Multivariate Modeling:
Multivariate Workflows

Lecture outline . . .

- **Cosimulation**
- **Spatial Updating**

Multivariate Modeling:
Multivariate Workflows

Lecture outline . . .

- Cosimulation

Introduction

Prerequisites

Probability

Multivariate Analysis

Spatial Estimation

Feature Selection

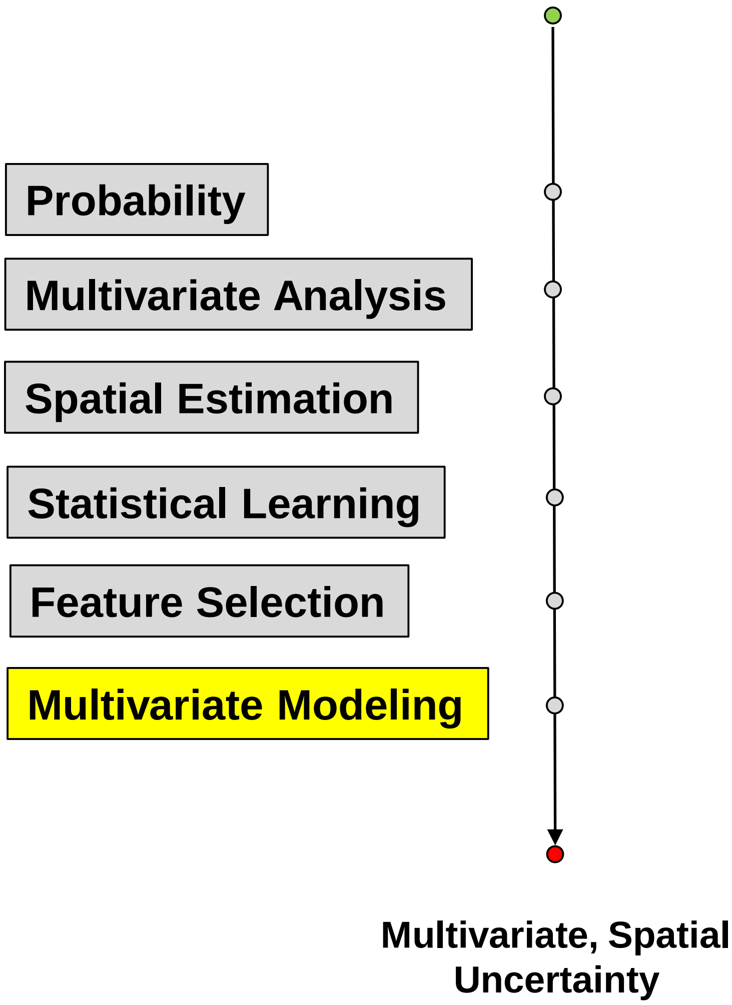
Multivariate Modeling

Conclusions

What Will You Learn?

Why Cover Multivariate Modeling?

- Introduce cokriging / cosimulation approaches.
- Demonstration of a multivariate modeling workflow.



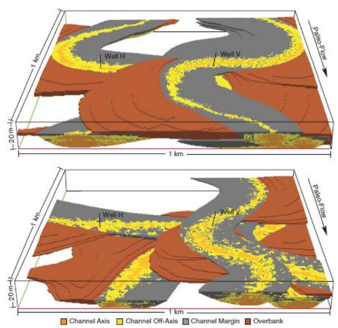
Flashback: A Common Modeling Workflow

- Facies Categorical, Porosity Continuous then Permeability Cosimulation – here's Some Context!

Categorical Simulation

Models Facies PDF, Trends and Spatial Continuity

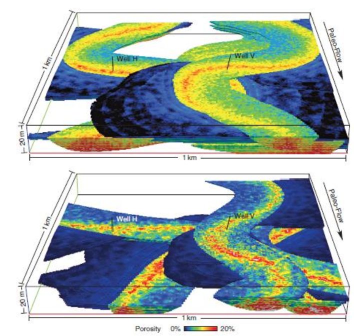
Multiple Facies Realizations



Continuous Simulation

Models Porosity PDF, Trends and Spatial Continuity within each facies

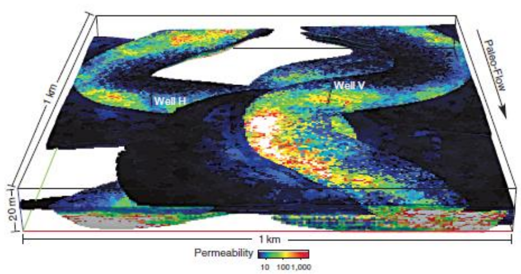
Multiple Porosity By-facies Realizations



Continuous Cosimulation

Models Permeability PDF, Trends and Spatial Continuity and Porosity Relationship within each facies

Multiple Permeability By-facies Realizations Correlated with Porosity



Note: only had 1 realization on hand (should be two in figure).

Motivation for Cosimulation

- We typically need to build reservoir models of more than one property of interest.

