Data Analytics and Geostatistics: Spatial Estimation

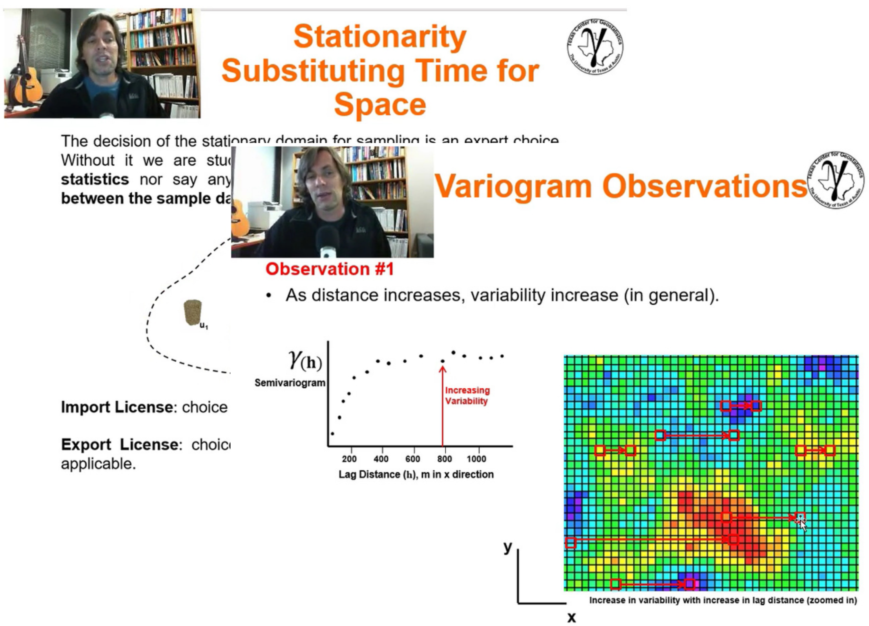
Lecture outline . . .

- **Variogram Interpretation**
- **Variogram Modeling**

Data Analytics and Geostatistics: Spatial Estimation

Other Resources:

- **Lectures recorded on YouTube.**



Stationarity, spatial calculation and modeling on YouTube.

- **Worked out examples on GitHub in Excel and Python**

Data Analytics and Geostatistics:
Spatial Estimation

Lecture outline . . .

- **Variogram Interpretation**

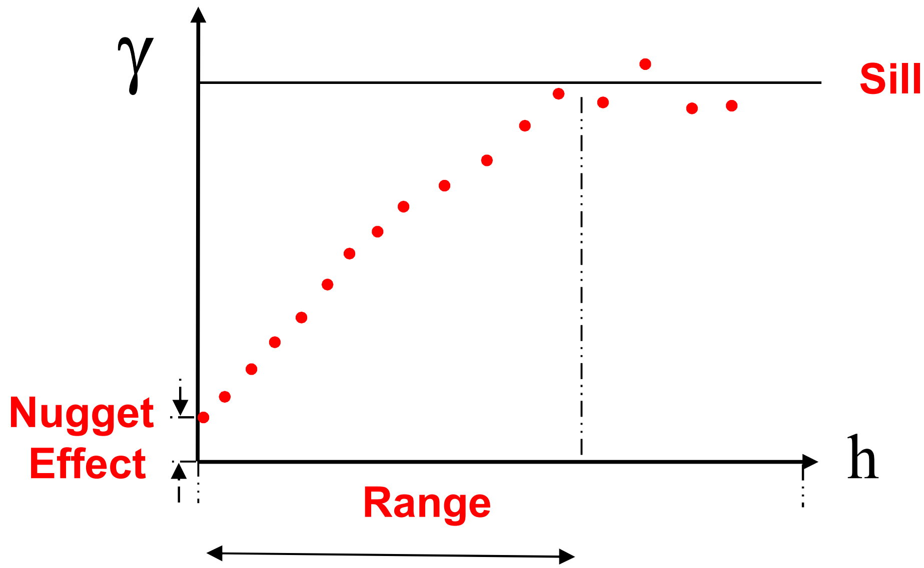
Variogram Terminology

- The ***sill*** is the variance of the data used for variogram calculation (1.0 if the data are normal scores)
- The ***range*** is the distance at which the variogram reaches the sill
- The ***nugget effect*** is the behavior at distances less than the smallest experimental lag:

geological microstructure + measurement error

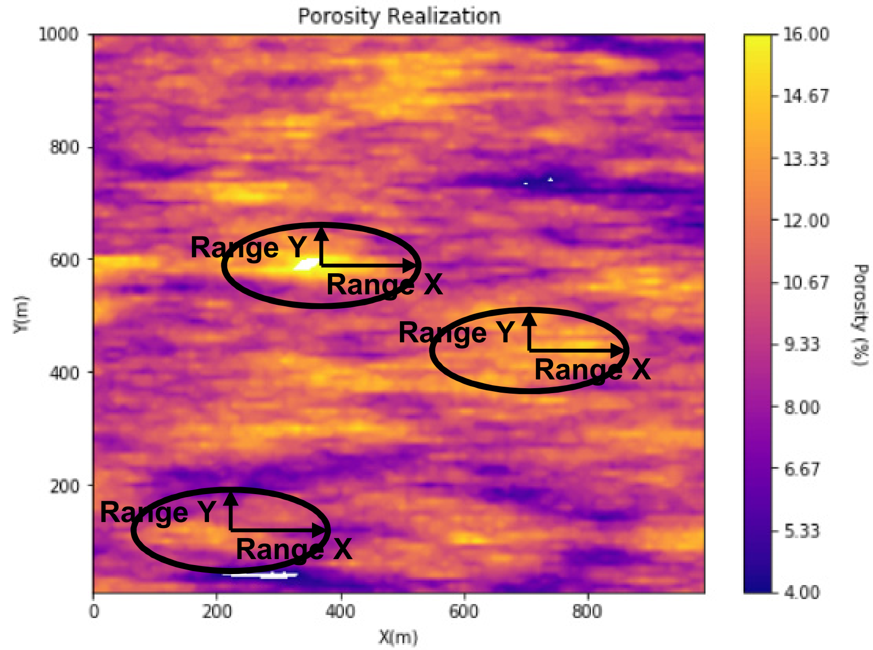
- Any error in the measurement value or the location assigned to the measurement translates to a higher nugget effect

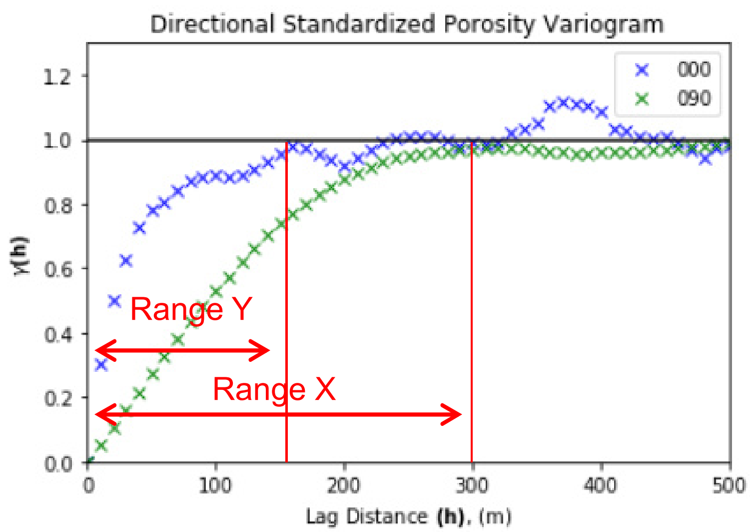
- Sparse data may also lead to a higher than expected nugget effect



Geometric Anisotropy

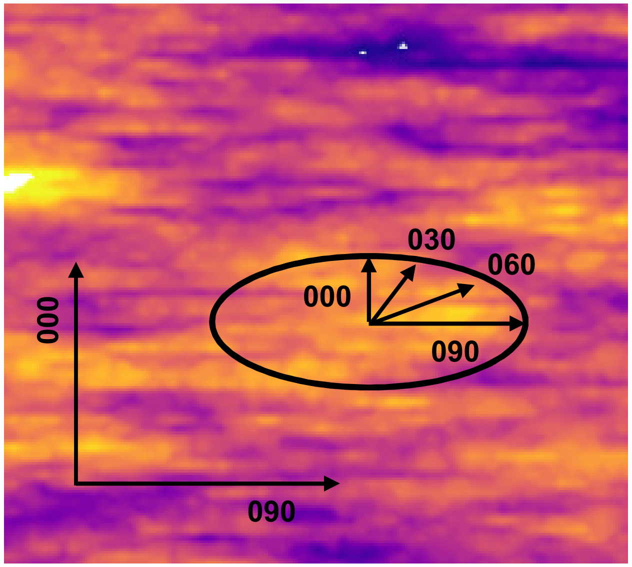
- The range of correlation / variogram range **depends on direction**
- Commonly, the vertical range of correlation is much less than horizontal range due to larger lateral distance during deposition.
- The ratio of the **horizontal: vertical range** is commonly known as the horizontal to vertical anisotropy ratio
- Geometric anisotropy is common in horizontal direction
- The ratio of horizontal **major direction: horizontal minor direction range** is commonly known as the horizontal major to minor anisotropy ratio

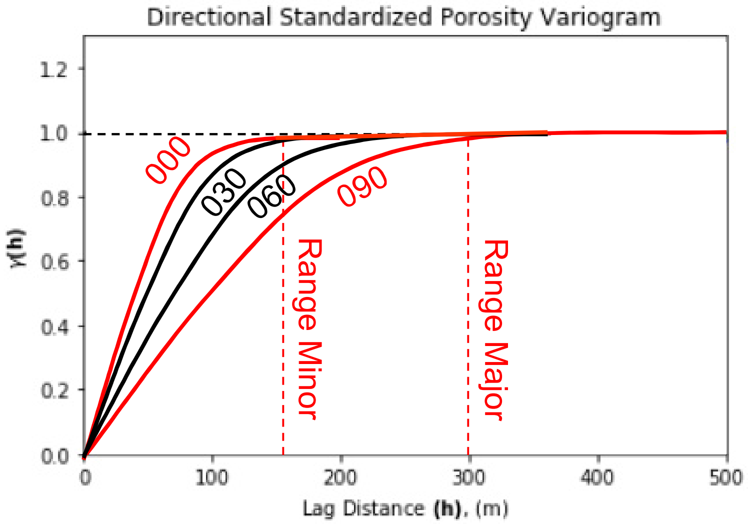




Geometric Anisotropy

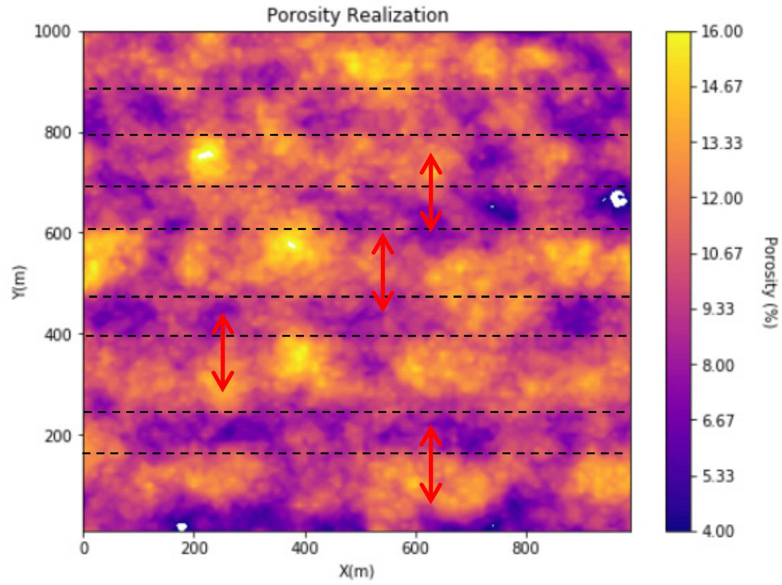
- We assume geometric anisotropy to model 2d and 3D variogram from experimental variograms calculated in primary directions.
- This model provides a valid interpolation of the variogram between the primary directions.
- When we start modeling this is critical to build nested variogram models with structures that:
    - Describe components of the variance
    - Act over all directions.

