PGE 338 Data Analytics and Geostatistics

Lecture 13: Simulation

Lecture outline . . .

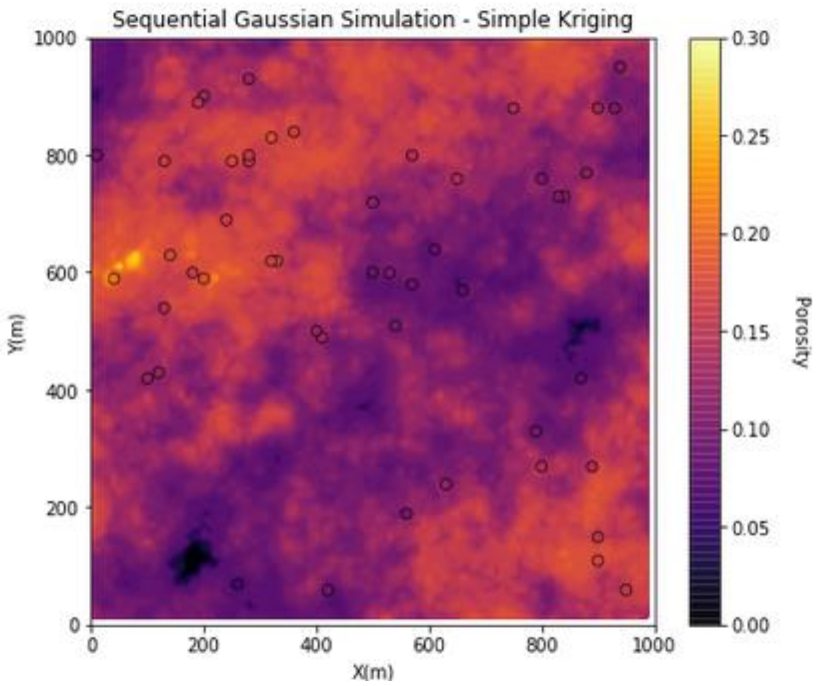
- Simulation

Michael Pyrcz, The University of Texas at Austin

Motivation

We need subsurface models that:

- Honor the spatial data, univariate distribution, and spatial continuity.



A geostatistically simulated realization of porosity.

PGE 338 Data Analytics and Geostatistics

Lecture 13: Simulation

Lecture outline . . .

- Simulation

Introduction

General Concepts

Univariate

Bivariate

Spatial

- Calculation
- Variogram Modeling
- Kriging
- Simulation

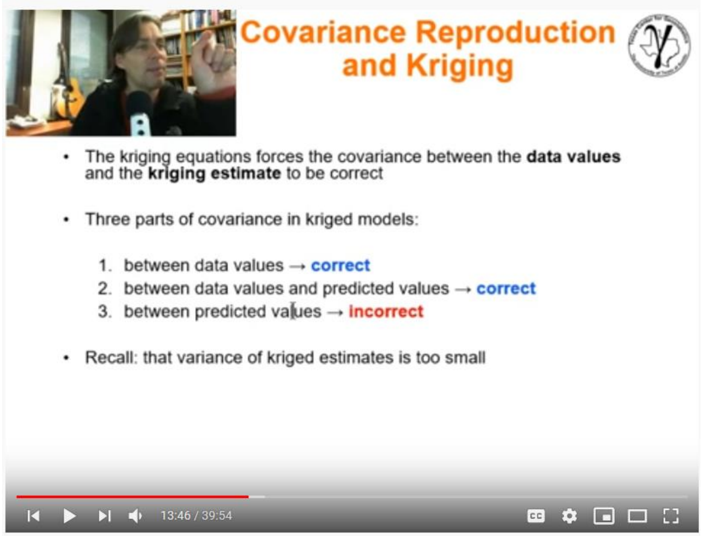
Time Series

Machine Learning

Uncertainty Analysis

Michael Pyrcz, The University of Texas at Austin

Recorded Lectures

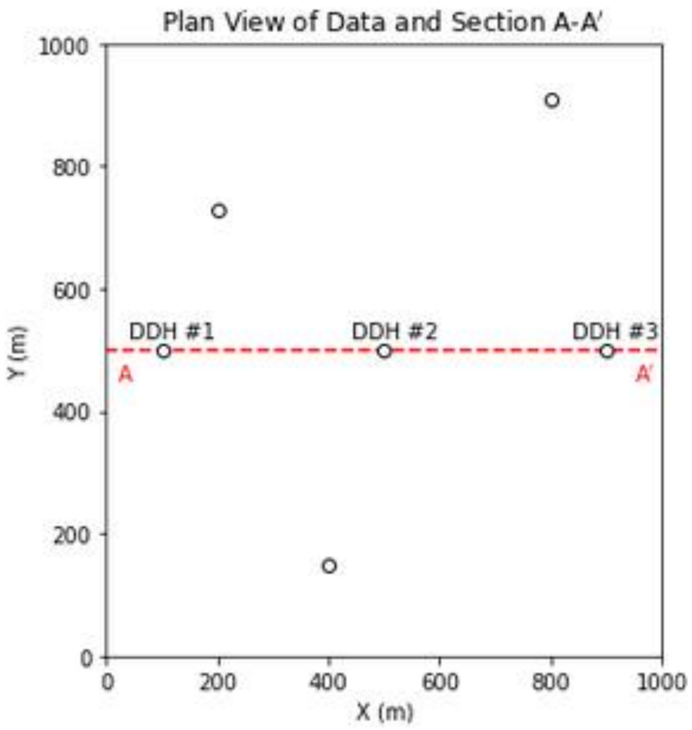


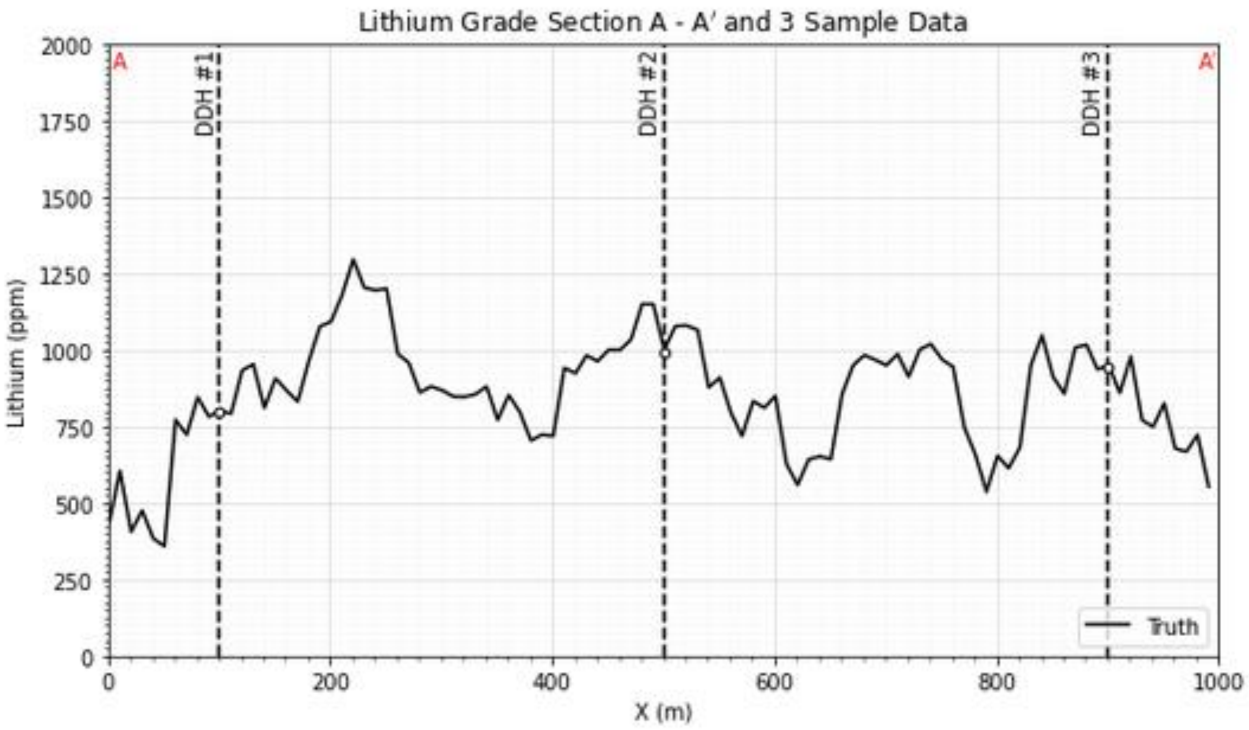
13 Data Analytics: Simulation

Simulation

Motivation for Simulation

Let's look at a section of 2D lithium grades. This is the inaccessible truth values along section $A - A'$.

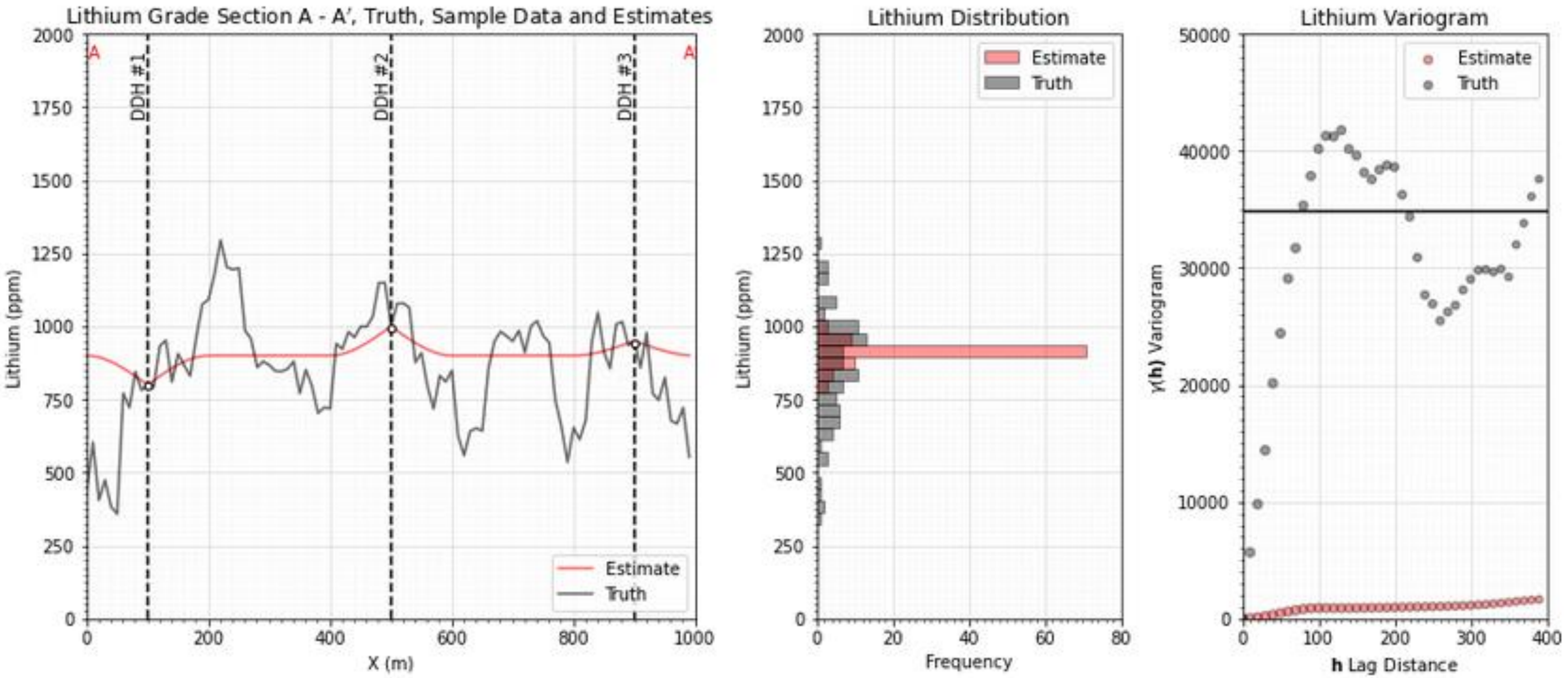




Plan view map with data and section A-A' (left) and section A-A' with lithium grades at and between the available data, file is GeostatsPy_estimation_vs_simulation.ipynb.

Motivation for Simulation

Recall estimation: assign the most accurate value at each location.



Section A-A' with lithium grades at and between the available data and an estimation model, the file is GeostatsPy_estimation_vs_simulation.ipynb.

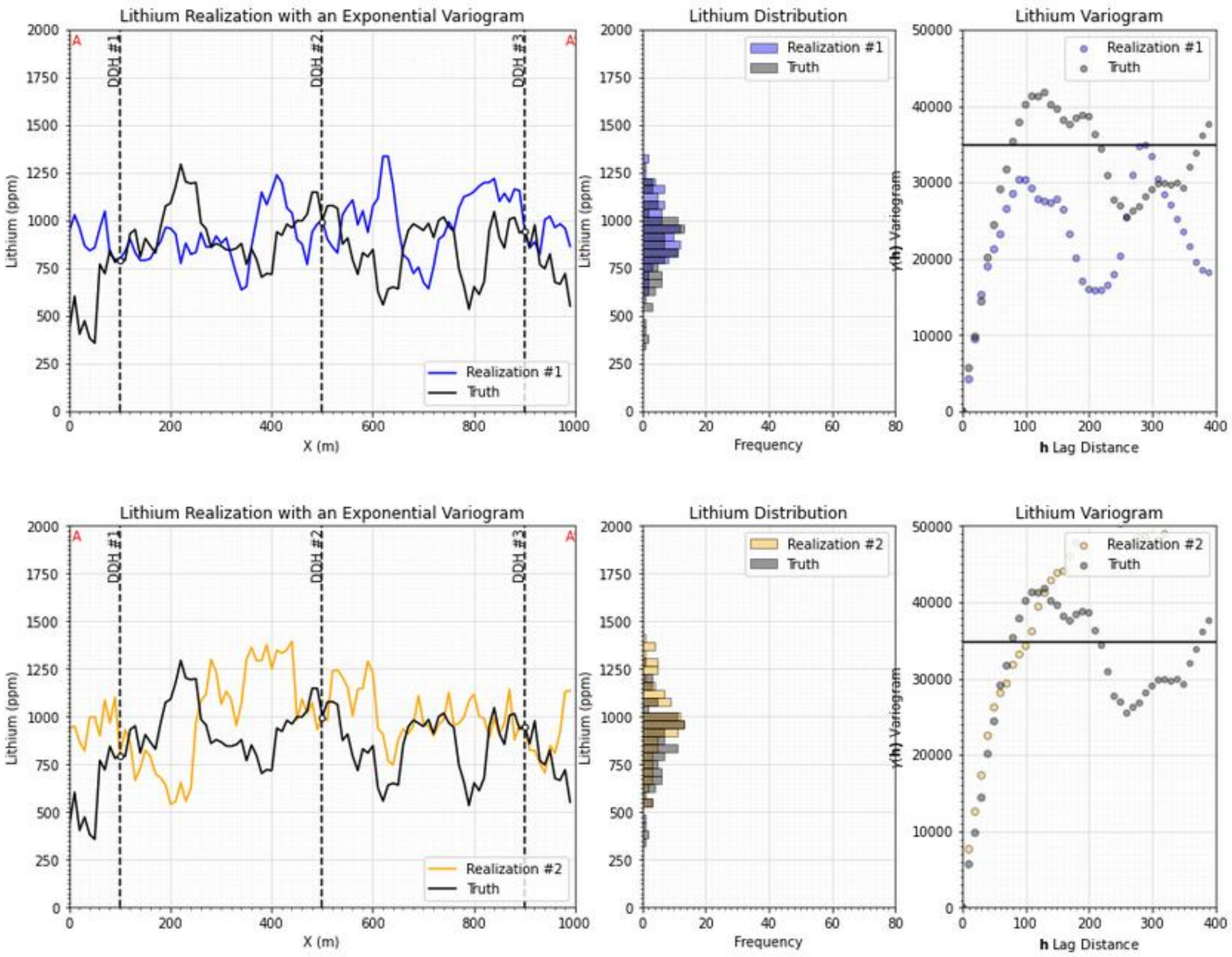
- This would not be appropriate for flow simulation!
  - Permeability variance is too low, Dykstra Parsons is incorrect, and too much spatial continuity.

Motivation for Simulation

We need models that:

- look like the truth
- have similar statistics with the truth
- behave similarly to the truth

This is the concept of simulation!



Section A-A’ with lithium grades at and between the available data and realizations, the file is GeostatsPy_estimation_vs_simulation.

Motivation for Simulation

What do we accomplish with simulation?

What does a simulated dill pickle potato chip taste like?

What does a simulated reservoir model look like?

What does a simulated reservoir flow like?

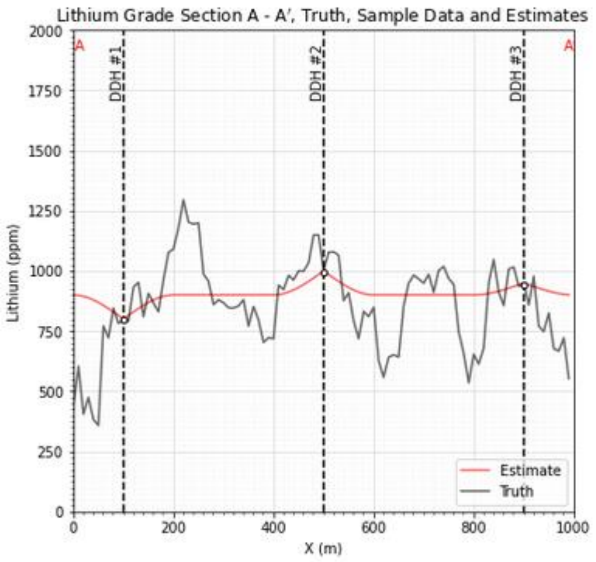
Estimation vs. Simulation

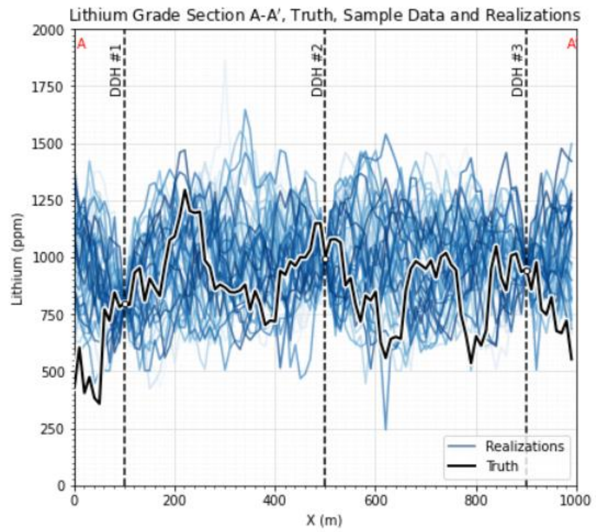
Estimation:

- honors local data
- locally accurate, primary goal of estimation is 1 estimate!
- too smooth, appropriate for visualizing trends
- too smooth, inappropriate for flow simulation
- one model, no assessment of global uncertainty

Simulation:

- honors local data
- sacrifices local accuracy, reproduces histogram
- honors spatial variability, appropriate for flow simulation
- alternative realizations, change random number seed
- many models (realizations), assessment of global uncertainty





Truth, estimation model (upper) and simulated realizations (lower), the file is GeostatsPy_estimation_vs_simulation.ipynb.