- For example a typical workflow is:

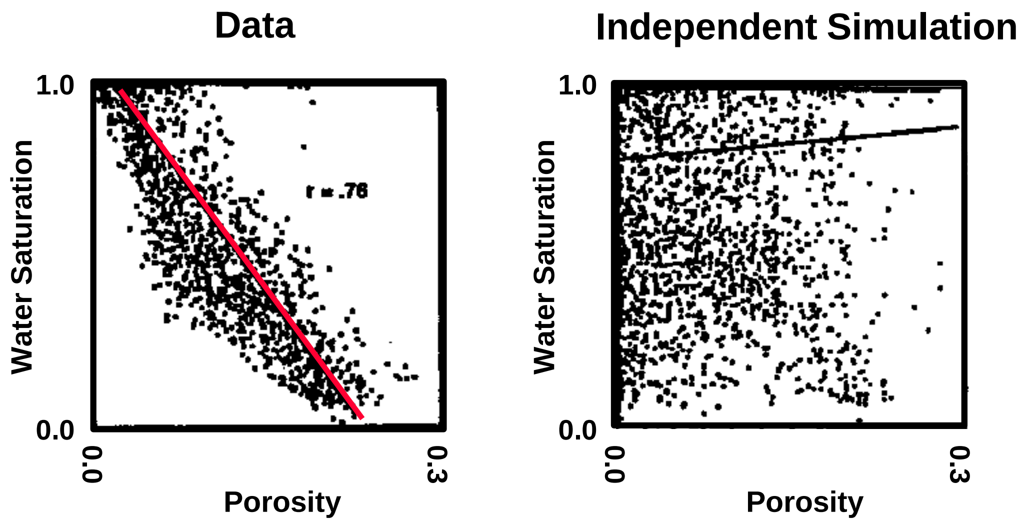
  – Build facies models realizations
  – Within the facies realizations simulate the porosity
  – Within the facies realizations cosimulate the permeability (primary variable) correlated with the previously simulated porosity realization (secondary variable)

- We will only cover the commonly applied cosimulations methods

  – collocated cokriging and cloud transform
  – we limit ourselves to simulating one property correlated to one other (bivariate)

Motivation for Cosimulation

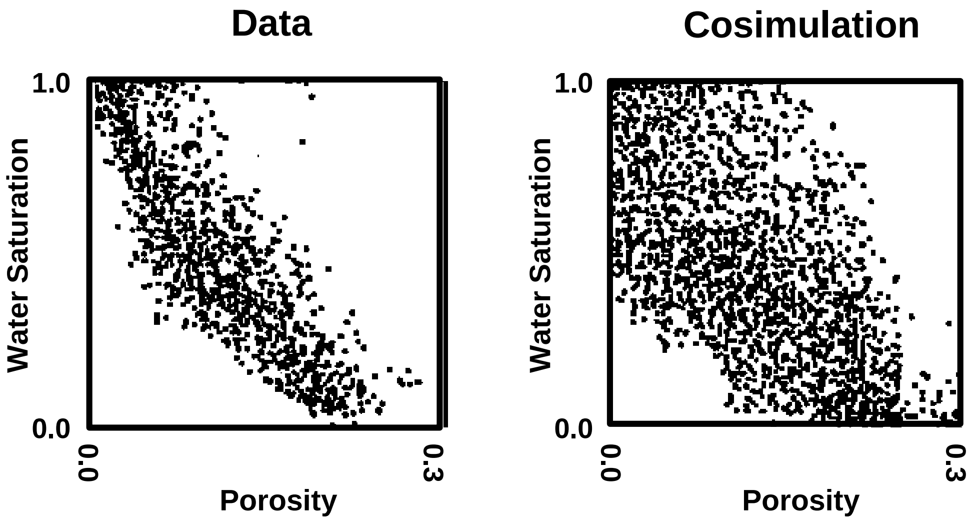
- Cosimulation is not perfect, but the alternative is to simulate each property independently



- No correlation or some minor correlation due to data conditioning (correlation at the data locations get propagated away in each variable). This would result in implausible combinations away from data and too high uncertainty. The variables have information about each other.

Motivation for Cosimulation

- Cosimulation is not perfect, but the alternative is to simulate each property independently

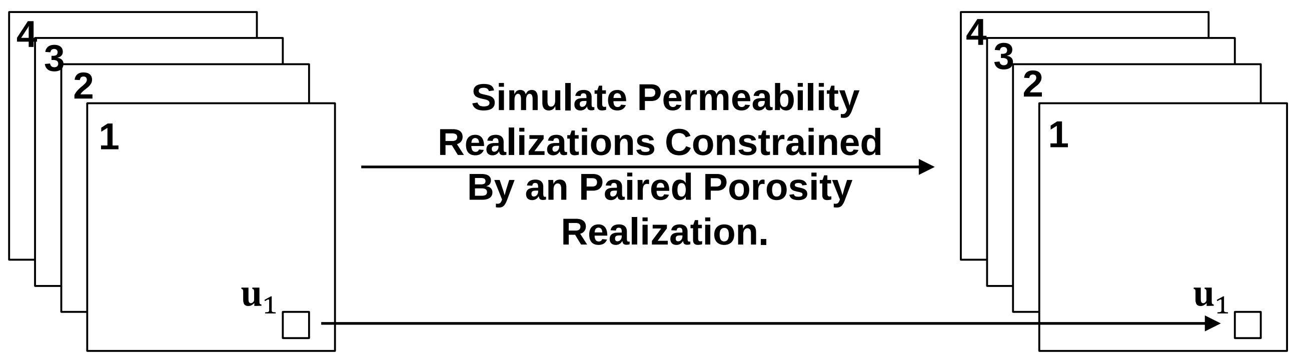


- An example with some reproduction of bivariate features based on collocated cokriging
- Capture the general trend with this method. Much improved.