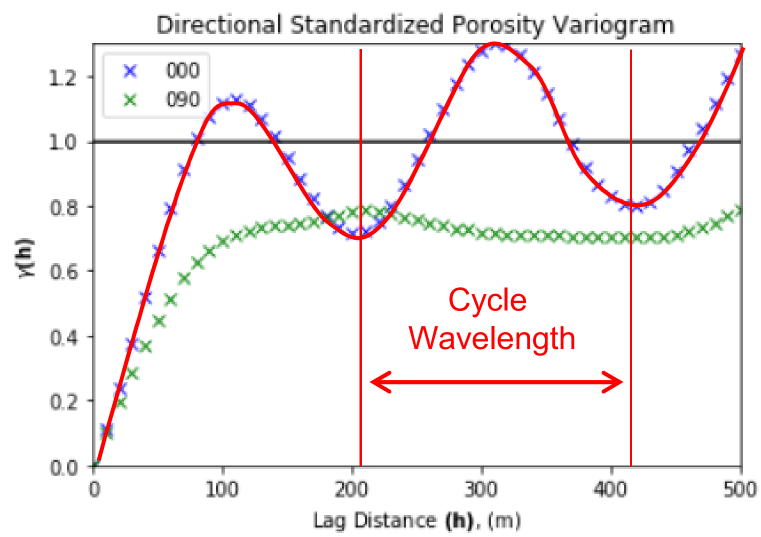




Cyclicity

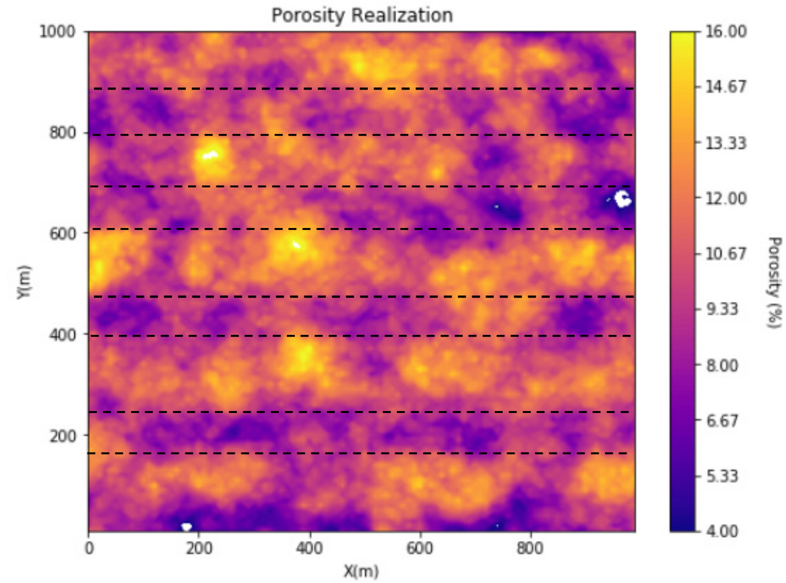
- Cyclicity may be linked to underlying geological periodicity, cycles in the deposition
- Sometimes noise in the experimental variogram due to too few data is mistaken as cyclicity
- The wavelength of the cycles in the experimental variogram is the wavelength of the spatial cycles.

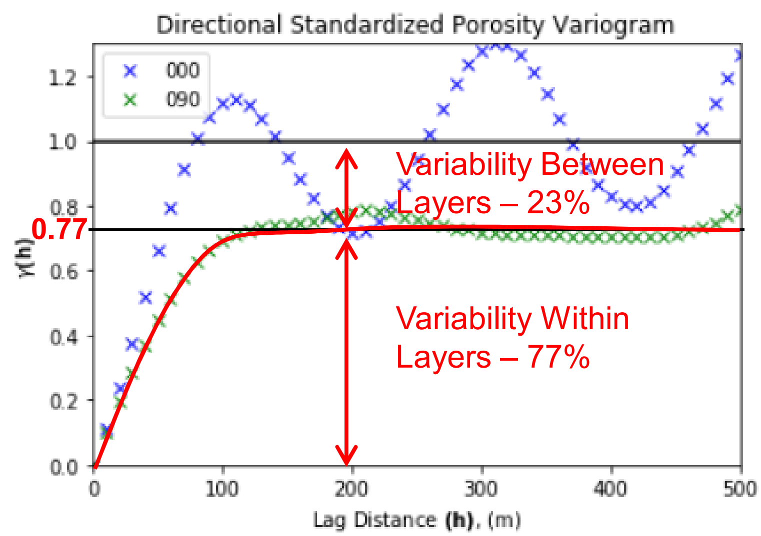




Zonal Anisotropy

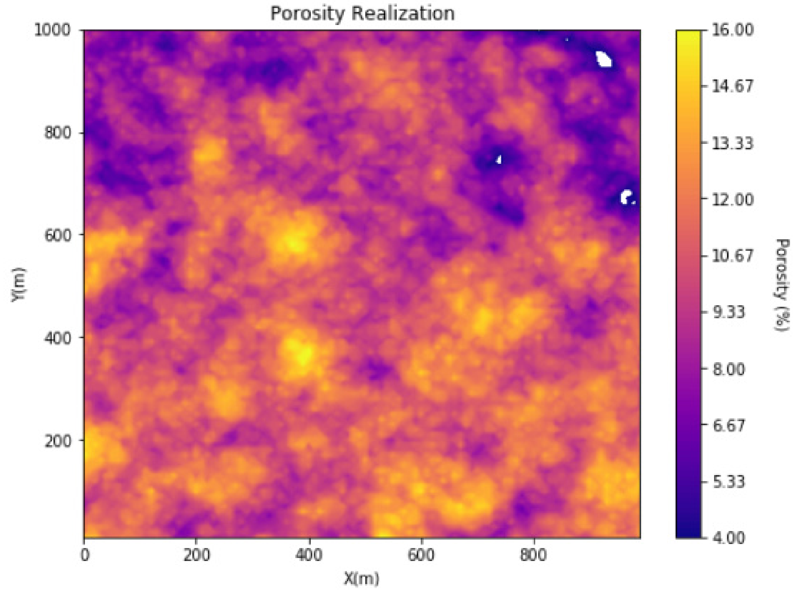
- When the experimental variogram does not reach the sill in a direction
- Often paired with cyclicity or trend in the other (orthogonal) direction.
- The variance at which the variogram levels off is called an apparent sill.

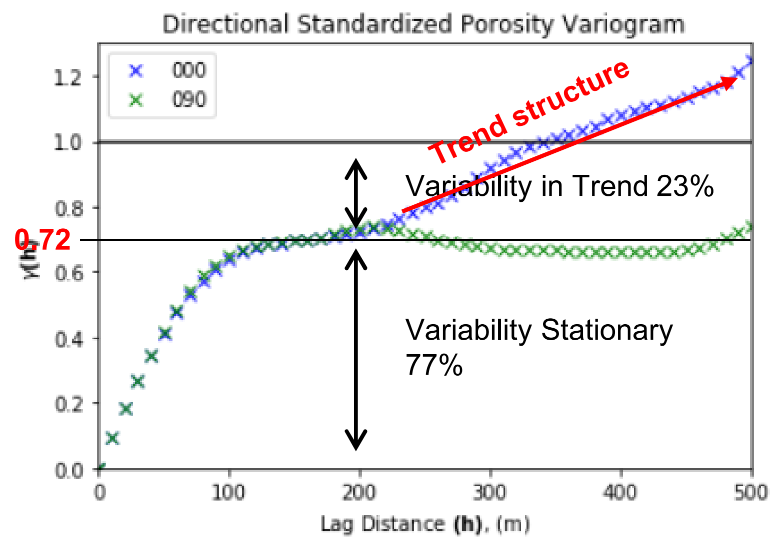




Trend

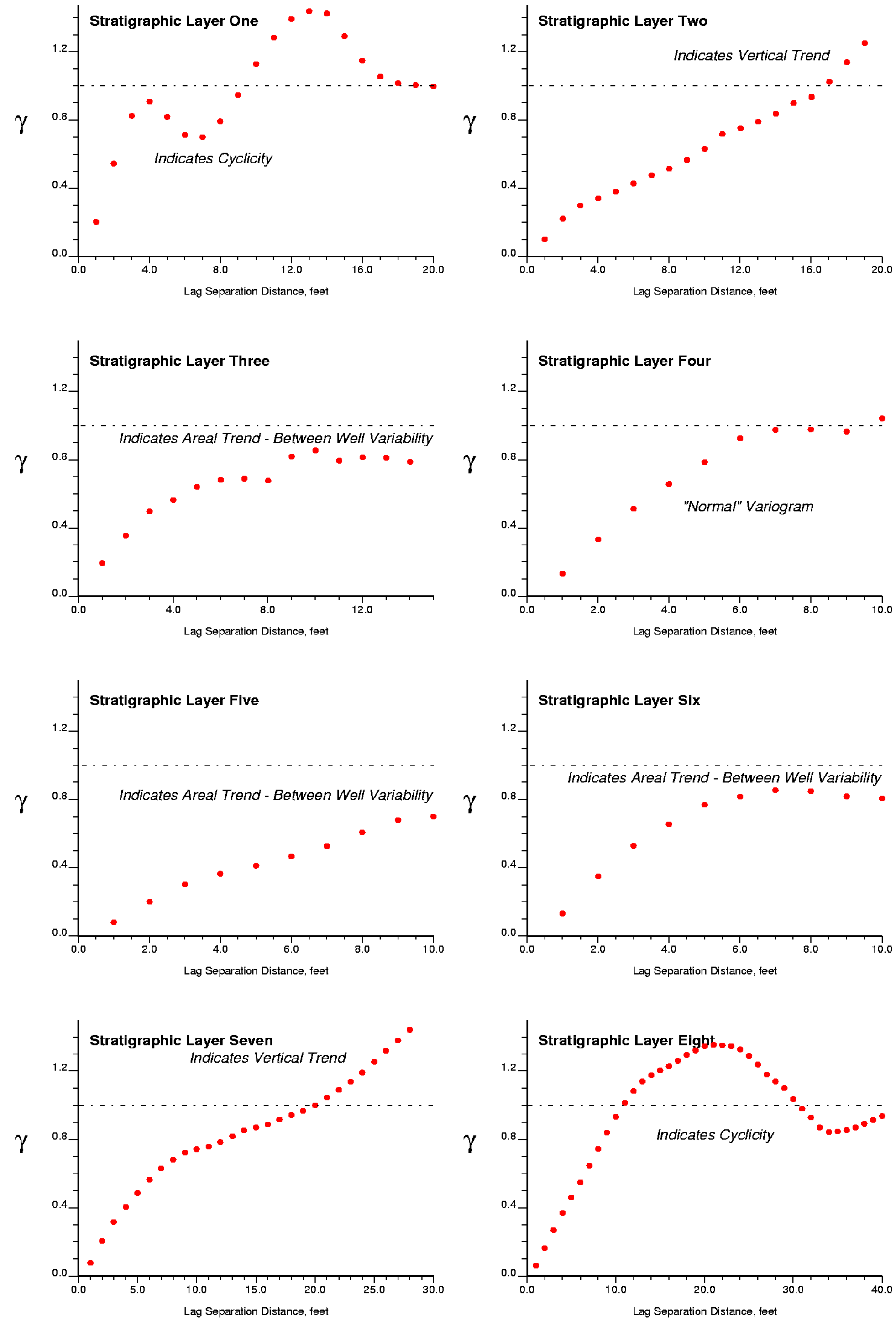
- Indicates a trend (fining upward, compacting with depth etc.)
- Could be interpreted as a fractal, fit a power law function
- May have to explicitly account for the trend in later simulation/modeling
  - Model, remove trend, work with the residual
  - If trend is removed the residual variogram will plateau at the sill





Some Experimental Variograms

- Superposition of the four interpretation principles
  - Trend
  - Cyclicity
  - Zonal anisotropy
  - Geometric anisotropy
- Very few variograms are ‘textbook variograms’, natural settings a complicated!



