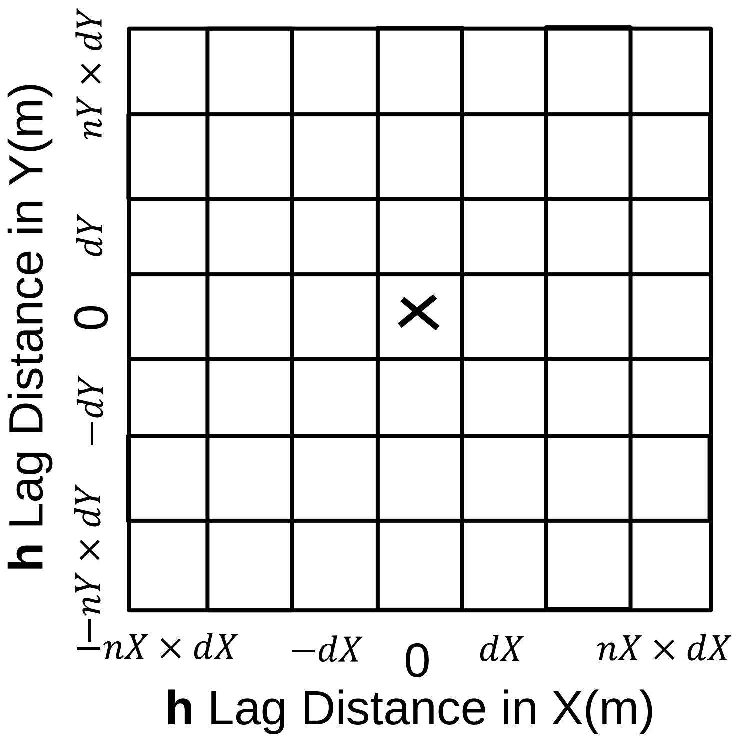
Calculating the variogram over all distances and directions at once with variogram maps

- Use a mesh template, cell size controls resolution like lag distance (assuming lag tolerance is ½ lag distance), number of cells controls the extent of calculation

- Can be useful to determine directions.