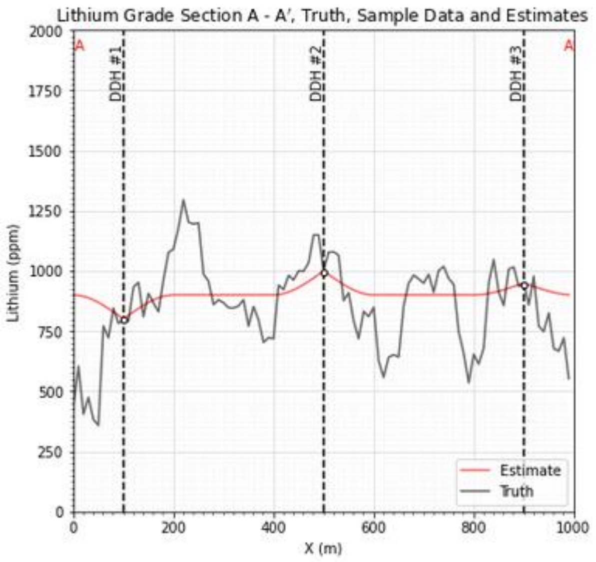
Estimation vs. Simulation

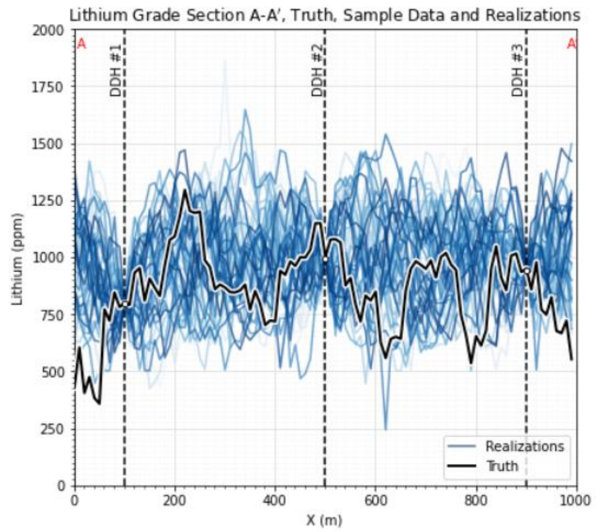
Estimation:

- a method to calculate the best estimate at each location
- focus on local accuracy, globally too smooth

Simulation:

- a method to calculate a good/reasonable estimate at each location
- focus on global accuracy, sacrifice local accuracy





Truth, estimation model (upper) and simulated realizations (lower), the file is GeostatsPy_estimation_vs_simulation.ipynb.

Global and Local Definition

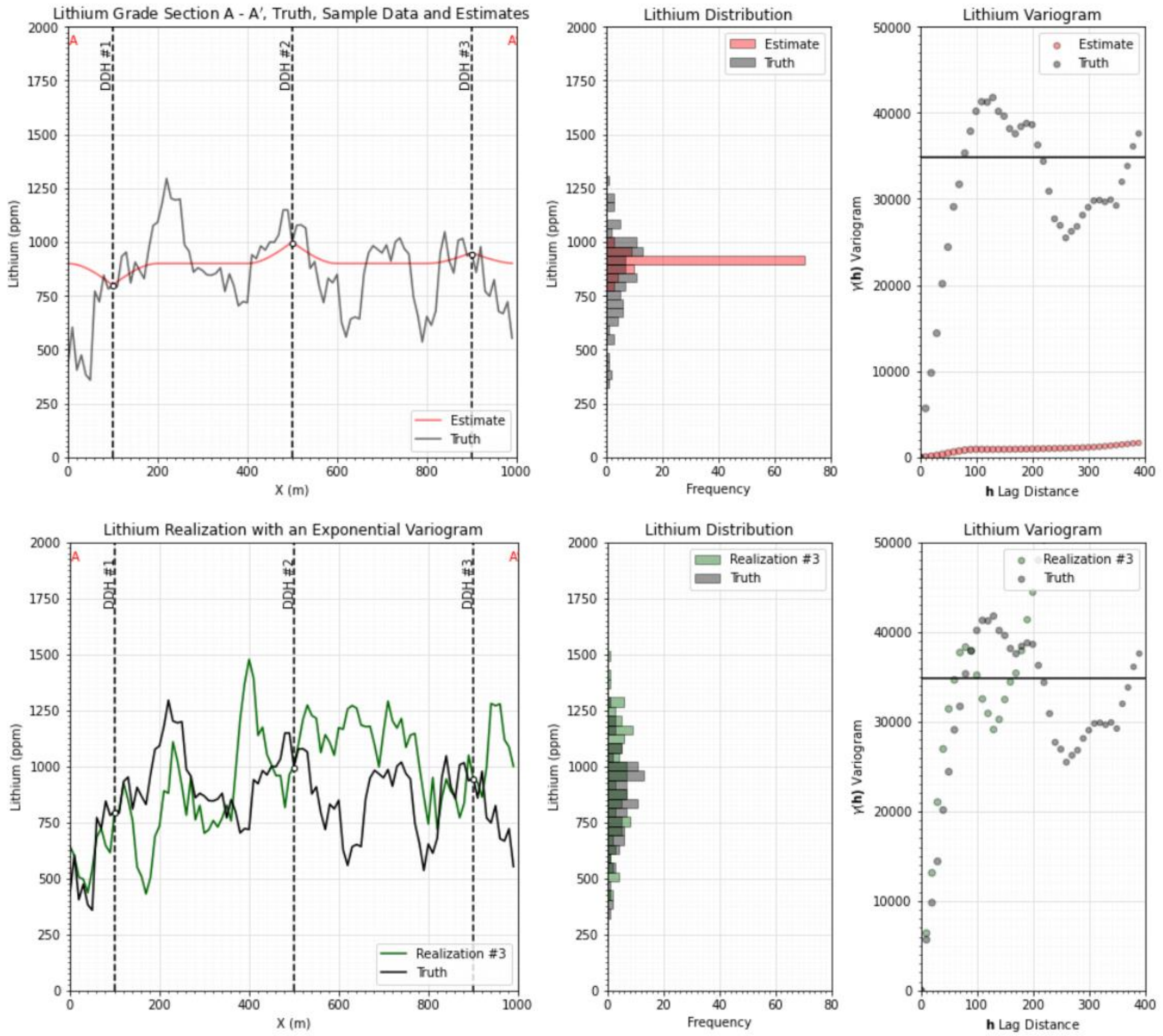
Estimation is locally accurate and **simulation** is globally accurate.

Global vs. Local Measures

- Local Measure – feature values considered location by location.
- Global Measure – a statistical summary over many local measures within the volume of interest.

Global vs. Local Accuracy

- Global Accuracy – we honor a statistic calculated over the volume of interest (e.g., variogram / distribution)
- Local Accuracy – we have an estimate that minimizes the estimation variance, the most likely value.



Truth, estimation model (upper), and 1 simulated realization (lower) with histograms and variograms, the file is GeostatsPy_estimation_vs_simulation.ipynb.

Smoothing Effect of Kriging

- Kriging is locally accurate and smooth, appropriate for visualizing trends, and is inappropriate for any applications where heterogeneity is important

- The "variance" of the kriged estimates is too small:

$$Var\{Y^*(\mathbf{u})\} = \sigma^2 - \sigma_{sk}^2(\mathbf{u})$$

Simple Kriging Estimation Variance

σ^2 is variance of the feature, $Var\{Y^*(\mathbf{u})\}$ is the variance of the estimates and $\sigma_{sk}^2(\mathbf{u})$ is the kriging / estimation variance.

Consider the following:

- $\sigma_{sk}^2(\mathbf{u})$ is zero at the data locations → no smoothing, σ^2 is correct!
- $\sigma_{sk}^2(\mathbf{u})$ is variance σ^2 beyond range of data locations → complete smoothing, estimate with the global mean.
- spatial variations of $\sigma_{sk}^2(\mathbf{u})$ depend on the variogram and data spacing

Proposal to Correct Kriging

1. To Correct the Missing Variance / Get the Right Histogram

- Missing variance in the estimates, $Var\{Y^*(\mathbf{u})\}$, *is* the kriging variance, $\sigma_{sk}^2(\mathbf{u})$
- Monte Carlo *Simulation* to add/correct the variance (get right histogram)

Add random residual $Y_s(\mathbf{u}) = Y^*(\mathbf{u}) + R(\mathbf{u})$ where *R*(u) is random residual that adds back in the missing variance, $\sigma_{sk}^2(\mathbf{u})$.

2. To Correct the Spatial Continuity / Get the Right Variogram

and correct the covariance (get the right variogram)

Sequential simulation – add simulated values to data

- Simulation reproduces histogram, honors spatial variability (variogram) → appropriate for process evaluation where heterogeneity is important
- Allows an assessment of uncertainty with alternative realizations

We know we can add the missing variance in, but will that 'wreck' the variogram?

- Let's look at the covariance reproduction with kriging.

Correct the Missing Variance in Kriging

We correct the missing variance by adding a random residual, a realization from a random variable, to our estimate.

- The variance of our random function: $\sigma^2 = C(0)$
- Stationary variance should be constant everywhere:

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

AOI is area of interest.

- But the variance of the kriged estimates is too small:

$$Var\{Y^*(\mathbf{u})\} = \sigma^2 - \sigma_{sk}^2(\mathbf{u})$$

the *missing variance* is the kriging variance $\sigma_{sk}^2(\mathbf{u})$!

We can add in the missing variance to the kriging estimate with: $y_s(\mathbf{u}_\alpha) = y^*(\mathbf{u}_\alpha) + r(\mathbf{u}_\alpha)$

$y_s(\mathbf{u}_\alpha)$ is our simulated value, $y^*(\mathbf{u}_\alpha)$ is our kriging estimate and $r(\mathbf{u})$ is a MCS realization from the residual random variable,

$\sim F_{y_s}[0, \sigma_{sk}^2(\mathbf{u}_\alpha)]$

Alternatively, we can directly perform Monte Carlo Simulation

$$Y_s(\mathbf{u}_\alpha) \sim F_{y_s}[y^*(\mathbf{u}_\alpha), \sigma_{sk}^2(\mathbf{u}_\alpha)]$$

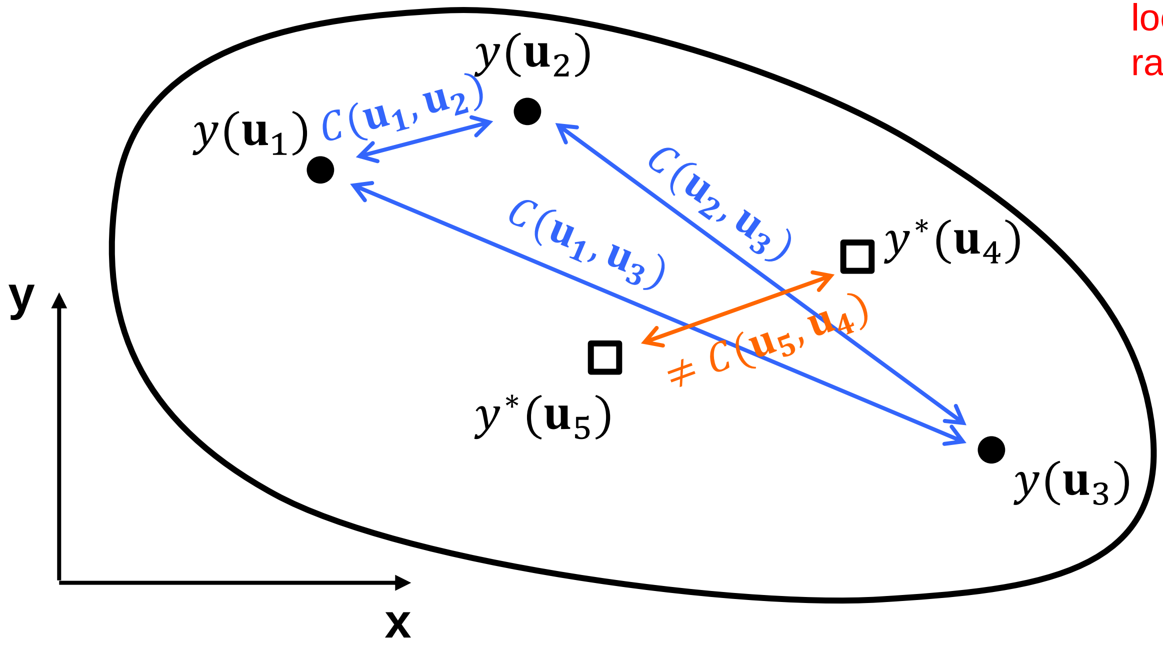
Covariance Reproduction and Kriging

Is the covariance reproduced with simple kriging?

1. between data values → **correct!**
2. between data values and predicted values → **?**
3. between predicted values → **incorrect!**

We calculated the variogram from the data!

This is apparent.
E.g., the simple kriging estimates are the feature mean for all locations outside the range from any data!



Covariance Reproduction and Kriging

Is the covariance reproduced with simple kriging between data and estimates?

- Recall the simple kriging estimator:

$$Y^*(\mathbf{u}) = \sum_{\beta=1}^{n} \lambda_\beta \cdot Y(\mathbf{u}_\beta)$$

  and the corresponding simple kriging system:

$$\sum_{\beta=1}^{n} \lambda_\beta C(\mathbf{u}_\alpha, \mathbf{u}_\beta) = C(\mathbf{u}, \mathbf{u}_\alpha), \quad \forall\, \mathbf{u}_\alpha$$

- Let's calculate the covariance between the kriged estimate and one of the data values:

$$Cov\{Y^*(\mathbf{u}), Y(\mathbf{u}_\alpha)\} = E\{Y^*(\mathbf{u}), Y(\mathbf{u}_\alpha)\}$$

Since residual mean is zero we can neglect the mean term.

$$= E\left\{\left[\sum_{\beta=1}^{n} \lambda_\beta \cdot Y(\mathbf{u}_\beta)\right] \cdot Y(\mathbf{u}_\alpha)\right\}$$

Substitute the simple kriging estimator

$$= \sum_{\beta=1}^{n} \lambda_\beta \cdot E\{Y(\mathbf{u}_\beta) \cdot Y(\mathbf{u}_\alpha)\}$$

Expectation of a constant is a constant.

$$= \sum_{\beta=1}^{n} \lambda_\beta C(\mathbf{u}_\alpha, \mathbf{u}_\beta)$$

Again, we can neglect the mean term.

$$= C(\mathbf{u}, \mathbf{u}_\alpha)$$

Substitute the simple kriging estimator

The covariance between a kriging estimate and a data point is the covariance between their locations.

The covariance / spatial continuity between the data and the kriging estimates is correct!

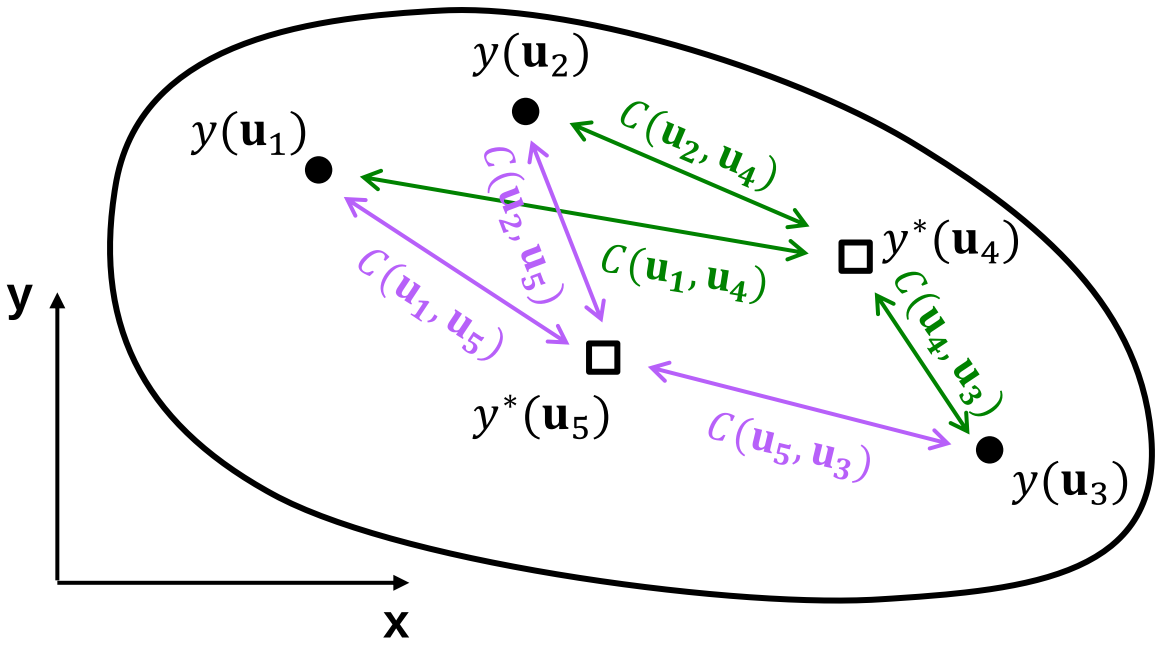
Proof is from Pyrcz and Deutsch, 2014, p. 122

Covariance Reproduction and Kriging

Is the covariance reproduced with simple kriging?

1. between data values → correct
2. **between data values and predicted values → correct!**
3. between predicted values → incorrect

We proved this on the previous slide.



Addition of Missing Variance

Does the addition of a random residual change the covariance between data and estimates?

