Data Analytics and Geostatistics: Sparse Data

Lecture outline . . .

- Spatial Continuity
- Variogram Concepts
- Variogram Calculation

Data Analytics and Geostatistics:
Sparse Data

Texas Center for Geostatistics
The University of Texas at Austin

Lecture outline . . .

- Spatial Continuity

Introduction

Modeling Prerequisites

Spatial Estimation

- Stationarity and Trends
- Spatial Continuity Calculation
- Spatial Continuity Modeling
- Spatial Continuity Estimation

Spatial Uncertainty

Multivariate, Spatial

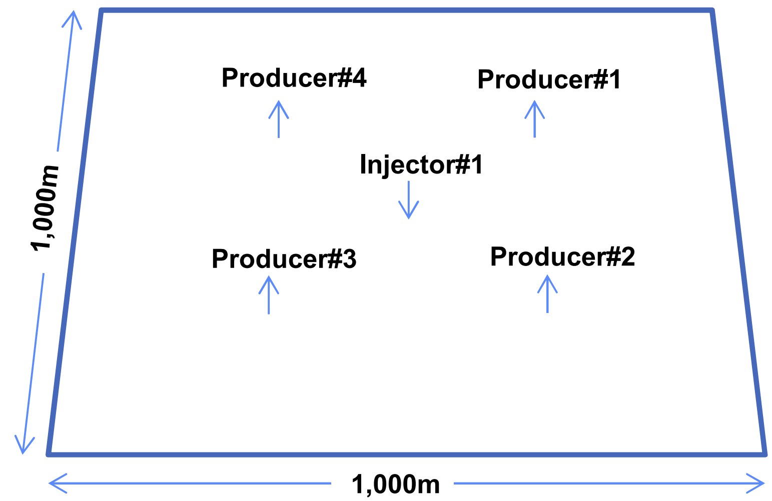
Novel Workflows

Conclusions

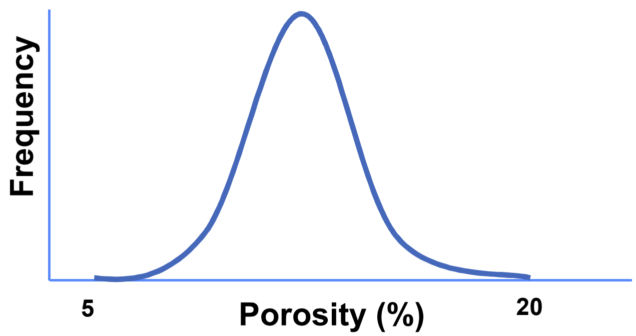
Motivation for Measuring Spatial Continuity

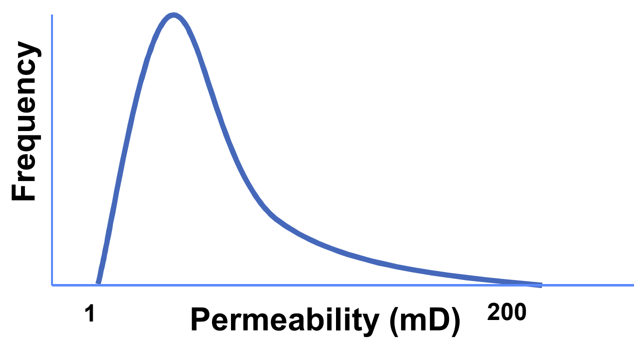
Simple Example

- Area of interest
- 1 Injector and 4 producers



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

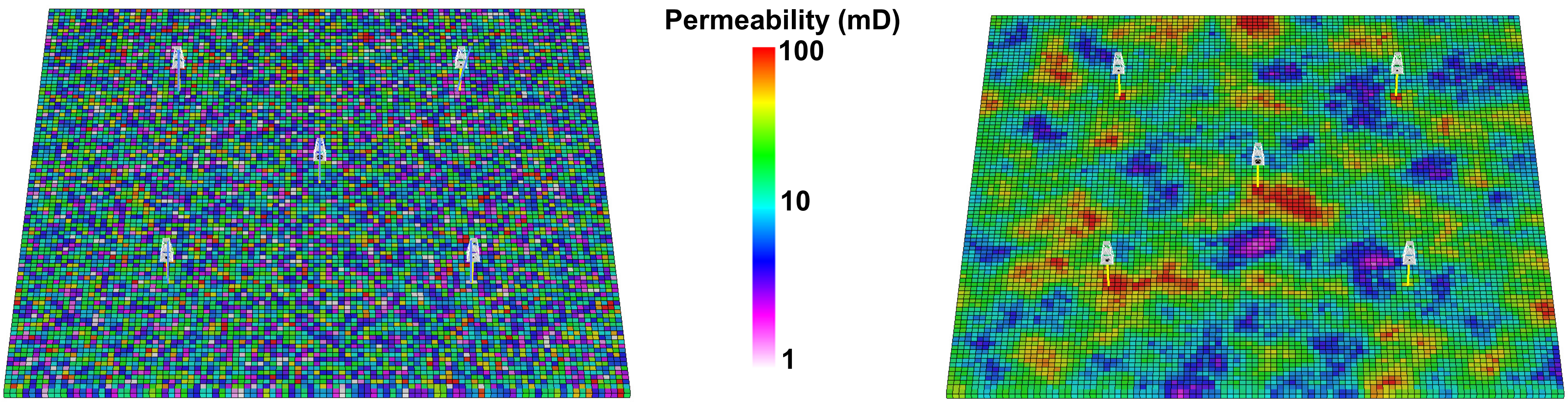




Motivation for Measuring Spatial Continuity

Does spatial continuity of reservoir properties matter?

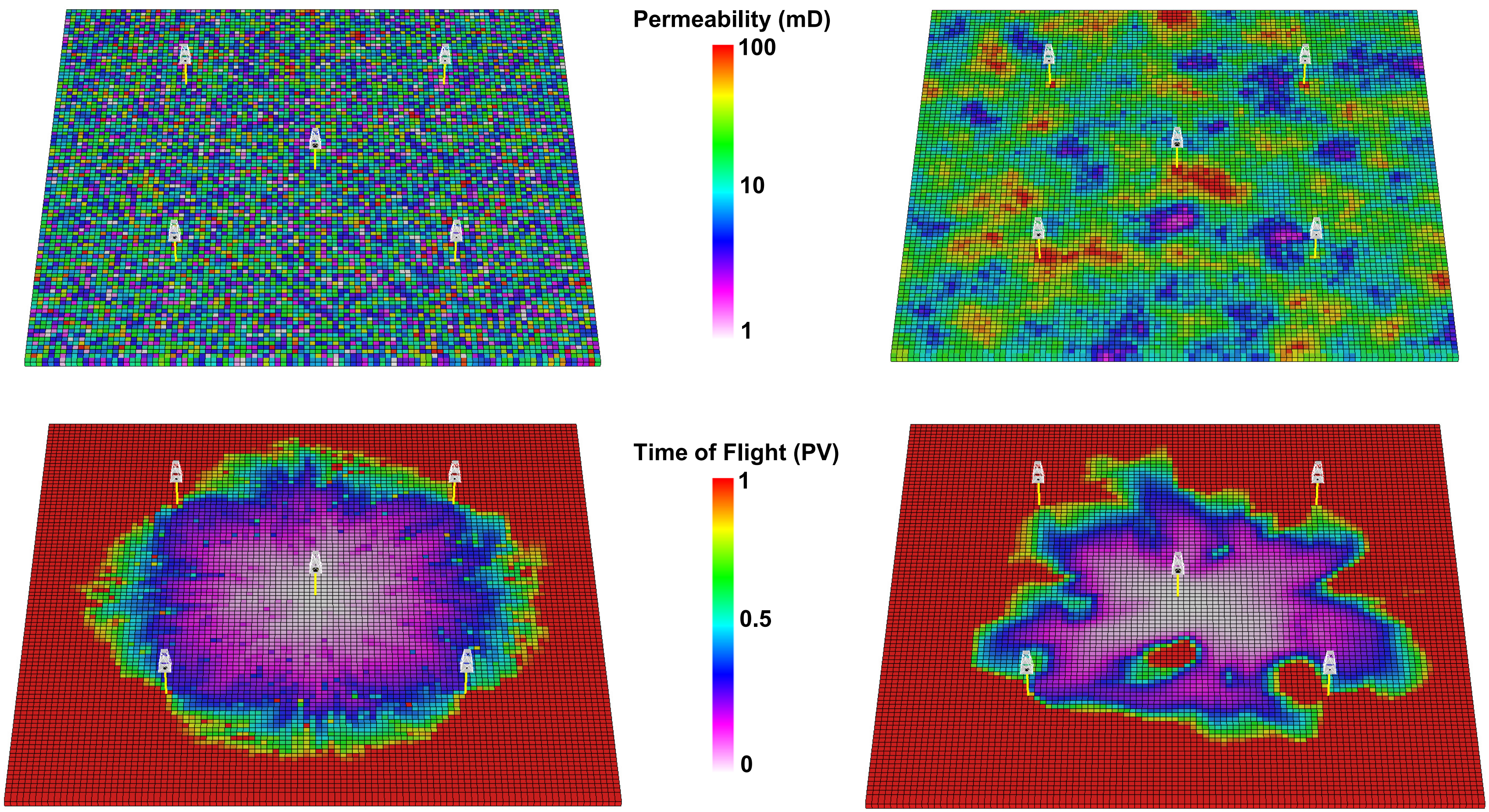
Consider these models of permeability



Recall – all models have the same porosity and permeability distributions

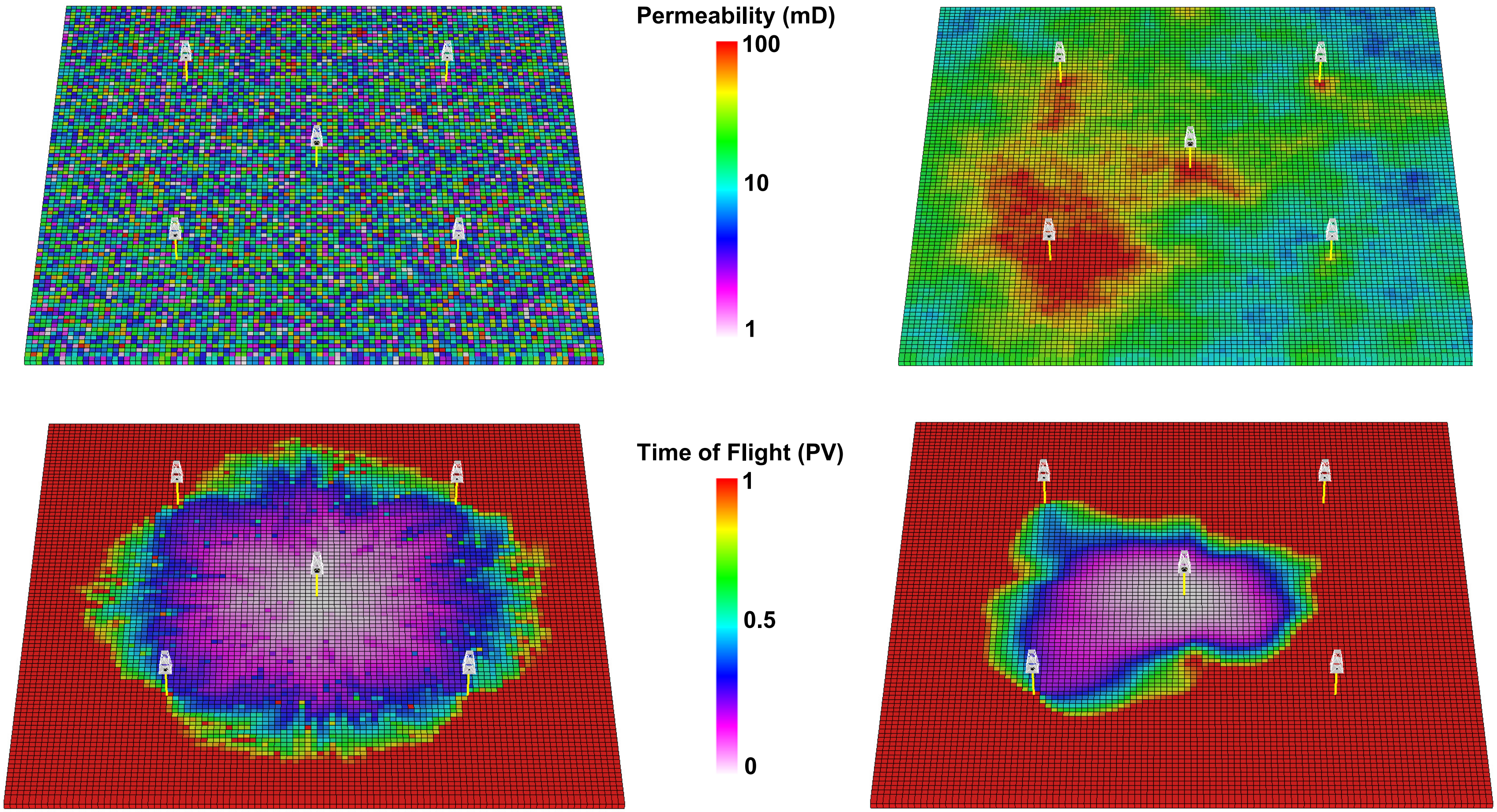
- Mean, variance, P10, P90 …
- Same static oil in place!