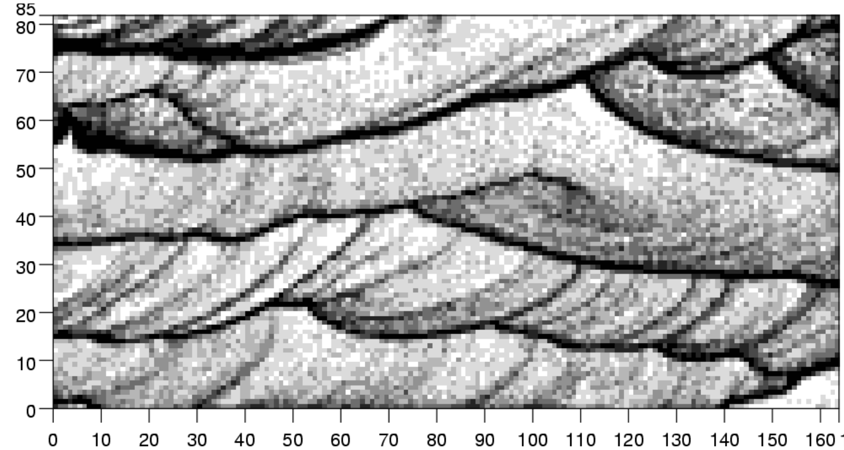
Variogram examples from Pyrcz and Deutsch, 2014.

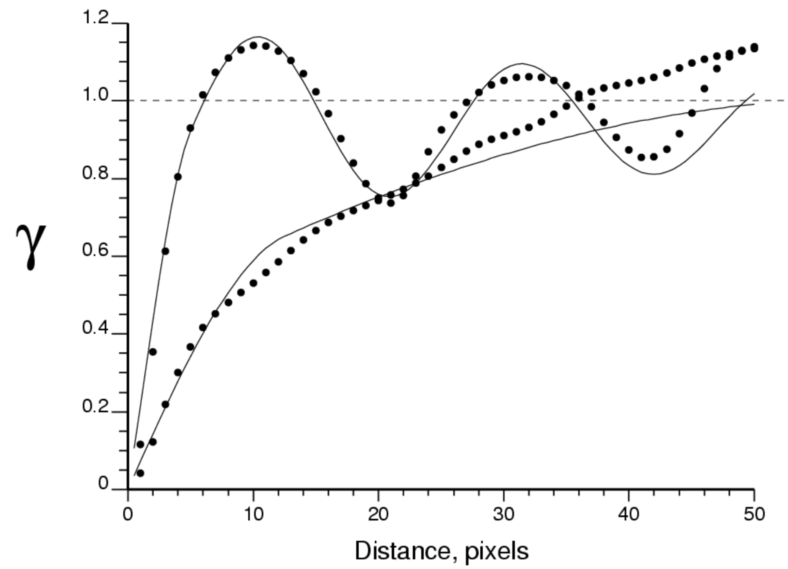
Variogram Interpretation

- Variogram interpretation consists of explaining the variability over different distance scales.

- The variogram is a chart of *variance versus distance* or *geological variability versus direction and Euclidean distance.*

- Establish the reasonableness of the variogram…





Slide from Clayton Deutsch's MinE 612 Class

Review of Variogram Interpretation

- Variogram is very important in geostatistical study; measure of geological distance
- Initial coordinate and data transformation may be required.
- Interpretation Principles:
  - Trend
  - Cyclicity
  - Geometric Anisotropy
  - Zonal Anisotropy
- Short scale structure is most important
  - nugget due to measurement error should not be modeled
  - size of geological modeling cells
- Vertical direction is typically well informed
  - can have artifacts due to spacing of core data
  - handle vertical trends and areal variations
- Horizontal direction is not well informed
  - take from analog field or outcrop
  - typical horizontal vertical anisotropy ratios

Slide from Clayton Deutsch’s MinE 612 Class

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

Complete

2. **Valid spatial model**
   – Fit with a couple different, nested (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

Variogram Modeling

Data Analytics and Geostatistics:
Spatial Estimation

Lecture outline . . .

- **Variogram Modeling**

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

Complete

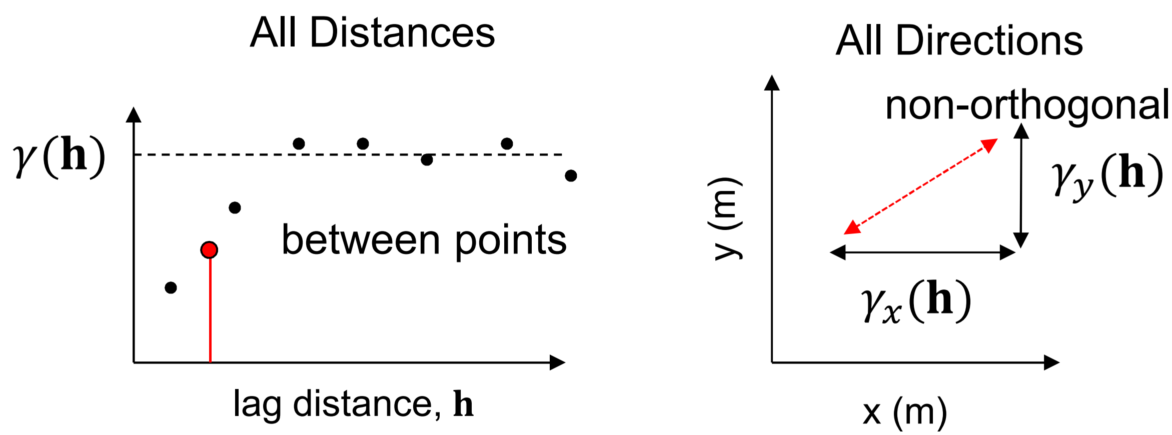
2. **Valid spatial model**
   – Fit with a couple different, nest (additive) spatial continuity models e.g. nugget, spherical, exponential and Gaussian

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

Variogram Modeling

Reasons for Variogram Modeling

1. Need to know the variogram for *all* possible **h** lags, distances and directions – not just the ones calculated

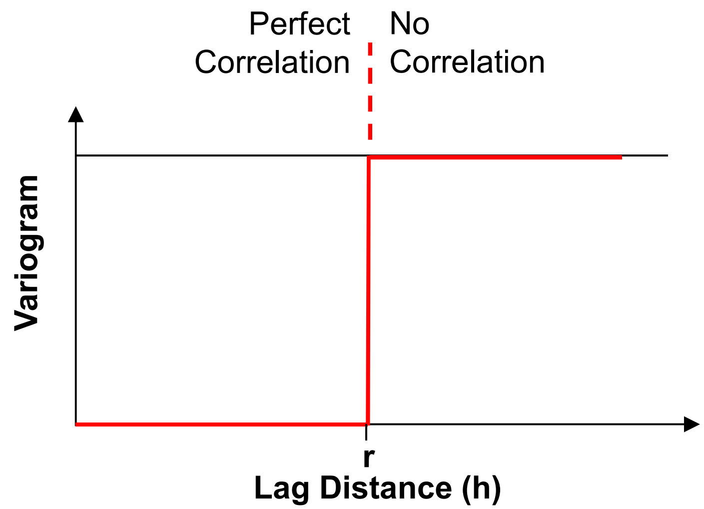


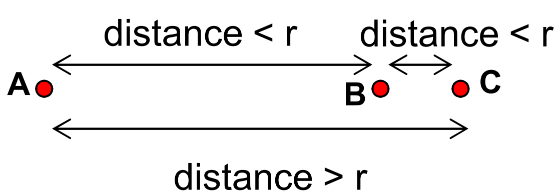
a specific lag offset may not exist in our data!

2. Incorporate additional geological knowledge (such as analogue information or information on directions of continuity …)
3. The variogram model must be **positive definite** (a legitimate measure of distance), that is, the variance of any linear combination must be positive

Reasons for Variogram Modeling

Extreme Example to Demonstrate the Need for Using Positive Definite Variogram





A and B and B and C are perfectly correlated, but A and C are not correlated!

-Spatial paradox!

Positive definite variogram models ensure for all possible spatial configurations there are no paradoxes.

Reasons for Variogram Modeling

- The variogram will be used in Kriging (next section)
- Kriging uses variogram (covariance) values in a linear system of equations
- Solves for estimate and estimation variance (uncertainty in the estimate)

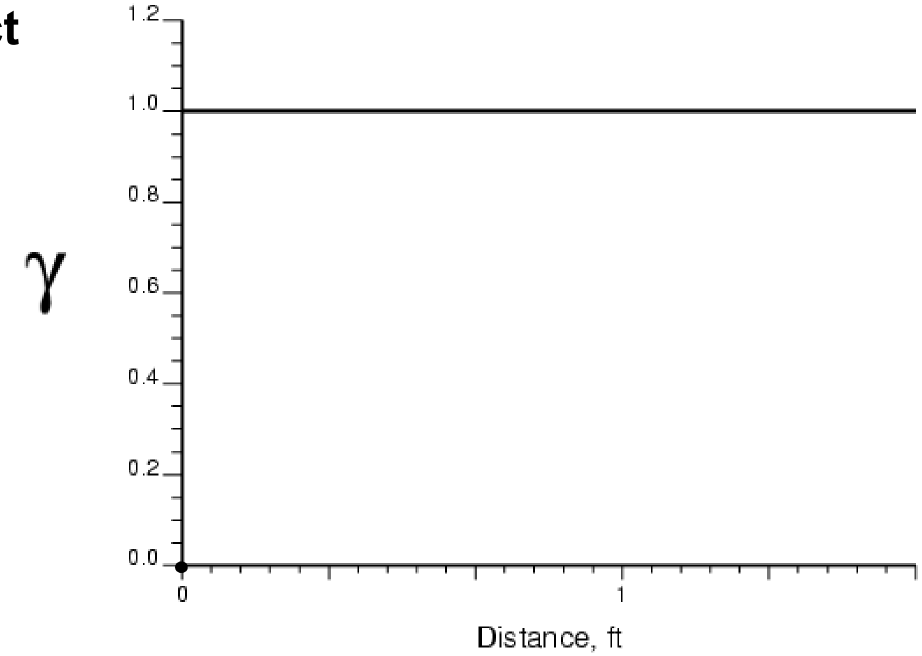
estimation variance = $\sigma_{x^*}^2 = \sigma_x^2 - \sum_{\alpha=1}^{n} \lambda_\alpha C_x(\mathbf{u}_\alpha - \mathbf{u}_0) \geq 0$

- Cannot just fit a smooth interpolation to the experimental variogram points.
  - It will result in unrealistic values for estimation variance

Slide modified from Clayton Deutsch's MinE 612 Class

Common Variogram Models

Nugget Effect

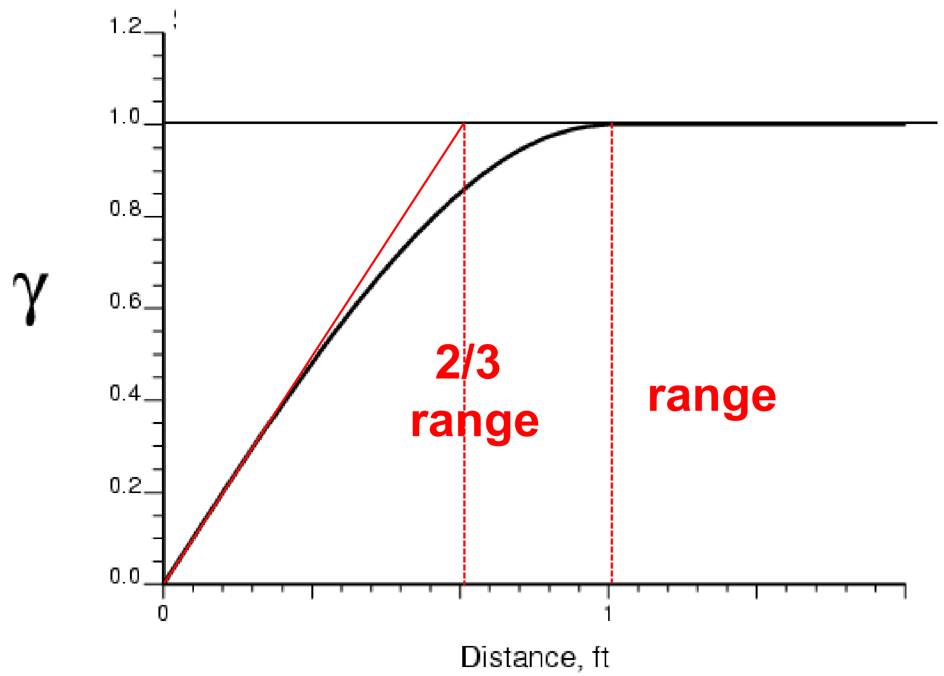


$$\gamma(\mathbf{h}) = C_1 \cdot Nugget = \begin{cases} 0 & , if\ h = 0 \\ C_1 & , if\ h > 0 \end{cases}$$