Directional variogram search template, from Pyrcz and Deutsch, 2014

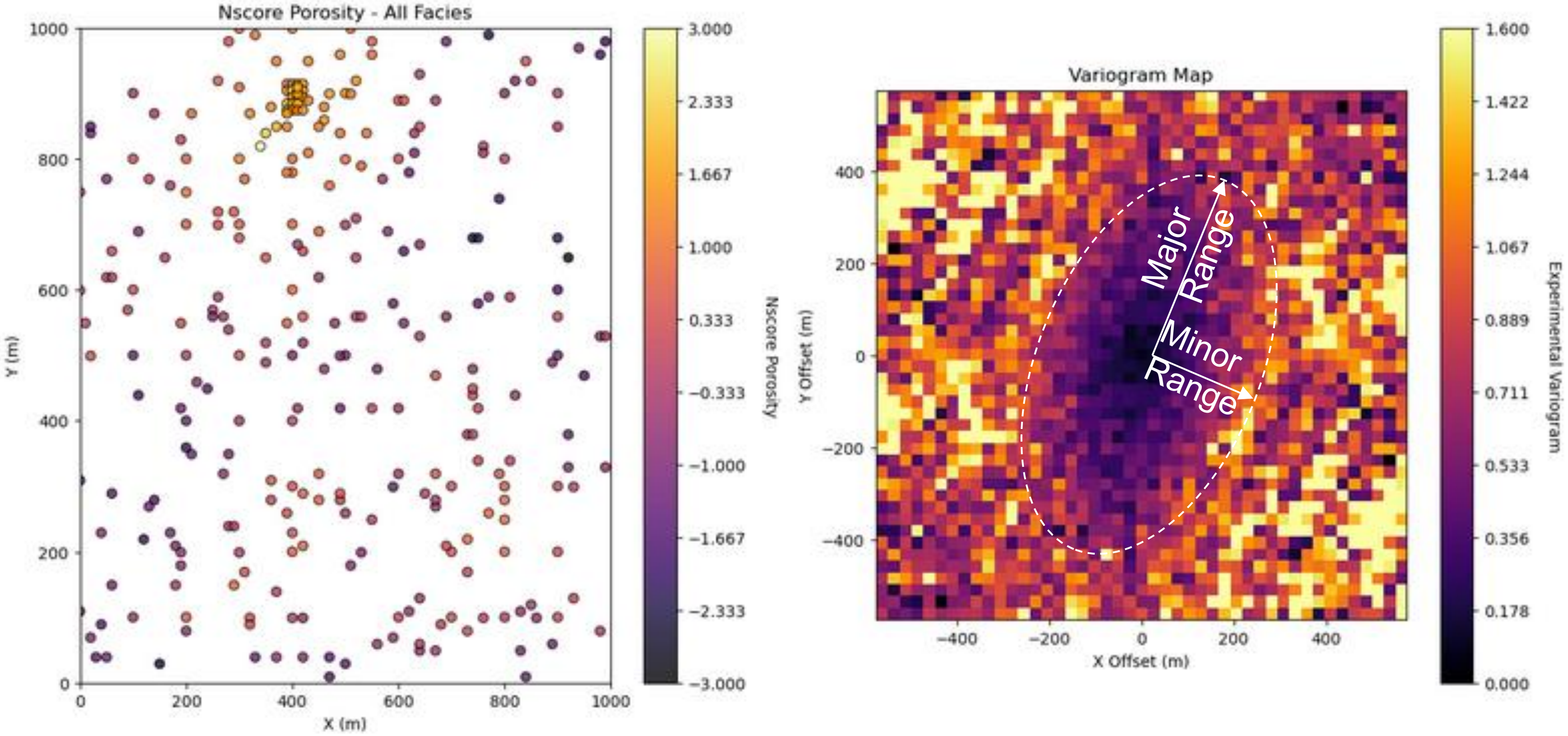


Variogram map search template

Variogram Maps

Calculating the variogram over all distance and directions at once with variogram maps

- Directly observe spatial continuity direction and approximate the major and minor ranges.



Normal scores, spatial, porosity data (left), and resulting variogram map and estimate of geometric anisotropy, direction and major / minor ranges (right).
From workflow GeostatsPy_spatial_continuity_directions.ipynb.

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 of 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

Up Next

Some Options

Other Variogram Considerations

- 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
  - Could also transform to treat outliers or to correct for an expected theoretical distribution (**data cleaning**)
- 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

Method to Choose the Variogram Major and Minor 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:

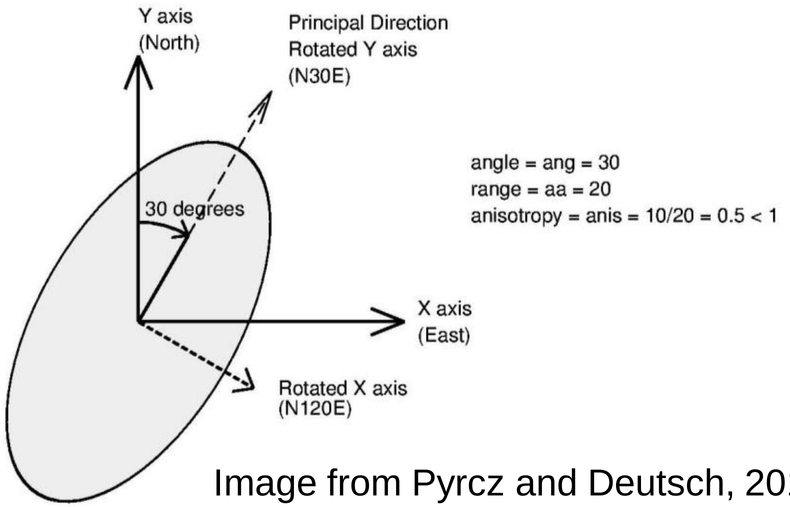
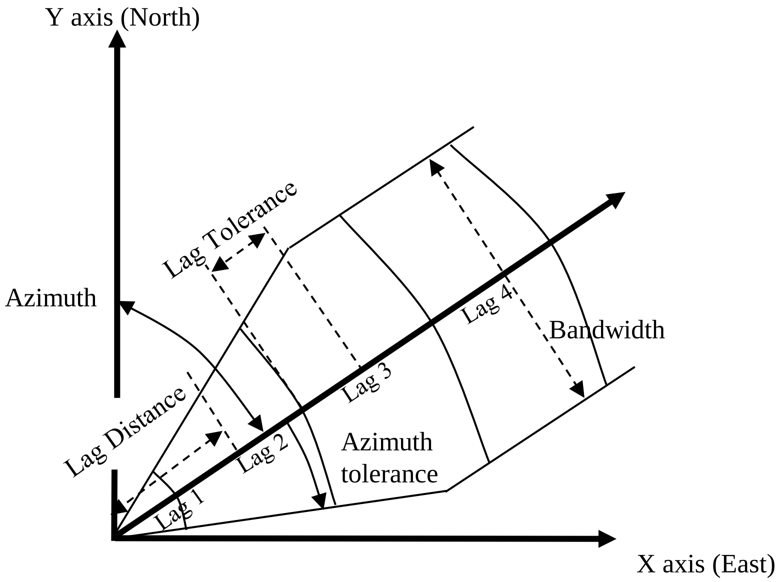


Image from Pyrcz and Deutsch, 2014

Choosing the Lag Distances and Tolerances

Guidance for Variogram Calculation Parameters:

- Lag separation distance should coincide with data spacing
- Lag tolerance is 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

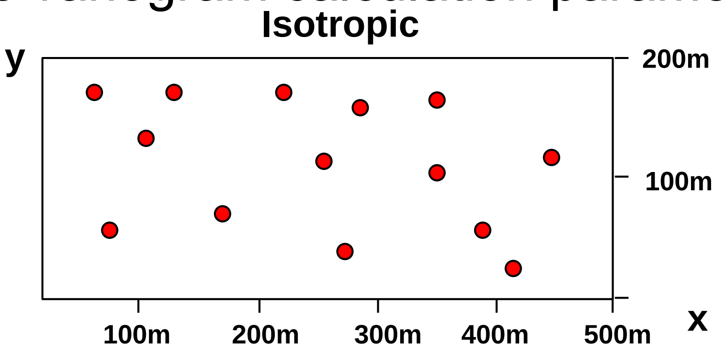


Directional variogram search template.

Image from Pyrcz and Deutsch, 2014

Choosing the Lag Distances and Tolerances

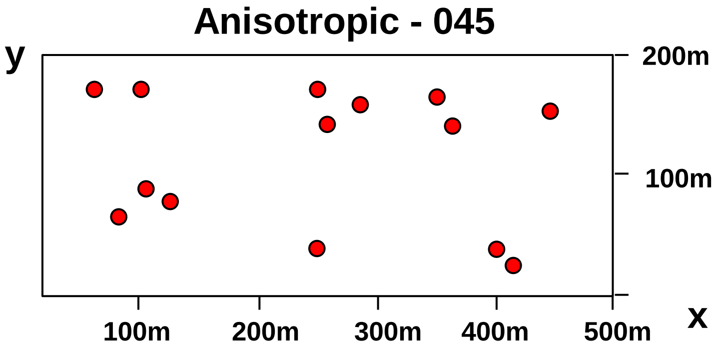
Let's pick some variogram calculation parameters in groups (estimate):