Motivation for Measuring Spatial Continuity



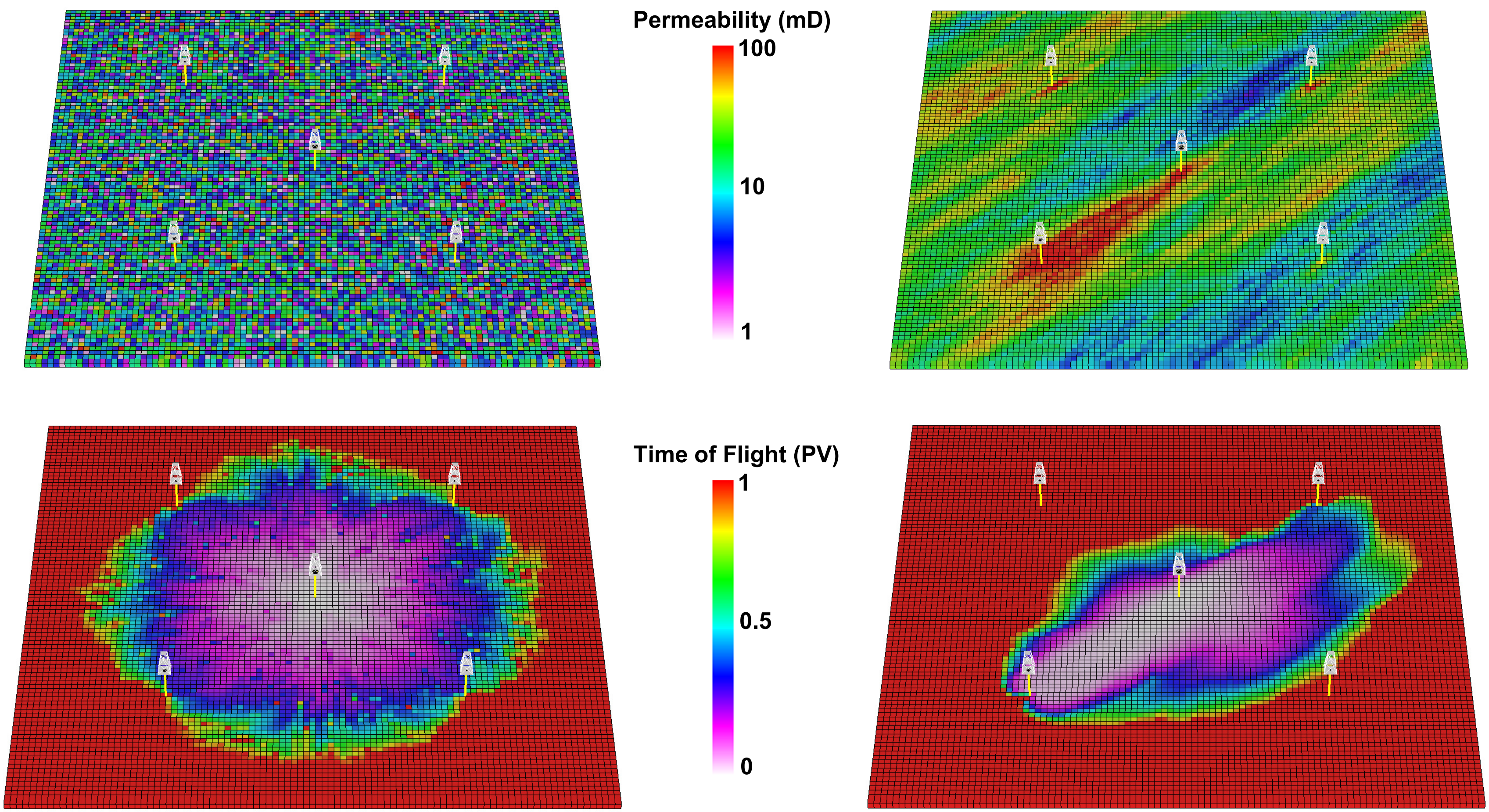
How does heterogeneity impact recovery factor?

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

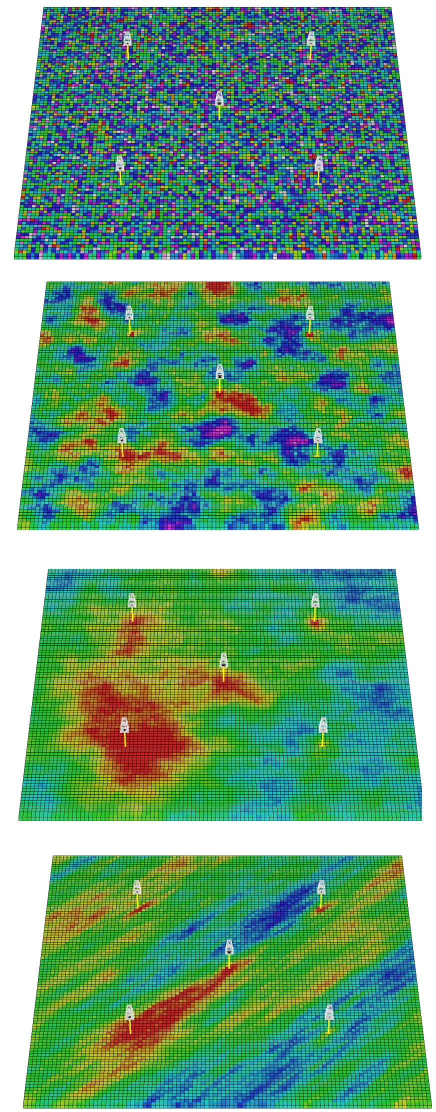
- For the same reservoir property distributions a wide range of spatial continuities are possible.
- Spatial continuity often impacts reservoir forecasts.
- Need to be able to:

Characterize / Quantify

↓ Spatial Continuity Measures

Impose in Reservoir Models

Spatial Continuity



"Very Short"

"Medium"

"Long"

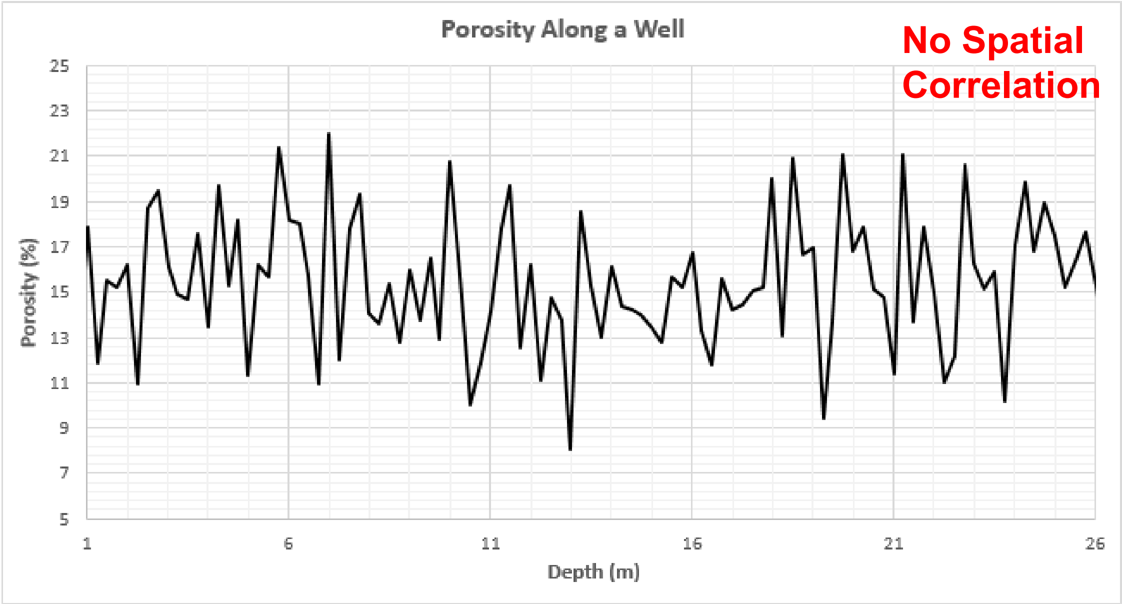
"Anisotropic "

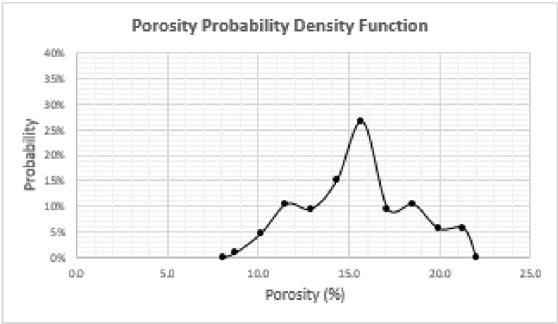
Spatial Continuity Definition

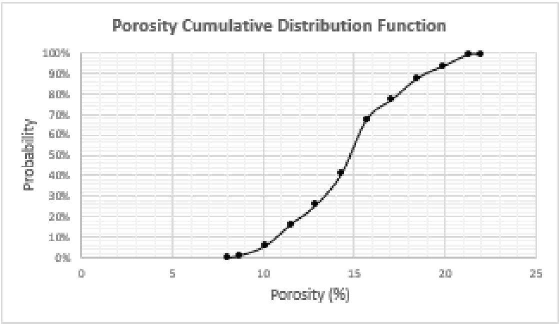
- **Spatial Continuity** – correlation between values over distance.
  - No spatial continuity – no correlation between values over distance, random values at each location in space regardless of separation distance.
  - Homogenous phenomenon have perfect spatial continuity, since all values as the same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