- No spatial correlation
- Should be a small component of the overall variance

Common Variogram Models

Spherical



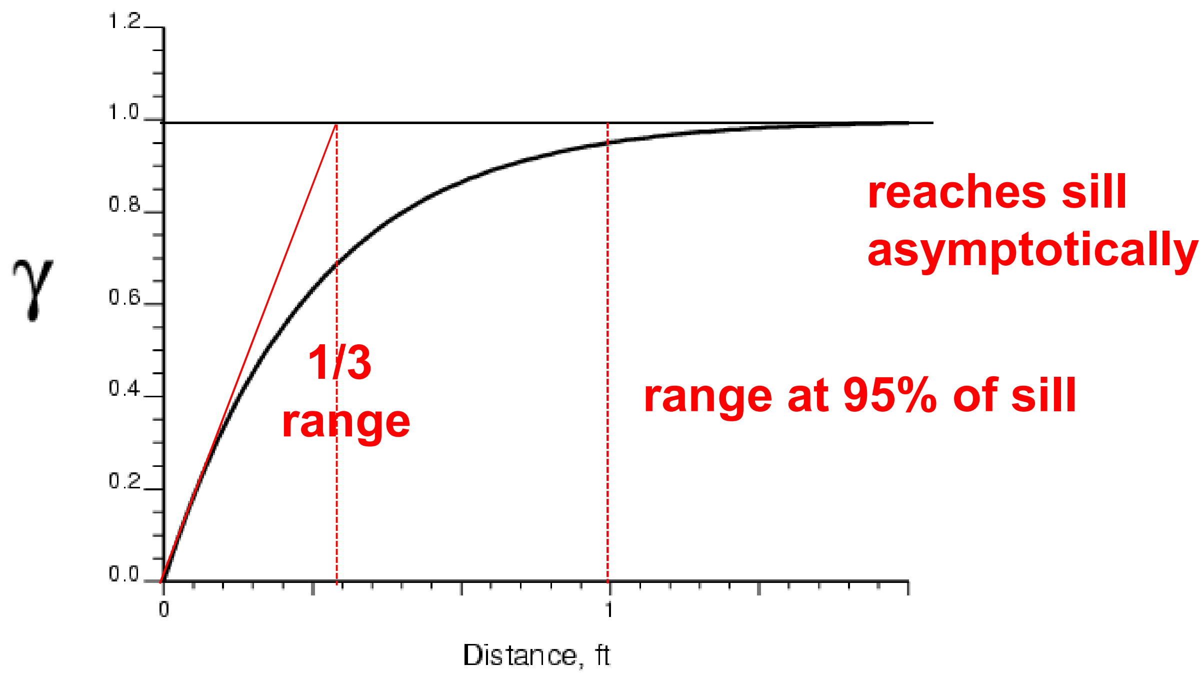
Commonly encountered variogram shape

$$\gamma(\mathbf{h}) = C_1 \cdot Sph\left(\frac{\mathbf{h}}{a}\right) = \begin{cases} C_1 \cdot \left[1.5\left(\frac{\mathbf{h}}{a}\right) - 0.5\left(\frac{\mathbf{h}}{a}\right)^3\right] & , if\ h < a \\ C_1 & , if\ h \geq a \end{cases},$$

where a is the range.

Common Variogram Models

Exponential



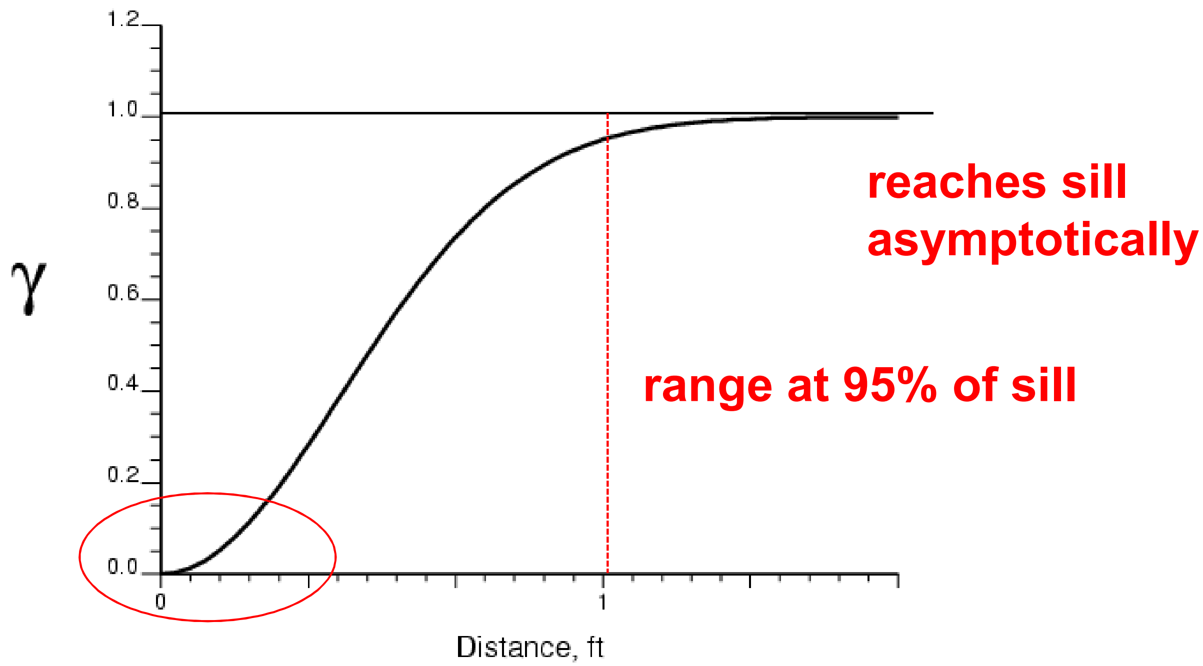
Similar to spherical but rises more steeply and reaches the sill asymptotically

$$\gamma(\mathbf{h}) = C_1 \cdot Exp\left(\frac{\mathbf{h}}{a}\right) = C_1 \cdot \left[1.0 - exp\left(-\frac{\mathbf{h}}{a}\right)\right]$$

where a is the range.

Common Variogram Models

Gaussian



Implies short scale continuity; parabolic behavior at the origin, instead of linear

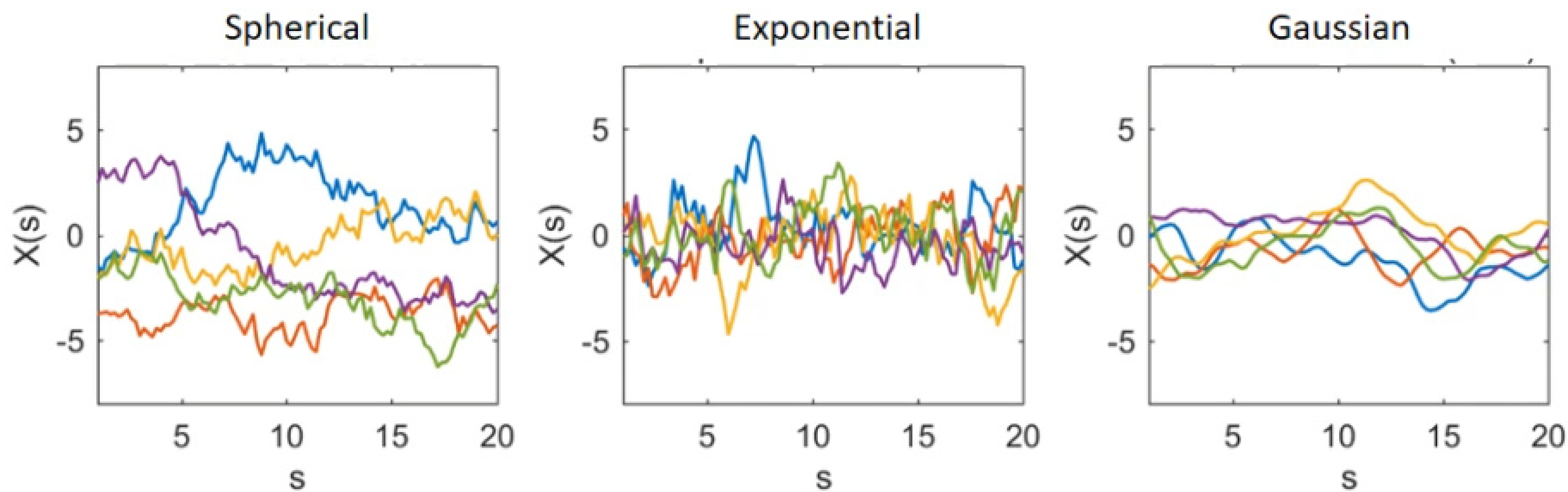
$$\gamma(\mathbf{h}) = C_1 \cdot Gaus\left(\frac{\mathbf{h}}{a}\right) = C_1 \cdot \left[1.0 - \exp\left(-3\frac{\mathbf{h}^2}{a^2}\right)\right]$$

where a is the range.

Common Variogram Models

Behavior over short distances

Linear —— **Exponential** —— **Gaussian**



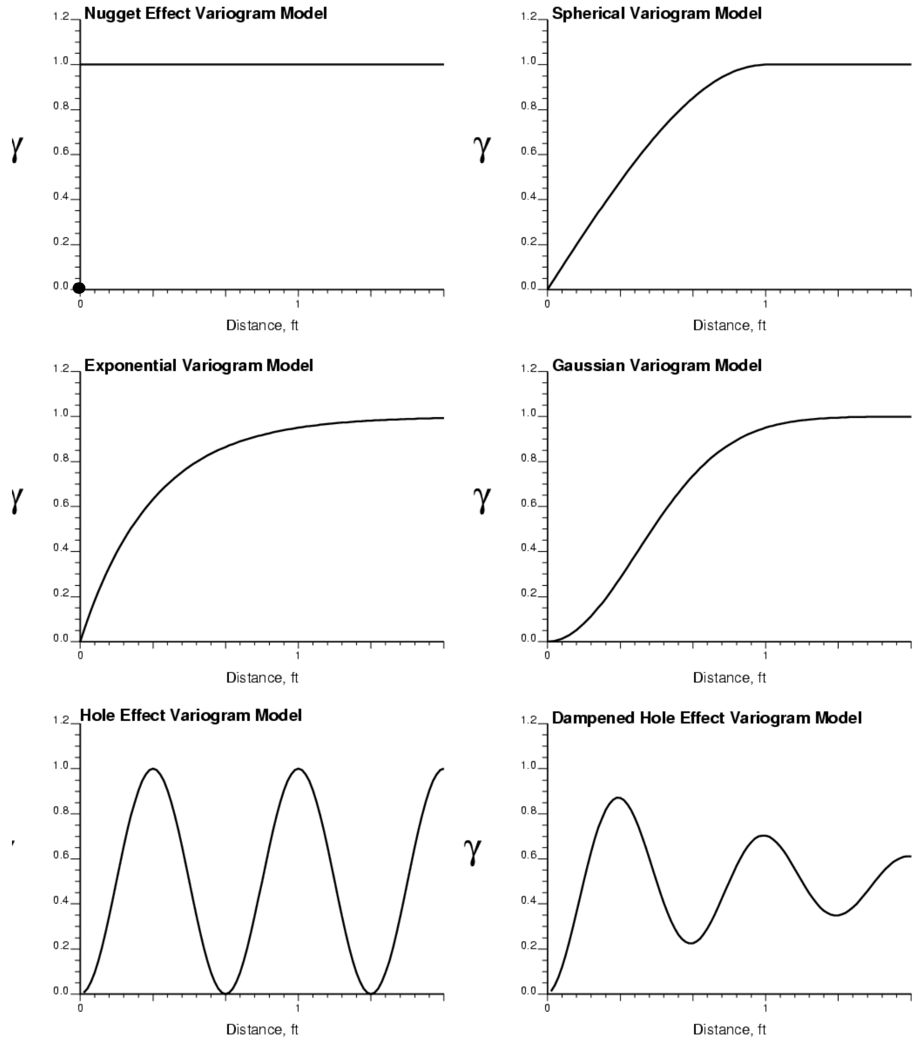
Slide from Prof. Jef Caers (https://www.youtube.com/watch?v=OxjrnVHO64c)