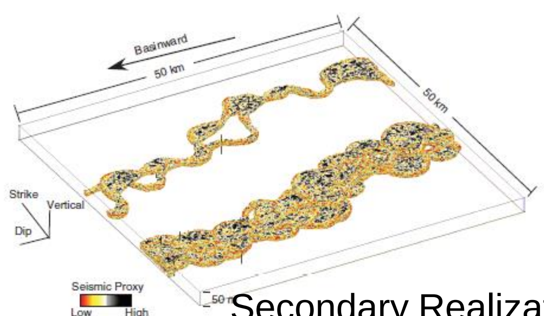
Cosimulation

- Cosimulation is a simulation method that imposes correlation with a previously simulated property.
  - The realizations are paired.

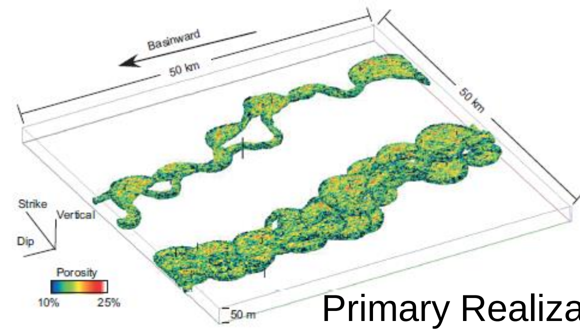
Porosity Realizations **Permeability Realizations**



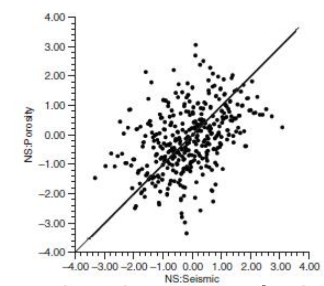
Honor the bivariate relationship between pair values at the same locations, $\mathbf{u}_\alpha$



Secondary Realization



Primary Realization



Bivariate Realationship

Limitations of Cosimulation

- Each method will have a priority
  - **Collocated Cokriging** prioritizes the histogram and variogram and may honor the correlation coefficient between the two variables
  - **Cloud transform** will honor the specific form of the bivariate relationship (cloud) between the two variables, but may not honor the histogram nor the variogram.
- Both of these methods start with a completed realization of the secondary variable
  - e.g. porosity, if we are cosimulated permeability constrained by porosity

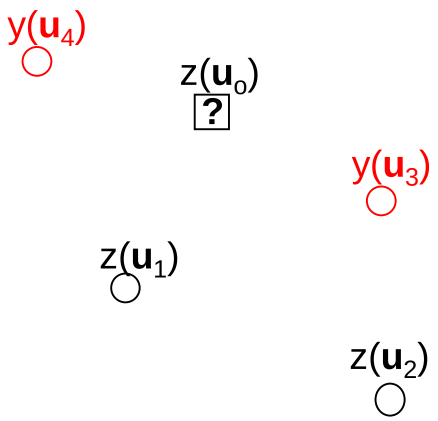
Full Cokriging

- We can extend the simple kriging system to integrate other variables!

Primary Variable at $\mathbf{u}_1$ and $\mathbf{u}_2$

Secondary Variable at $\mathbf{u}_3$ and $\mathbf{u}_4$

Estimate Primary Variable at $\mathbf{u}_0$



Redundancy **Weights** **Closeness**

Primary, Primary | Primary, Secondary | Primary | Primary

$$
\begin{bmatrix}
C_z(\mathbf{u}_1,\mathbf{u}_1) & C_z(\mathbf{u}_1,\mathbf{u}_2) & C_{z,y}(\mathbf{u}_1,\mathbf{u}_3) & C_{z,y}(\mathbf{u}_1,\mathbf{u}_4) \\
C_z(\mathbf{u}_2,\mathbf{u}_1) & C_z(\mathbf{u}_2,\mathbf{u}_2) & C_{z,y}(\mathbf{u}_2,\mathbf{u}_3) & C_{z,y}(\mathbf{u}_2,\mathbf{u}_4) \\
C_{z,y}(\mathbf{u}_3,\mathbf{u}_1) & C_{z,y}(\mathbf{u}_3,\mathbf{u}_2) & C_{y,y}(\mathbf{u}_3,\mathbf{u}_3) & C_y(\mathbf{u}_3,\mathbf{u}_4) \\
C_{z,y}(\mathbf{u}_4,\mathbf{u}_1) & C_{z,y}(\mathbf{u}_4,\mathbf{u}_2) & C_{y,y}(\mathbf{u}_4,\mathbf{u}_3) & C_y(\mathbf{u}_4,\mathbf{u}_4)
\end{bmatrix}
\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}
=
\begin{bmatrix}
C_z(\mathbf{u}_o,\mathbf{u}_1) \\
C_z(\mathbf{u}_o,\mathbf{u}_2) \\
C_{z,y}(\mathbf{u}_o,\mathbf{u}_3) \\
C_{z,y}(\mathbf{u}_o,\mathbf{u}_4)
\end{bmatrix}
$$

Seondary, Primary | Seondary, Secondary | Secondary | Secondary

Cross Variogram and Cross Covariance

- Cross Variogram, measure of how two variables differ together over distance.