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

horizontal bandwidth ___

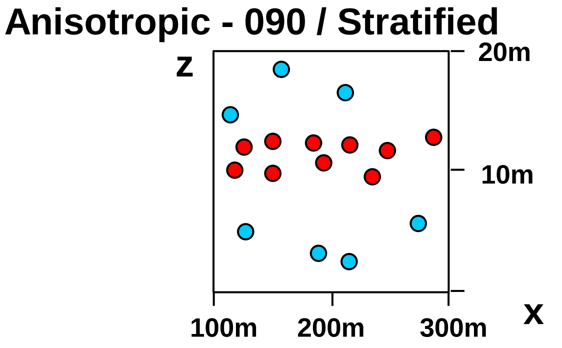
azimuth ____, azimuth tolerance ______



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

horizontal bandwidth ___

azimuth ____, azimuth tolerance ______



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

vertical bandwidth ___

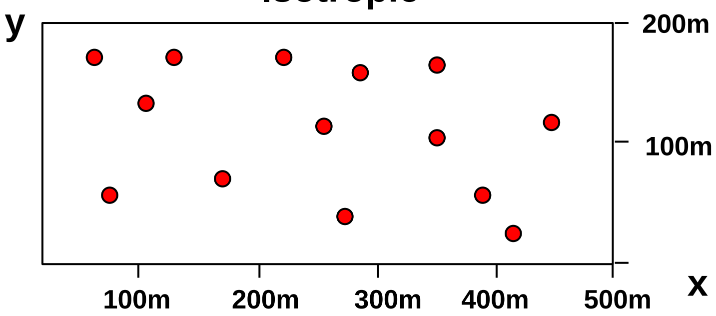
azimuth **90**, azimuth tolerance **22.5**

Along x

Choosing the Lag Distances and Tolerances

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

Isotropic



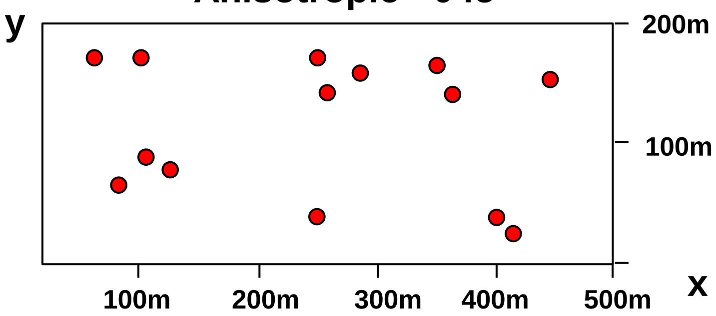
lag size **50**, lag tolerance **25**, number of lags **5**