Common Variogram Models

No spatial correlation

Should be a small component of the overall variance

Commonly encountered variogram shape



Similar to spherical but rises more steeply and reaches the sill asymptotically

Implies short scale continuity; parabolic behavior at the origin, instead of linear

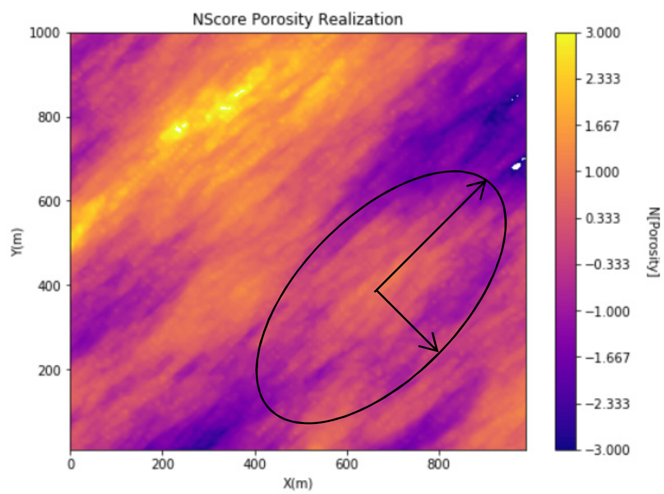
For periodic variables

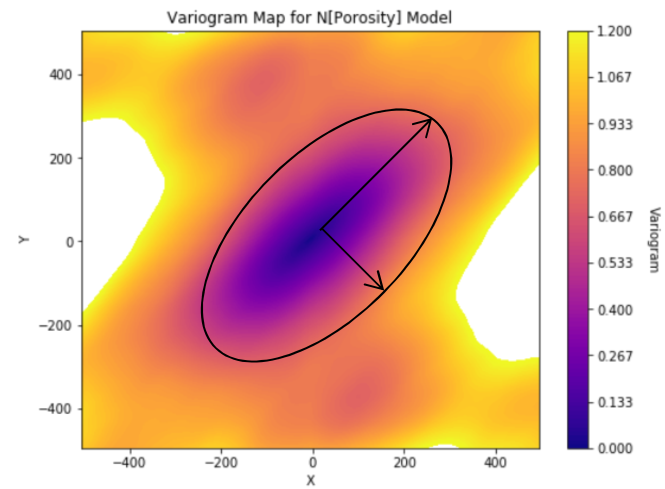
For periodic variables, when the period is not regular

Variogram examples from Pyrcz and Deutsch, 2014.

2D Variogram Models

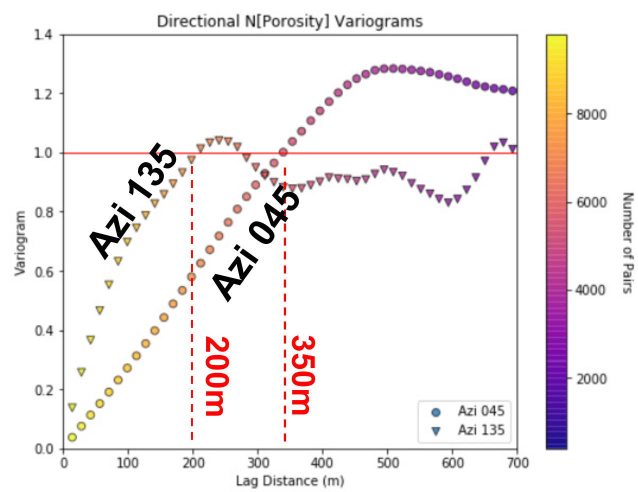
Calculate the variogram for all possible distances and directions (variogram map).



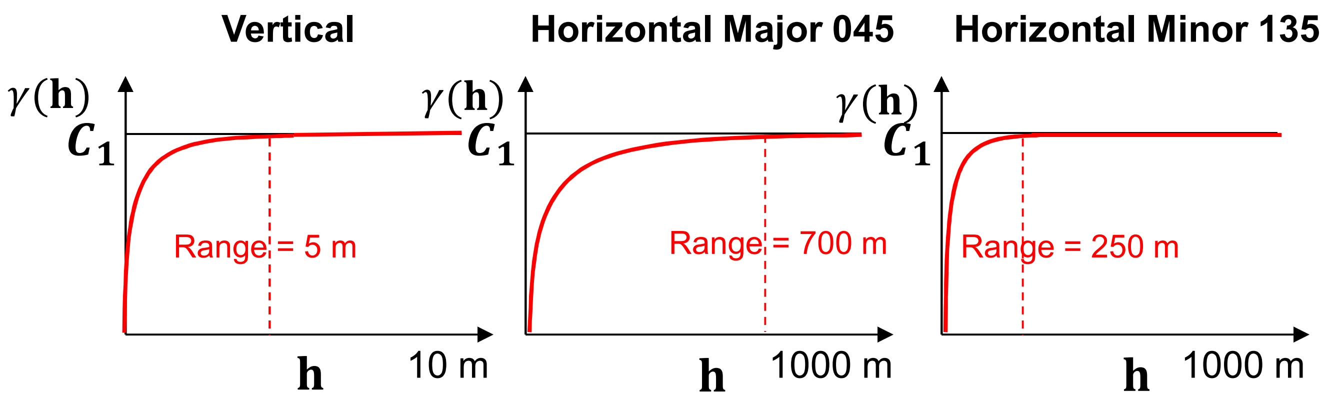


There is an ellipsoidal variation in continuity (geometric anisotropy):

- Parameters for a 2D variogram model:
  - direction, major and minor range, type of variogram



Directional Variograms



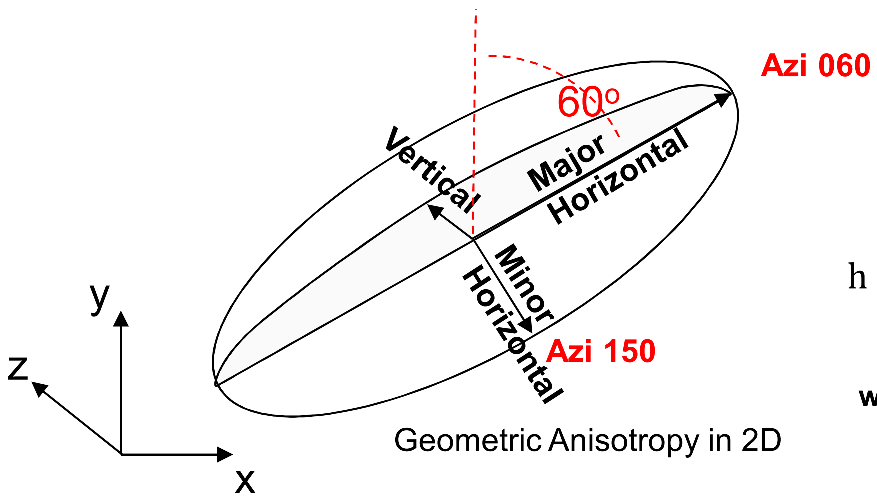
Variable range in each principal direction of a single variogram structure.

For each structure you determine the

- Contribution
- Orientation
- Type (nugget, spherical etc.)
- Range in the primary directions (geometric anisotropy).

2D / 3D Variogram Models

The variation of range along different directions is modeled using an ellipse in 2D and an ellipsoid in 3D



Geometric Anisotropy in 2D

Geometric Distance

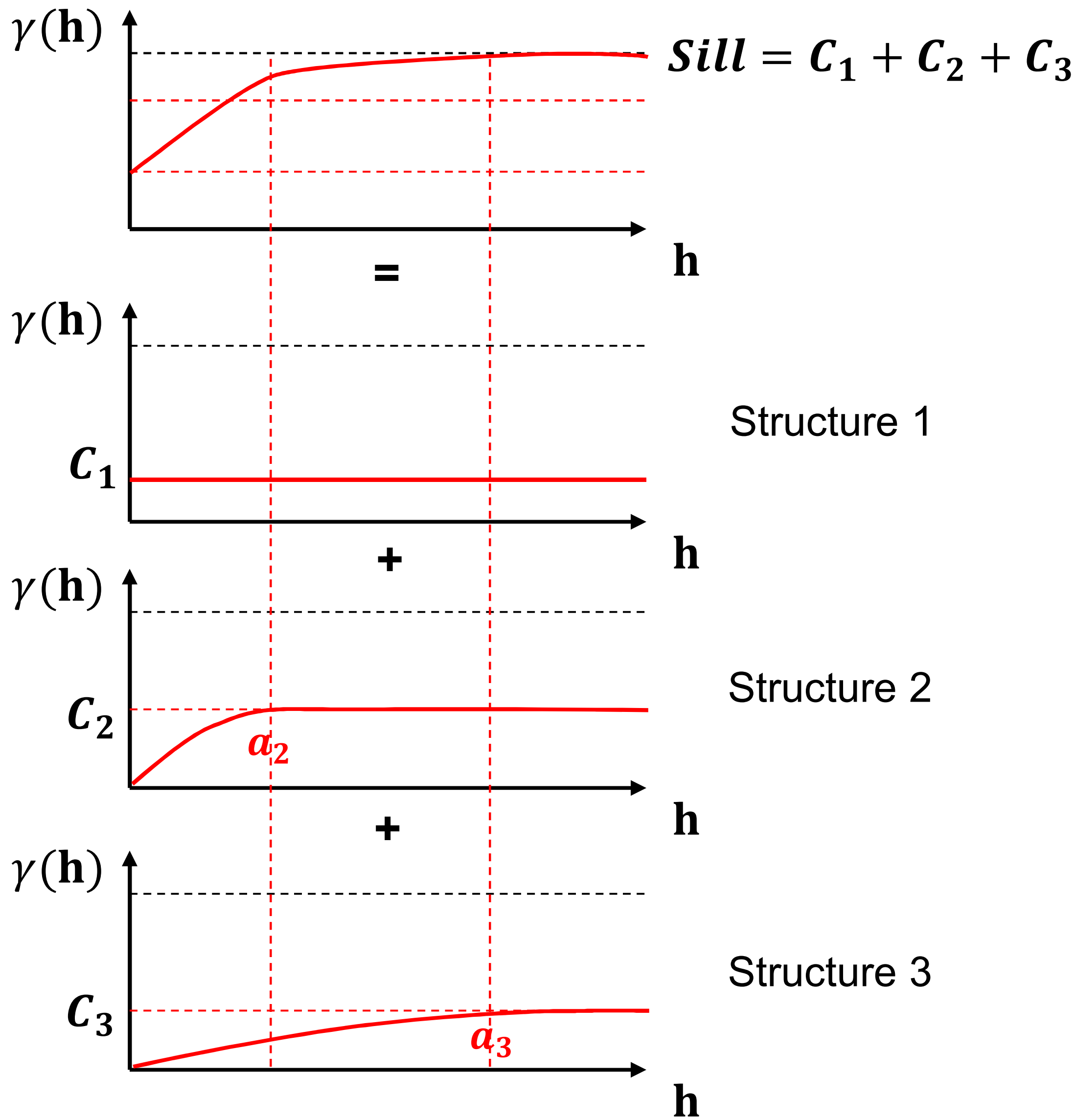
$$h = \sqrt{\left(\frac{h_{maj}}{a_{maj}}\right)^2 + \left(\frac{h_{min}}{a_{min}}\right)^2 + \left(\frac{h_{vert}}{a_{vert}}\right)^2}$$

where a_{maj}, a_{min} and a_{vert} are range parameters

- There is an ellipsoidal variation in continuity (geometric anisotropy):
  - Parameters for a 2D variogram model:
    - direction, major, minor, contribution and type of variogram
  - Parameters for a 3D variogram model:
    - direction, dip, major, minor and vertical range, contribution and type of variogram

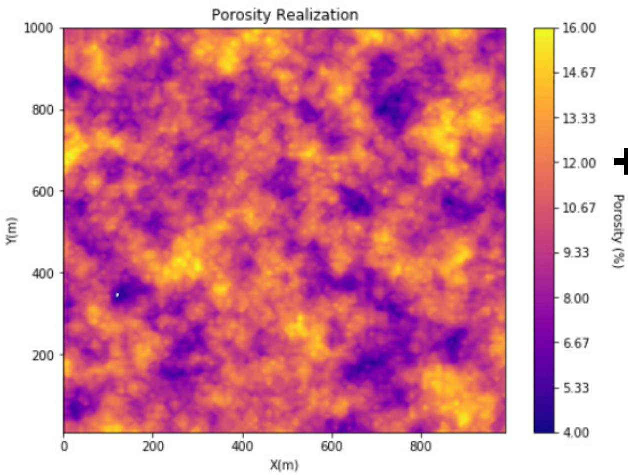
Nested Structures

- The addition of positive definite variogram structures is positive definite.
- Each structure covers a proportion of the sill.
- For each structure we can change the:
  - orientation
  - range in major and minor
- We are spatially explaining parts of the variance!