$$\gamma_{z,y}(\mathbf{h}) = \frac{1}{2}E\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})][Y(\mathbf{u}) - Y(\mathbf{u}+\mathbf{h})]\} \quad \forall\, \mathbf{u}$$

- Cross Covariance, measure of how two variables vary together over distance.

$$C_{z,y}(\mathbf{h}) = E\left\{[Z(\mathbf{u}) - m_z]\left[Y(\mathbf{u}+\mathbf{h}) - m_y\right]\right\}$$

$$= E\{[Z(\mathbf{u})][Y(\mathbf{u}+\mathbf{h})]\} - m_z \cdot m_y \qquad \forall\, \mathbf{u}$$

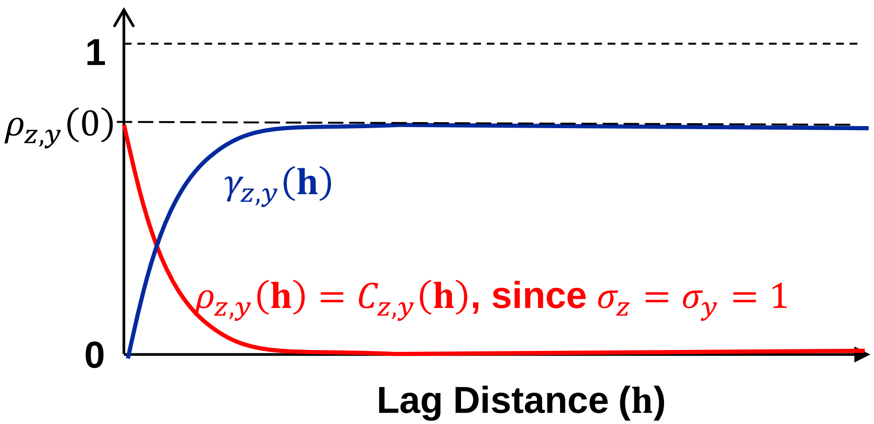
- Cross correlogram is the standardized cross covariance. Correlation coefficient vs. lag distance!

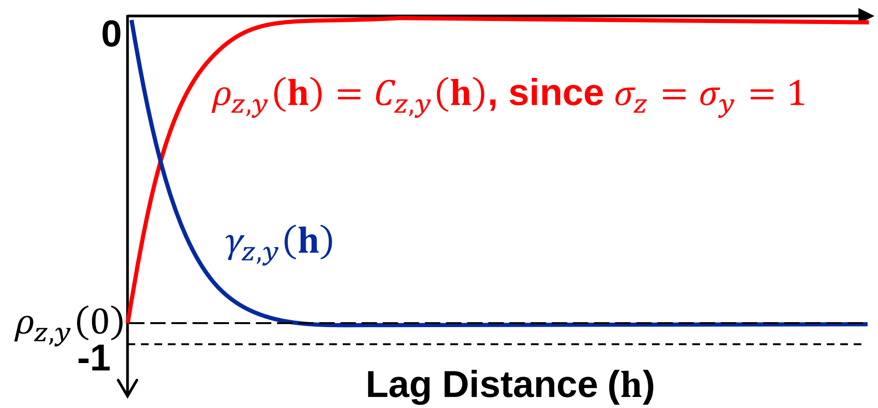
$$\rho_{z,y}(\mathbf{h}) = \frac{C_{z,y}(\mathbf{h})}{\sigma_z \cdot \sigma_y} \qquad \rho_{z,y}(\mathbf{h}) = C_{z,y}(\mathbf{h}), \text{ if } \sigma_z = \sigma_y = 1.0$$

Cross Variogram and Cross Covariance

- Cross variogram starts at 0.0, $\gamma_{z,y}(\mathbf{0}) = \mathbf{0}$, and then at the range reaches the correlation coefficient, $\gamma_{z,y}(\mathbf{h}) \rightarrow \rho_{z,y}(0)$, as $\mathbf{h} \rightarrow$ range.
  - if the correlation coefficient is less than zero, $\rho_{z,y}(0) < 0$, then the cross variogram is negative!

- Cross correlogram (equal to cross covariance if $\sigma_z = \sigma_y = 1$), starts at the correlation coefficient, $\rho_{z,y}(\mathbf{0}) = \rho_{z,y}$, and then at the range reaches the 0, $\rho_{z,y}(\mathbf{h}) \rightarrow \mathbf{0}$, as $\mathbf{h} \rightarrow$ range.
  - if the correlation coefficient is less than zero, $\rho_{z,y}(0) < 0$, then the cross correlogram is negative!





Collocated Cokriging Simulation

- Collocated Cokriging makes two simplifications of full cokriging:
  - Only one (the collocated) secondary variable is considered
  - Cross covariance $C_{z,y}(\mathbf{h})$ is assumed to be a linear scaling of $C_z(\mathbf{h})$
- The collocated secondary value is surely the most important and likely screens the influence of multiple secondary data
- Consider the implications for the cokriging system with only the collocated secondary data value included (below).