- Add an independent component with zero mean and the correct variance to the estimate:

- Is the covariance still correct between data and unknown locations?:

$$Y_s(\mathbf{u}) = Y^*(\mathbf{u}) + R(\mathbf{u})$$

$$Cov\{Y_s(\mathrm{u}), Y(\mathrm{u}_\alpha)\} = E\{Y_s(\mathbf{u}) \cdot Y(\mathbf{u}_\alpha)\}$$

Recall: $Y^*(\mathbf{u}) = \sum_{\beta=1}^{n} \lambda_\beta Y(\mathbf{u}_\beta)$

$$= E\left\{\left[\sum_{\beta=1}^{n} \lambda_\beta Y(\mathbf{u}_\beta) + R(\mathbf{u})\right] \cdot \boldsymbol{Y}(\boldsymbol{u}_\alpha)\right\}$$

Distributive property of expectation.

$$= \sum_{\beta=1}^{n} \lambda_\beta E\{Y(\mathbf{u}_\beta) \cdot \boldsymbol{Y}(\boldsymbol{u}_\alpha)\} + E\{R(\mathbf{u}) \cdot \boldsymbol{Y}(\boldsymbol{u}_\alpha)\}$$

$Cov\{Y(\mathbf{u}_\beta), Y(\mathbf{u}_\alpha)\}$, since $\bar{y} = 0$.

$$Cov\{Y_s(u), Y(u_\alpha)\} = \sum_{\beta=1}^{n} \lambda_\beta Cov\{Y(\mathbf{u}_\beta), Y(\mathbf{u}_\alpha)\} + E\{R(\mathbf{u}) \cdot Y(\mathbf{u}_\alpha)\}$$

Addition of Missing Variance

Does the addition of a random residual change the covariance between data and estimates?

Continue from last slide:

$$Cov\{Y_s(\mathbf{u}), Y(\mathbf{u}_\alpha)\} = \sum_{\beta=1}^{n} \lambda_\beta \, Cov\{Y(\mathbf{u}_\beta), Y(\mathbf{u}_\alpha)\} + E\{R(\mathbf{u}) \cdot Y(\mathbf{u}_\alpha)\}$$

- note that $E\{R(\mathbf{u}) \cdot Y(\mathbf{u}_\alpha)\} = E\{R(\mathbf{u})\} \cdot E\{Y(\mathbf{u}_\alpha)\}$, since $R(\mathbf{u})$ is random and independent from the data values, $Y(\mathbf{u}_\alpha)$.

- since the residual mean is 0, $E\{R(\mathbf{u})\} = 0.0, E\{R(\mathbf{u})\} \cdot E\{Y(\mathbf{u}_\alpha)\}$ = 0.0.

$$Cov\{Y_s(\mathbf{u}), Y(\mathbf{u}_\alpha)\} = \sum_{\beta=1}^{n} \lambda_\beta C(\mathbf{u}_\alpha, \mathbf{u}_\beta) = Cov\{\mathbf{u}, \mathbf{u}_\alpha\} \longleftarrow$$

given the simple kriging system

$$\sum_{\beta=1}^{n} \lambda_\beta C(\mathbf{u}_\alpha, \mathbf{u}_\beta) = C(\mathbf{u}, \mathbf{u}_\alpha), \quad \forall \mathbf{u}_\alpha$$

The addition of the random residual does not change the covariance between the data and the unknown location. **Still correct!**

The trick is to simulate values and sequentially add them to the data, since we know the covariance between estimates and data are correct.

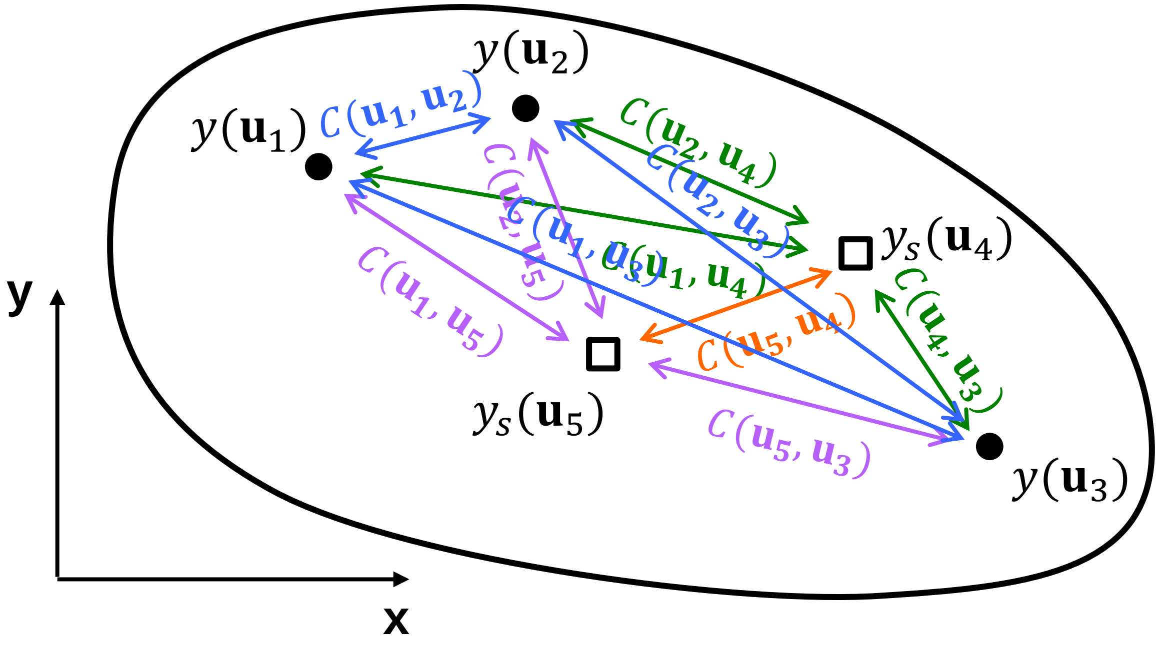
Proof is from Pyrcz and Deutsch, 2014, p. 123

We Corrected Kriging with Sequential Simulation

The variance is correct with the added random residual.

and is the covariance reproduced with simple kriging?

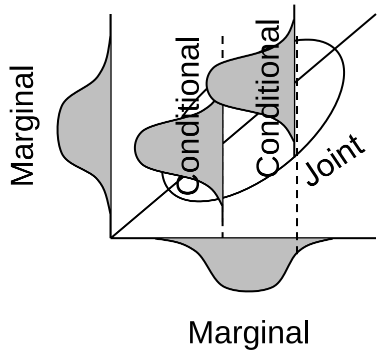
1. between data values → **correct**
2. between data values and predicted values → **correct**
3. **between predicted values → correct if simulated sequentially!**



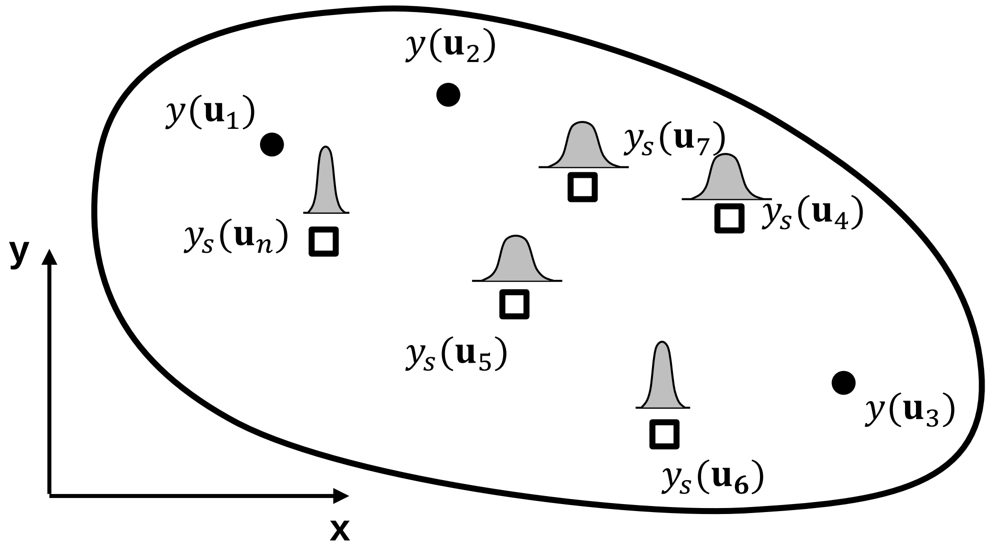
What About the Global Histogram?

We need the global histogram of simulated values to be correct, $\boldsymbol{F_x(x)}$.

$$y_s(\mathbf{u}_\alpha)\ \forall\ \alpha = 1, \dots, n \sim \boldsymbol{F_x(x)}$$



In Gaussian space, all distributions are Gaussian!



- If we draw from Gaussian (conditional) distributions at each simulation location, the global marginal distribution will be Gaussian also!

- Also, Gaussian is parameterized by the mean and variance ∴

$$Y_s(\mathbf{u}_\alpha) \sim N[y^*(\mathbf{u}_\alpha),\ \sigma_{sk}^2(\mathbf{u}_\alpha)]$$

Original Data and Gaussian Space

We transform the data, work in Gaussian values and then back-transform to original data values for forecasting.

- Since kriging is performed in Gaussian space, we calculate variograms of Gaussian transformed features.

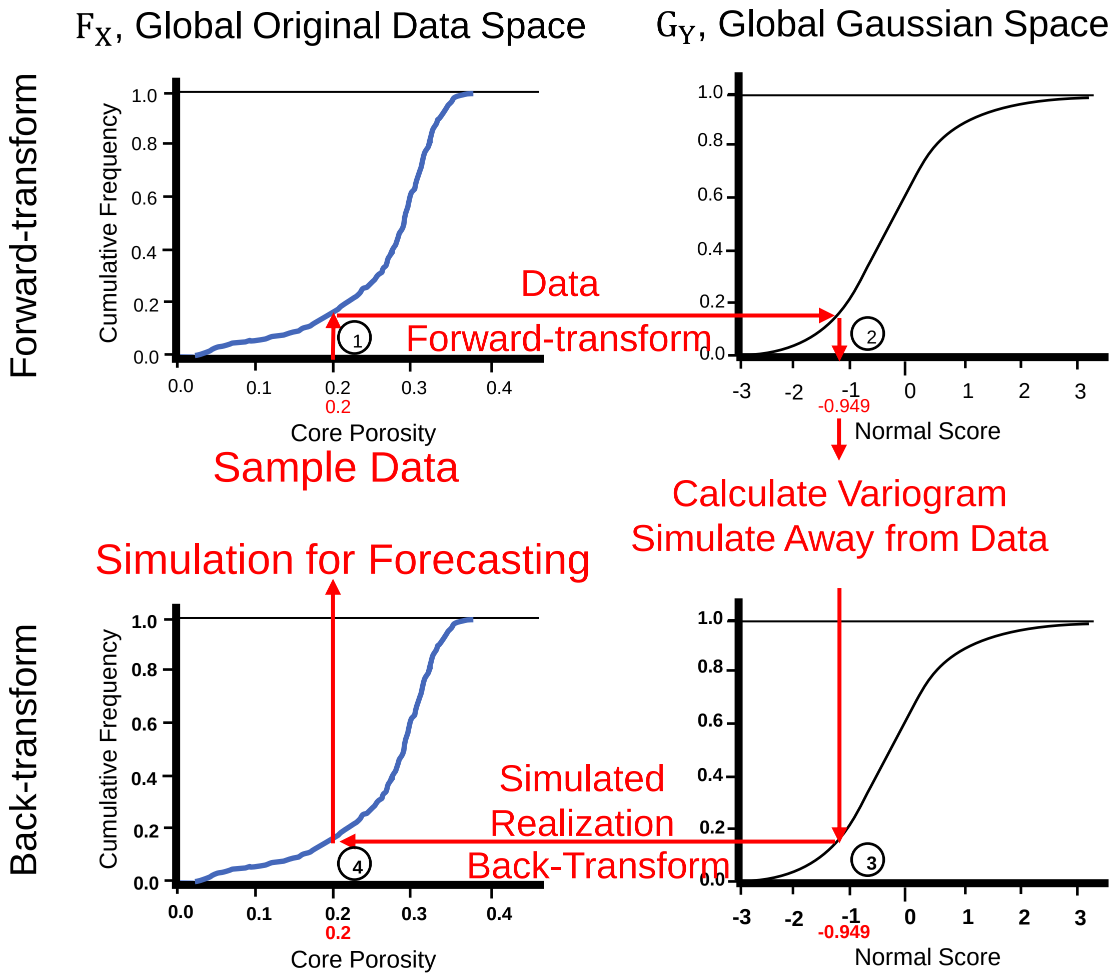


Illustration of simulation in Gaussian space.

Sequential Gaussian Simulation Definition

The most common method for calculating continuous feature simulated realizations for subsurface modeling.

- **Sequential** – sequential inclusion of simulated values to impose the correct spatial correlation between the simulated values.

- **Gaussian** – work in Gaussian space since the local conditional distribution shape is known and can be parameterized by mean (kriging estimate) and variance (kriging variance). Also, the global distribution will be correct (Gaussian), so we can back-transform after simulation to the original data space and get the correct distribution.

- **Simulation** – simulation with Monte Carlo simulation from the local distributions to add in the missing variance and construction of multiple, equiprobable realizations.

Sequential Simulation Workflow

- Transform data to standard normal distribution (all work will be done in "Gaussian" space)

- Go to a random location and perform kriging to get mean and corresponding kriging variance:

$$Y^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha Y(\mathbf{u}_\alpha)$$

$$\sigma_{sk}^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n} \lambda_\alpha Y(\mathbf{u}_\alpha)$$

- Draw a random residual, $R(\mathbf{u})$, that follows a normal distribution with mean of 0.0 and variance of $\sigma_{sk}^2(\mathbf{u})$.

- Add the kriged estimate, $Y^*(\mathbf{u})$, and residual, $R(\mathbf{u})$, to get simulated value:

$$Y_s(\mathbf{u}) = Y^*(\mathbf{u}) + R(\mathbf{u})$$

- Note that $Y_s(\mathbf{u})$ could be equivalently obtained by drawing from a normal distribution with mean $Y^*(\mathbf{u})$ and variance, $\sigma_{sk}^2(\mathbf{u})$.

Sequential Simulation Workflow

- Add $Y_s(\mathbf{u})$ to the set of data to ensure that the covariance with this value and all future predictions is correct

- A key idea of sequential simulation is to use previously kriged/simulated values as data so that we reproduce the covariance between *all* of the simulated values!

- Visit all locations in random order (to avoid artifacts of limited search)

- Back-transform all data values and simulated values when model is populated

- Create another equiprobable realization by repeating with different random number seed

Why Gaussian Simulation?

- Local estimate is given by kriging
- Mean of residual is zero and variance is given by kriging; however, what “shape” of distribution should we consider?
- Advantage of normal / Gaussian distribution is that the global $N(0,1)$ distribution will be preserved if we always use Gaussian distributions
- Transform data to normal scores in the beginning (before variography)
- Simulate 3-D realization in “normal space” then back transform to original space at the end.

Why Gaussian Simulation?

- Price of mathematical simplicity is the characteristic of maximum spatial entropy, i.e., low and high values are disconnected, e.g., not appropriate for permeability.

- Of all unbounded distributions of finite variance, the Gaussian distribution maximizes entropy:

$$\mathrm{H} = -\int_{-\infty}^{+\infty} ln\left(f_z(z)\right) f_z(z) dz$$

- Consequences:
  - maximum spatial disorder beyond the variogram
  - maximum disconnectedness of extreme values
  - median values have the greatest spatial ‘connectedness’
  - symmetric disconnectedness of extreme low/high values

Steps in Sequential Gaussian Simulation, Take 2

1. Establish grid network and coordinate system, flatten system
2. Assign data to the grid (account for scale change)
3. Transform data to “normal space”
4. Calculate variogram
5. Determine a random path through all of the grid nodes, at each node:
   a) find nearby data and previously simulated grid nodes
   b) construct the conditional distribution by kriging
   c) draw simulated value from conditional distribution
   d) assign the simulated value to the grid as data
6. *Check realization (could also check after back transform).*
   a) honor data?
   b) honor histogram: N(0,1) standard normal with a mean of zero and a variance of one?
   c) honor variogram?
7. Back transform from “normal space”
8. Restore to original framework.
9. *Check honor concept of geology? geophysics and production data?*
10. Calculate multiple realizations

Loop over model nodes

Loop over realizations

Steps in Sequential Gaussian Simulation, Take 2

Why you should know the sequential Gaussian (SGS) simulation workflow?

1. >90% of reservoir models use SGS in the workflow!

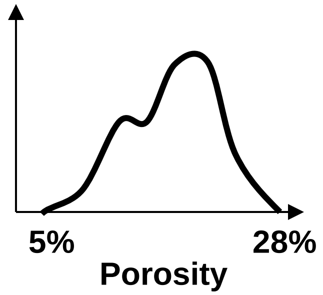
2. The workflow links directly to theory.
   a) Correcting for missing variance with addition of a random residual / kriging variance
   b) Ensuring correct covariance between simulated values with sequential approach
   c) Use of a random path and Monte Carlo simulation to calculate realizations.