Spatial Continuity Definition

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

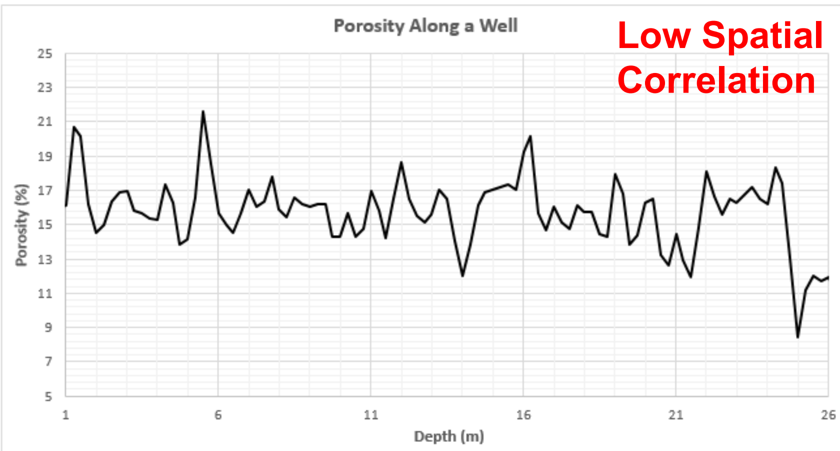


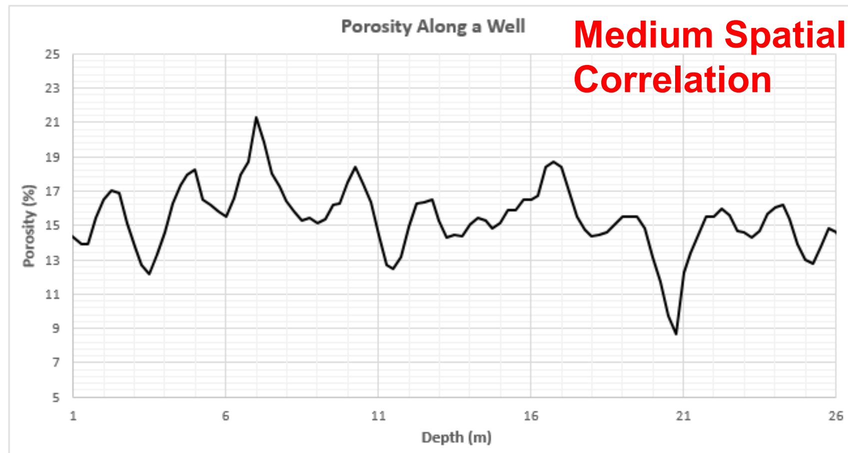


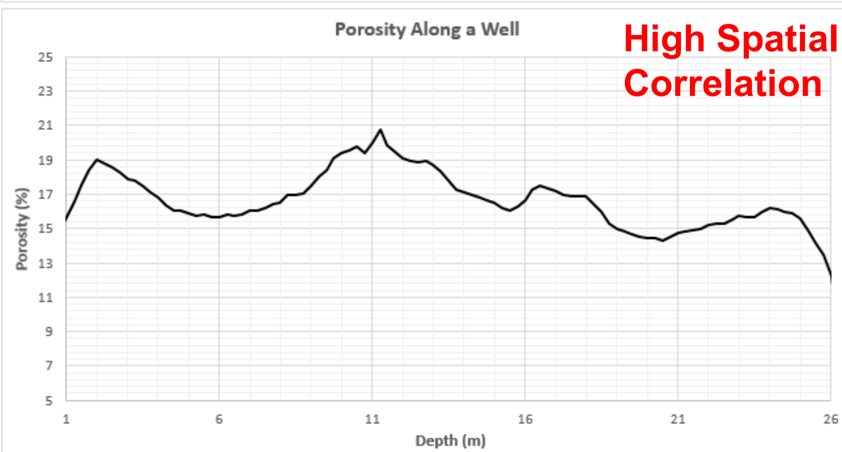


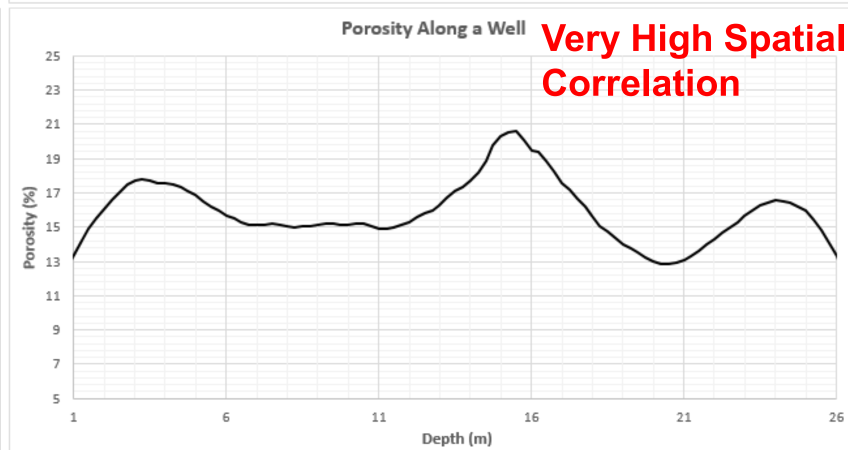
Spatial Continuity Definition

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



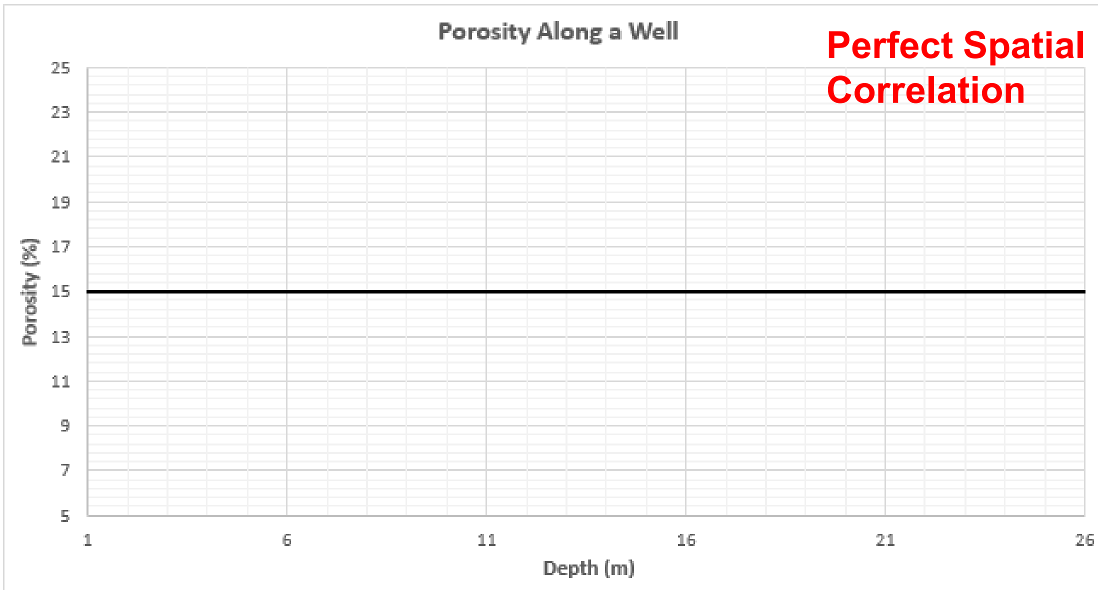


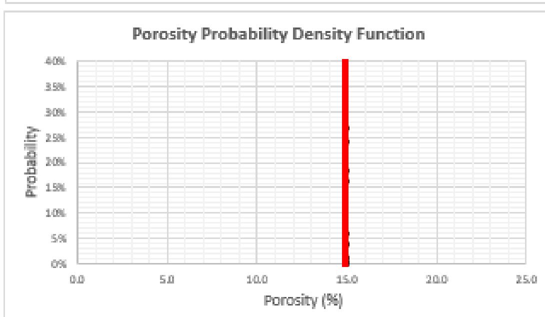


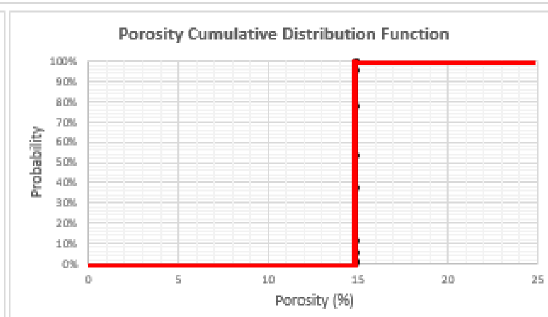


Spatial Continuity Definition

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







Data Analytics and Geostatistics: Sparse Data

Lecture outline . . .

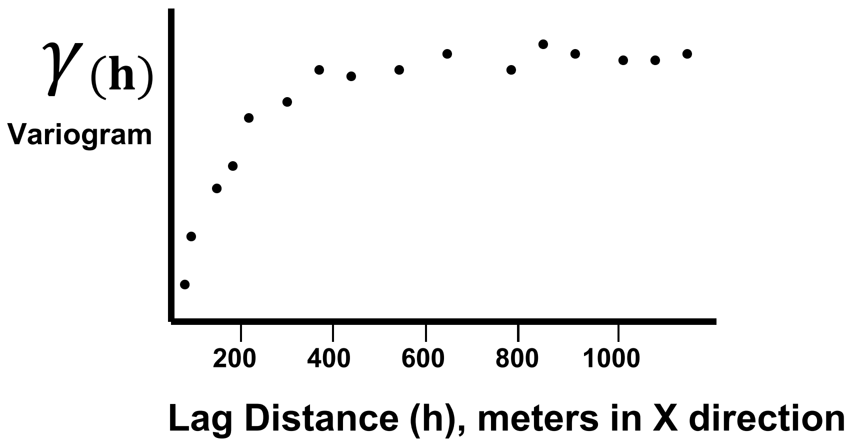
- Variogram Concepts

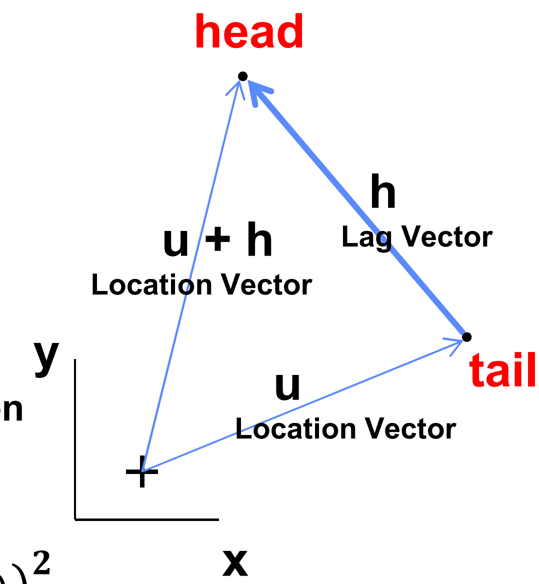
Measuring Spatial Continuity

We need a statistic to quantify spatial continuity!

The Semivariogram:

- Function of difference over distance.





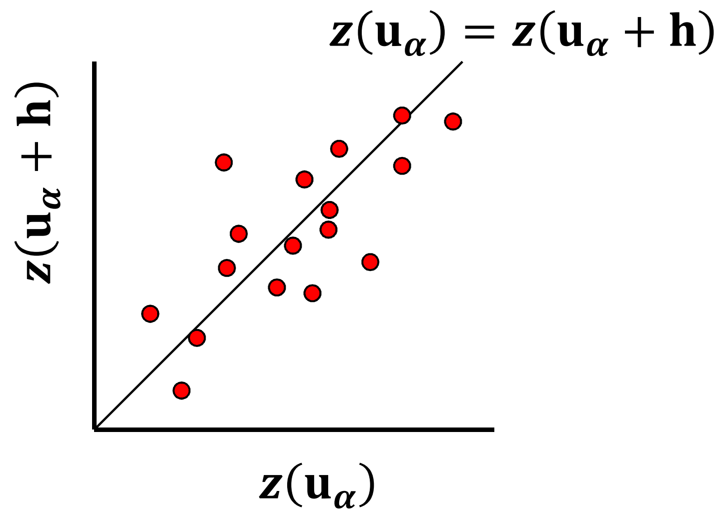
- The equation:

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

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

Modified from Deutsch, MinE310

"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:

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

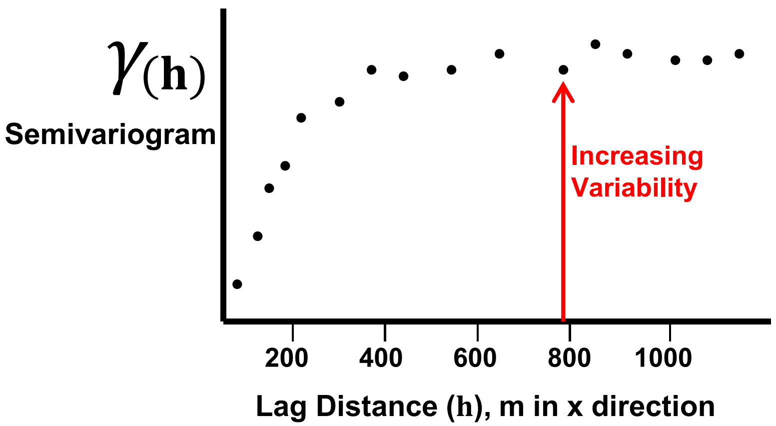
  - Note the correlogram is related to the covariance function as:

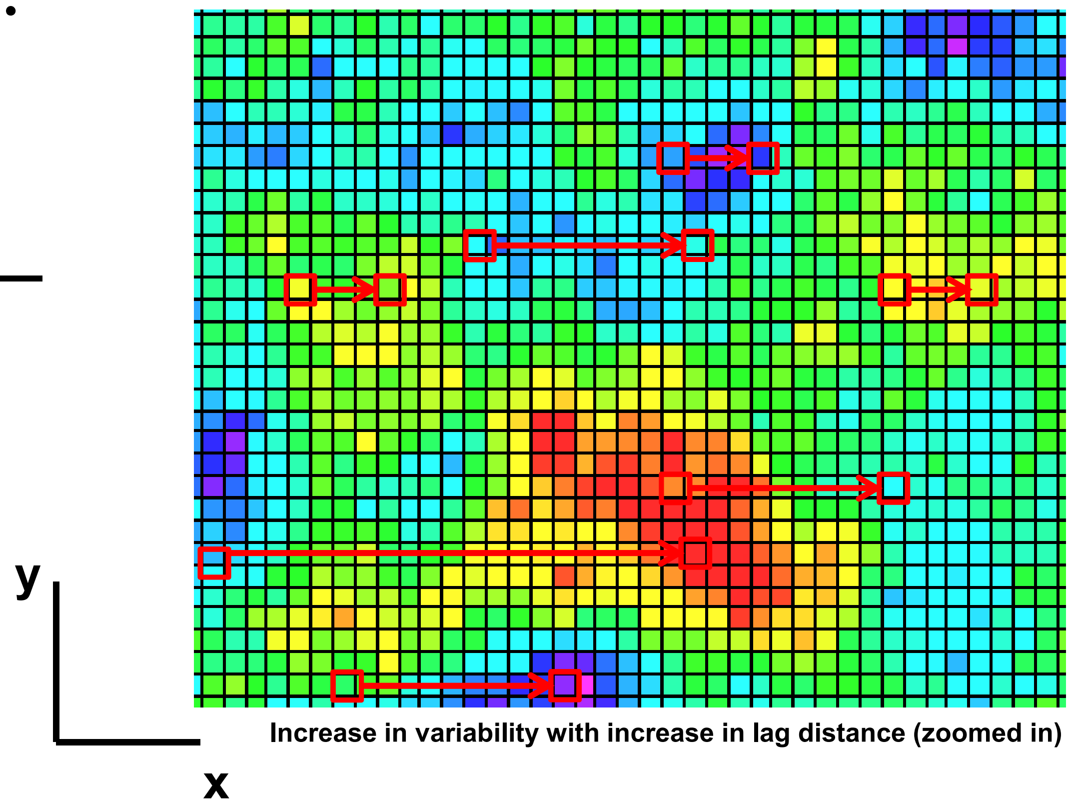
$$\rho_x(h) = \frac{C_x(h)}{\sigma_x^2}$$, h-scatter plot correlation vs. lag distance

Variogram Observations

Observation #1

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



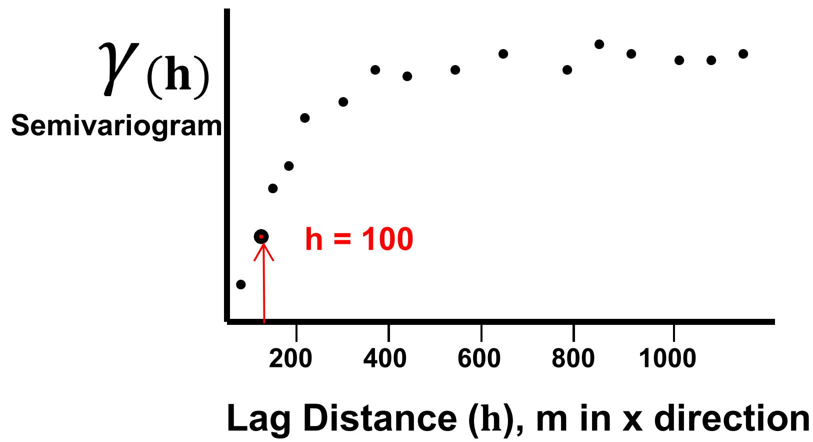


Increase in variability with increase in lag distance (zoomed in)