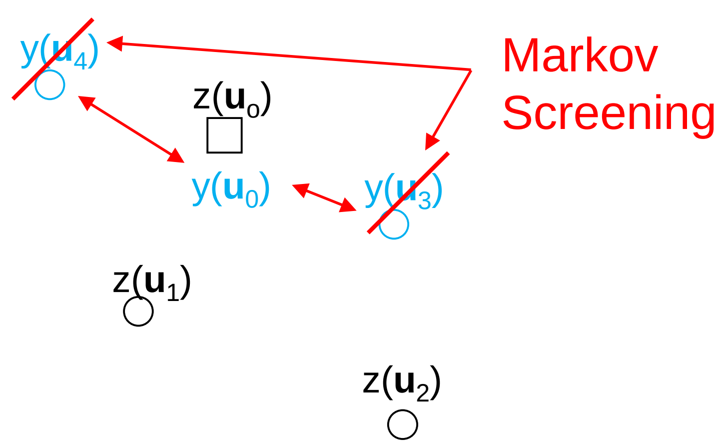
$$\begin{bmatrix} C_z(\mathbf{u}_1,\mathbf{u}_1) & C_z(\mathbf{u}_1,\mathbf{u}_n) & C_{z,y}(\mathbf{u}_1,\mathbf{u}_o) \\ C_z(\mathbf{u}_n,\mathbf{u}_1) & C_z(\mathbf{u}_n,\mathbf{u}_n) & C_{z,y}(\mathbf{u}_n,\mathbf{u}_o) \\ C_{z,y}(\mathbf{u}_0,\mathbf{u}_1) & C_{z,y}(\mathbf{u}_0,\mathbf{u}_n) & C_y(0) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_n \\ \lambda_y \end{bmatrix} = \begin{bmatrix} C_z(\mathbf{u}_o,\mathbf{u}_1) \\ C_z(\mathbf{u}_o,\mathbf{u}_n) \\ C_{z,y}(0) \end{bmatrix}$$

Collocated Cokriging Simulation

- The collocated value is surely the most important and likely screens the influence of multiple secondary data
- Therefore secondary variogram is not needed
  - Never need correlation between secondary data

$$\left[\begin{array}{cc|c} C_z(\mathbf{u}_1,\mathbf{u}_1) & C_z(\mathbf{u}_1,\mathbf{u}_2) & C_{z,y}(\mathbf{u}_1,\mathbf{u}_o) \\ C_z(\mathbf{u}_2,\mathbf{u}_1) & C_z(\mathbf{u}_2,\mathbf{u}_2) & C_{z,y}(\mathbf{u}_2,\mathbf{u}_o) \\ \hline C_{z,y}(\mathbf{u}_0,\mathbf{u}_1) & C_{z,y}(\mathbf{u}_0,\mathbf{u}_2) & \boxed{\sigma_y^2} \end{array}\right] \left[\begin{array}{c} \lambda_1 \\ \lambda_n \\ \hline \lambda_y \end{array}\right] = \left[\begin{array}{c} C_z(\mathbf{u}_o,\mathbf{u}_1) \\ C_z(\mathbf{u}_o,\mathbf{u}_2) \\ \hline C_{z,y}(0) \end{array}\right]$$

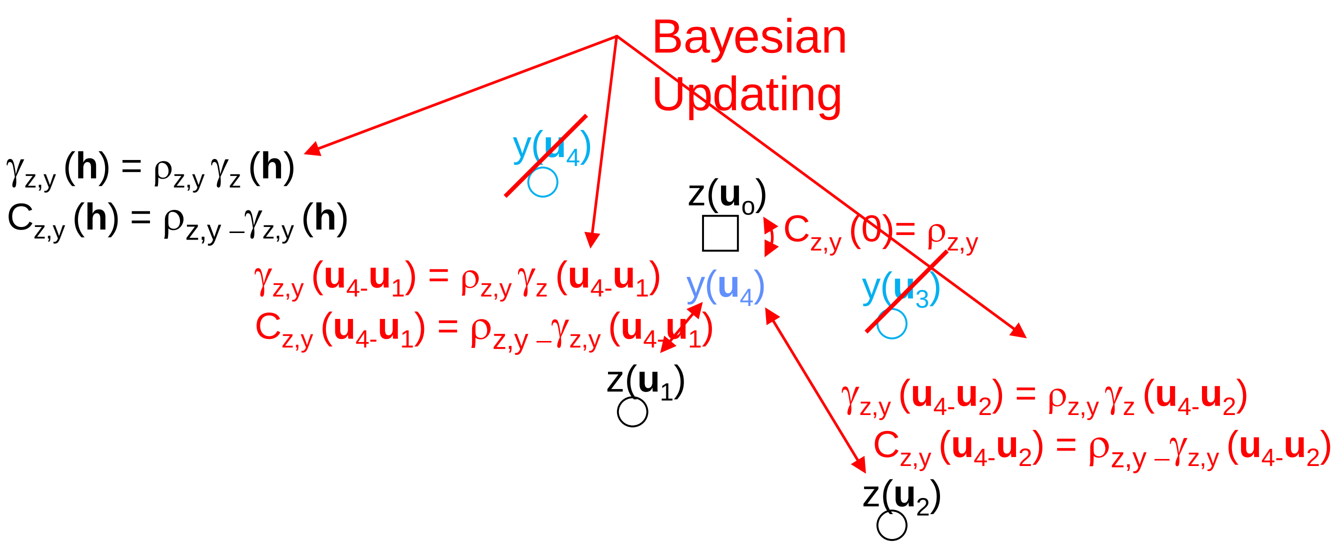
No need for $\gamma_y(\mathbf{h})$ nor $\gamma_{z,y}(\mathbf{h})$.
Only need $C_y(0) = \sigma_y^2$