horizontal bandwidth **∞**

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

Any

Anisotropic - 045

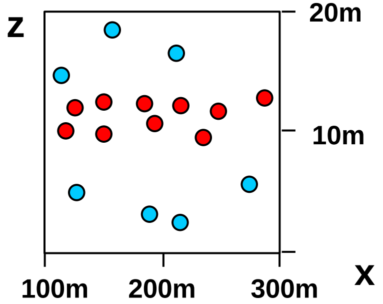


lag size **20**, lag tolerance **10**, number of lags **12**

horizontal bandwidth **200**

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

Anisotropic - 090 / Stratified



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

vertical bandwidth **5**

azimuth **90**, azimuth tolerance **22.5** **dip** **0**

Along x

Variogram Calculation Exercise in Excel

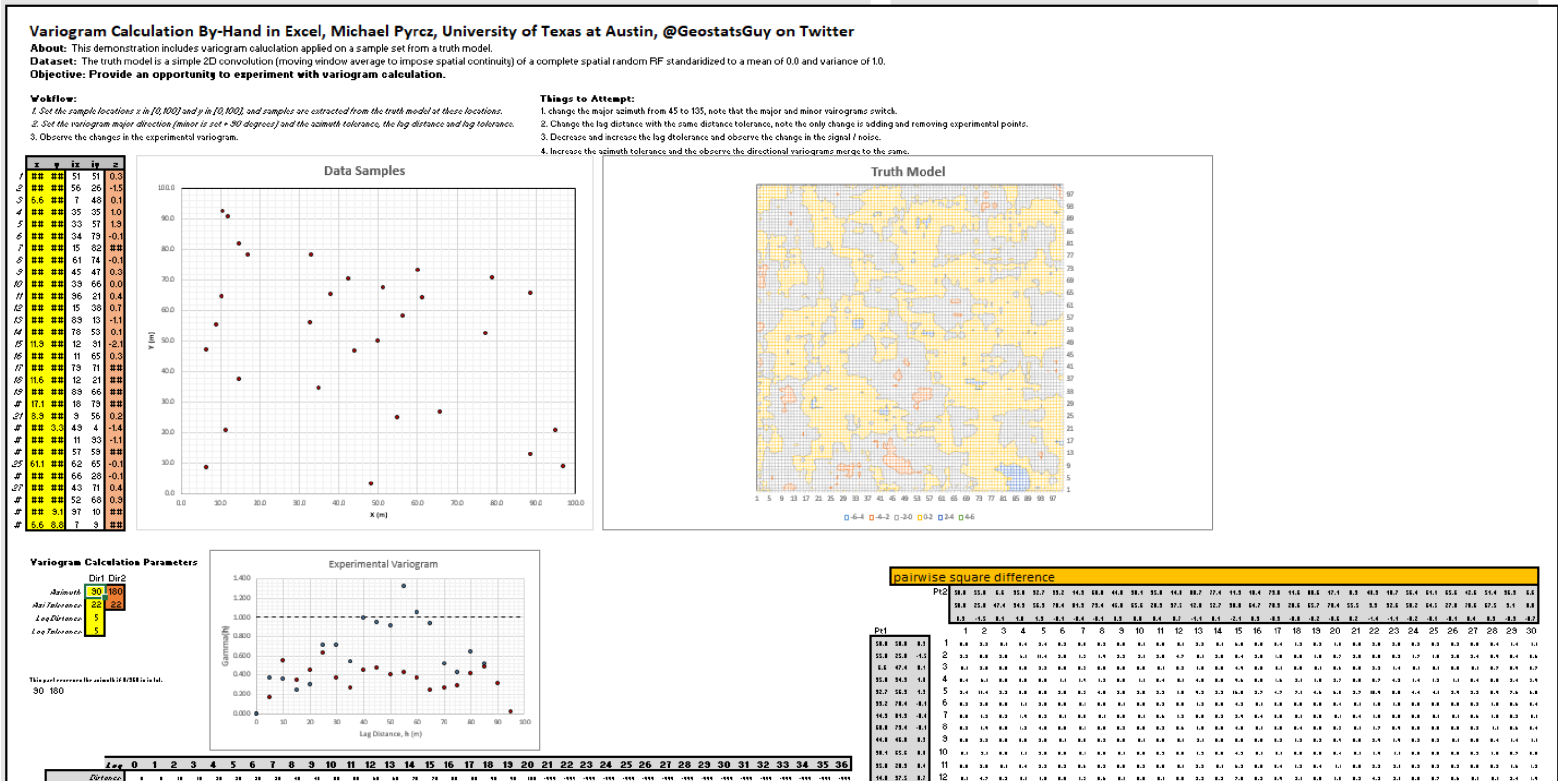
Variogram Calculation:

Things to try:

1. Calculate an omnidirectional variogram.

2. Calculate directional variograms, major and minor.

Calculate well-behaved interpretable experimental variograms.