Variogram Observations

Observation #2

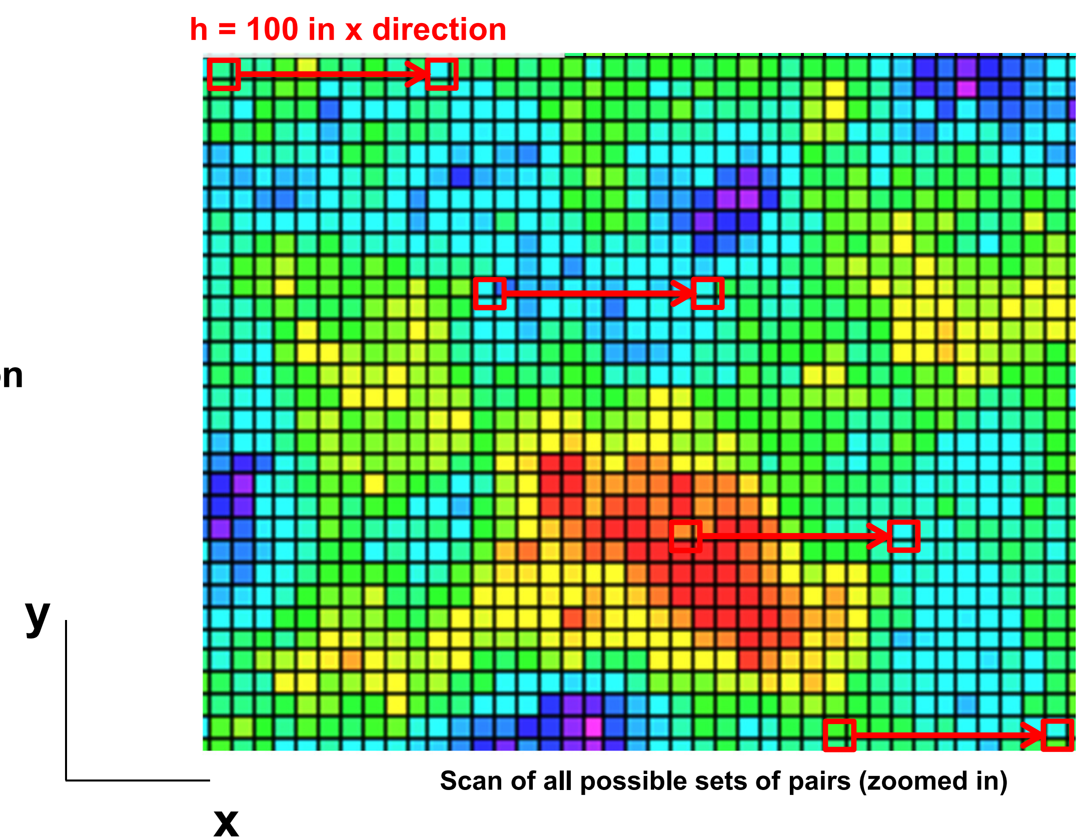
- Calculated with over all possible pairs separated by lag vector, $\mathbf{h}$.



- **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).



Scan of all possible sets of pairs (zoomed in)

Variogram Observations

Observation #3

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

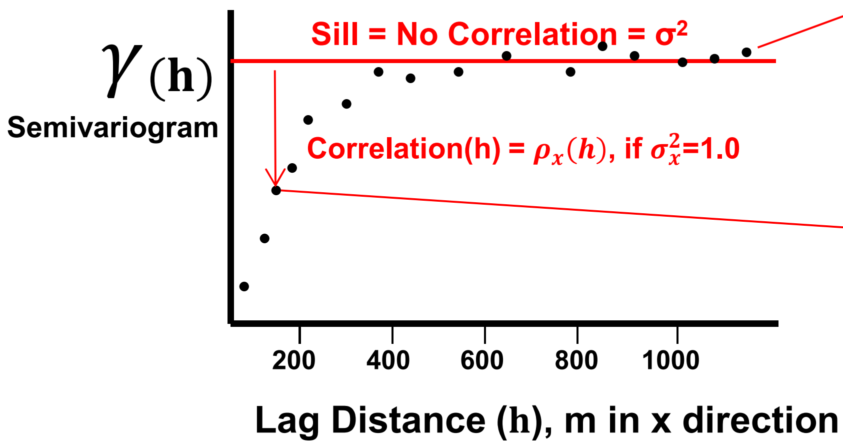
  Sill is the Variance, σ^2

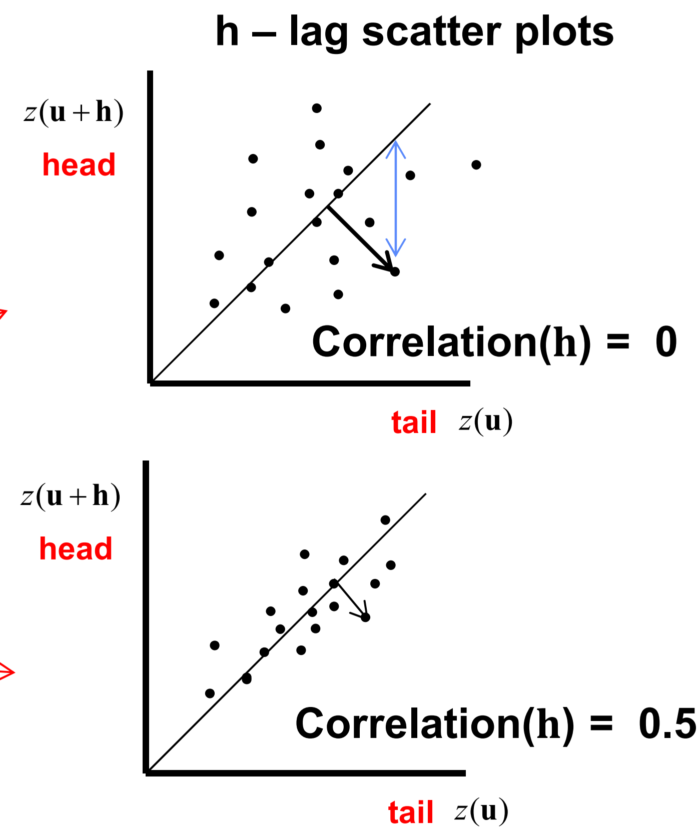
- Given stationarity of the variance and $\gamma_x(\mathbf{h})$:

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.

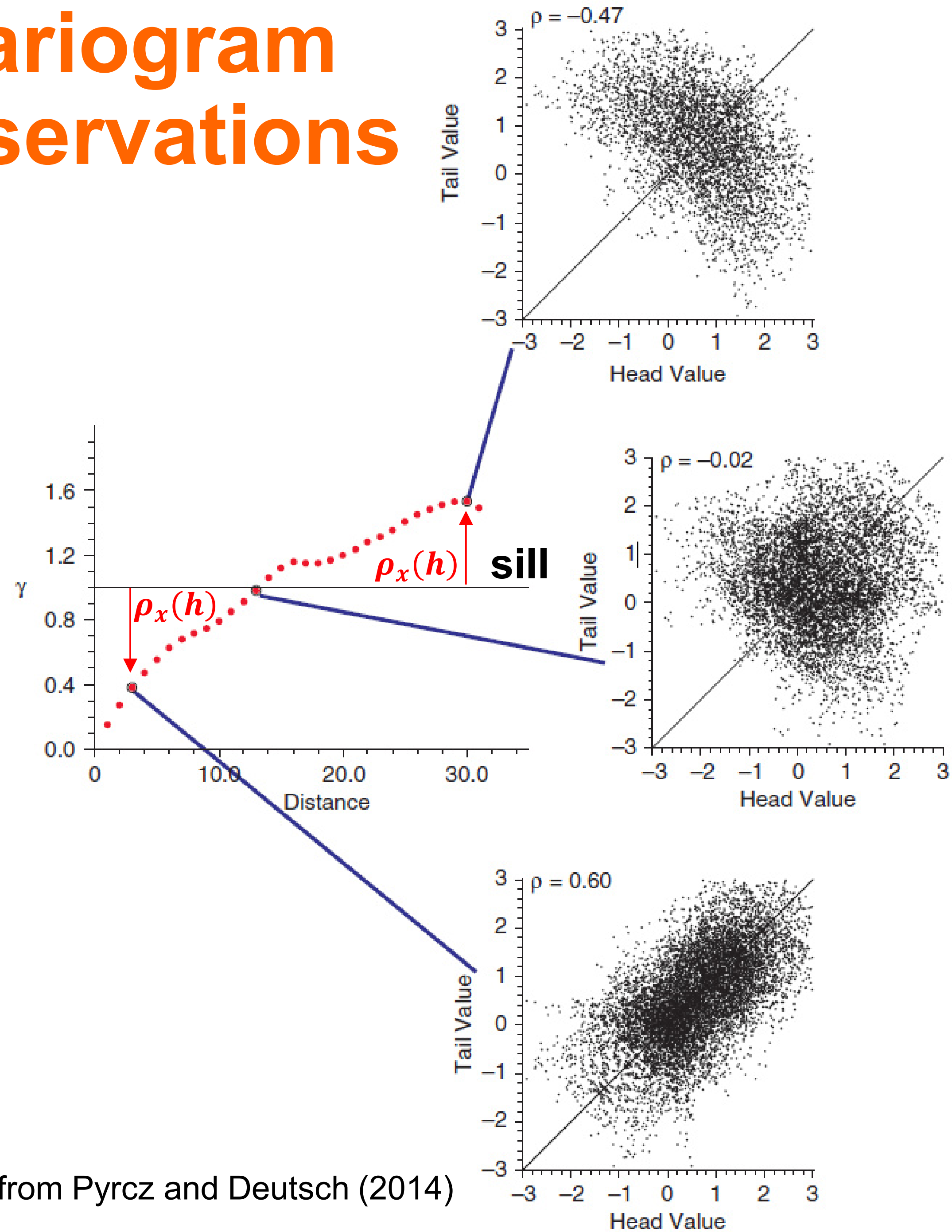


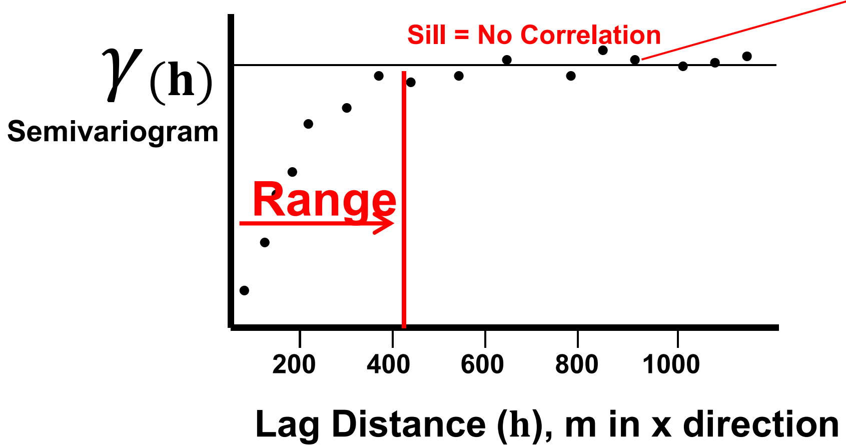
Image modified from Pyrcz and Deutsch (2014)

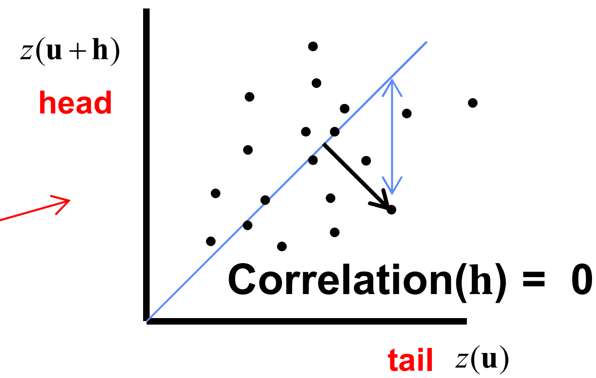
Variogram Observations

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.