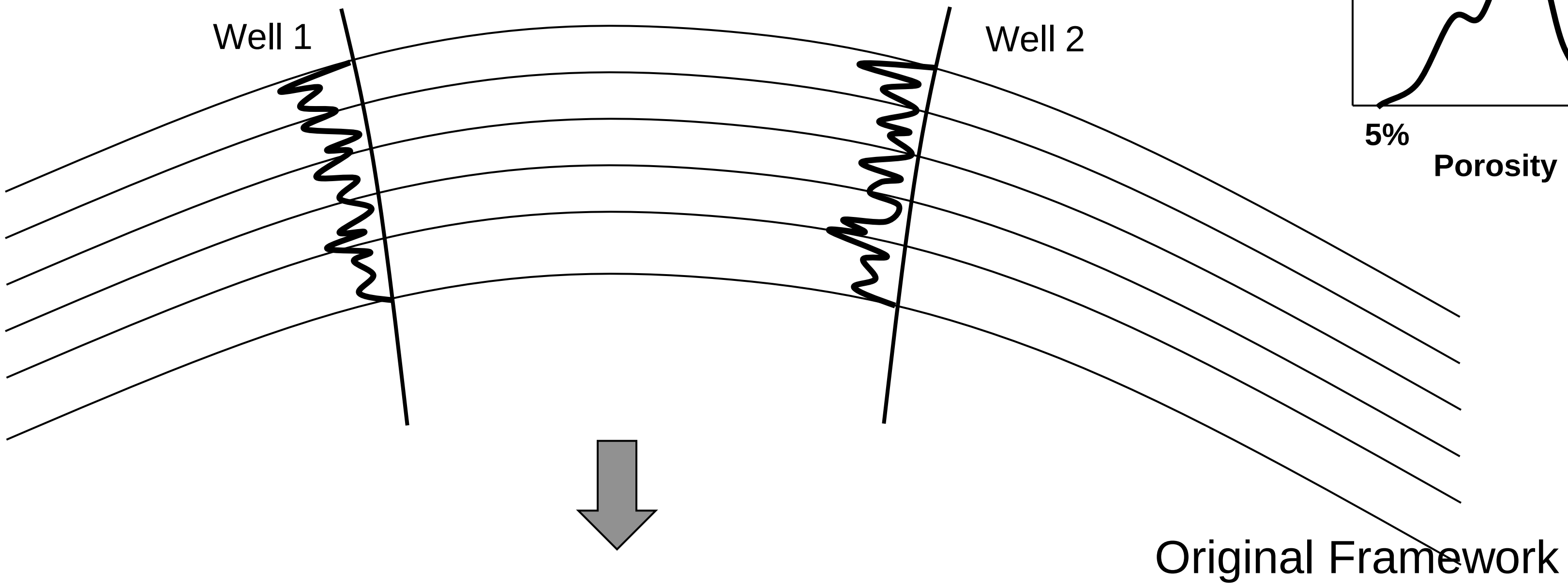
3. You'll understand when things go wrong!
   a) Limited search artifacts
   b) Mismatch with input statistics / ergodic fluctuations
   c) String effect

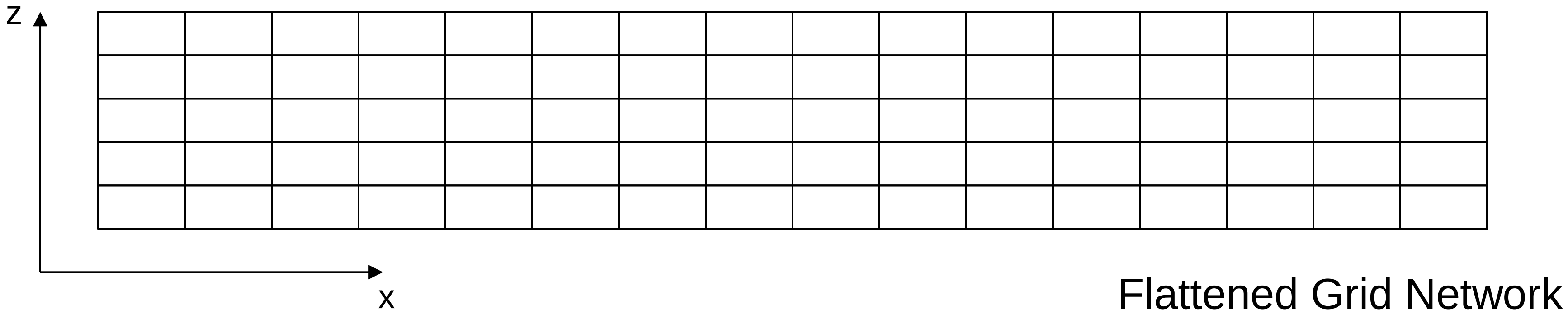
4. The sequential simulation framework is used in many other methods
   a) cosimulations, indicator simulation, truncated Gaussian, MPS etc.

Steps in Sequential Gaussian Simulation, Take 2

1. Establish grid network and coordinate system (-space)

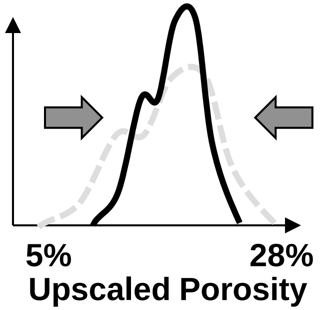


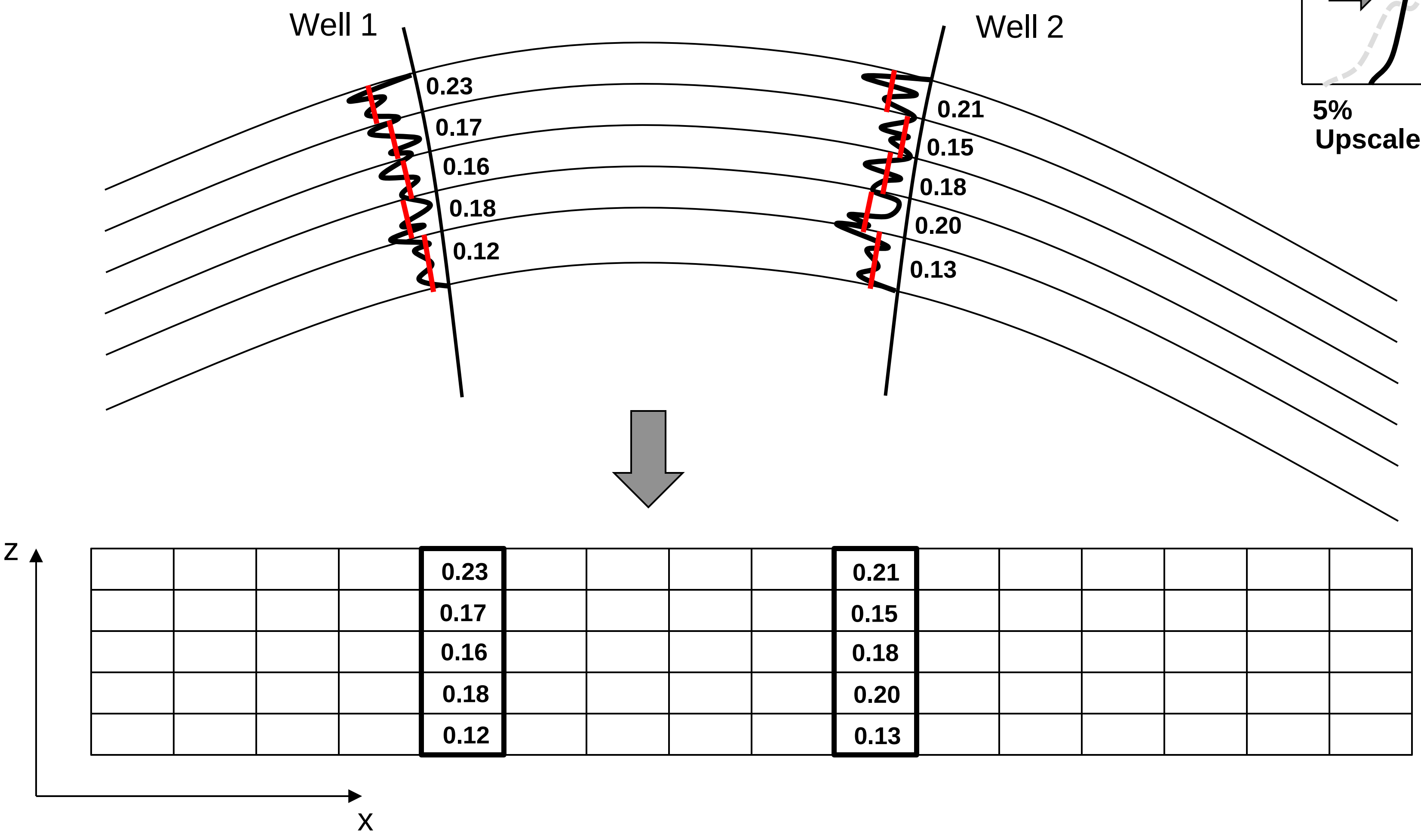




Steps in Sequential Gaussian Simulation, Take 2

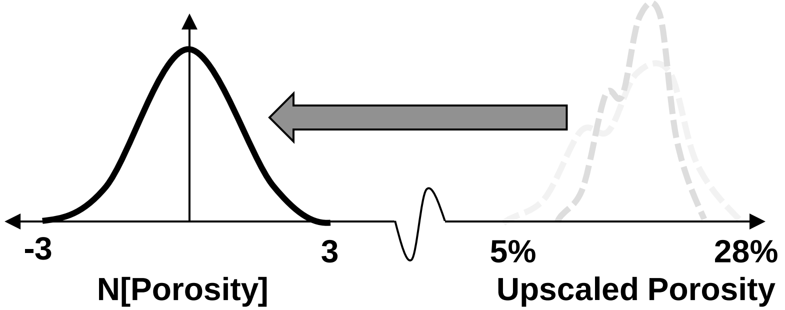
2. Assign data to the grid (account for scale change)

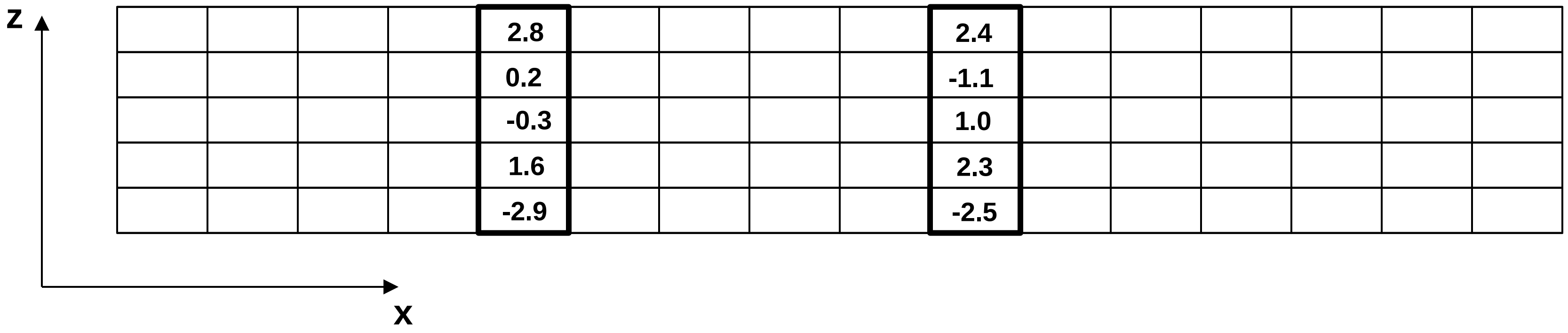




Steps in Sequential Gaussian Simulation, Take 2

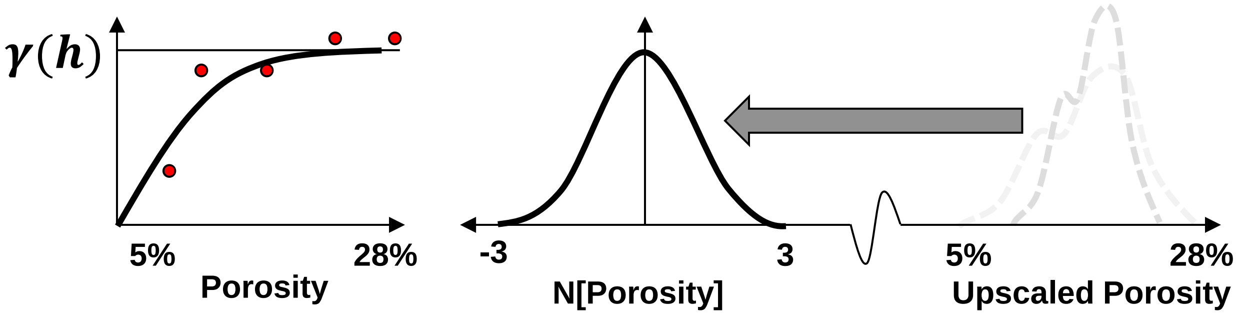
3. Transform data to "normal space"

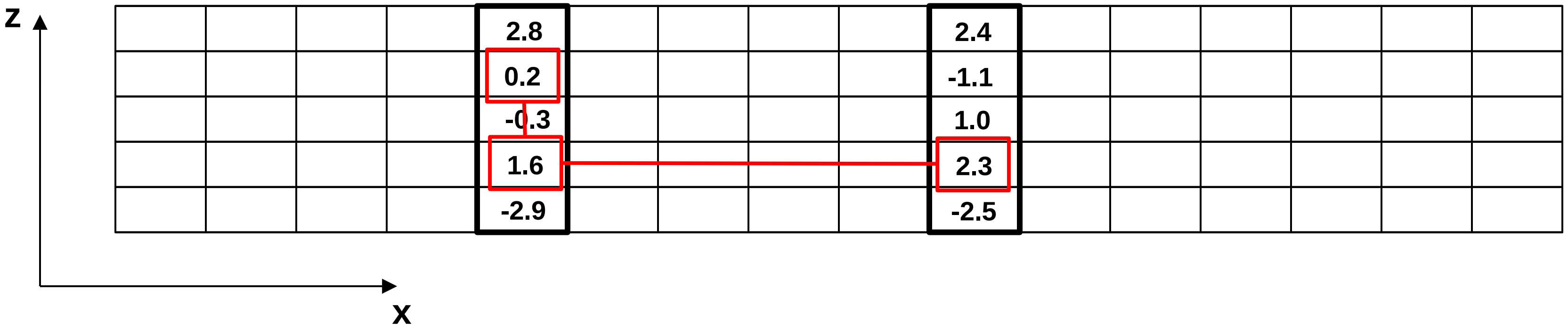




Steps in Sequential Gaussian Simulation, Take 2

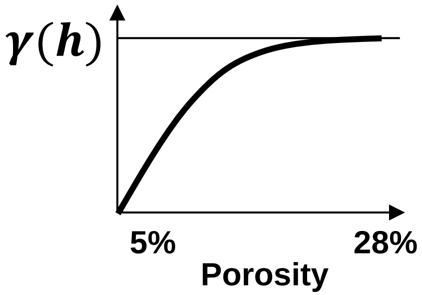
4. Calculate and model variogram

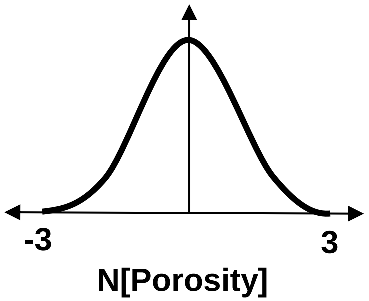


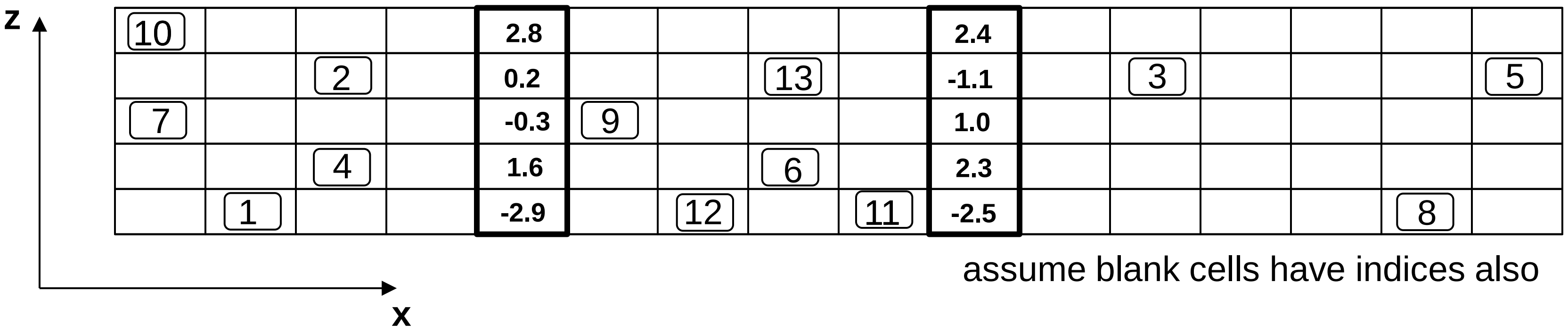


Steps in Sequential Gaussian Simulation, Take 2

5. Determine a random path through all of the grid nodes
   - only included 1st 13 nodes



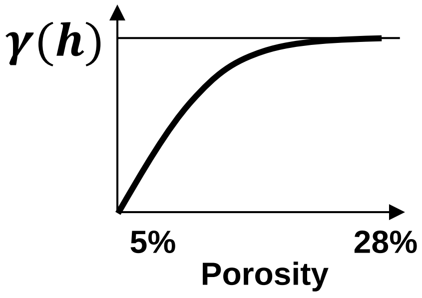


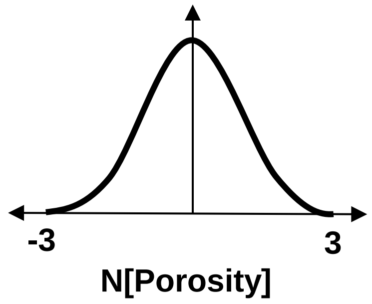


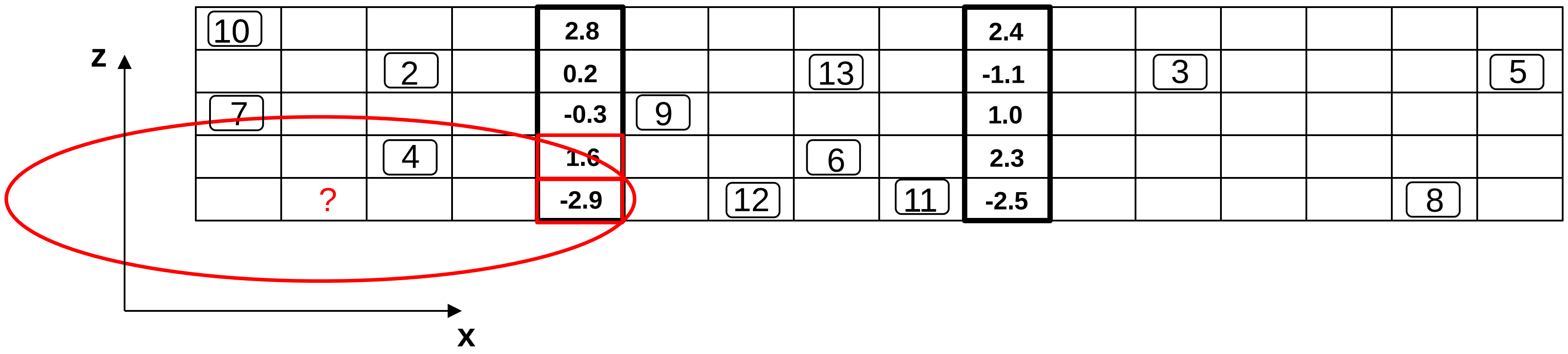
assume blank cells have indices also

Steps in Sequential Gaussian Simulation, Take 2

5. For each node on random path:
   a) find nearby data and previously simulated grid nodes

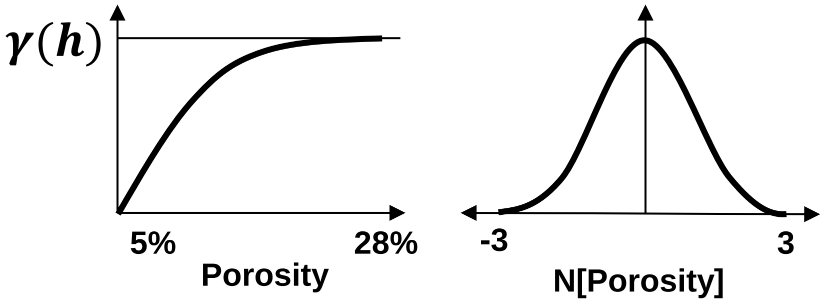


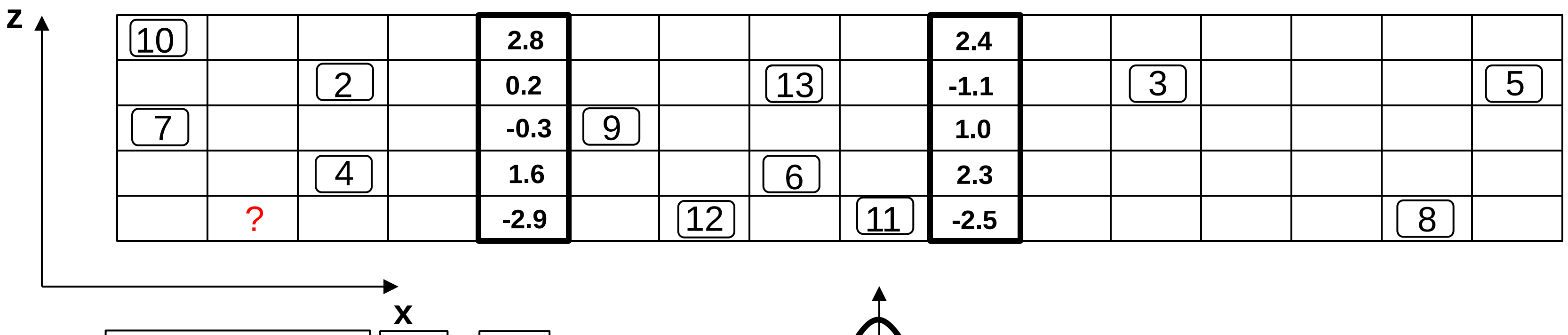




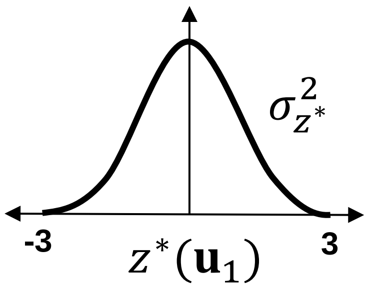
Steps in Sequential Gaussian Simulation, Take 2