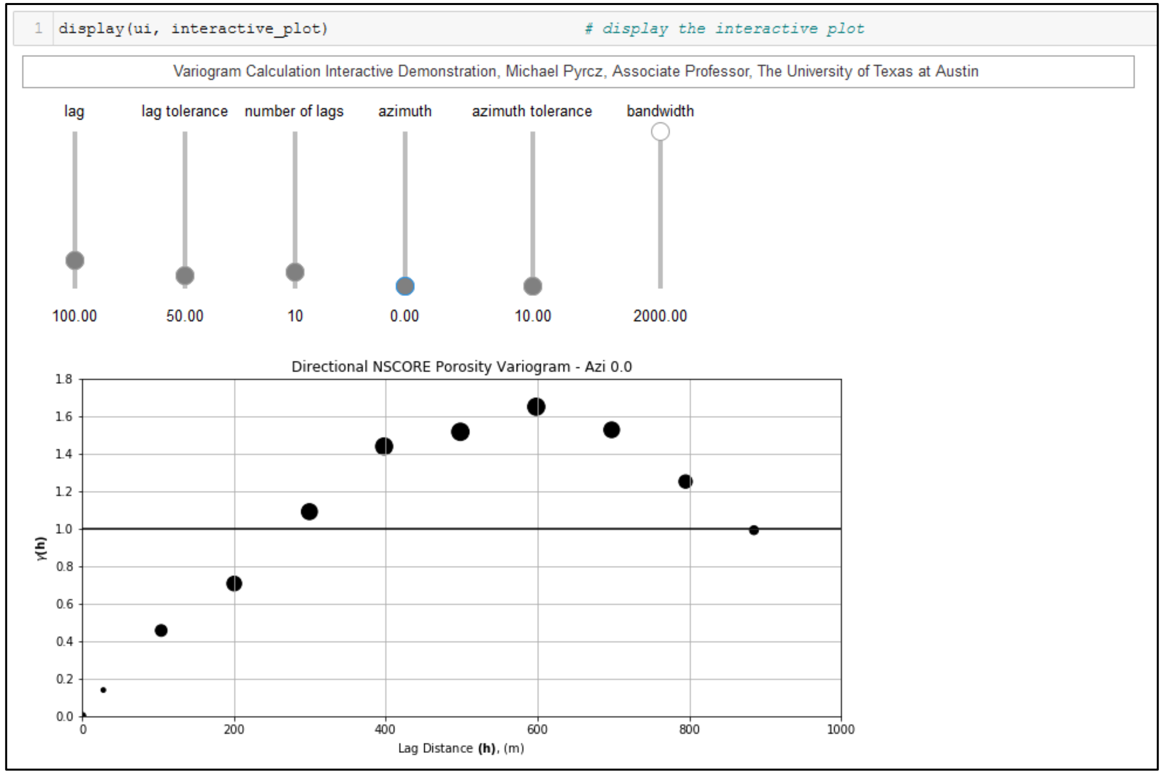
Variogram calculation dashboard in Excel, file is Variogram_Calc_Model_Demo_v2.0.xlsx.

Variogram Calculation in Python

Interactive Variogram Calculation in Python

Try these and make observations:

- Change the isotropic lag, lag tolerance and azimuth
- Change the tolerances and to calculate anisotropic experimental variograms.



Variogram calculation dashboard in Python, file is Interactive_Variogram_Calculation.ipynb.