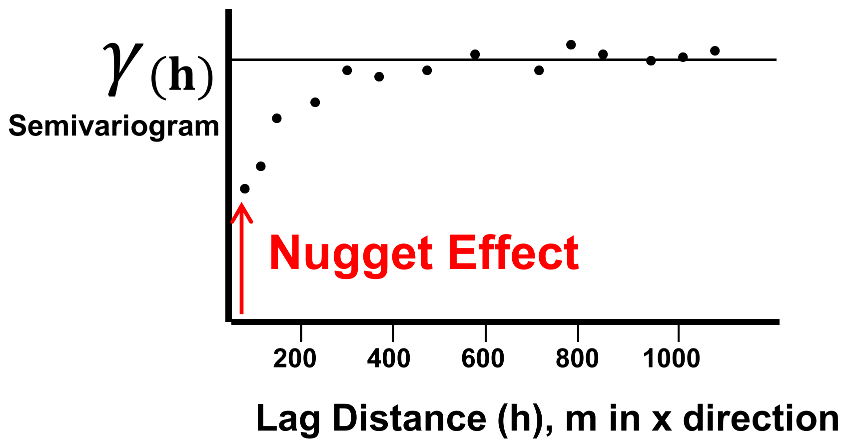




Variogram Observations

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

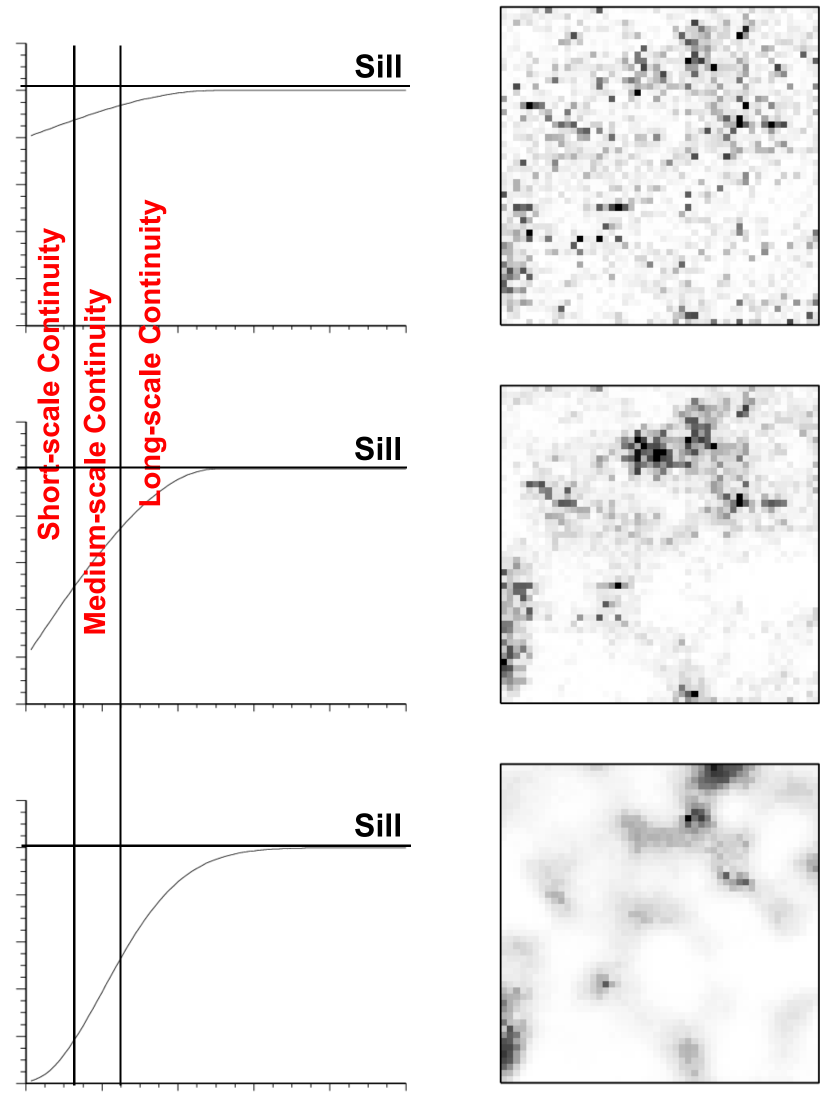


Image modified from Pyrcz and Deutsch (2014)

Data Analytics and Geostatistics: Sparse Data

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:

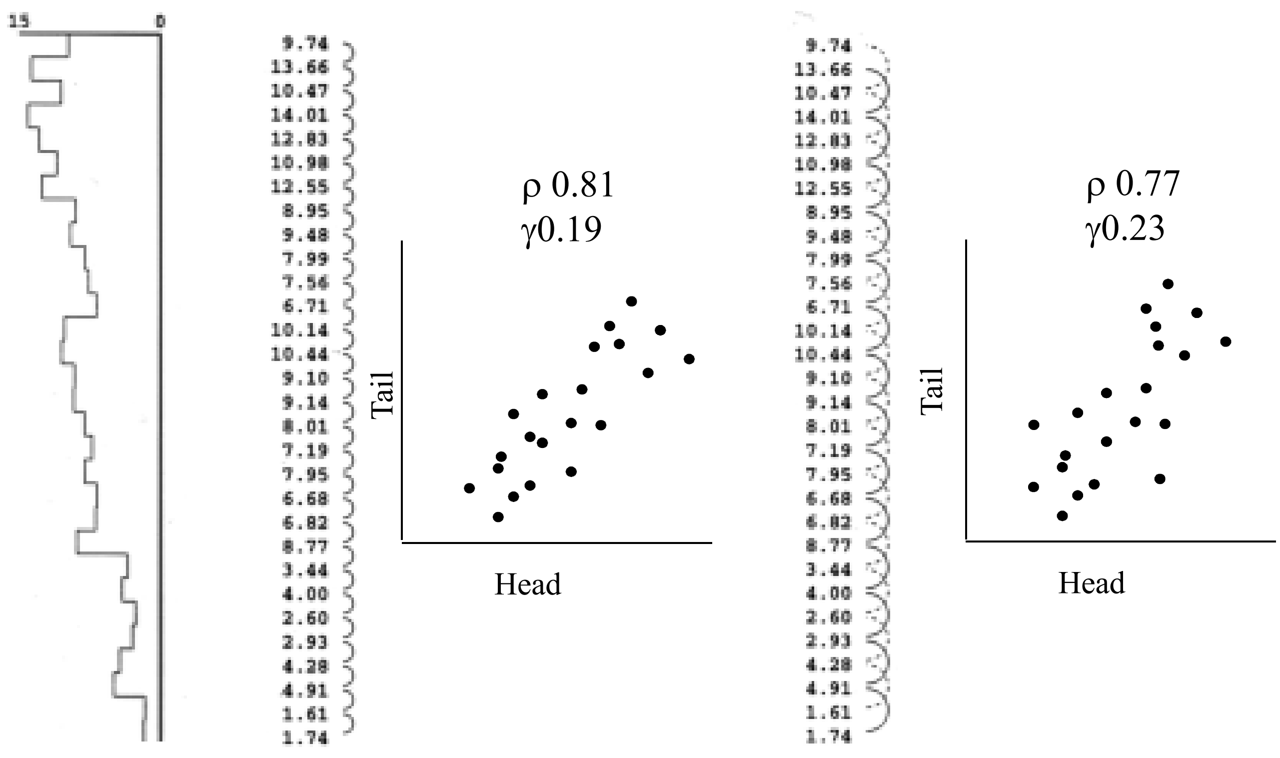
$$C_x(h) = \sigma_x^2 - \gamma_x(\mathbf{h})$$

  - 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

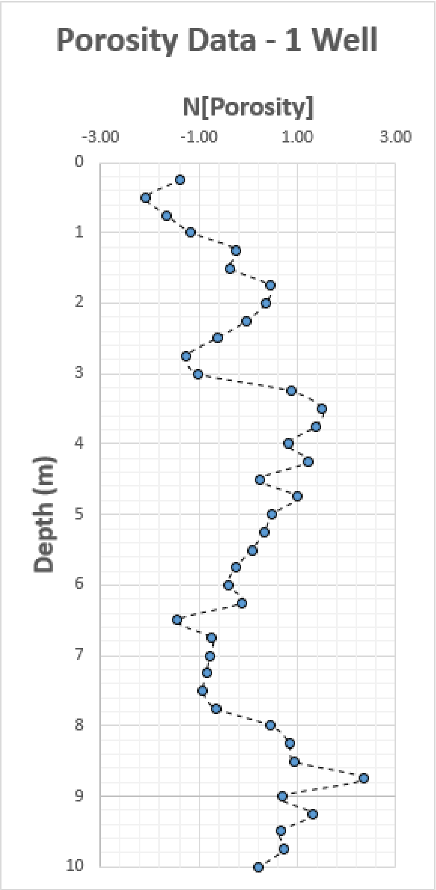
- Consider data values separated by *lag* vectors (the h values)
- Here are two examples of a lag vector equal to the data spacing and then twice the data spacing:



Variogram Calculation Example

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

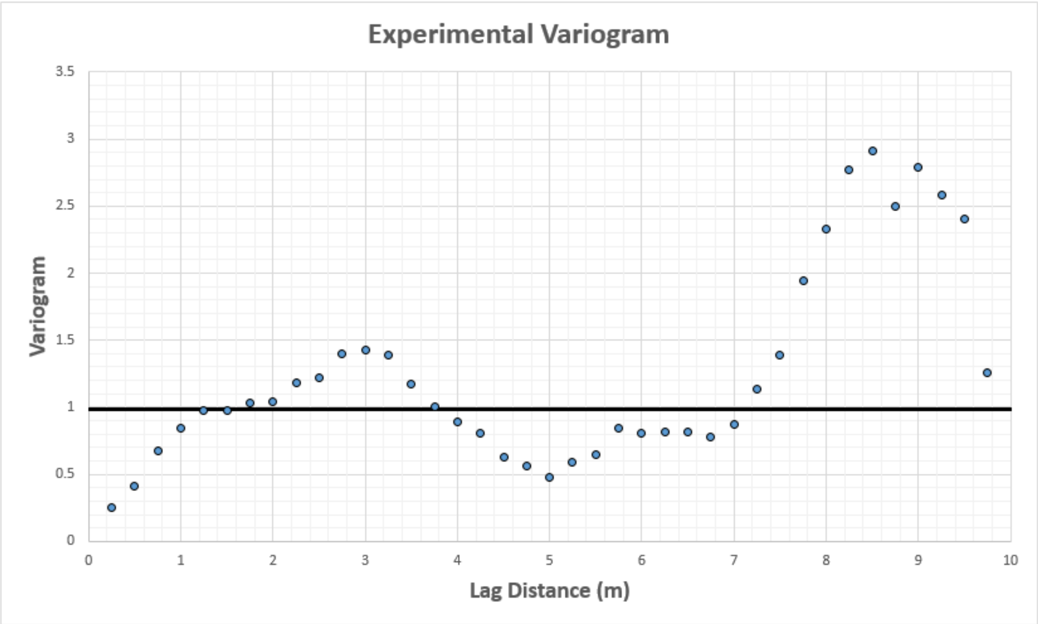
- Dataset is at GitHub/GeostatsGuy
- GeoDataSets/1D_Porosity.xlsx



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):

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

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

where:

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

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

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

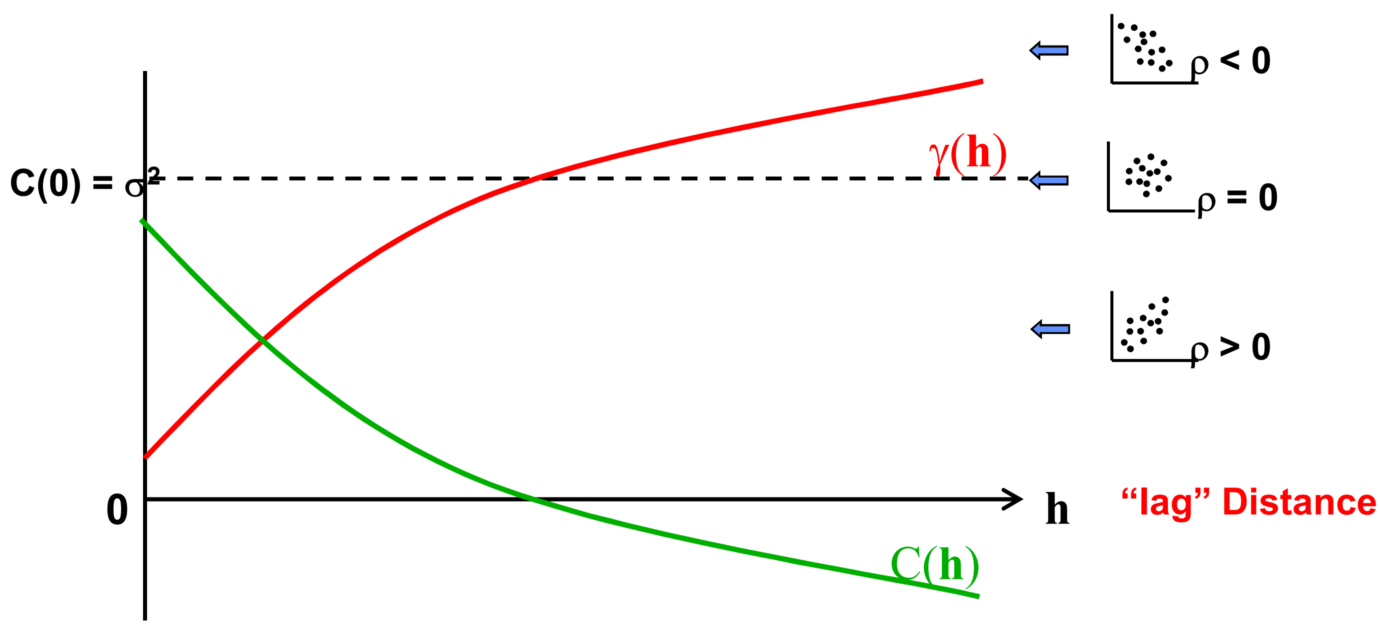
Stationarity entails that:

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

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

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.