5. For each node on random path:
   a) find nearby data and previously simulated grid nodes
   b) construct the conditional distribution by kriging



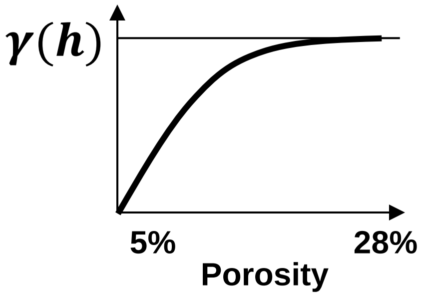


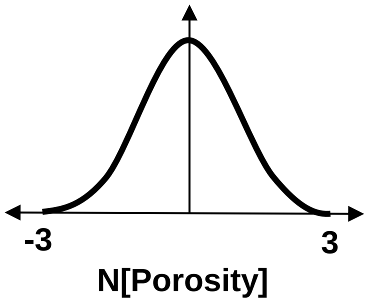
$$\begin{bmatrix} C_{1,1} & & \\ & \ddots & \\ & & C_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{0,1} \\ \vdots \\ C_{0,n} \end{bmatrix}$$

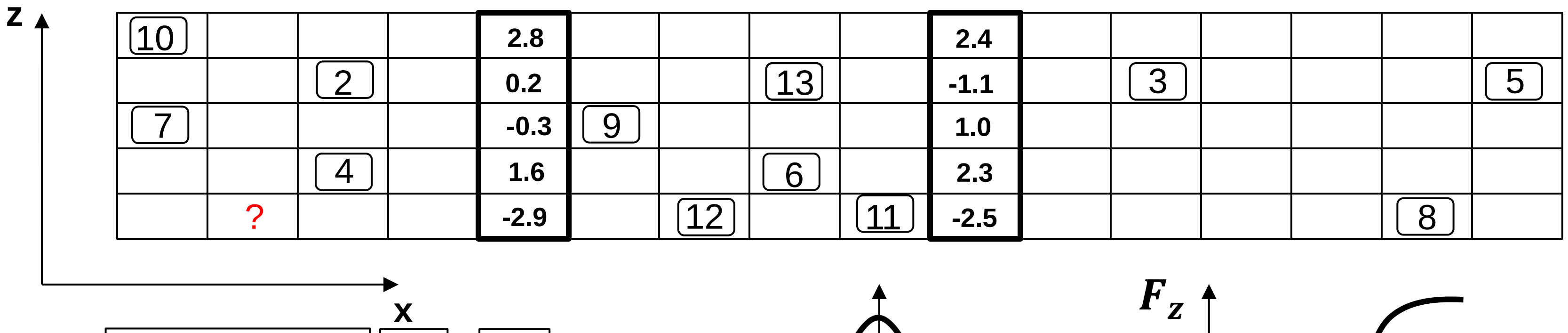


Steps in Sequential Gaussian Simulation, Take 2

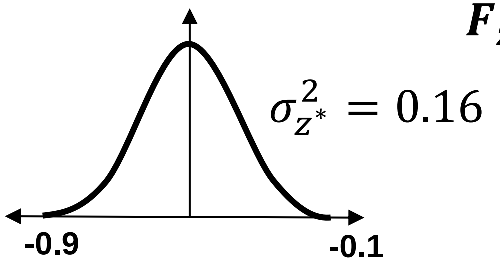
5. For each node on random path:
   a) find nearby data and previously simulated grid nodes
   b) construct the conditional distribution by kriging
   c) draw simulated value from conditional distribution



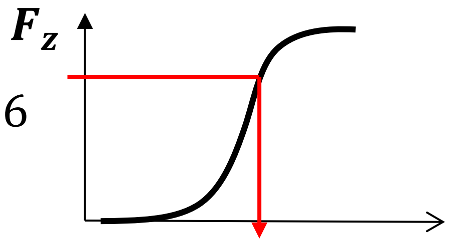




$$\begin{bmatrix} C_{1,1} & & \\ & & \\ & & C_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{0,1} \\ \\ C_{0,n} \end{bmatrix}$$



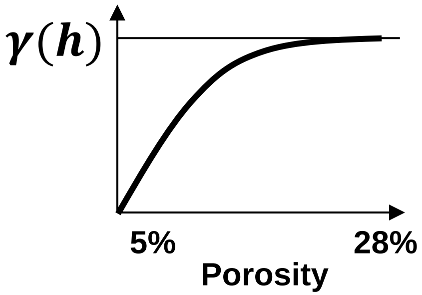
$z^*(\mathbf{u}_1) = -0.5$

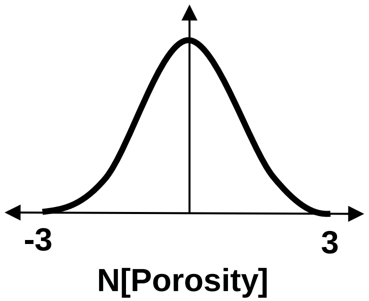


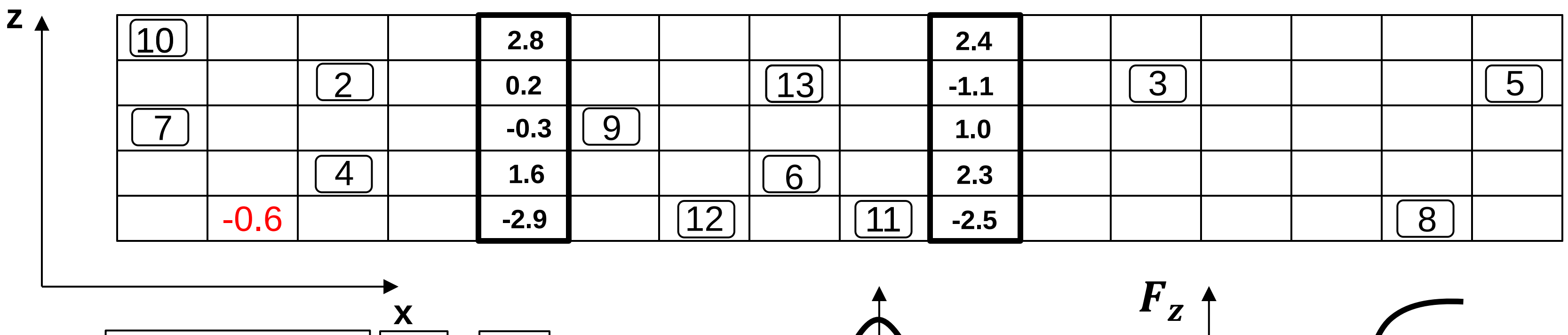
$z_s^{\ell=1}(\mathbf{u}_1) = -0.6$

Steps in Sequential Gaussian Simulation, Take 2

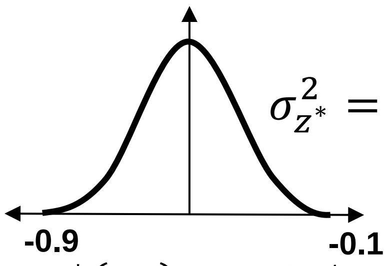
5. For each node on random path:
   a) find nearby data and previously simulated grid nodes
   b) construct the conditional distribution by kriging
   c) draw simulated value from conditional distribution
   d) assign the simulated value to the grid as data





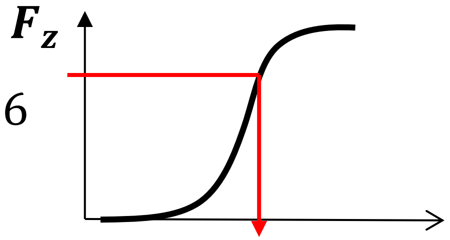


$$\begin{bmatrix} C_{1,1} & \\ & C_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{0,1} \\ C_{0,n} \end{bmatrix}$$



$\sigma^2_{z^*} = 0.16$

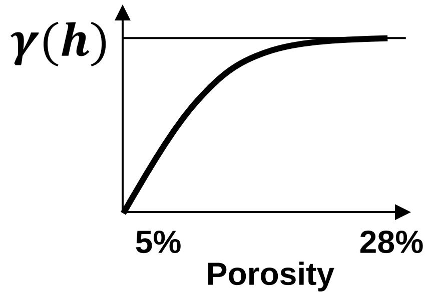
$z^*(\mathbf{u}_1) = -0.5$

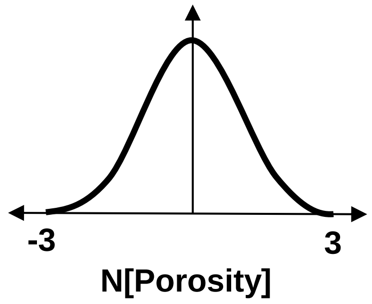


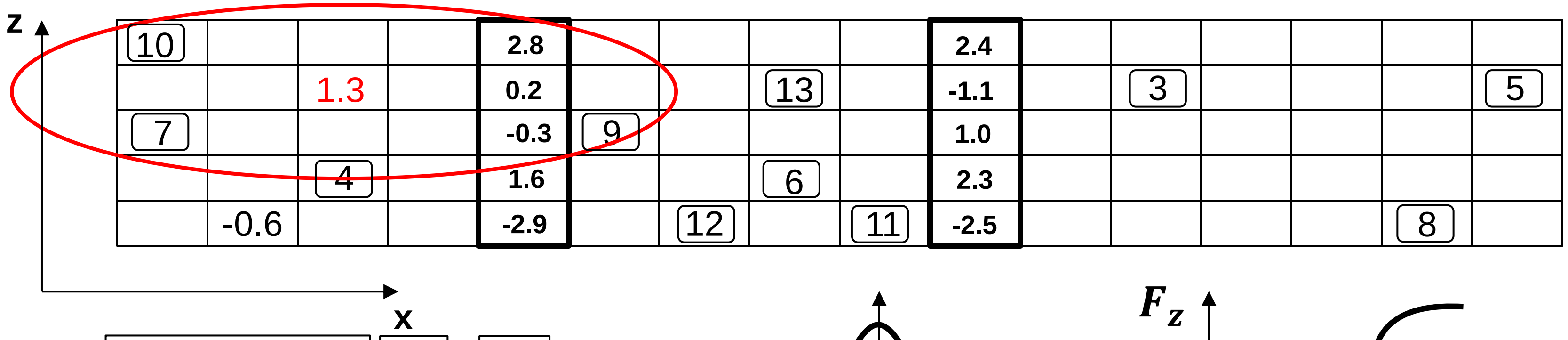
$z_s^{\ell=1}(\mathbf{u}_1) = -0.6$

Steps in Sequential Gaussian Simulation, Take 2

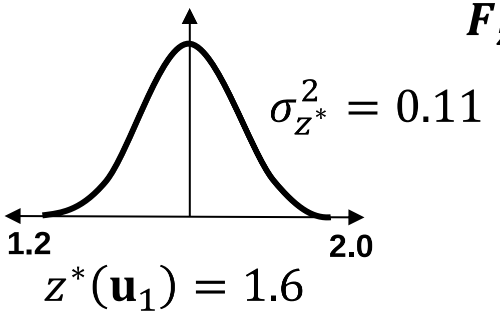
5. For each node on random path:
   a) find nearby data and previously simulated grid nodes
   b) construct the conditional distribution by kriging
   c) draw simulated value from conditional distribution
   d) assign the simulated value to the grid as data

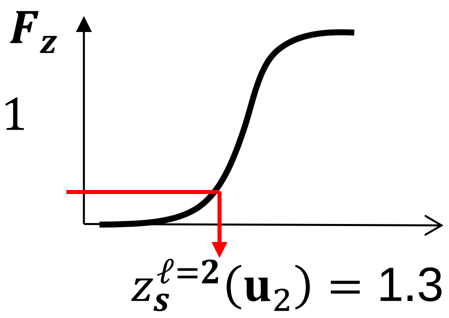






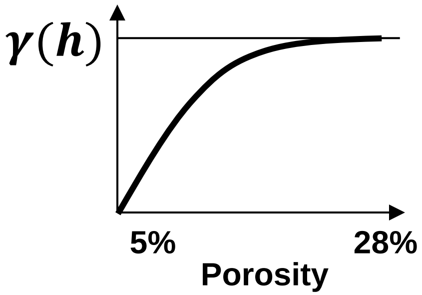
$$\begin{bmatrix} C_{1,1} & & \\ & \ddots & \\ & & C_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{0,1} \\ \vdots \\ C_{0,n} \end{bmatrix}$$

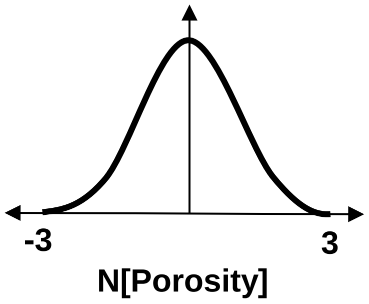


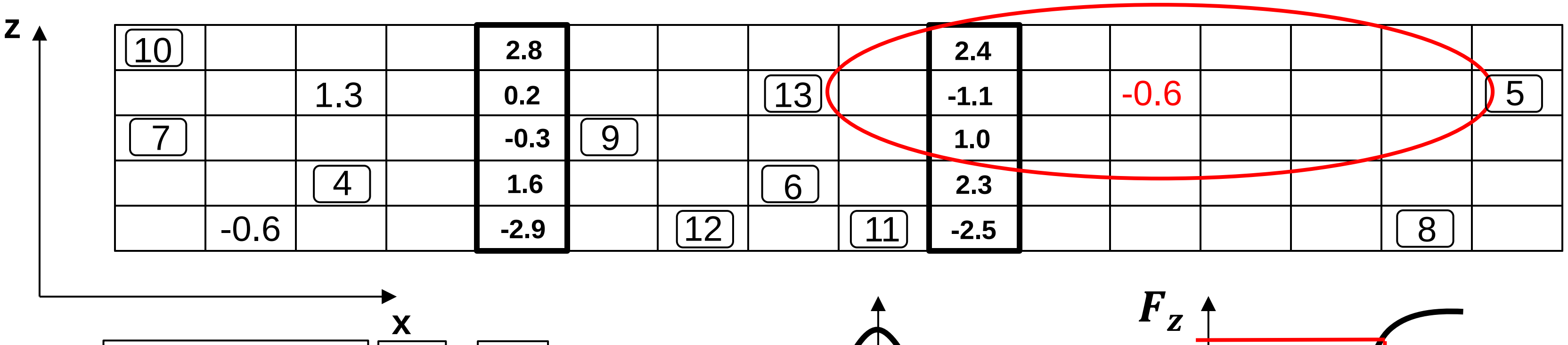


Steps in Sequential Gaussian Simulation, Take 2

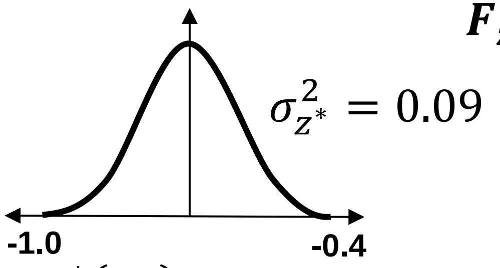
5. For each node on random path:
   a) find nearby data and previously simulated grid nodes
   b) construct the conditional distribution by kriging
   c) draw simulated value from conditional distribution
   d) assign the simulated value to the grid as data



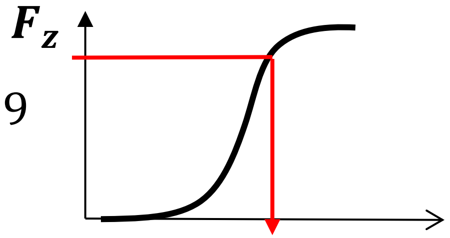




$$\begin{bmatrix} C_{1,1} & & \\ & \ddots & \\ & & C_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{0,1} \\ \vdots \\ C_{0,n} \end{bmatrix}$$