Collocated Cokriging Simulation

- Bayesian updating to calculate the cross variogram from the primary variogram x correlation coefficient.
  - The cross variogram is not needed, No need for $\gamma_{z,y}(\mathbf{h})$ for h > 0.

$$\begin{bmatrix} C_z(\mathbf{u}_1,\mathbf{u}_1) & C_z(\mathbf{u}_1,\mathbf{u}_2) & \rho_{z,y}C_z(\mathbf{u}_1,\mathbf{u}_o) \\ C_z(\mathbf{u}_2,\mathbf{u}_1) & C_z(\mathbf{u}_2,\mathbf{u}_2) & \rho_{z,y}C_z(\mathbf{u}_2,\mathbf{u}_o) \\ \rho_{z,y}C_z(\mathbf{u}_0,\mathbf{u}_1) & \rho_{z,y}C_z(\mathbf{u}_0,\mathbf{u}_2) & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_n \\ \lambda_y \end{bmatrix} = \begin{bmatrix} C_z(\mathbf{u}_o,\mathbf{u}_1) \\ C_z(\mathbf{u}_o,\mathbf{u}_2) \\ \rho_{z,y} \end{bmatrix}$$



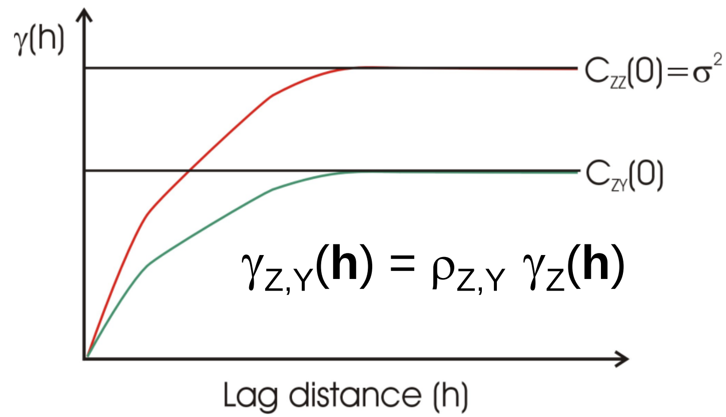
Collocated Cokriging Simulation Markov Assumption

- Simplification due to retaining the collocated secondary variable:
  - no need for the variogram of the secondary variable
  - no need for a cross variogram
  - if the secondary data are smooth then considering more than the collocated variable *should* not help

- There is a ***serious potential problem*** with excess variance of the results when used in simulation mode
  - resulting mean is often biased if variance is too high
  - may require a variance reduction factor

Collocated Cokriging Simulation Bayesian Updating

- Eliminate the need for calculating a cross variogram:
  - no need for a cross variogram
  - If the secondary data are smooth then considering more than the collocated variable *should* not help



Collocated Cokriging Simulation Hands-on in Excel

- Hands-on example of collocated cokriging:

Interactive Collocated cokriging Demonstration, Multivariate Spatial Estimates with Variogram Models

Michael Pyrcz, Geostatistics at Petroleum and Geosystems Engineering, University of Texas at Austin (mpyrcz@austin.utexas.edu) @GeostatsGuy

0. Representative Distribution and Transform

	Primary	Secondary
mean	0.15	0.13
st. deviation	0.03	0.02
correlation coefficent	0.6	

1. Data

Point	x	y	value	Gaussian
1	10	5	0.1	-1.667
2	5	3	0.07	-2.667
3	30	1	0.2	1.667
secondary	20	0	0.11	-1.000
unknown	100	0		
mean			0.120	0.00

2. Distance Matrix

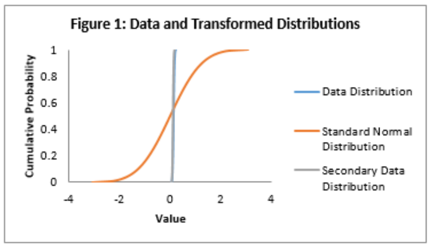
0.00	5.39	20.40	90.14
5.39	0.00	25.08	95.05
20.40	25.08	0.00	70.01

3. Variogram Model

Nugget	0		
Spherical	1	Range	300

4. Variogram Matrix

0.000	0.027	0.102	0.437
0.027	0.000	0.125	0.459
0.102	0.125	0.000	0.344



5. Covariance Matrix

1.000	0.973	0.898	0.338	0.563
0.973	1.000	0.875	0.324	0.541
0.898	0.875	1.000	0.394	0.656
0.338	0.324	0.394	1.000	0.6

6. Inverse Left Side

22.848	-18.241	-4.561	-0.003
-18.241	18.836	-0.136	0.104
-4.561	-0.136	5.435	-0.556
-0.003	0.104	-0.556	1.186

7. Weights

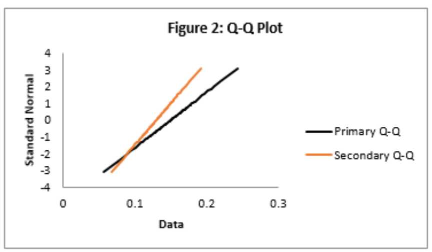
	0.003
	-0.111
	0.593
	0.402
Sum Weights	0.886