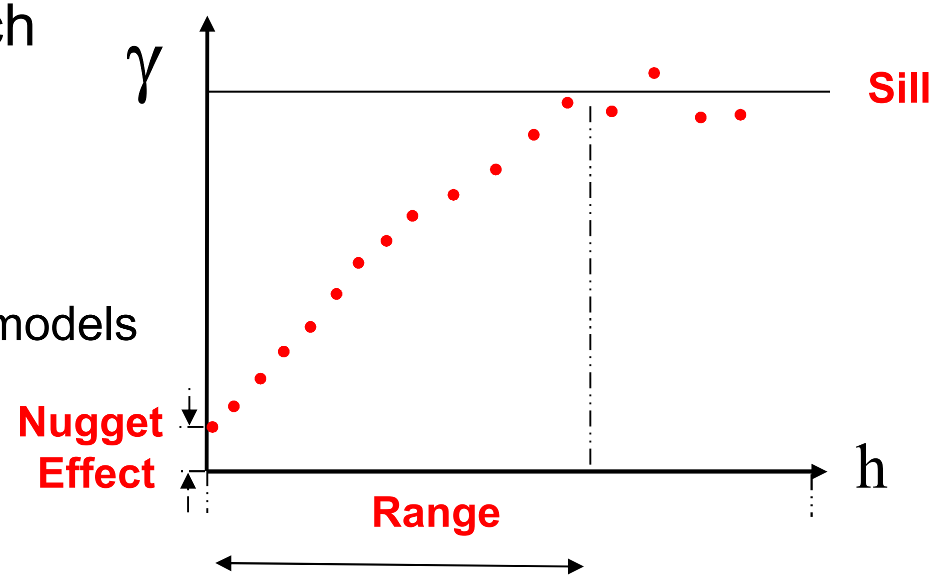
$$\boldsymbol{C_x}(\mathbf{h}) = \frac{\sum_{\alpha=1}^{n} x(\mathbf{u}_\alpha) \cdot x(\mathbf{u}_\alpha + \mathbf{h})}{n} - (\overline{x})^2, \text{ if stationary mean}$$

  – 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

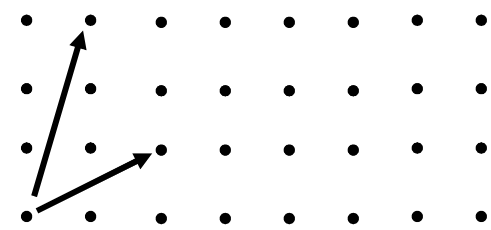
2. Valid spatial model
   – Fit with a couple different, nest (additive) spatial continuity models e.g. nugget, spherical, exponential and Gaussian

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



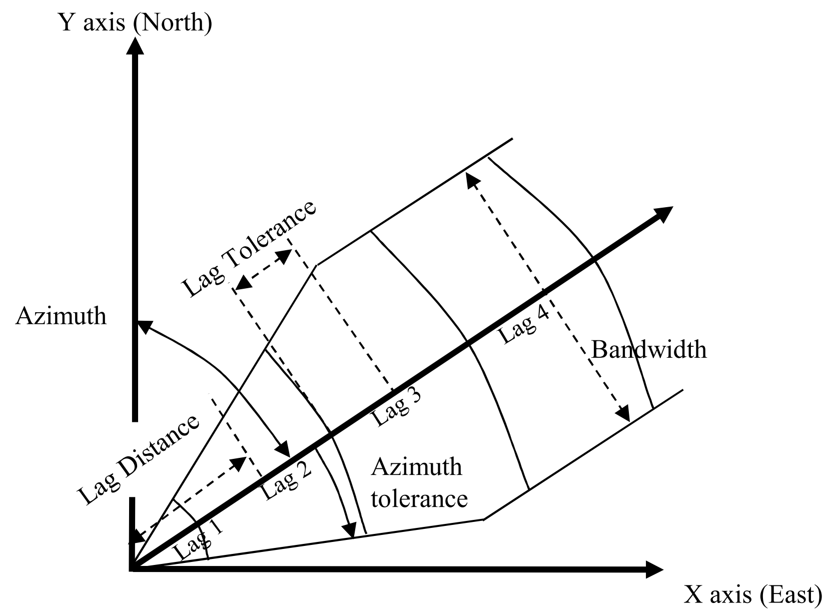
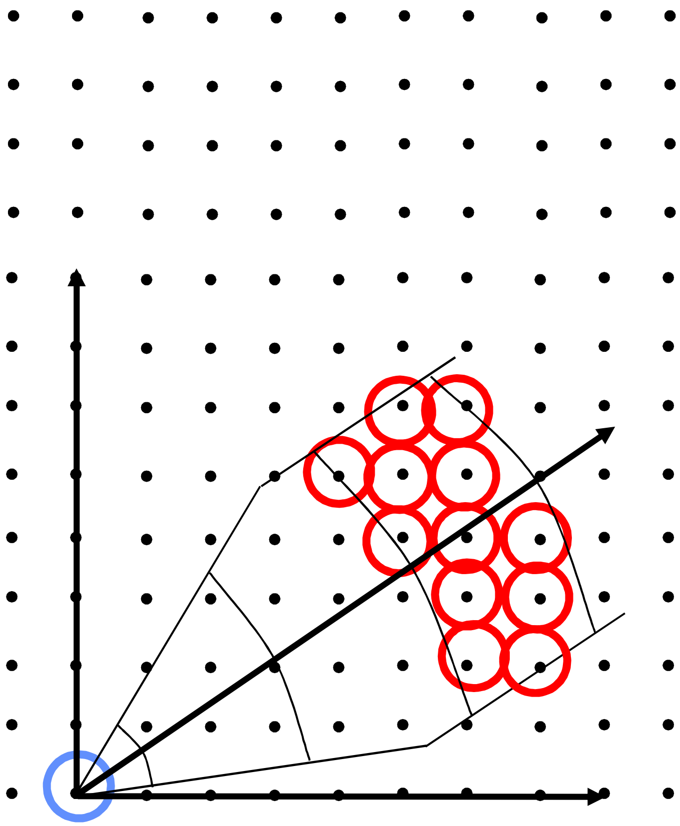


Image from Pyrcz and Deutsch, 2014

Calculating Experimental Variograms

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

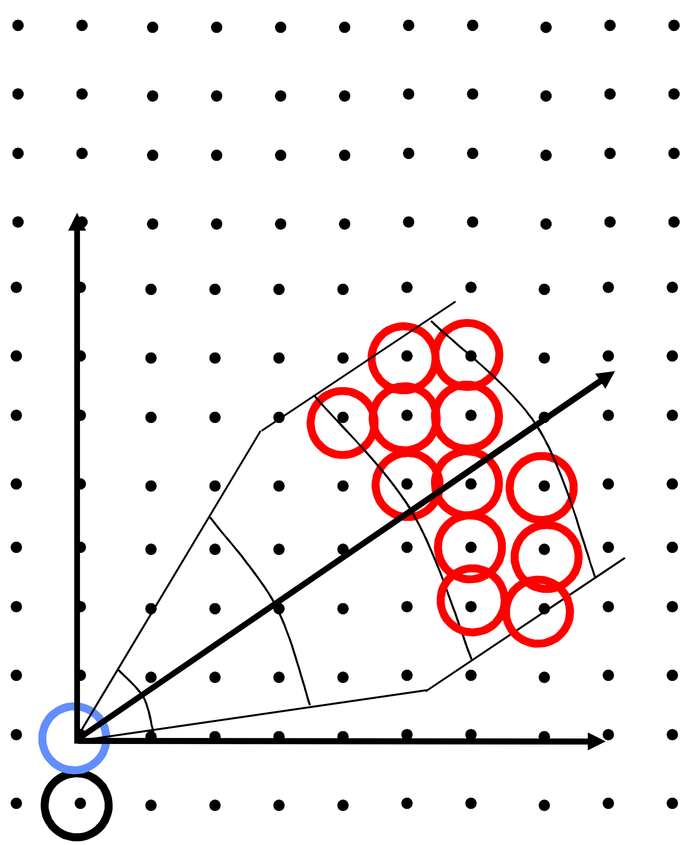


$$2\gamma(h) = \frac{1}{N(h)} \sum_{N(h)} [z(u) - z(u+h)]^2$$

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

...

Calculating Experimental Variograms



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

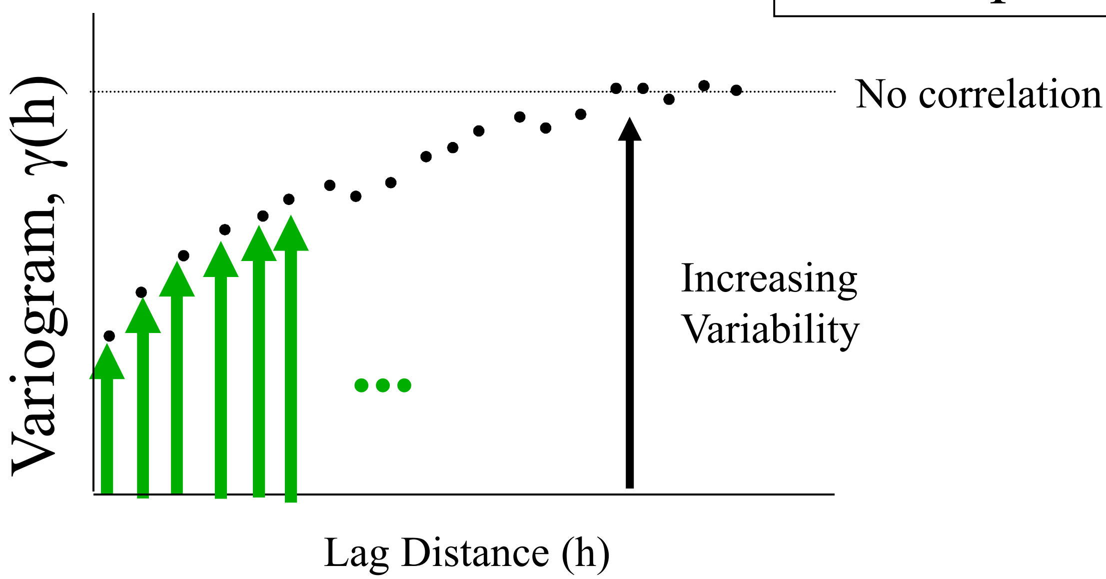
Move to next node.

...

Calculating Experimental Variograms

Now Repeat for All Nodes

And Repeat for All Lags



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

Remainder of Today

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:

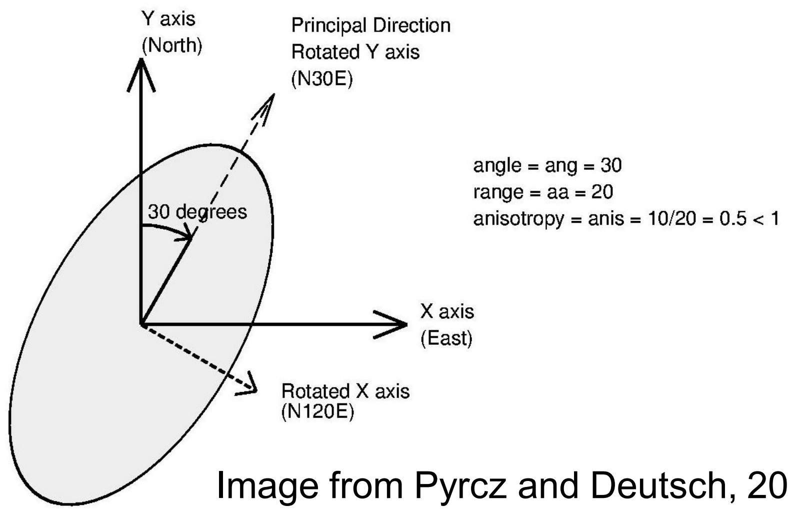


Image from Pyrcz and Deutsch, 2014

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

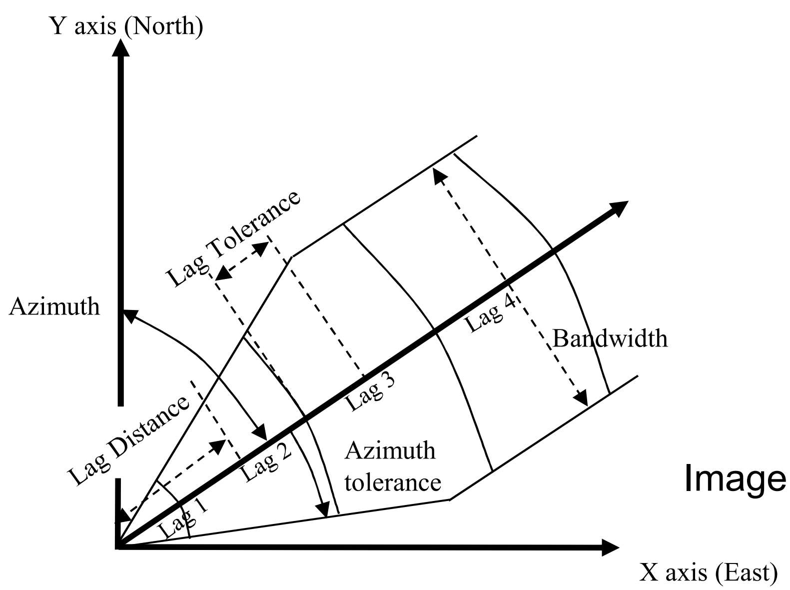


Image from Pyrcz and Deutsch, 2014

Variogram Calculation

Parameter File

```
                    Parameters for GAMV
                    *******************

START OF PARAMETERS:
1D_Porosity.dat                   -file with data
1   2   3                           - columns for X, Y, Z coordinates
1   4   0                           - number of variables,col numbers
-1.0e21 1.0e21                    - trimming limits
gamv.out                          -file for variogram output
20                                -number of lags
0.25                                -lag separation distance
0.125                               -lag tolerance
1                                 -number of directions
0.0 90.0 50.0  90.0 22.5 50.0       -azm,atol,bandh,dip,dtol,bandv
0                                 -standardize sills? (0=no, 1=yes)
1                                 -number of variograms
1   1   1                         -tail var., head var., variogram type
```