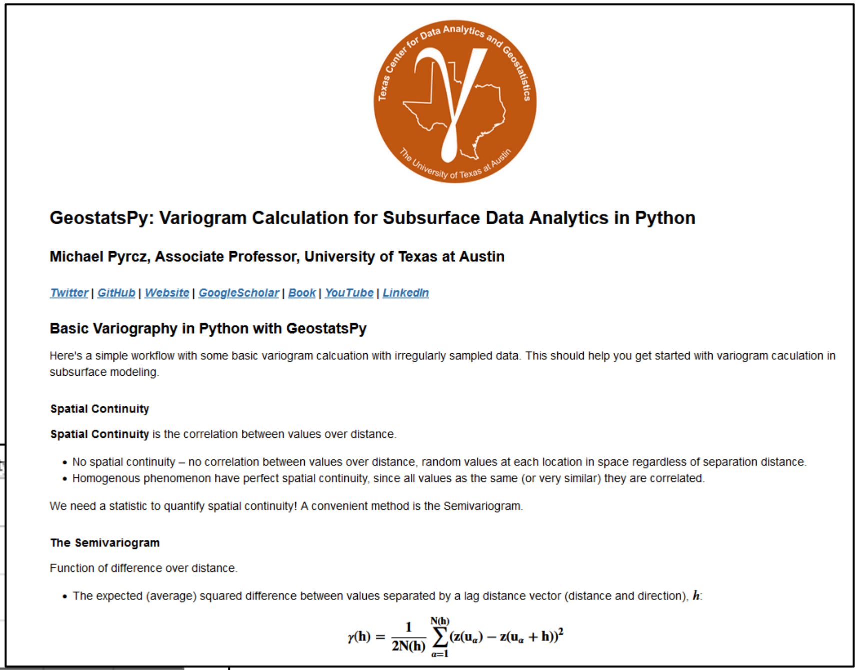
Variogram Calculation in Python

Variogram Calculation Workflow in Python

Walkthrough and try to:

- Change the tolerances and to calculate anisotropic experimental variograms.

- File is: GeostatsPy_variogram_calculation.ipynb



GeostatsPy: Variogram Calculation for Subsurface Data Analytics in Python

Michael Pyrcz, Associate Professor, University of Texas at Austin

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

Basic Variography in Python with GeostatsPy

Here's a simple workflow with some basic variogram calcuation with irregularly sampled data. This should help you get started with variogram caculation in subsurface modeling.

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.

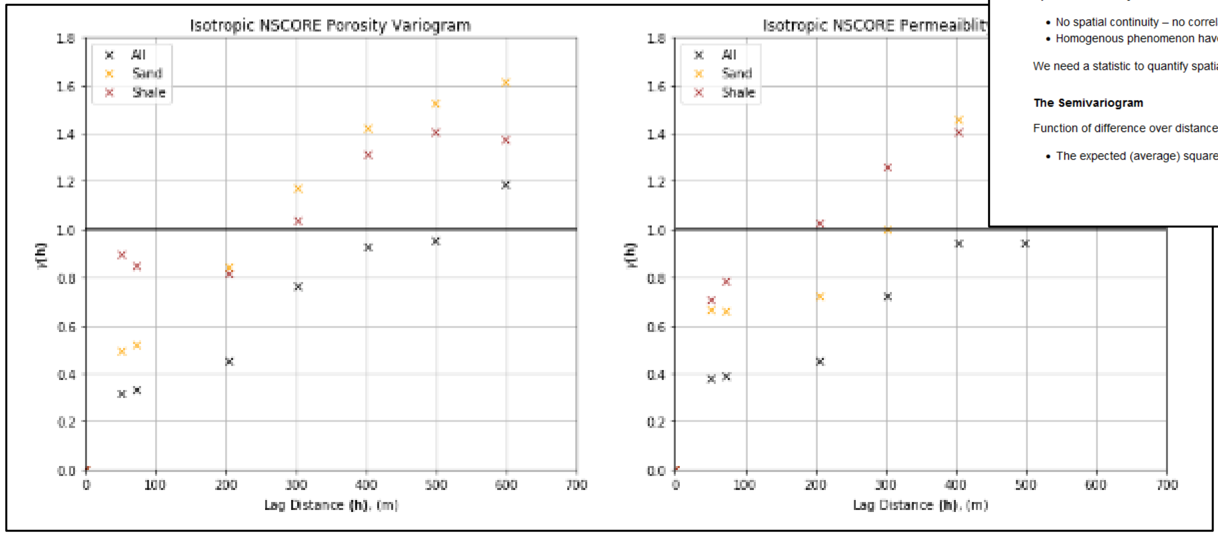
We need a statistic to quantify spatial continuity! A convenient method is the Semivariogram.

The Semivariogram

Function of difference over distance.

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

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



PGE 338 Data Analytics and Geostatistics

Lecture 10: Spatial Continuity

Lecture outline . . .

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