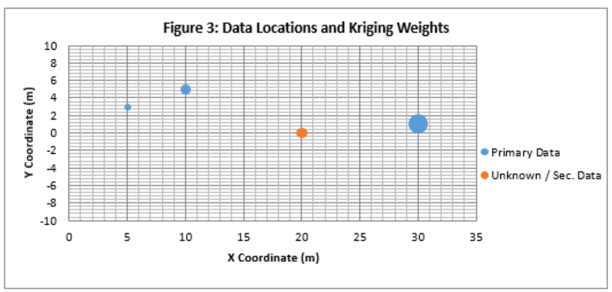
8. Kriging Results

Kriging Estimate	0.877
Kriging Variance	0.428

9. Simulation Results

Realization	Simulation	Back Transformed
1	0.919	0.178
2	1.250	0.187
3	1.680	0.200
4	2.041	0.211
5	1.345	0.190
6	1.191	0.186
7	2.178	0.215
8	0.310	0.159
9	1.173	0.185
10	0.556	0.167





Legend

- Information
- User Input
- Calculation
- Redundancy Measures
- Closeness Measures

Description

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

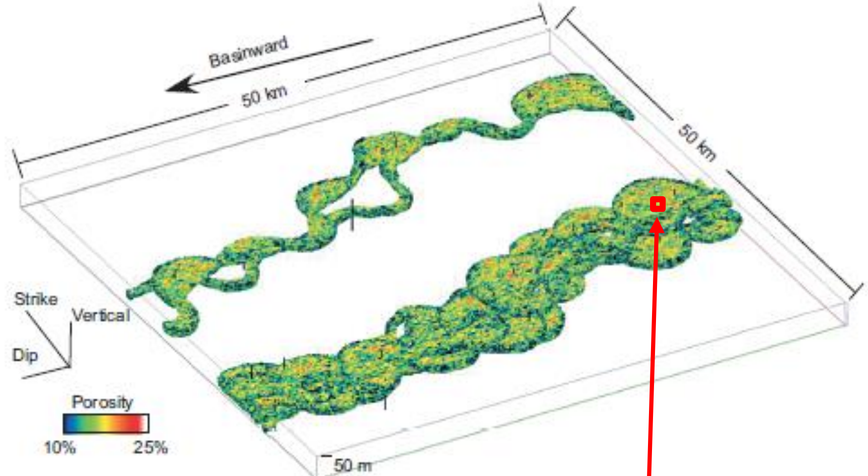
File is: Collocated_Cokriging_Demo.xlsx

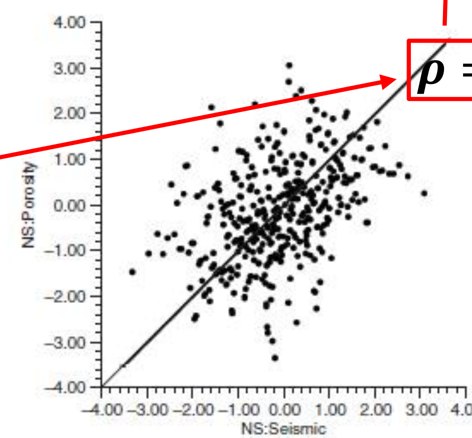
Collocated Cokriging Simulation Example

- Workflow:

1. Simulate realization of secondary variable (e.g. acoustic impedance)
2. Include collocated secondary realization at each location with correlation coefficient.
3. Check model histogram, vairogram and bivariate plot.