Nested Spatial Frequencies

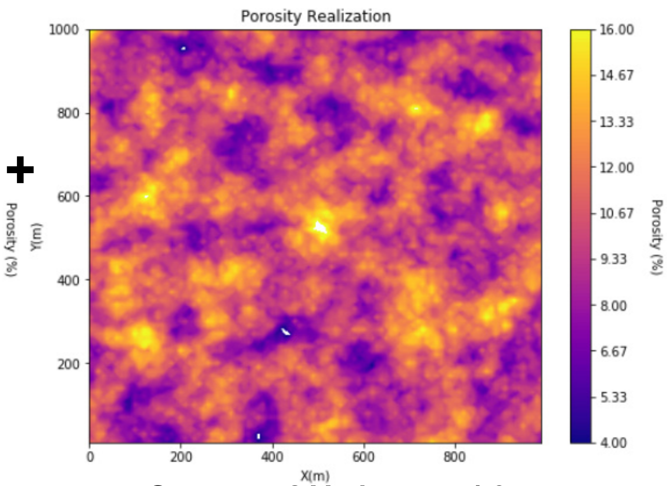
Isotropic 100 m Range, Sill = 4



Structure 1 Variance = 4.0

+

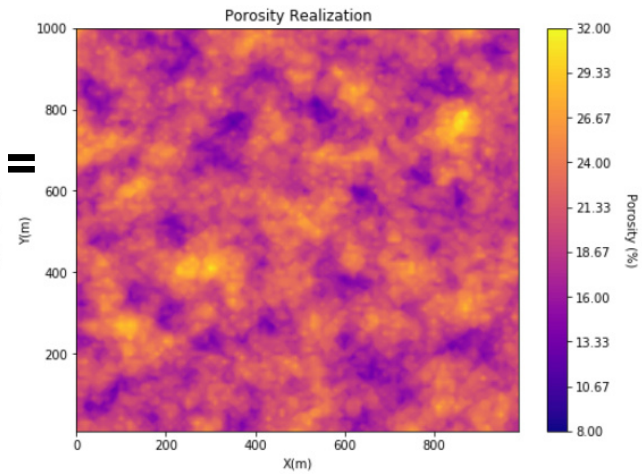
Isotropic 100 m Range, Sill = 4



Structure 2 Variance = 4.0

=

Isotropic 100m Range, Sill = 8



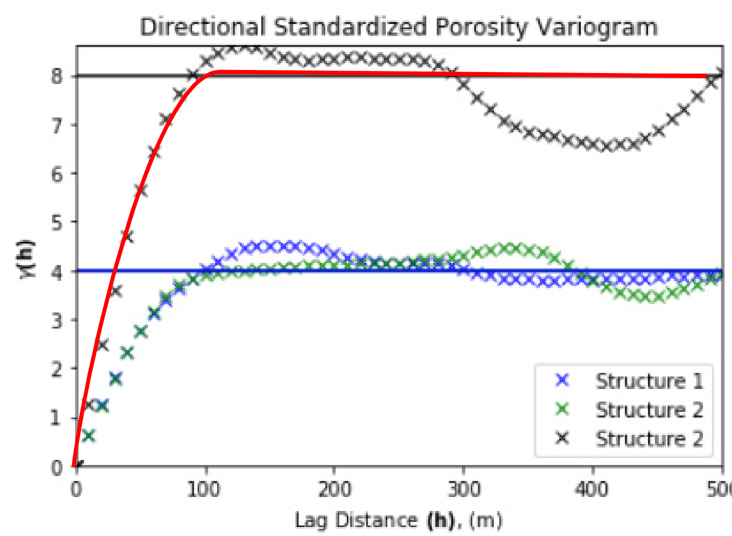
Structure 1+2 Variance = 8.0

Superposition of the same structure only changes the sill.

- The range stays the same:

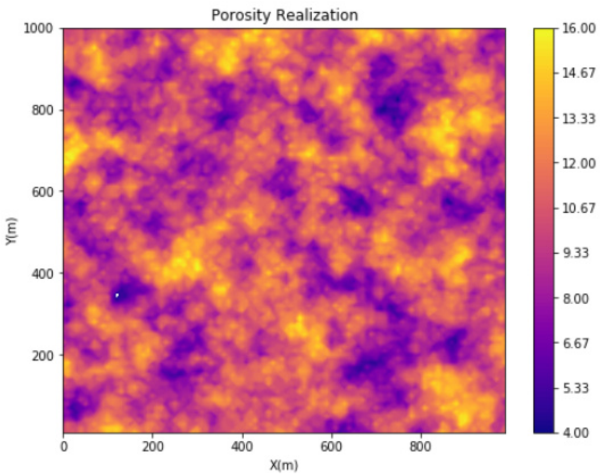
$$(C_0+C_1)\,\gamma_0(h) = C_1\gamma_0(h) + C_2\gamma_0(h)$$

- If you have a single structure in a direction just use the same range for all contributions.



Nested Spatial Frequencies

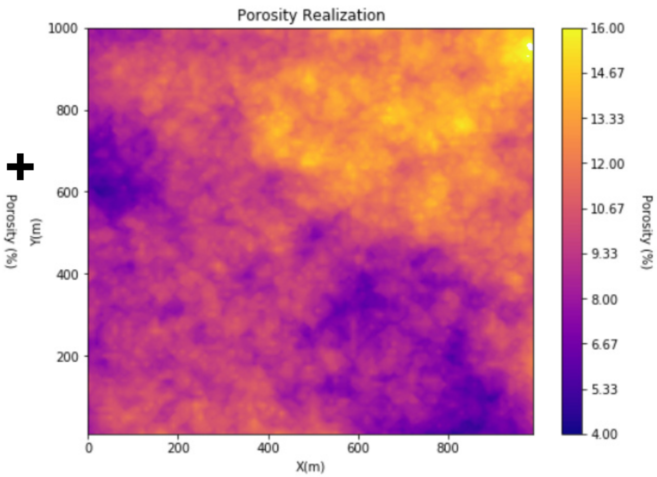
Isotropic 100 m Range, Sill = 4



Structure 1 Variance = 4.0

+

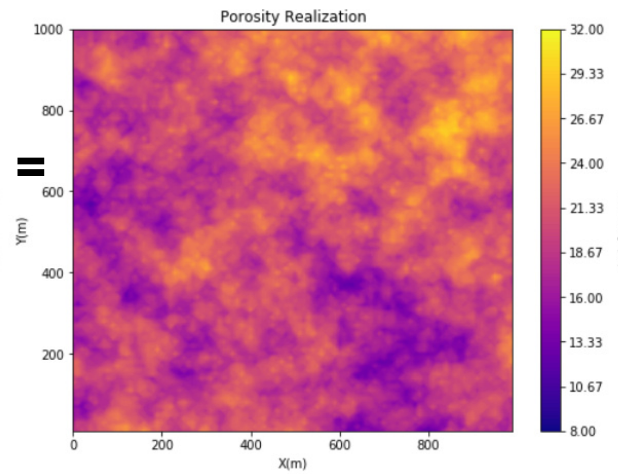
Isotropic 400 m Range, Sill = 4



Structure 2 Variance = 4.0

=

Isotropic Nested 400m Range, Sill = 8



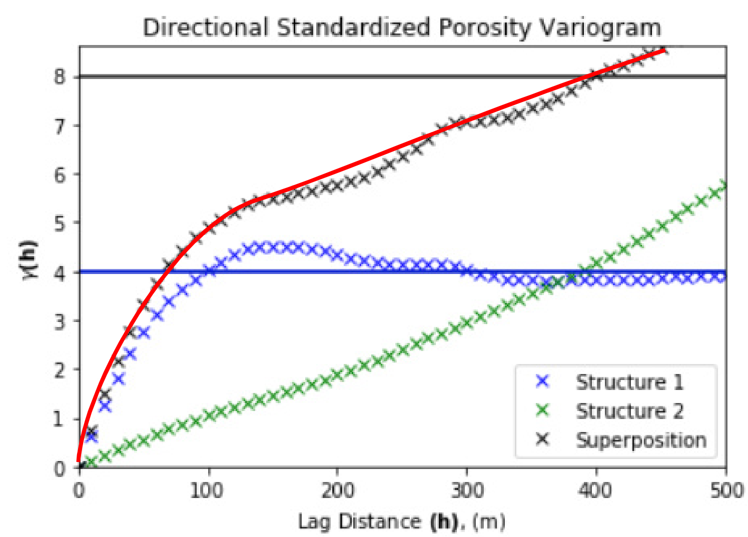
Structure 1+2 Variance = 8.0

Superposition results if different ranges are applied.

- New nested model:

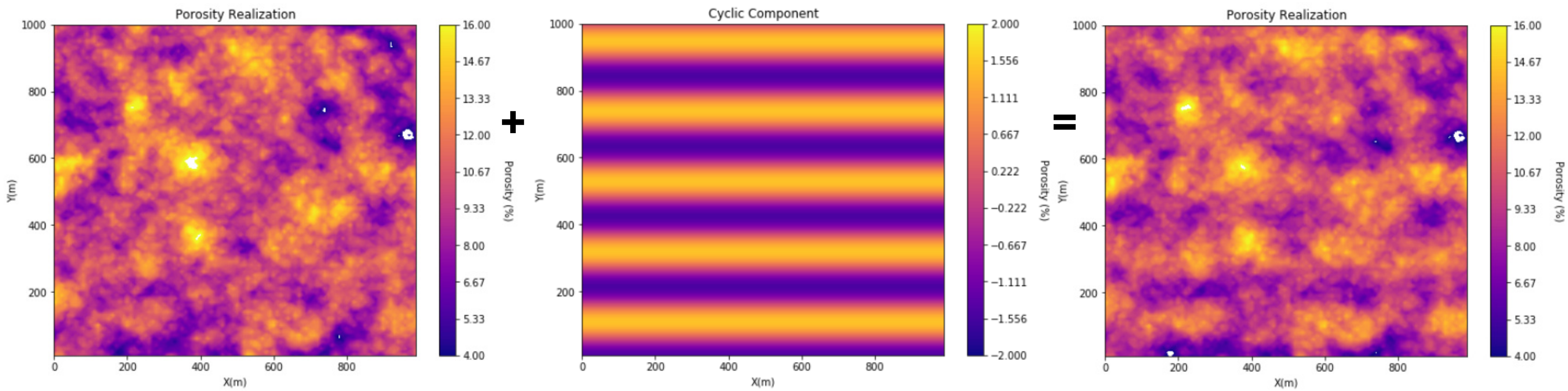
$$\gamma_{tot}(h) = C_1\gamma_1(h) + C_2\gamma_2(h)$$

- Forms an inflection point due to the combination of dissimilar ranges.