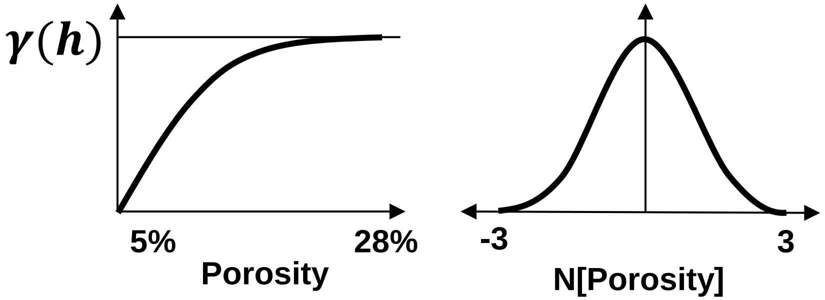
$z^*(\mathbf{u}_1) = -0.7$



$z_s^{\ell=3}(\mathbf{u}_3) =$ -0.6

Steps in Sequential Gaussian Simulation, Take 2

5. For each node on random path:
   a) find nearby data and previously simulated grid nodes
   b) construct the conditional distribution by kriging
   c) draw simulated value from conditional distribution
   d) assign the simulated value to the grid as data



z

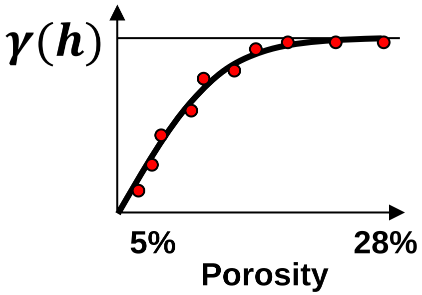
2.2				**2.8**					**2.4**						
		1.3		**0.2**			-0.5		**-1.1**		-1.3				-0.8
-1.2				**-0.3**	-0.1				**1.0**						
		-0.1		**1.6**			2.1		**2.3**						
	-0.6			**-2.9**		-2.3		-2.2	**-2.5**					-1.7	

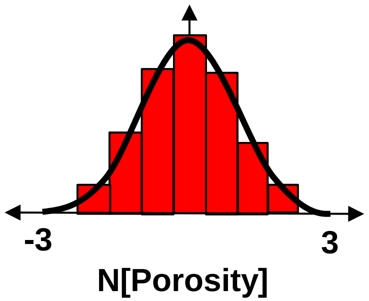
x

assume blank cells are simulated also

Steps in Sequential Gaussian Simulation, Take 2

6. Check results
   a) honor data?
   b) honor histogram: N(0,1) standard normal with a mean of zero and a variance of one?
   c) honor variogram?





z

2.2				**2.8**					**2.4**						
		1.3		**0.2**			-0.5		**-1.1**		-1.3				-0.8
-1.2				**-0.3**	-0.1				**1.0**						
		-0.1		**1.6**			2.1		**2.3**						
	-0.6			**-2.9**		-2.3		-2.2	**-2.5**					-1.7	

x

assume blank cells are simulated also

Steps in Sequential Gaussian Simulation, Take 2

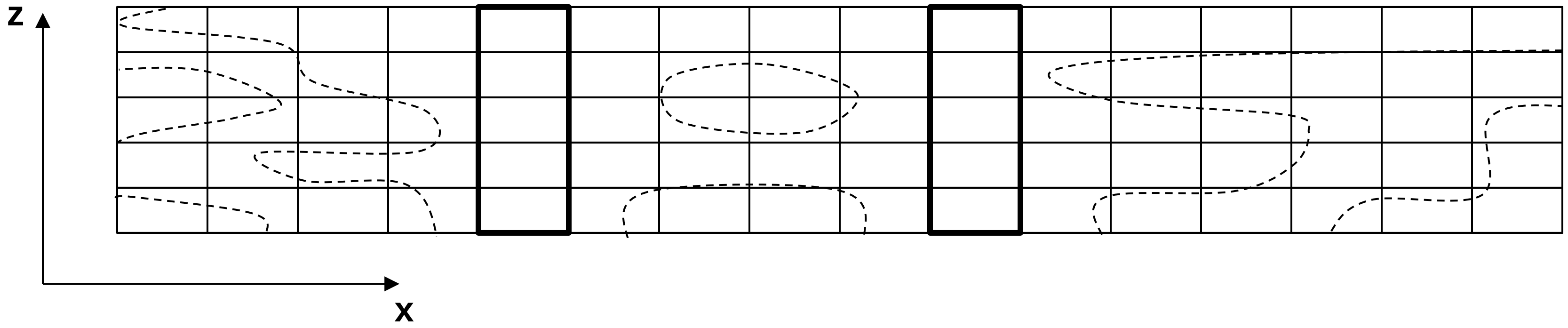
7. Back transform from “normal space”

0.21				**0.23**					**0.21**						
		0.20		**0.17**			0.16		**0.15**		0.13				0.15
0.13				**0.16**	0.18				**0.18**						
		0.18		**0.18**			0.22		**0.20**						
	0.15			**0.12**		0.12		0.14	**0.13**					0.14	

z, x

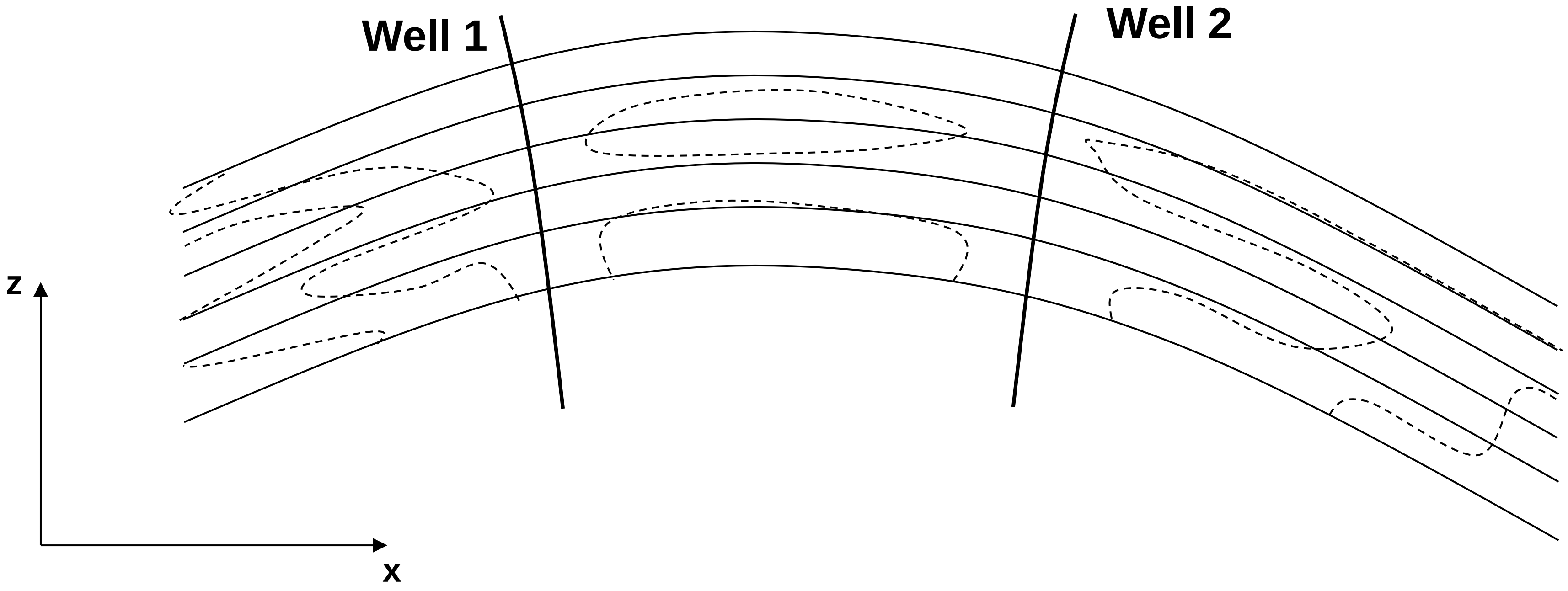
assume blank cells are simulated also

Reservoir Property Model



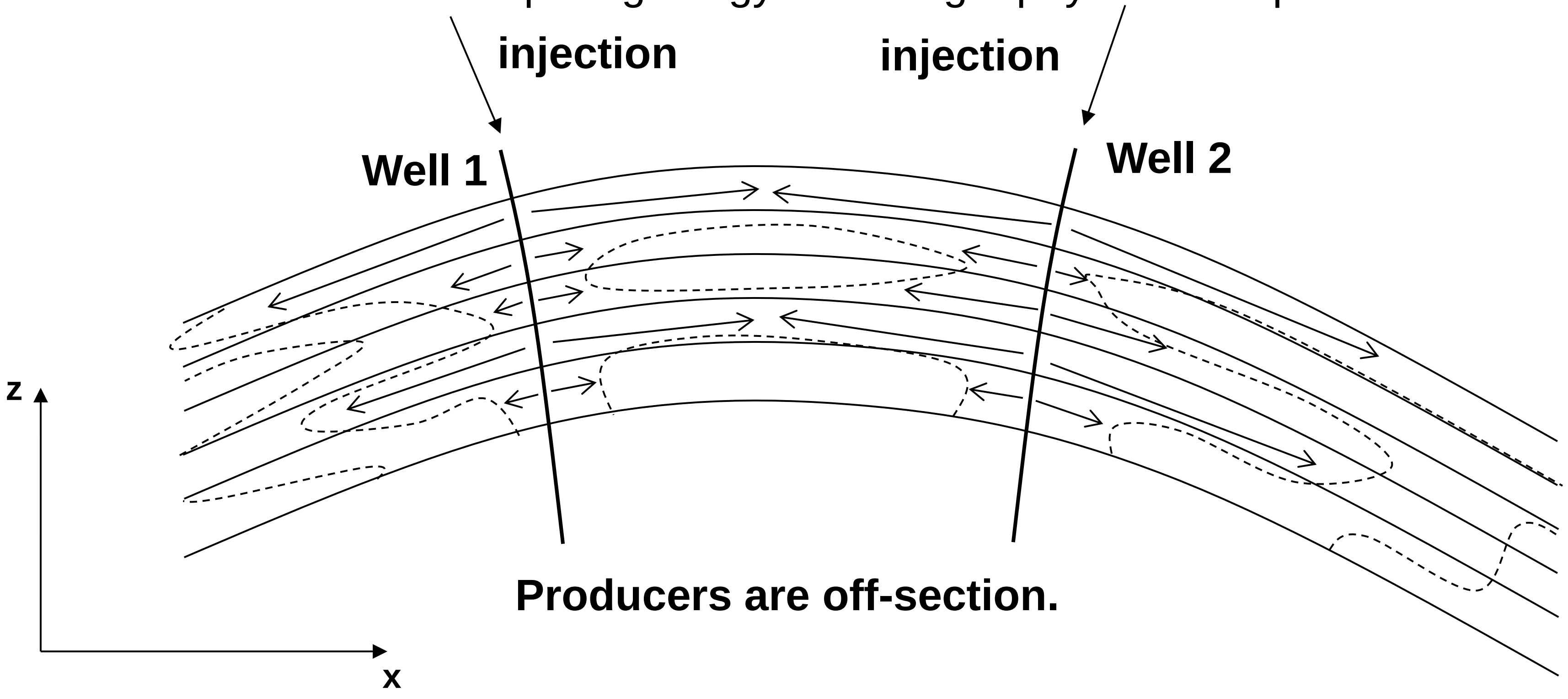
Steps in Sequential Gaussian Simulation, Take 2

8. Restore to original framework



Steps in Sequential Gaussian Simulation, Take 2

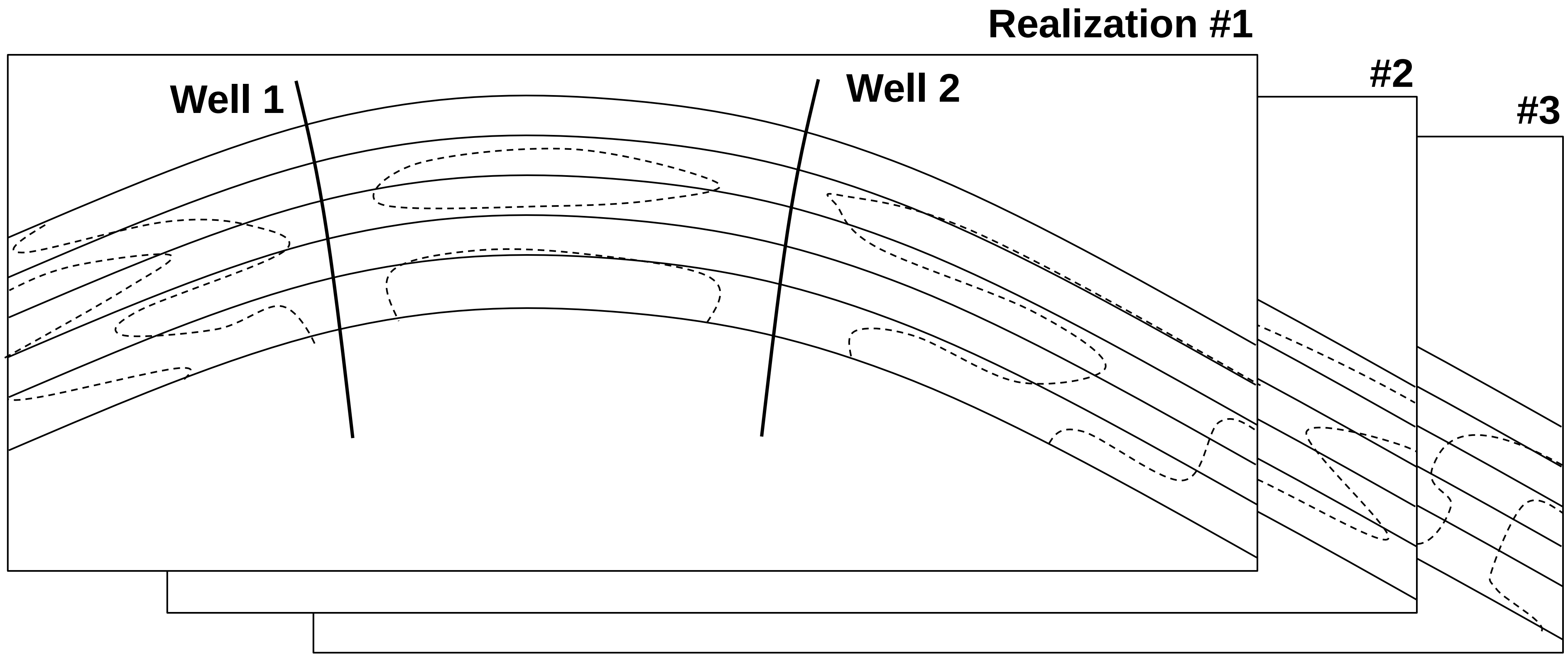
9. Check honor concept of geology? Match geophysical and production data?



Producers are off-section.

Steps in Sequential Gaussian Simulation, Take 2

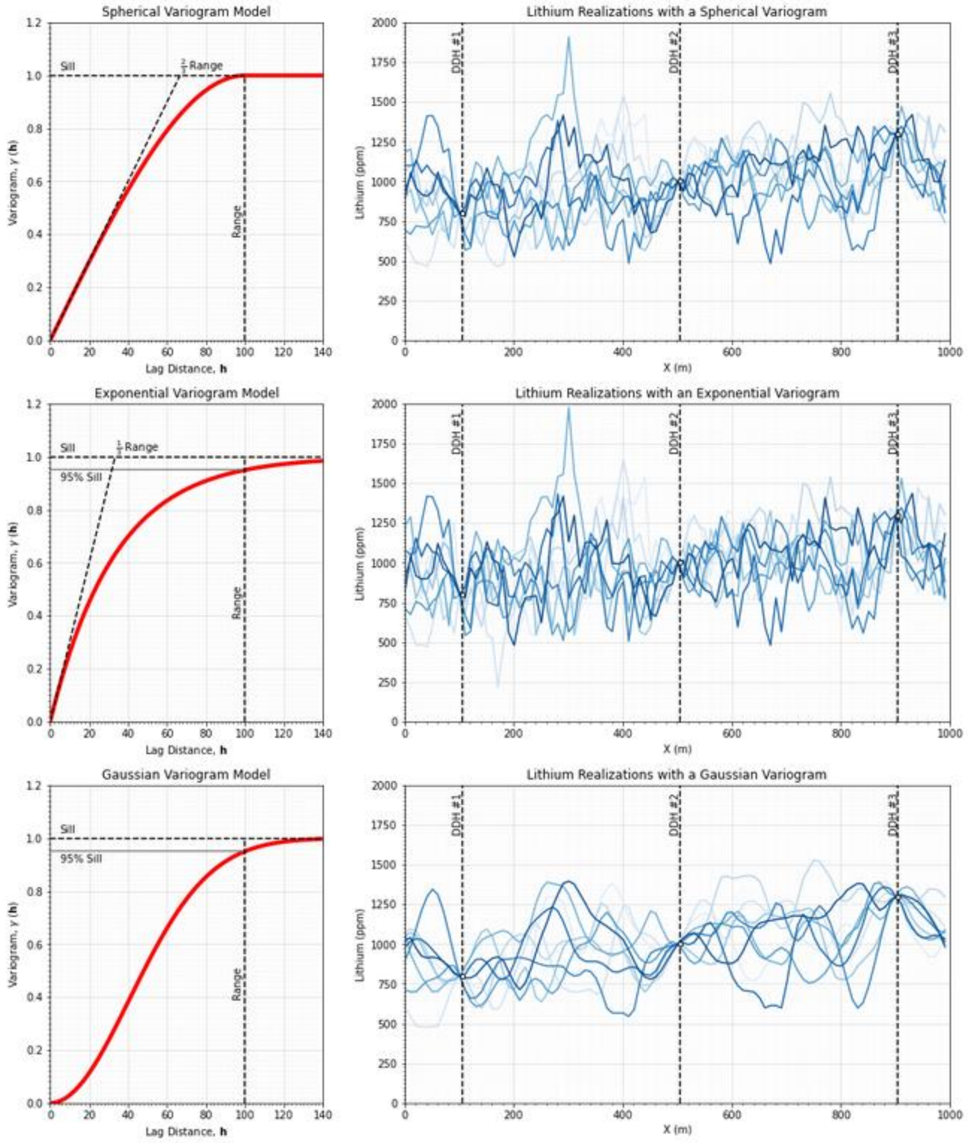
10. Calculate multiple realizations to represent uncertainty.