Output File

Semivariogram		tail:Nporosity			head:Nporosity
1	.000	.00000	80	.02225	.02225
2	.000	.00000	0	.00000	.00000
3	[illegible]	.24947	78	.03769	.03769
4	[illegible]	.41112	76	.05658	[illegible]
5	.750	.67790	74	.07176	.07176
6	[illegible]	.84266	[illegible]	.07167	[illegible]
7	1.250	.97437	70	.06729	.06729
8	[illegible]	.97806	68	.04000	.04000
9	1.750	1.03163	[illegible]	.02015	.02015
10	2.000	[illegible]	64	.00188	.00188
11	2.250	[illegible]	62	[illegible]	-.00532
12	2.500	1.21635	60	.01600	.01600
13	2.750	1.39850	58	.05414	.05414
14	3.000	1.42226	56	[illegible]	.08964
15	3.250	1.38802	54	.09111	.09111
16	3.500	1.17446	52	.08000	.08000
17	3.750	.99823	50	.08460	.08460
18	4.000	.89012	48	.07417	.07417
19	4.250	.80637	46	.06000	.06000
20	4.500	.62689	44	.06318	.06318
21	4.750	.56242	42	.04095	.04095
22	5.000	.47498	40	.02225	.02225

Lag # lag distance h

γ(h)

number of pairs

mean tail

mean head

Data File

```
1DPorosity
4
X
Y
Z
Nporosity
```

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

Variogram Calculation

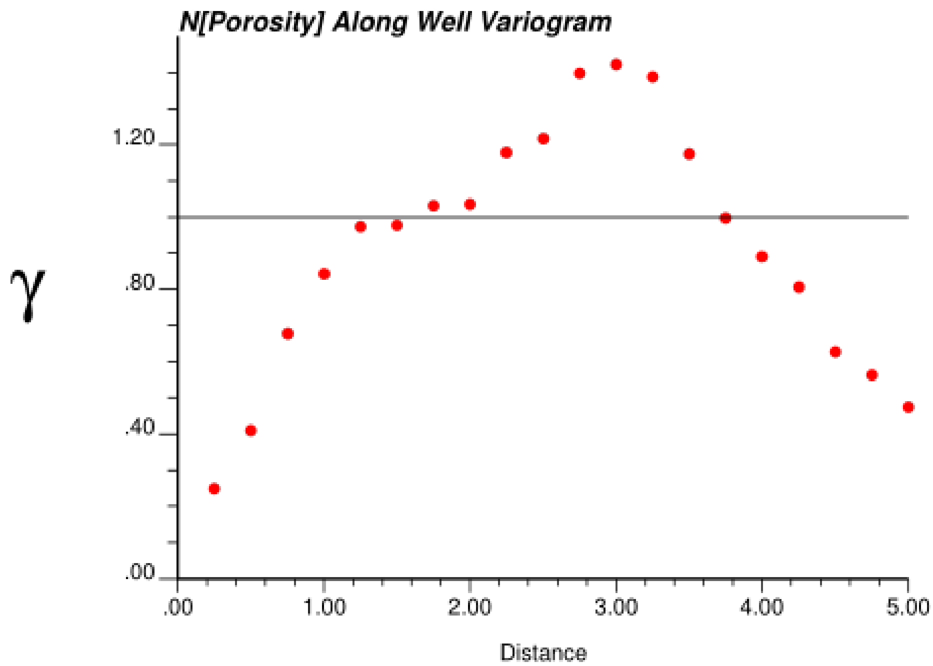
First Walkthrough Together with GSLIB

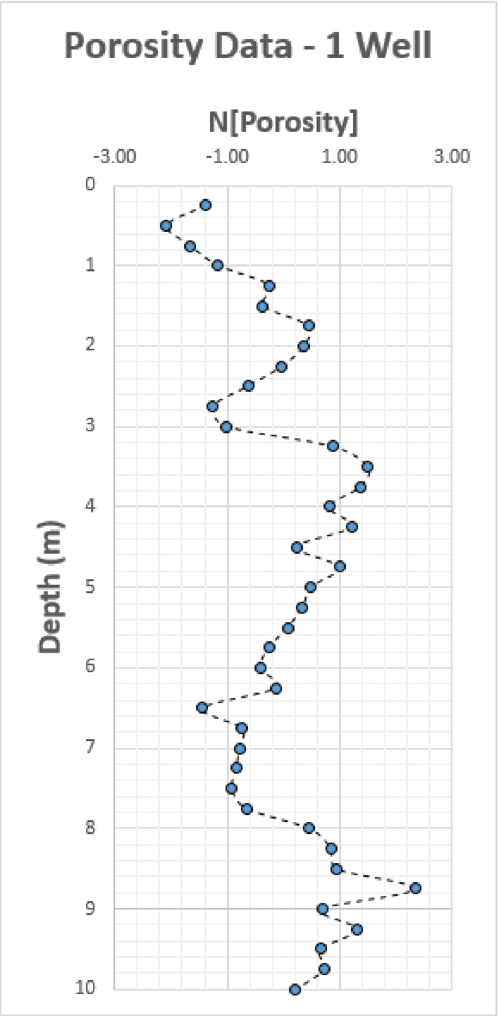
Steps:

1. Copy data file (1D_Porosity.dat), parameter files (gamv.par and vargplt.par) and executables (gamv.exe and vargplt.exe – windows / linux available) to your working directory.
2. Review the data file.
3. Review the parameter files.
4. Open a command prompt.
5. Run variogram calculation. [gamv gamv.par]
6. Review output file.
7. Run variogram visualization. [vargplt vargplt.par]. Requires a program to open postscript (e.g. ghostview).

Variogram Calculation

First Walkthrough Together with GSLIB

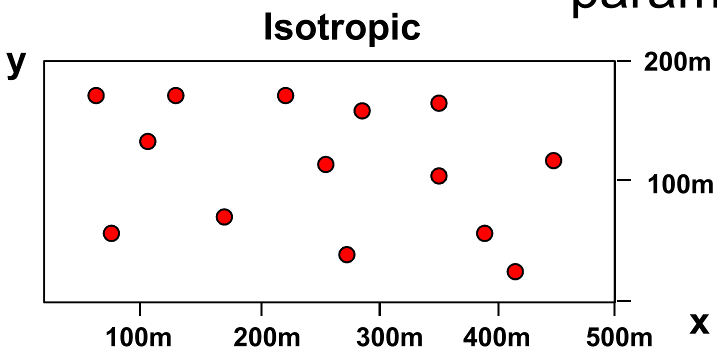




How would you interpret this result?

Choosing the Lag Distances and Tolerances

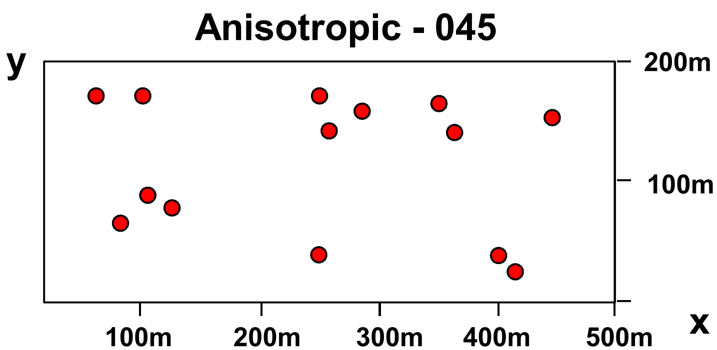
Let's pick some variogram calculation parameters in groups (estimate):



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

horizontal bandwidth ___

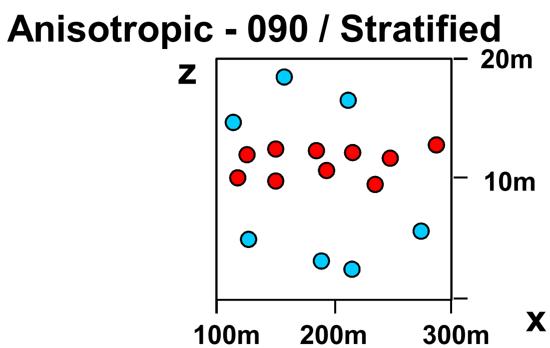
azimuth ____, azimuth tolerance ______



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

horizontal bandwidth ___

azimuth ____, azimuth tolerance ______



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

vertical bandwidth ___

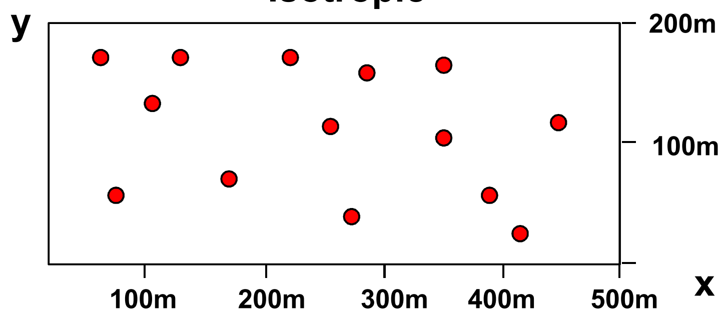
azimuth **90**, azimuth tolerance **22.5**

Along x

Choosing the Lag Distances and Tolerances

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

Isotropic



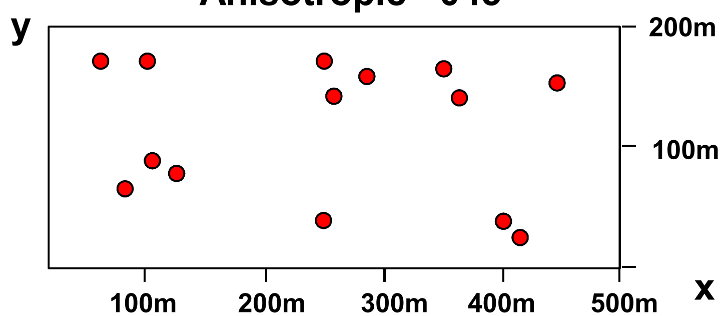
lag size **50**, lag tolerance **25**, number of lags **5**

horizontal bandwidth **∞**

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

Any

Anisotropic - 045

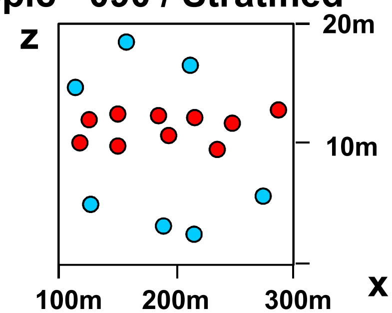


lag size **20**, lag tolerance **10**, number of lags **12**

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 Hands-on
Omnidirection and Directional

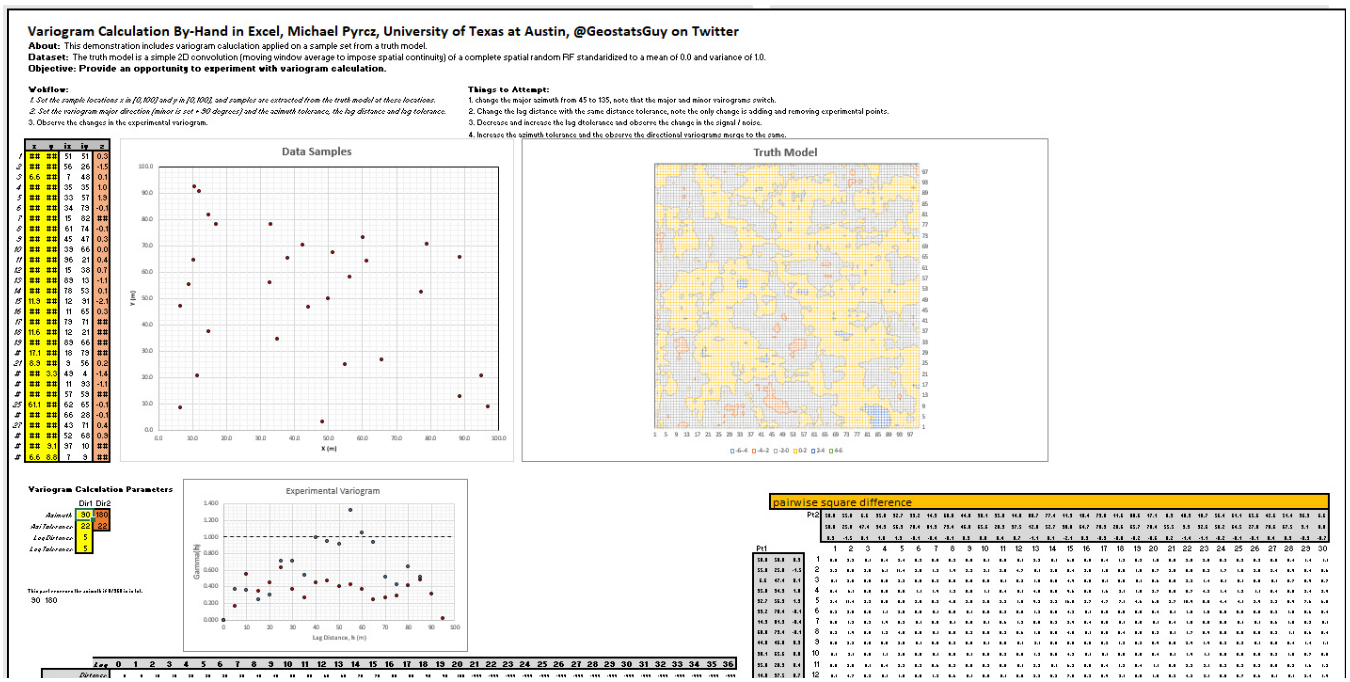
Variogram Calculation:

Things to try:

1. Calculate an omnidirectional variogram.

2. Calculate directional variograms, major and minor.

Calculate well-behaved interpretable experimental variograms.