Nested Spatial Frequencies

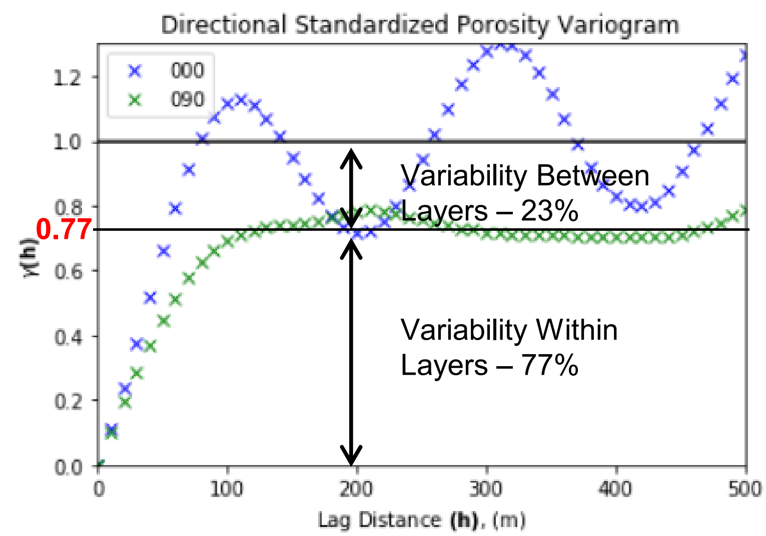
Isotropic 100 m Range | Cyclic Trend in Y 200 m Wavelength | Isotropic 100m Range + Cyclic 200m



77% Variance | **23% Variance** | **100% Variance**

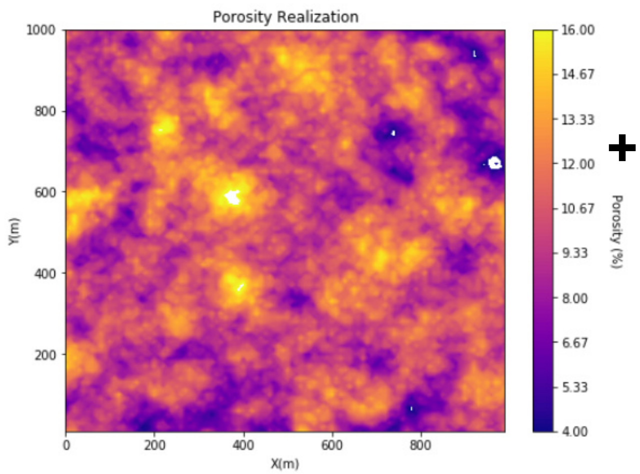
Variance within each spatial component is the contribution of each variogram structure.

- Superposition of multiple spatial structures each describing a proportion of the total variability



Nested Spatial Frequencies

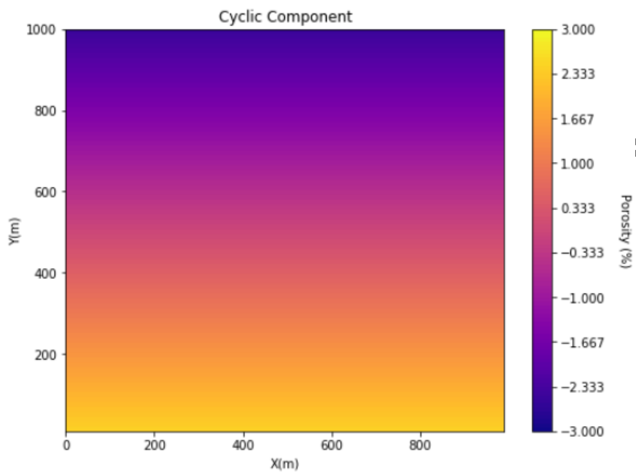
Isotropic 100 m Range



72% Variance

\+

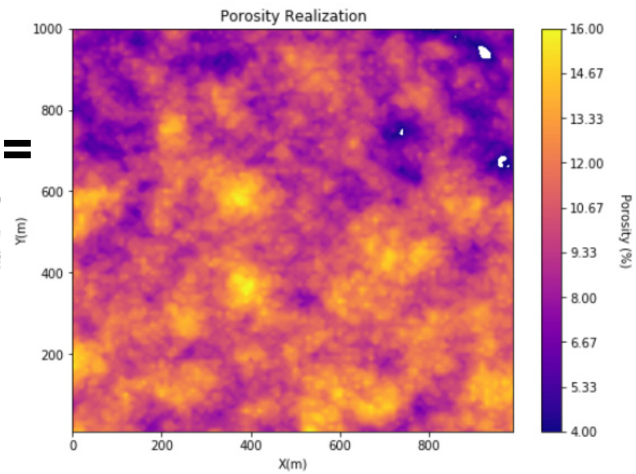
Trend in Y



28% Variance

=

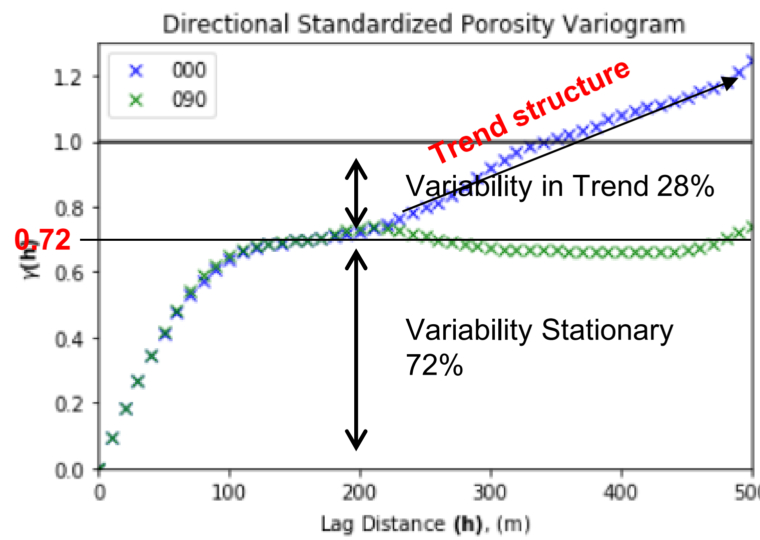
Isotropic 100m Range + Trend in Y



100% Variance

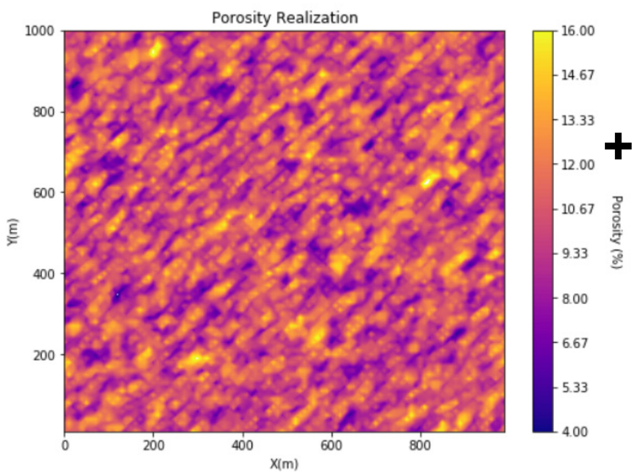
Variance within each spatial component is the contribution of each variogram structure.

- This illustrates the partitioning of spatial variance between trend and stationary, stochastic residual.



Nested Spatial Frequencies

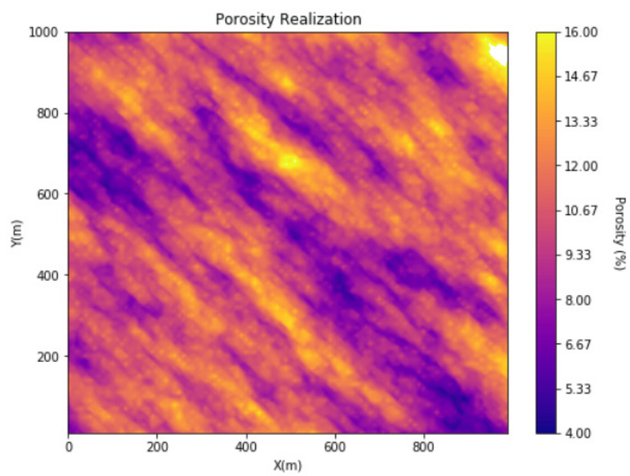
Anisotropic 045 50/20 m Range



50% Variance

+

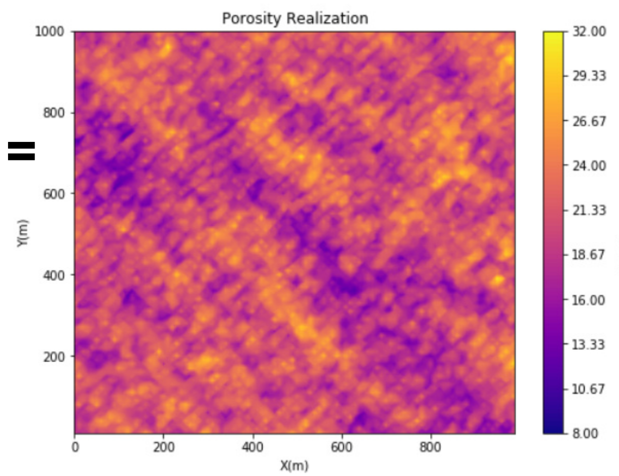
Anisotropic 135 300/100 m Range



50% Variance

=

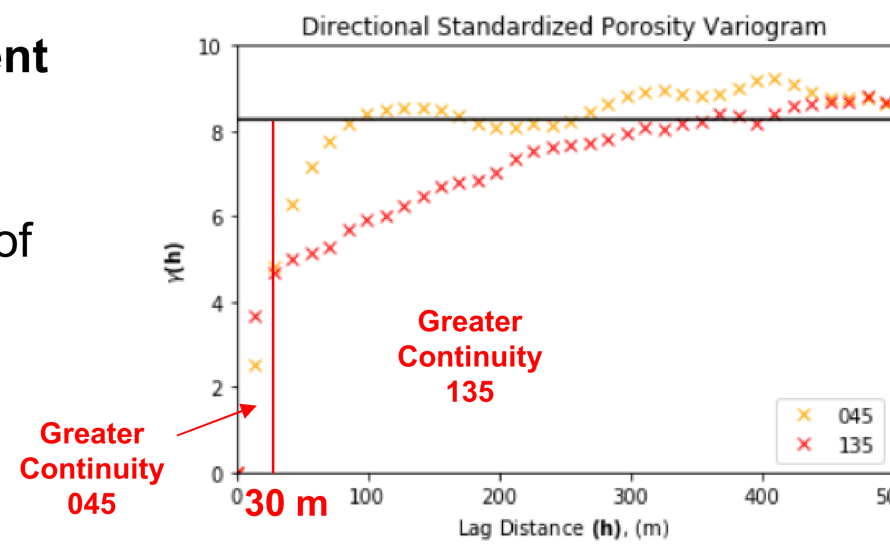
Multiscale, Orientation Variant



100% Variance

Variance within each spatial component is the contribution of each variogram structure.