Some SGSIM Realizations

Here are some realizations with sequential Gaussian simulation for the cases of:

- spherical variogram
- exponential variogram
- Gaussian variogram

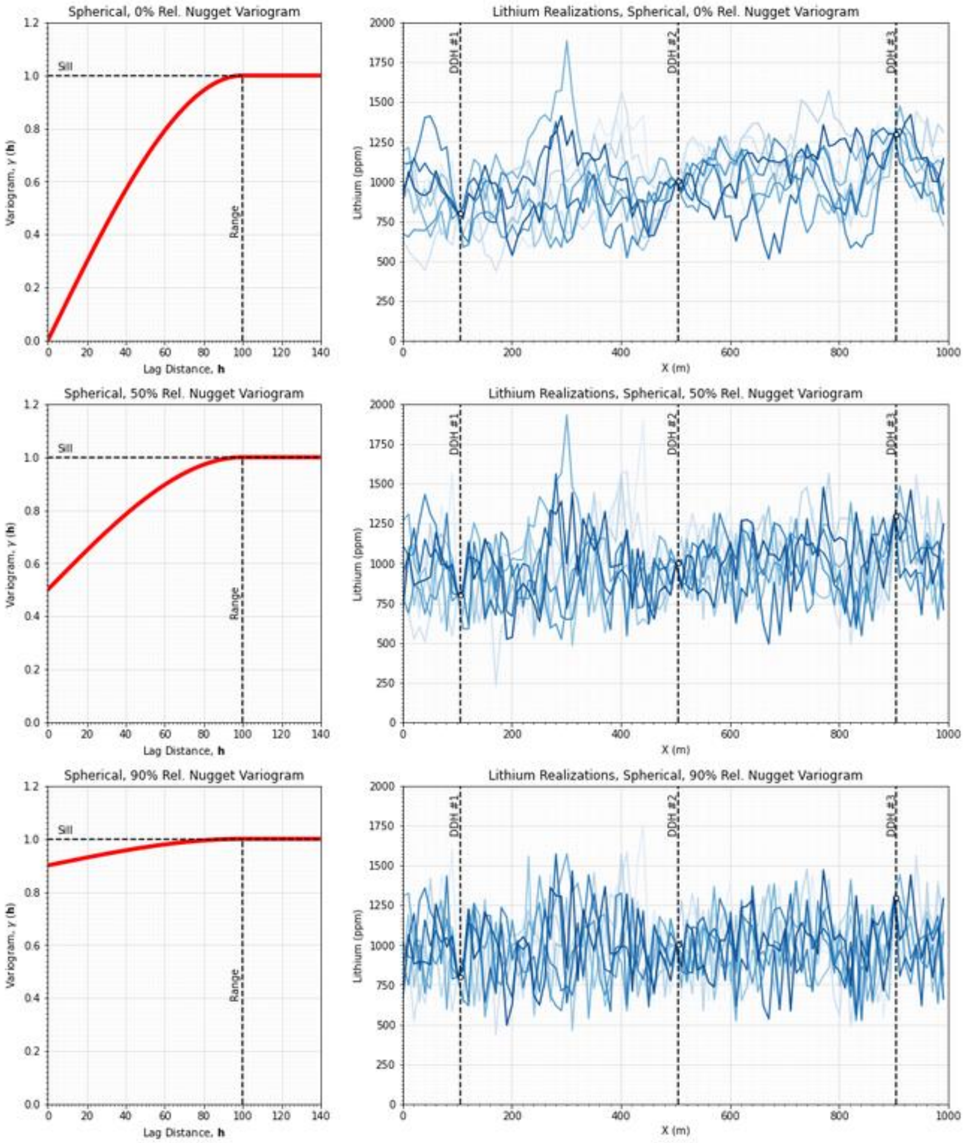


Simulations with different variogram model types, file is GeostatsPy_variogram_models_simulation.ipynb.

Some SGSIM Realizations

Here are some realizations with sequential Gaussian simulation for the cases of:

- spherical variogram
- spherical variogram, 50% relative nugget efffect
- spherical variogram, 90% relative nugget efffect

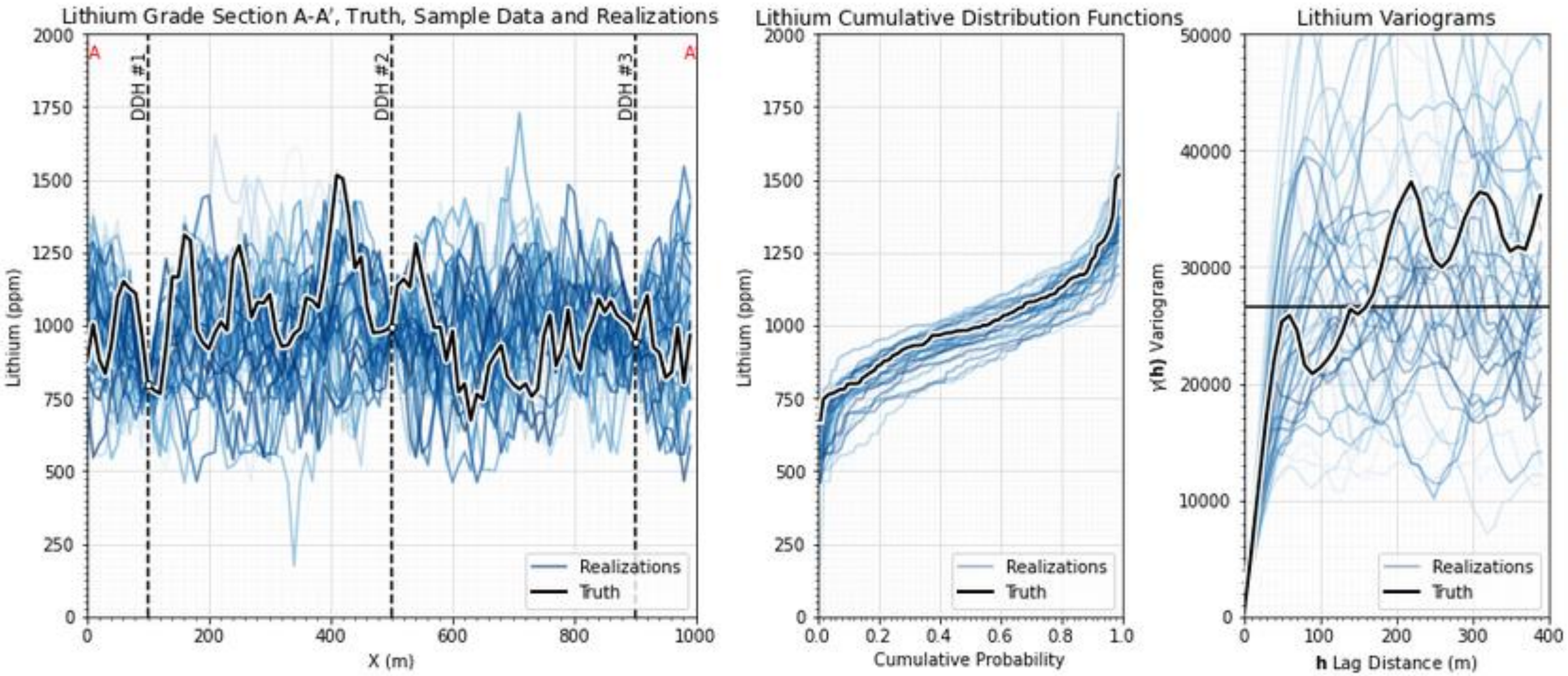


Simulations with different relative nugget effect, file is GeostatsPy_variogram_models_simulation.ipynb.

Ergodic Fluctuations

Expect some statistical fluctuation in the input statistics

- These are a function of the ratio of spatial continuity to the size of the model.
  - If model is large relative to spatial continuity range then fluctuations should minimal
  - If model is small relative to spatial continuity range then fluctuations may be extreme



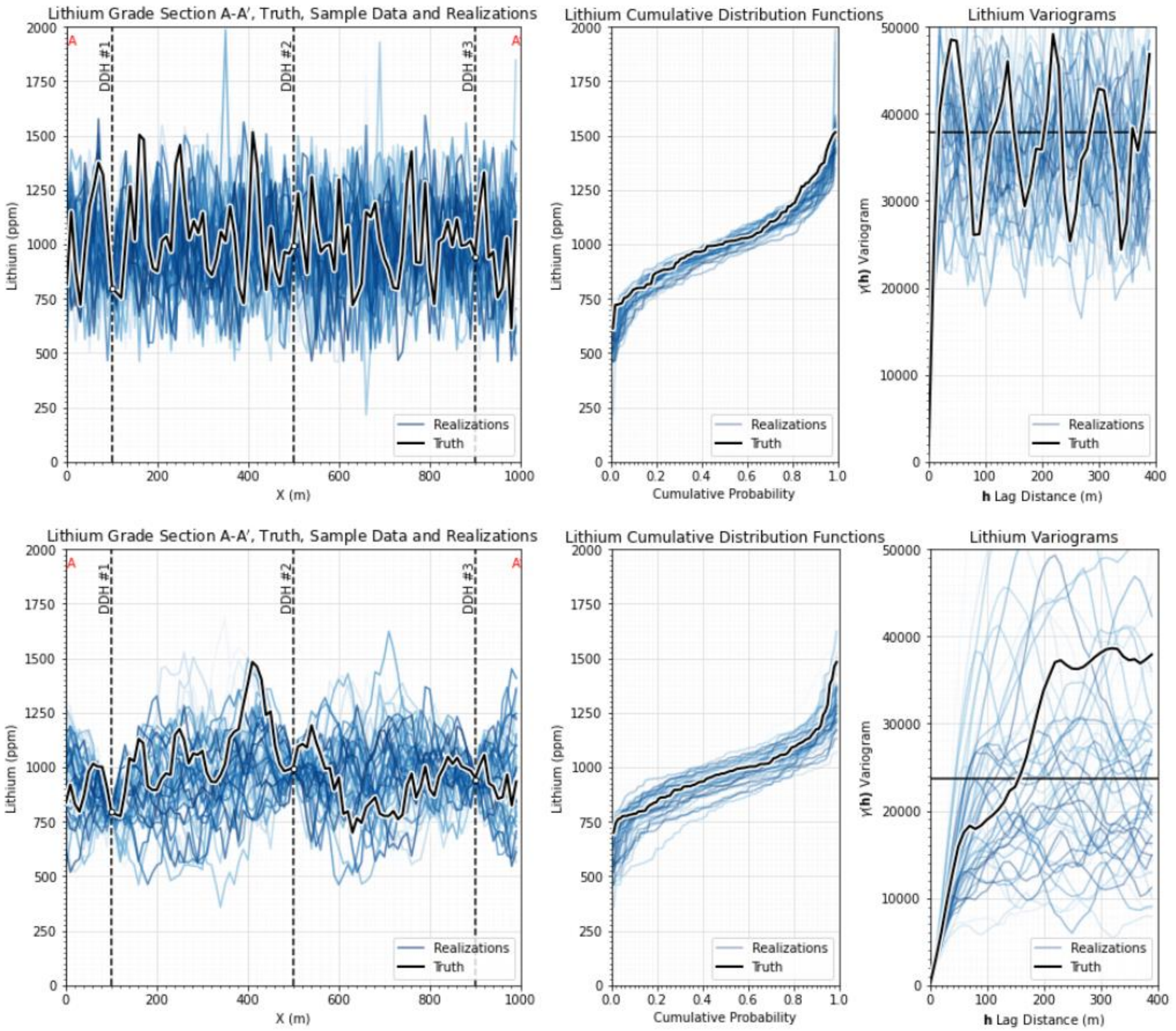
Lithium grades truth, samples, and realizations (left), cumulative distribution functions (center), and variograms (right), file is GeostatsPy_ergodic_fluctuations.ipynb.

Ergodic Fluctuations

These are a function of the ratio of spatial continuity to the size of the model.

Here we have a case with:

- Above: 20 m variogram range over 1,000 m AOI
- Below: 200 m variogram range over 1,000 m AOI

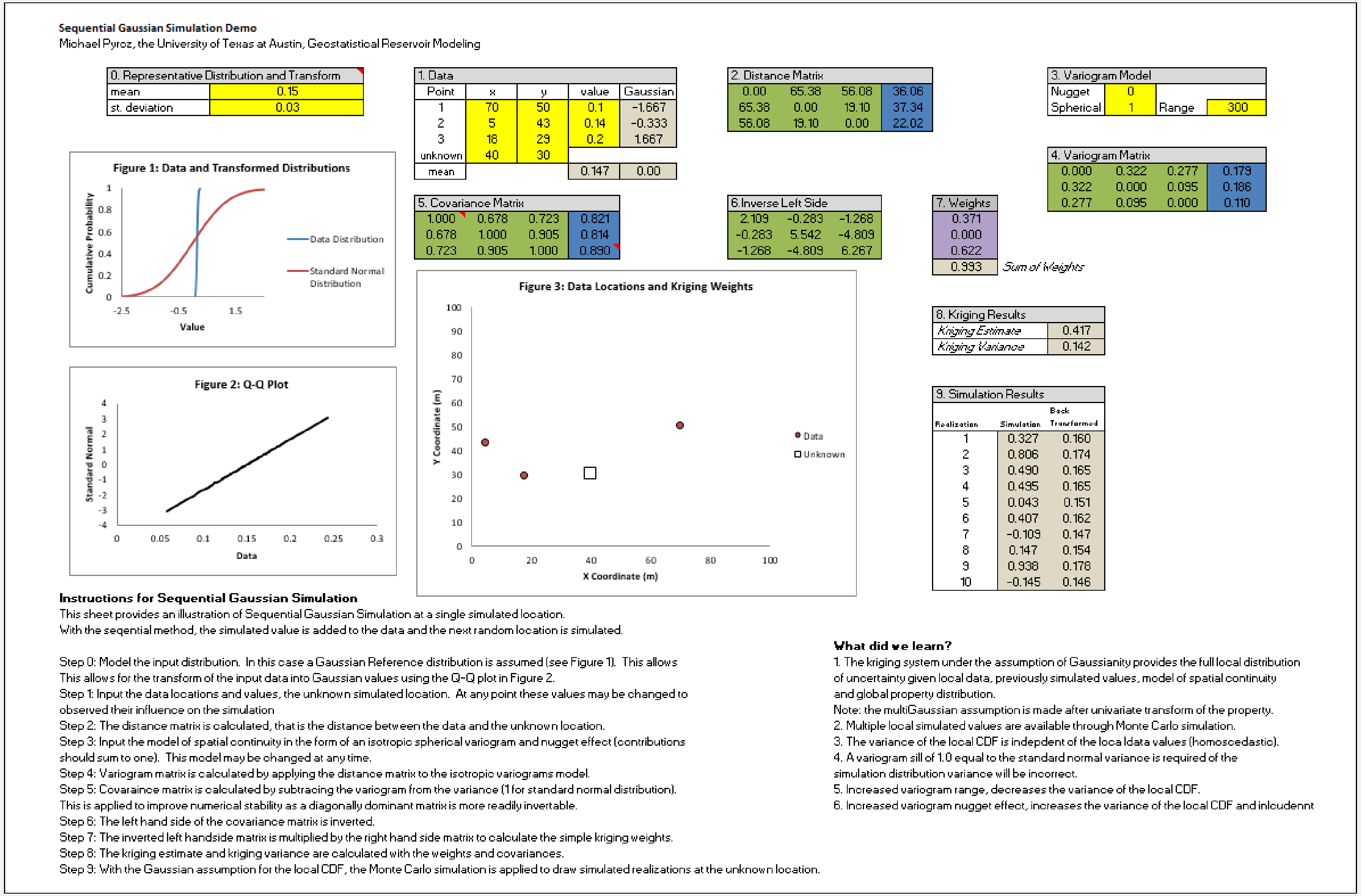


Lithium grades truth, samples, and realizations, cumulative distribution functions, and variograms for 20 m variogram range (above) and 200 m variogram range (below), file is GeostatsPy_ergodic_fluctuations.ipynb.

Gaussian Simulation Hands On in Excel

Demonstration of Gaussian simulation in Excel

- only for a single location, does not demonstrate random path, nor sequential simulation.



Sequential Gaussian Simulation Demo

Michael Pyrcz, the University of Texas at Austin, Geostatistical Reservoir Modeling

0. Representative Distribution and Transform	
mean	0.15
st. deviation	0.03

1. Data				
Point	x	y	value	Gaussian
1	70	50	0.1	-1.667
2	5	43	0.14	-0.333
3	18	29	0.2	1.667
unknown	40	30		
mean			0.147	0.00

2. Distance Matrix			
0.00	65.38	56.08	36.06
65.38	0.00	19.10	37.34
56.08	19.10	0.00	22.02

3. Variogram Model			
Nugget	0		
Spherical	1	Range	300

4. Variogram Matrix			
0.000	0.322	0.277	0.179
0.322	0.000	0.095	0.186
0.277	0.095	0.000	0.110

5. Covariance Matrix			
1.000	0.678	0.723	0.821
0.678	1.000	0.905	0.814
0.723	0.905	1.000	0.890

6.Inverse Left Side		
2.109	-0.283	-1.268
-0.283	5.542	-4.809
-1.268	-4.809	6.267

7. Weights	
0.371	
0.000	
0.622	
0.993	*Sum of Weights*

8. Kriging Results	
Kriging Estimate	0.417
Kriging Variance	0.142

9. Simulation Results		
Realization	Simulation	Back Transformed
1	0.327	0.160
2	0.806	0.174
3	0.490	0.165
4	0.495	0.165
5	0.043	0.151
6	0.407	0.162
7	-0.109	0.147
8	0.147	0.154
9	0.938	0.178
10	-0.145	0.146

Instructions for Sequential Gaussian Simulation

This sheet provides an illustration of Sequential Gaussian Simulation at a single simulated location.
With the seqential method, the simulated value is added to the data and the next random location is simulated.

Step 0: Model the input distribution. In this case a Gaussian Reference distribution is assumed (see Figure 1). This allows
This allows for the transform of the input data into Gaussian values using the Q-Q plot in Figure 2.
Step 1: Input the data locations and values, the unknown simulated location. At any point these values may be changed to
observed their influence on the simulation
Step 2: The distance matrix is calculated, that is the distance between the data and the unknown location.
Step 3: Input the model of spatial continuity in the form of an isotropic spherical variogram and nugget effect (contributions
should sum to one). This model may be changed at any time.
Step 4: Variogram matrix is calculated by applying the distance matrix to the isotropic variograms model.
Step 5: Covaraince matrix is calculated by subtracing the variogram from the variance (1 for standard normal distribution).
This is applied to improve numerical stability as a diagonally dominant matrix is more readily invertable.
Step 6: The left hand side of the covariance matrix is inverted.
Step 7: The inverted left handside matrix is multiplied by the right hand side matrix to calculate the simple kriging weights.
Step 8: The kriging estimate and kriging variance are calculated with the weights and covariances.
Step 9: With the Gaussian assumption for the local CDF, the Monte Carlo simulation is applied to draw simulated realizations at the unknown location.

What did we learn?