The file is at: https://git.io/fxhxr. The file is Variogram_Calc_Model_Demo_v2.0.xlsx

Variogram Calculation in Python

Variogram Calculation Workflow in Python

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*

PGE 383 Exercise: Basic Univariate Summary Statistics and Data Distribution Representativity Plotting 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), h:

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

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.

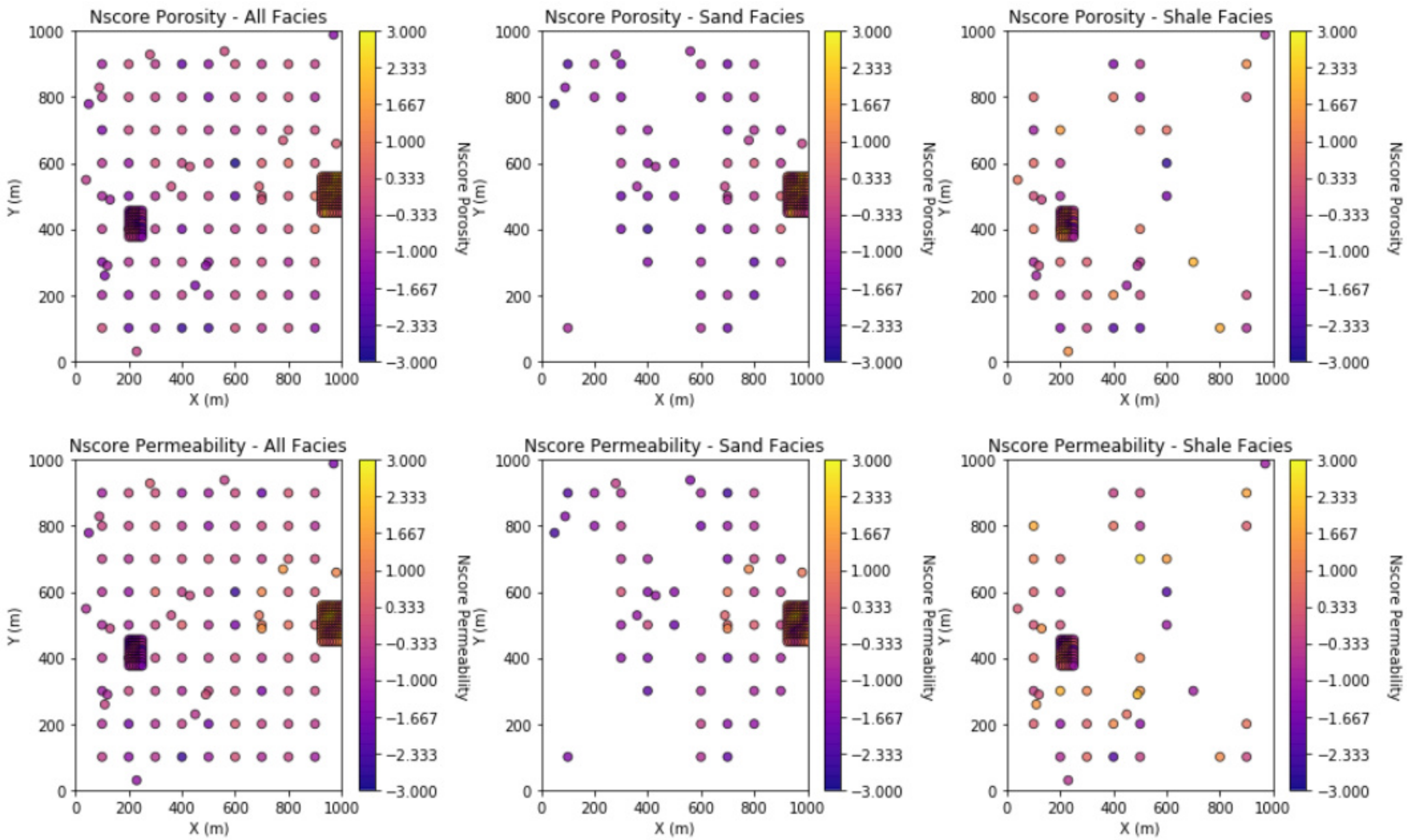
- Calculated over a suite of lag distances to obtain a continuous function.
- the $\frac{1}{2}$ term converts a variogram into a semivariogram, but in practice the term variogram is used instead of semivariogram.
- We prefer the semivariogram because it relates directly to the covariance function, $C_x(\mathbf{h})$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

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

https://github.com/GeostatsGuy/PythonNumericalDemos/blob/master/GeostatsPy_variogram_calculation.ipynb

Variogram Calculation in Python

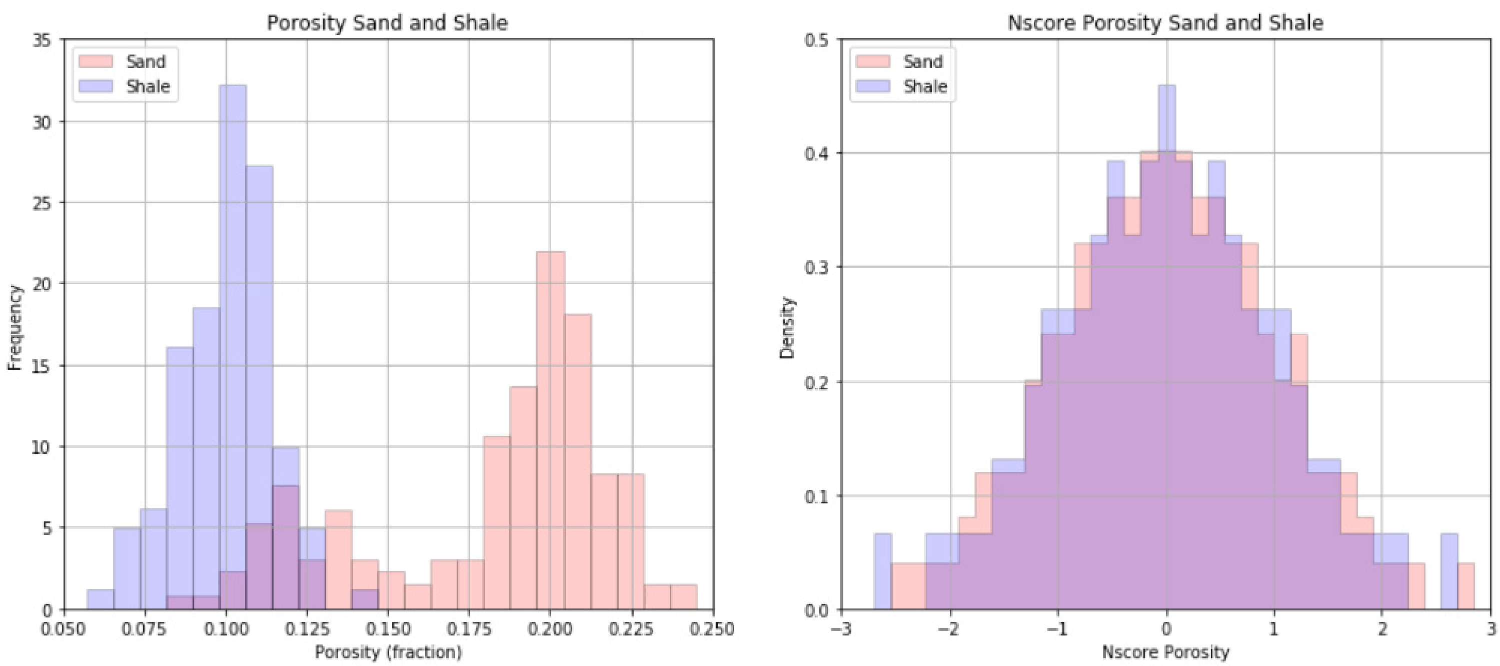
Variogram Calculation Workflow in Python



Visualize the data to check for spacing, coverage, trends etc.

Variogram Calculation in Python

Variogram Calculation Workflow in Python



Check data distributions and transform is needed.

Variogram Calculation in Python

Variogram Calculation Workflow in Python

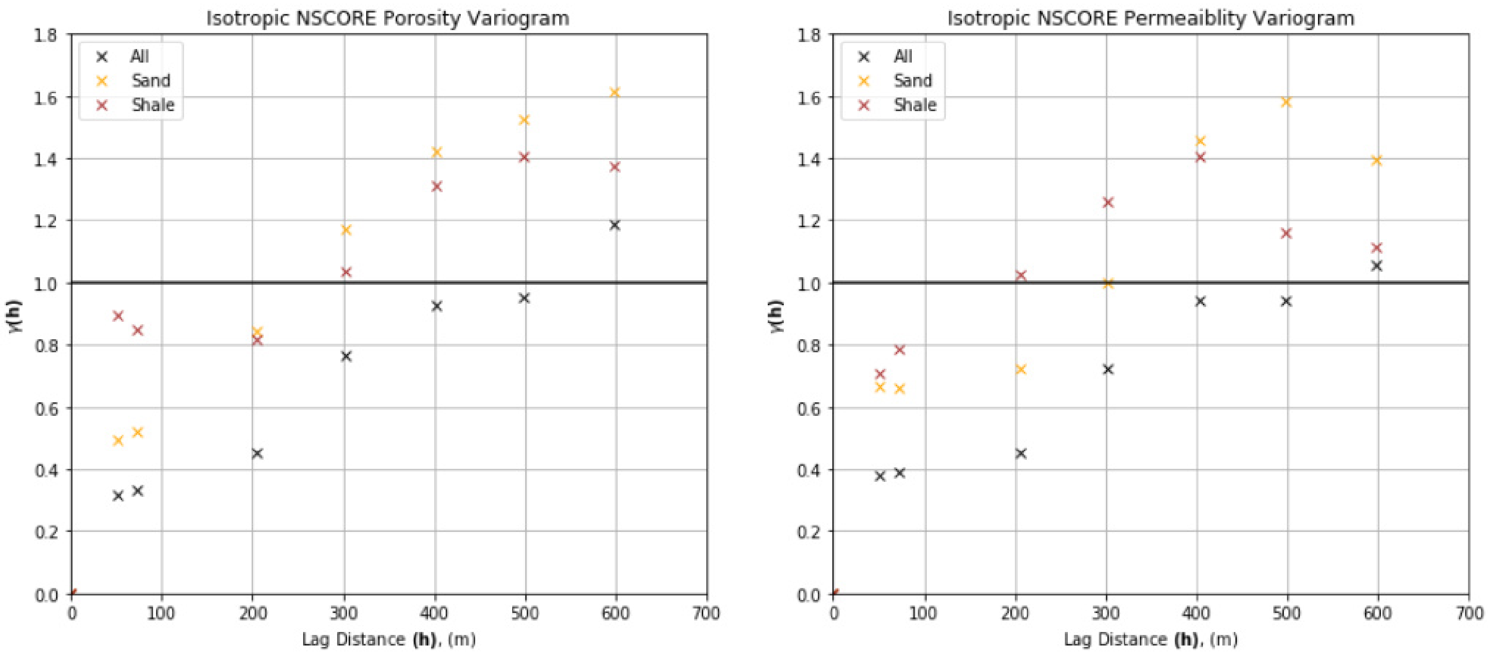
Lag Distance Lag Tolerance Number of Lags Band Width Azimuth Azimuth Tolerance Sandardize Sill

```
# Calculate Sample Data Isotropic Variograms
tmin = -9999.; tmax = 9999.;
lag_dist = 100.0; lag_tol = 100.0; nlag = 7; bandh = 9999.9; azi = 0; atol = 90.0; isill = 1

lag, por_sand_gamma, por_sand_npair = geostats.gamv(df_sand,"X","Y","NPor",tmin,tmax,lag_dist,lag_tol,nlag,azi,atol,bandh,isi
lag, por_shale_gamma, por_shale_npair = geostats.gamv(df_shale,"X","Y","NPor",tmin,tmax,lag_dist,lag_tol,nlag,azi,atol,bandh,
lag, por_gamma, por_npair = geostats.gamv(df,"X","Y","NPor",tmin,tmax,lag_dist,lag_tol,nlag,azi,atol,bandh,isill)

lag, perm_sand_gamma, perm_sand_npair = geostats.gamv(df_sand,"X","Y","NPerm",tmin,tmax,lag_dist,lag_tol,nlag,azi,atol,bandh,
lag, perm_shale_gamma, perm_shale_npair = geostats.gamv(df_shale,"X","Y","NPerm",tmin,tmax,lag_dist,lag_tol,nlag,azi,atol,ban
lag, perm_gamma, perm_npair = geostats.gamv(df,"X","Y","NPerm",tmin,tmax,lag_dist,lag_tol,nlag,azi,atol,bandh,isill)
```

Set the variogram calculation parameters.

Variogram Calculation in Python

Variogram Calculation Workflow in Python



Visualize the experimental variogram (the calculated variogram values at each lag). We say experimental to differentiate with the variogram model discussed in the next lecture.

Data Analytics and Geostatistics: Sparse Data

Lecture outline . . .

- Spatial Continuity
- Variogram Concepts
- Variogram Calculation