- In this example the primary direction of continuity shifts over distance!
- < 30 m 045 has greater continuity
- > 30 m 135 has greater continuity



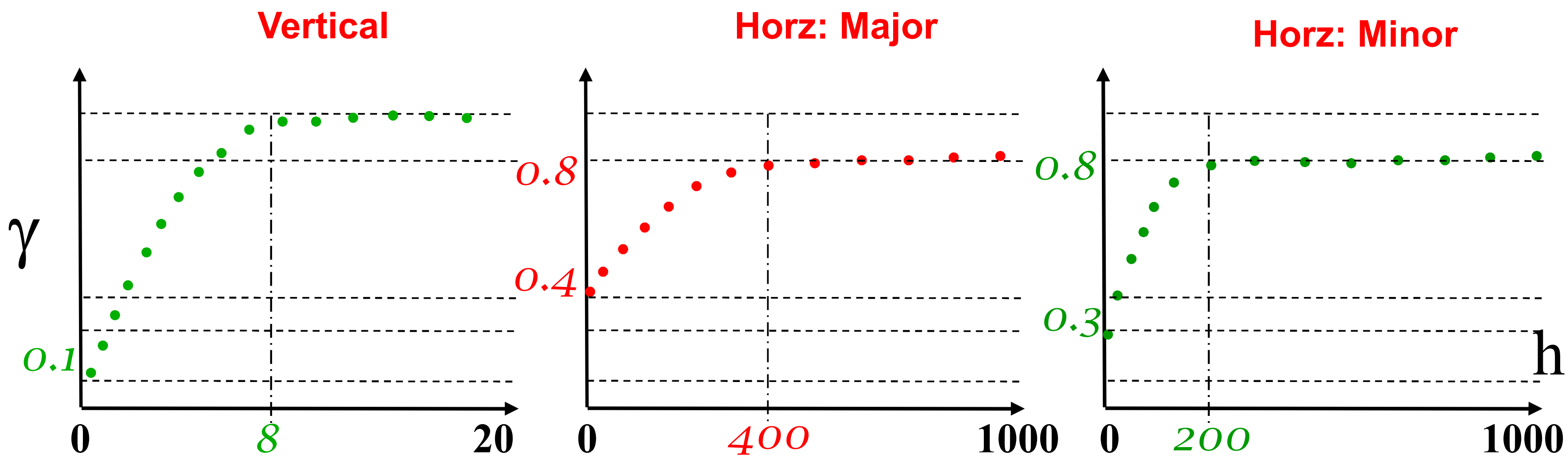
Variogram Modeling

- The following procedure is used to ensure a legitimate model:
  - Pick a single (lowest) isotropic nugget effect
  - Choose the same number of variogram structures for all directions based on most complex direction
  - Ensure that the same contribution parameter is used for all variogram structures in all directions
  - Allow a **different range parameter** in each direction
  - Model a zonal anisotropy by setting a very large range parameter in one or more of the principal directions

- The responsibility is yours, but most software helps a little
  - force same structure and contributions in all directions, let you modify the range and observe the result in the 3 primary directions.

Slide modified from Clayton Deutsch's MinE 612 Class

A Complicated Case



same ranges, behaves as one structure

nugget is isotropic over all distances

no spatial correlation in those directions over those structures

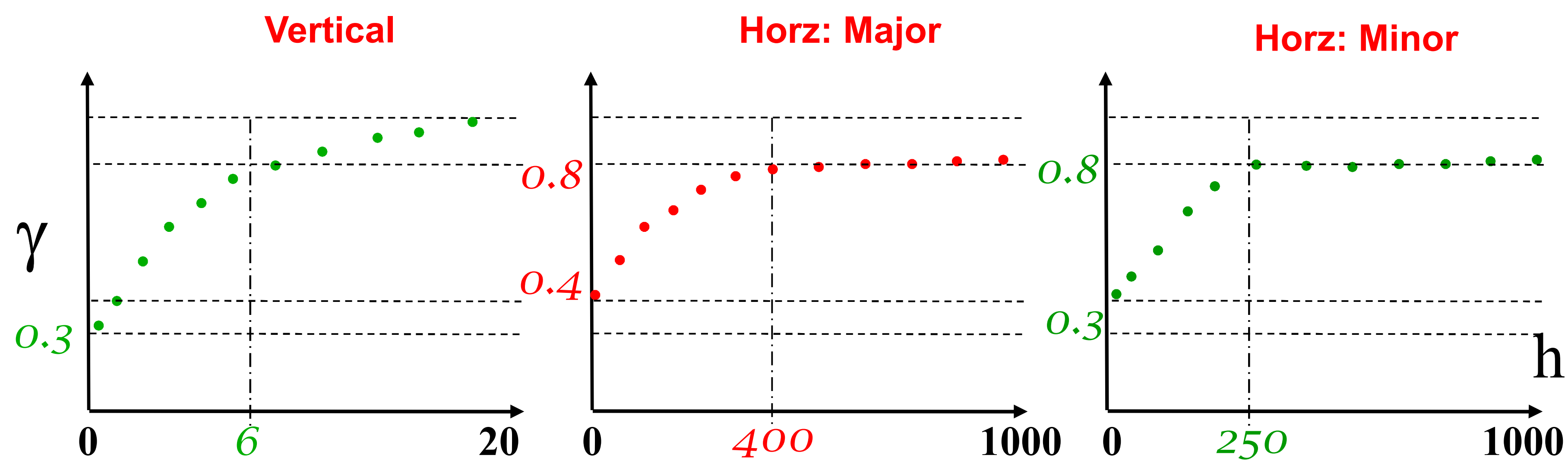
	Type	σ²	ν	Hmaj	Hmin
1.-	nugget	0.1			
2.-	Sph	0.2	8	0	0
3.-	Sph	0.1	8	0	200
4.-	Sph	0.4	8	400	200
5.-	Sph	0.2	8	9999	9999

1.0

all contributions sum to sill we are explaining each portion of the variance

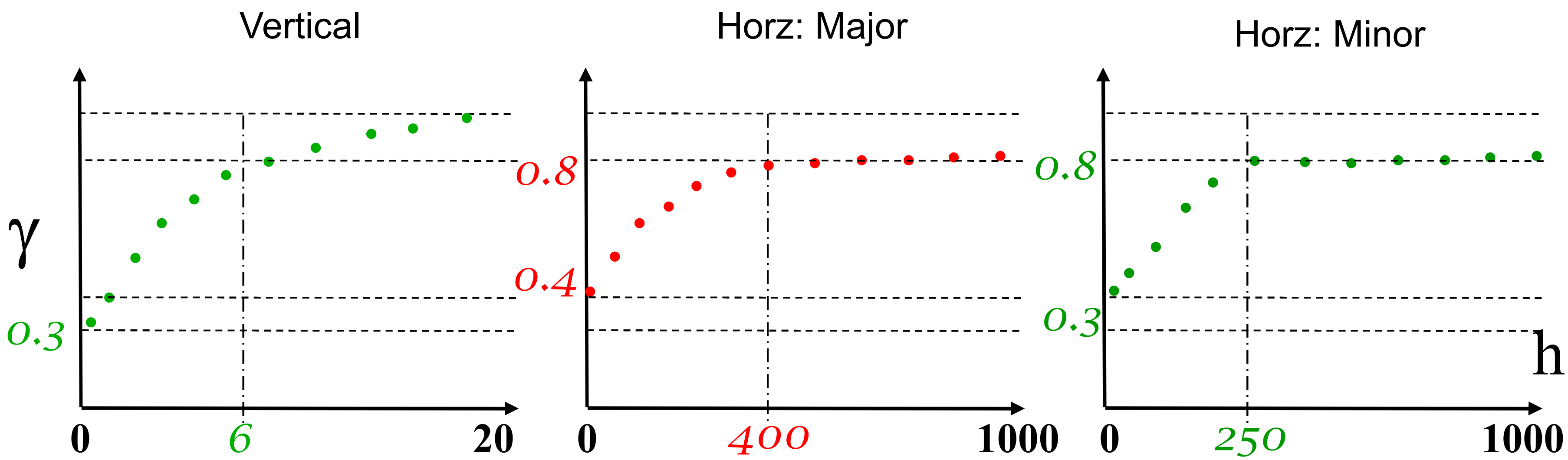
$$\gamma(h) = 0.1 + 0.2 \cdot sph_{\substack{av=8 \\ ah1=0 \\ ah2=0}} + 0.1 \cdot sph_{\substack{av=8 \\ ah1=0 \\ ah2=200}} + 0.4 \cdot sph_{\substack{av=8 \\ ah1=400 \\ ah2=200}} + 0.2 \cdot sph_{\substack{av=8 \\ ah1=\infty \\ ah2=\infty}}$$

Variogram Modeling Exercise



Structure	Type	σ^2	a_{maj}	a_{min}	a_z
1					
2					
3					
4					
5					

Variogram Modeling Exercise



Structure	Type	s^2	a_{major}	a_{minor}	a_{vert}
1	Nugget	0.3	-	-	-
2	Sph	0.1	0	0	6
3	Sph	0.4	400	250	6
4	Sph	0.2	9999	9999	20
5					

Inference in Presence of Sparse Data

- Most often there are inadequate data to infer a reliable horizontal variogram.

- Horizontal wells have not significantly helped with horizontal variogram inference (yet!):
  - *hard* core data and sophisticated well log measurements are rarely collected from horizontal wells
  - horizontal wells rarely track the stratigraphic “time lines”; they typically intersect the formation at some angle that undulates along the length of the wellbore

  Horizontal well data will become increasingly important

- At present, we depend on analogue data deemed relevant to the site being considered such as:
  - other, more extensively sampled, reservoirs,
  - geological process simulation, or
  - outcrop measurements

- In all cases, care must be taken to integrate global information from analogues with sparse local data

Variogram Modeling Hands-on Directional

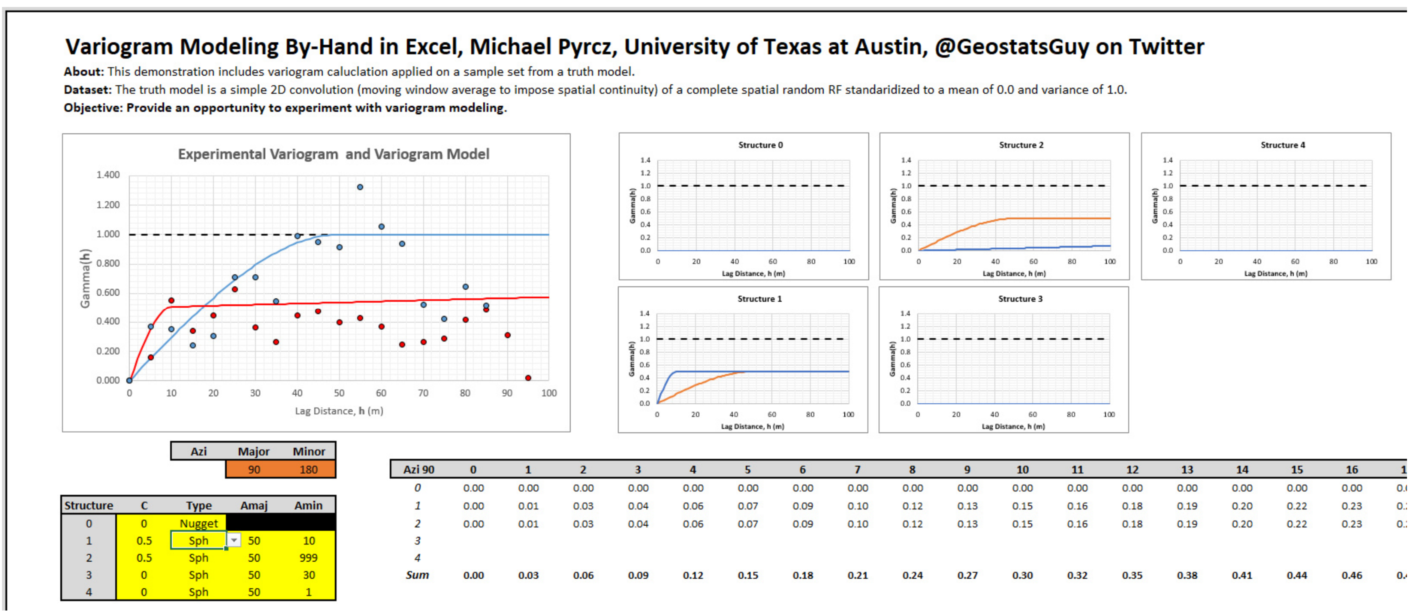
Variogram Modeling:

Things to try:

1. Calculate a 2D variogram model with 2 structures both spherical.

2. Vary structure type.

Formulate a 2D variogram model.



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

Variogram Modeling Hands-on

Here's an opportunity for experiential learning with Variogram Modeling.

- **Things to try:**

1. Modify and improve the variogram model by changing the ranges.

2. Modify and improve the variogram model by changing the types of structures.

GeostatsPy: Vairogram Modeling 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: Variogram Modeling 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 preciction 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), $\boldsymbol{h}$:

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

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 $\mathbf{h}$.

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

File Name: GeostatsPy_variogram_modeling.ipynb File is at: https://git.io/fhxYc

Data Analytics and Geostatistics: Spatial Estimation

Lecture outline . . .

- **Variogram Interpretation**
- **Variogram Modeling**