1. The kriging system under the assumption of Gaussianity provides the full local distribution of uncertainty given local data, previously simulated values, model of spatial continuity and global property distribution.
Note: the multiGaussian assumption is made after univariate transform of the property.
2. Multiple local simulated values are available through Monte Carlo simulation.
3. The variance of the local CDF is indepdent of the loca ldata values (homoscedastic).
4. A variogram sill of 1.0 equal to the standard normal variance is required of the simulation distribution variance will be incorrect.
5. Increased variogram range, decreases the variance of the local CDF.
6. Increased variogram nugget effect, increases the variance of the local CDF and inlcudennt

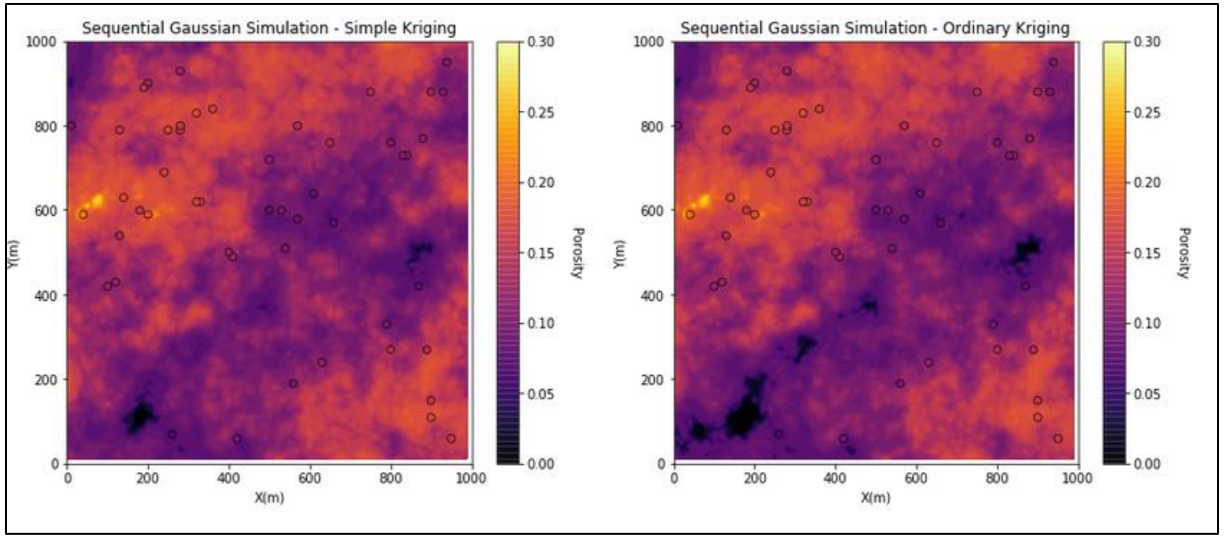
Simulation at a single location in Excel, file is Sequential_Gaussian_Simulation.ipynb.

Sequential Gaussian Simulation in Python

Sequential Simulation Workflow in Python

Walkthrough and try to:

- Change the variogram and search parameters.



Sequential Gaussian Simulation for Geostatistical Realizations

Michael Pyrcz, Associate Professor, University of Texas at Austin

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

Here's a simple workflow for spatial simulation with sequential Gaussian simulation. This step is critical for:

1. Prediction away from wells, e.g. pre-drill assessments, with uncertainty
2. Spatial uncertainty modeling.
3. Heterogeneity realizations ready for application to the transfer function.

First let's explain the concept of spatial simulation.

Spatial Simulation

For more information see this lecture on geostatistical simulation and sequential Gaussian simulation.

- Simulation

Getting Started

Here's the steps to get setup in Python with the GeostatsPy package:

1. Install Anaconda 3 on your machine (https://www.anaconda.com/download/).
2. From Anaconda Navigator (within Anaconda3 group), go to the environment tab, click on base (root) green arrow and open a terminal.
3. In the terminal type: pip install geostatspy.
4. Open Jupyter and in the top block get started by copy and pasting the code block below from this Jupyter Notebook to start using the geostatspy functionality.

You will need to copy the data file to your working directory. They are available here:

- Tabular data - sample_data_MV_biased.csv available at https://git.io/fhgu0.

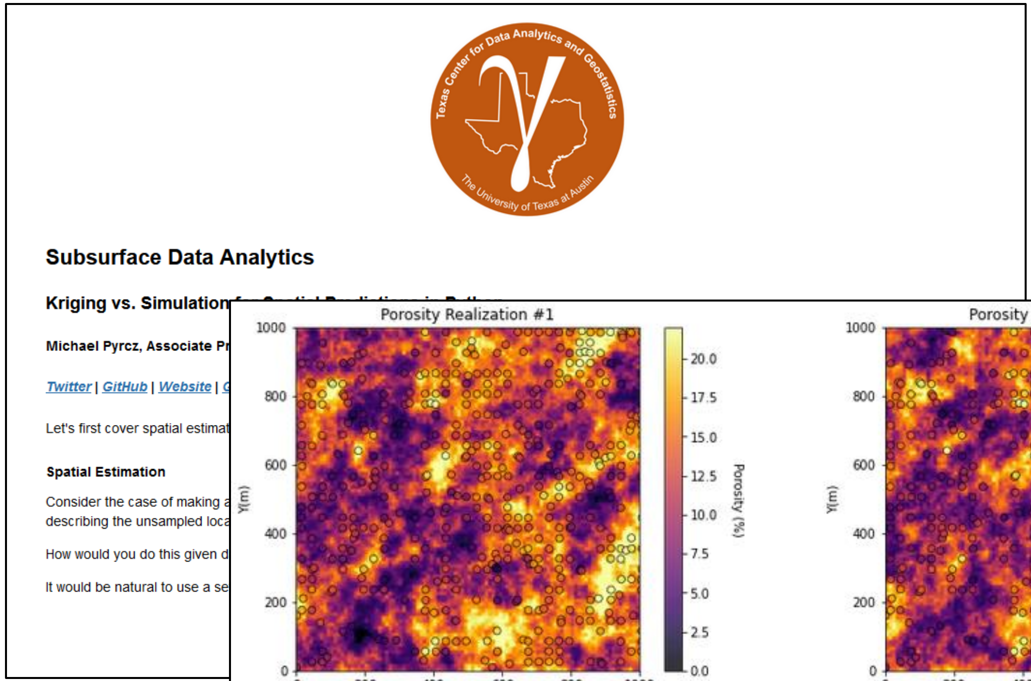
Simulation of a 2D porosity model with Python, file is GeostatsPy_simulation.ipynb.

Kriging vs. Simulation in Python

Sequential Simulation Workflow in Python

Walkthrough and try to:

- Observe the distribution and variogram reproduction of kriging and simulation.
- Change the variogram and repeat.


Subsurface Data Analytics

Kriging vs. Simulation

Michael Pyrcz, Associate Pr

Twitter | *GitHub* | *Website* |

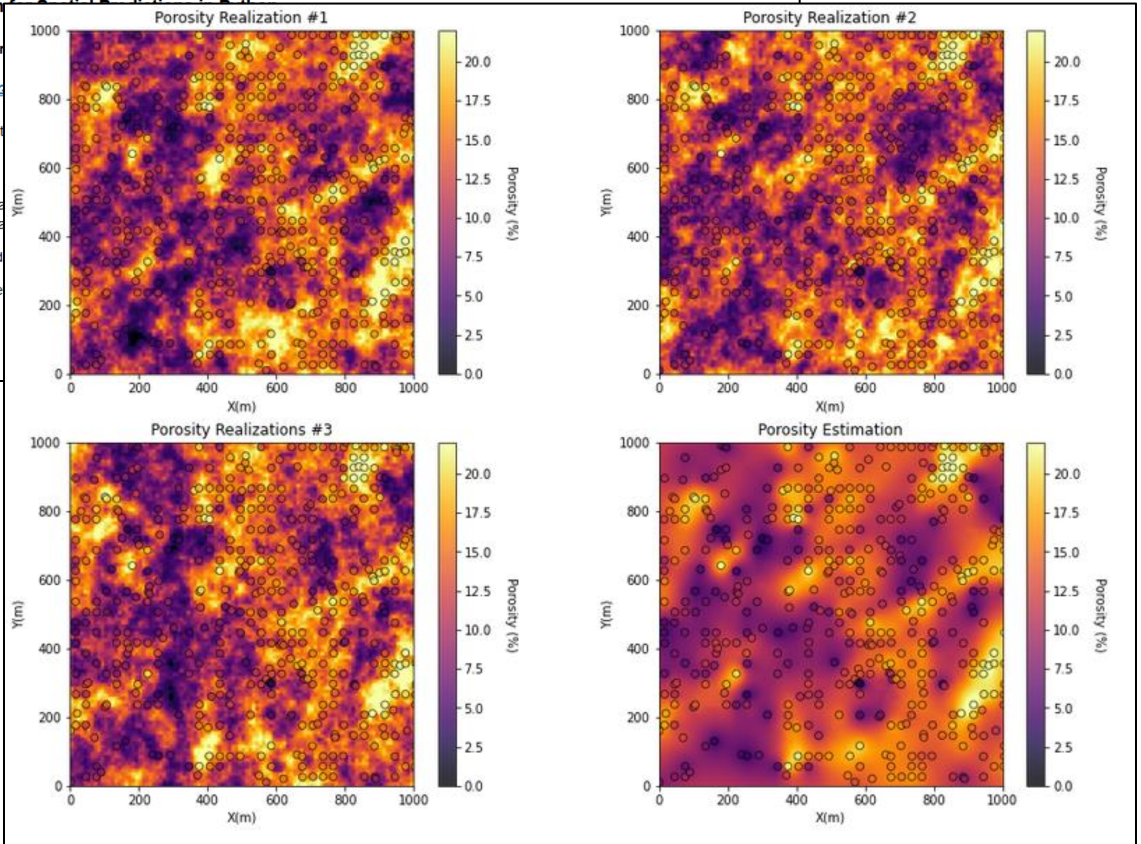
Let's first cover spatial estimat

Spatial Estimation

Consider the case of making a
describing the unsampled loca

How would you do this given d

It would be natural to use a se



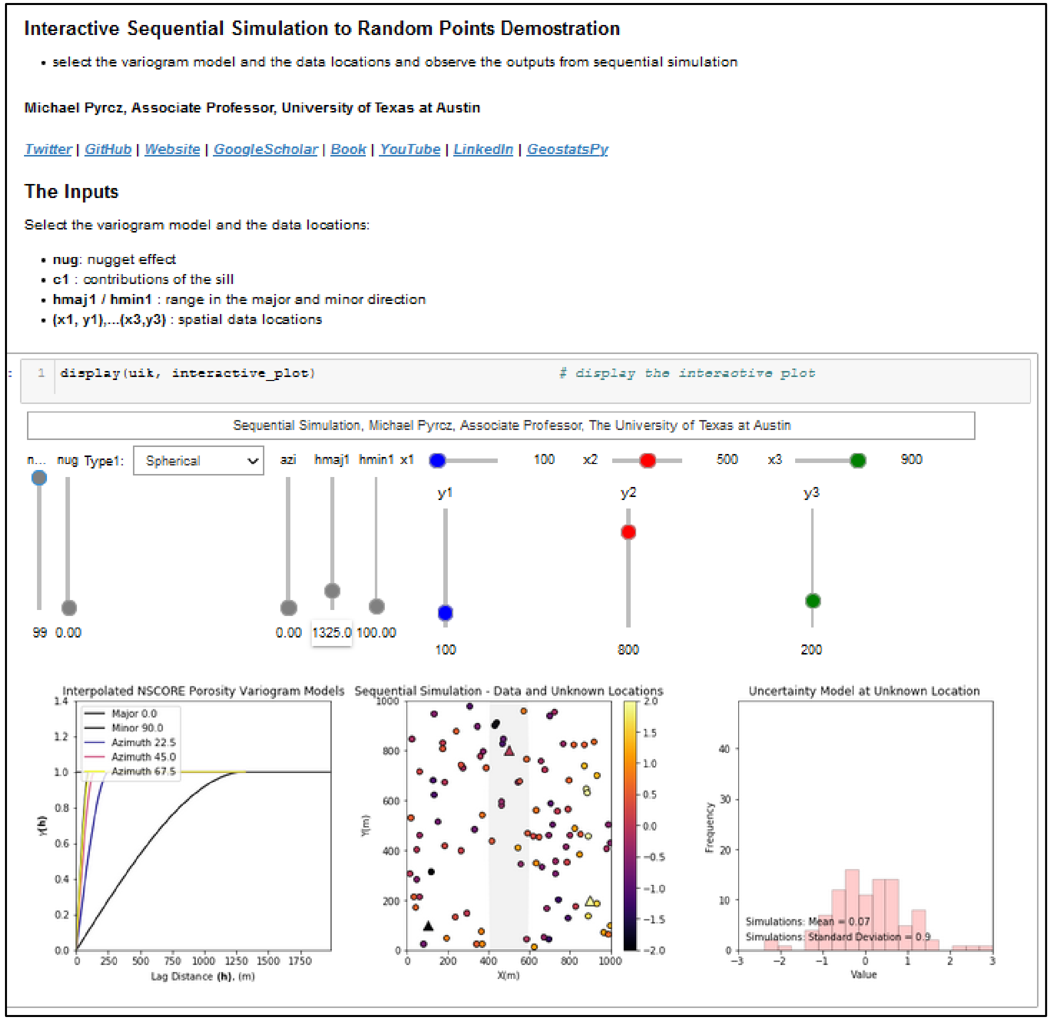
Kriging vs. sequential Gaussian simulation of a 2D porosity model with Python, file is GeostatsPy_kriging_vs_simulation.ipynb.

Interactive Simulation Python

Simulation Workflow in Python

Walkthrough and try to:

- Change the variogram
- Change the data locations
- Change the number of simulated locations



Interactive sequential Gaussian simulation demonstration, file is Interactive_Simulation.ipynb.

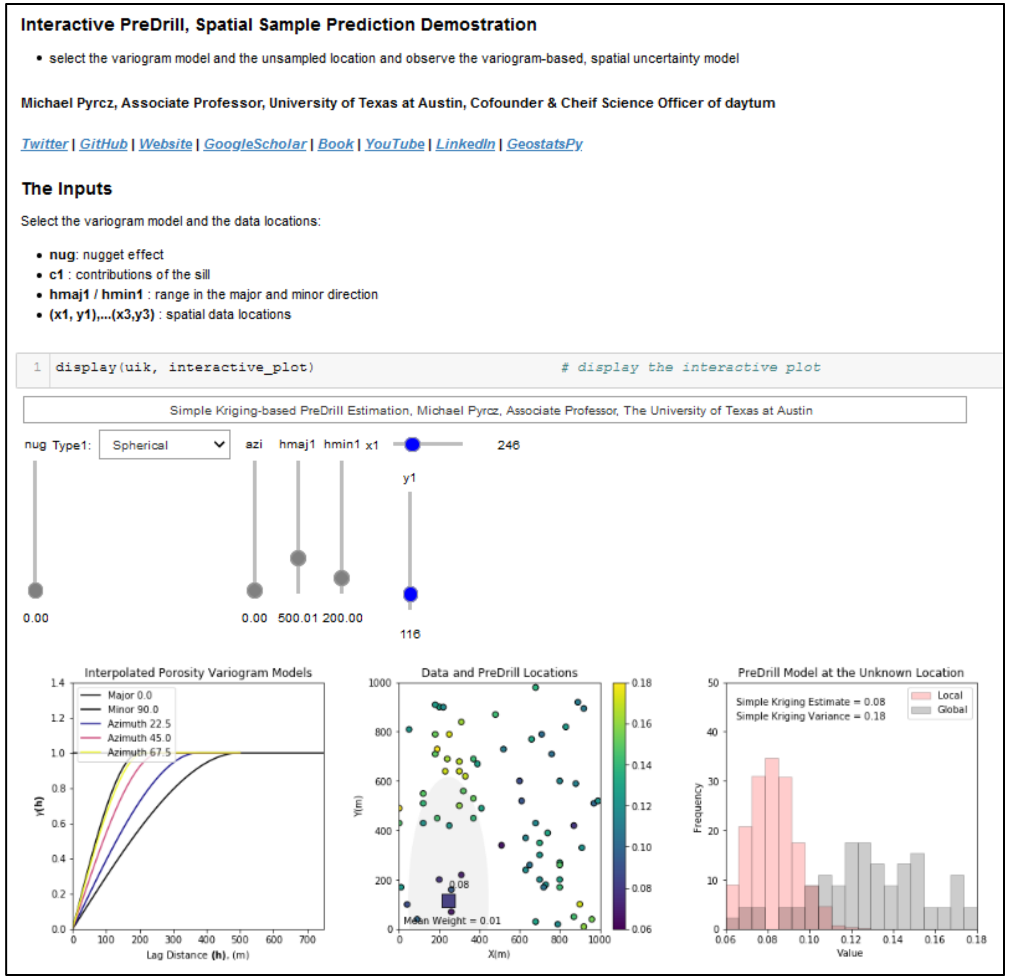
Simulation in Python

Simulation Workflow in Python for Pre-Drill Uncertainty Modelling

Walkthrough and try to:

- Change the variogram.

- Change the predrill location.



Interactive pre-drill uncertainty model with simulation, file is Interactive_PreDrill_Prediction.ipynb.

Bayesian Perspective on Sequential Simulation

- Consider N nodes in a simulation model, $A_{i,}\ i = 1, \dots, N$

- We need to sample a realization from the joint distribution, $P(A_1, \dots, A_1)$, this would not be possible, it is typically a vast solution space. We use Bayes to make it practical:

  From recursive application of Bayes law:

$$P(A_1, \dots, A_1) = P(A_N|A_1, \dots, A_{N-1}) \cdot P(A_1, \dots, A_{N-1})$$

$$P(A_N|A_1, \dots, A_{N-1}) \cdot P(A_{N-1}|A_1, \dots, A_{N-2}) \cdot P(A_1, \dots, A_{N-2})$$

$$P(A_N|A_1, \dots, A_{N-1}) \cdot P(A_{N-1}|A_1, \dots, A_{N-2}) \cdot \cdots \cdot P(A_2|A_1) \cdot P(A_1)$$

- Now we can proceed sequentially to jointly simulate the *N* events A_j:
  - Draw A_1 from the marginal, $P(A_1)$
  - Draw A_2 from the conditional, $P(A_2|A_1 = a_1)$
  - Draw A_3 from the conditional, $P(A_3|A_1 = a_1, A_2 = a_2)$
  - …
  - Draw A_N from the conditional, $P(A_N|A_1 = a_1, A_2 = a_2, \dots, A_{N-1} = a_{N-1})$
- This is theoretically valid with no approximations/assumptions

PGE 338 Data Analytics and Geostatistics

Lecture 13: Simulation

Lecture outline . . .

- Simulation

Michael Pyrcz, The University of Texas at Austin