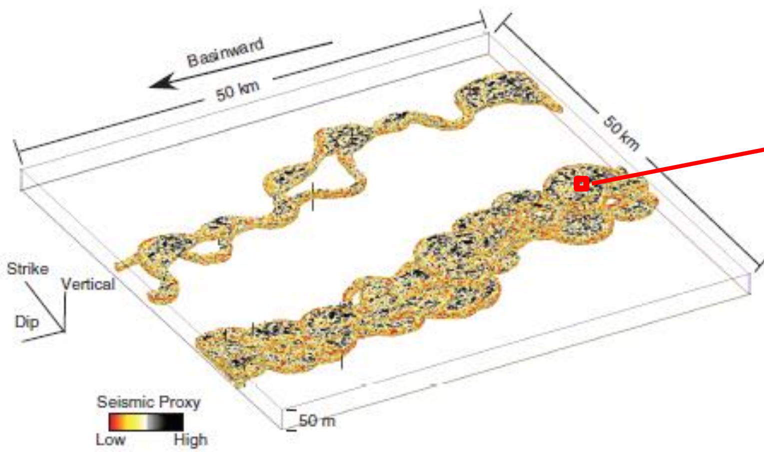
3. Exhaustive Primary Cosimulation





$\boxed{\rho = 0.41}$

2. Correlation coefficient after Gaussian transform of primary and secondary variable

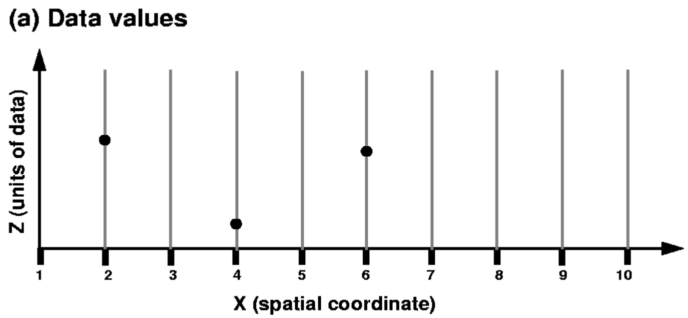


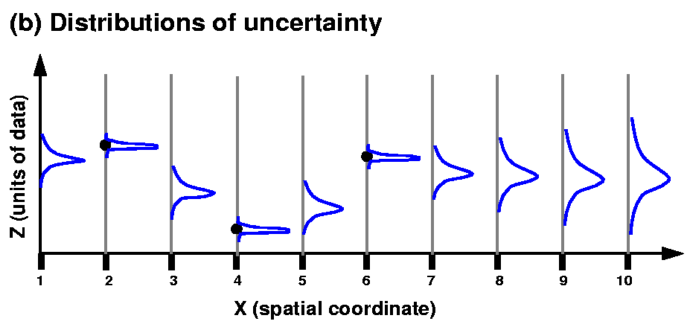
1. Exhaustive Secondary Simulation

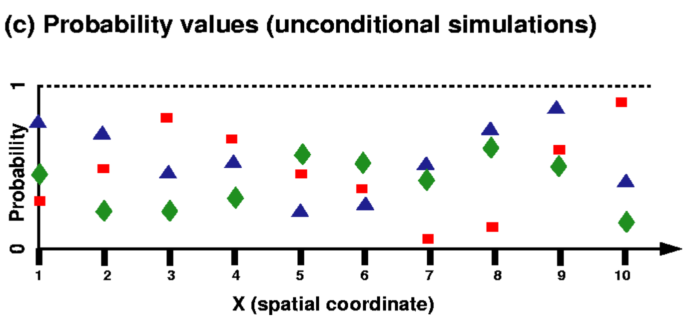
p-field / Cloud Transform

- The key idea of p-field is to perform the simulation in two separate steps:
  - Construct local distributions of uncertainty
  - Draw from those distributions simultaneously with correlated probabilities
- Separating the two steps has advantages: (1) the distributions of uncertainty can be constructed to honor all data and checked before any realizations are drawn, and (2) the simulations are consistent with the distributions of uncertainty
- Also, good reproduction of the primary to secondary data scatter plot.
- Some disadvantages include potentially poor reproduction of the histogram and variogram.
- The most commonly used approach is known as cloud transform, often used to simulate permeability from porosity

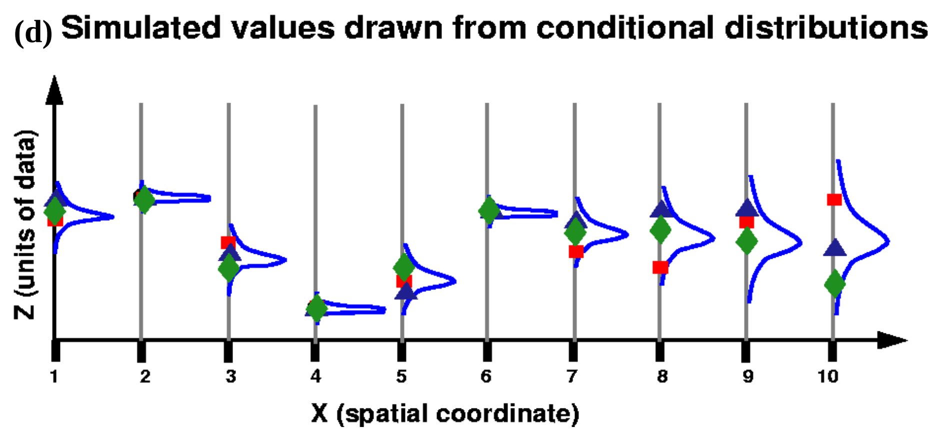
p-Field Approach





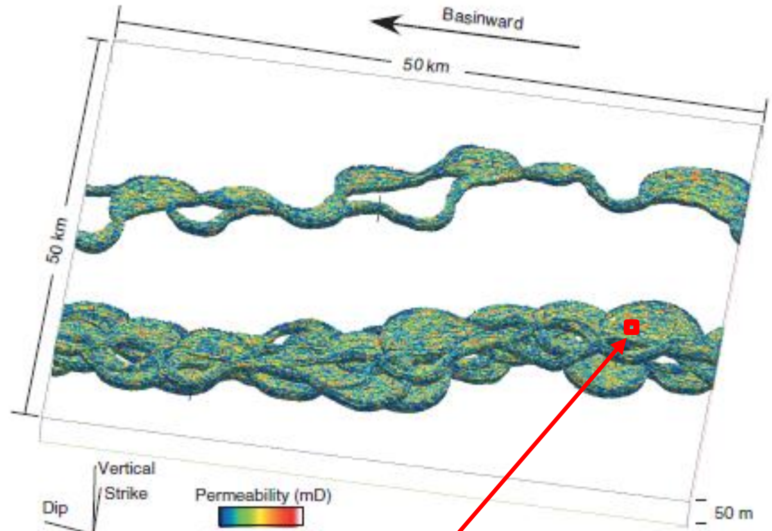


- Construct distributions of uncertainty with priors and likelihoods
- Generate spatially correlated probability values (usual with Gaussian simulation)
- Draw values simultaneously and keep together as a realization



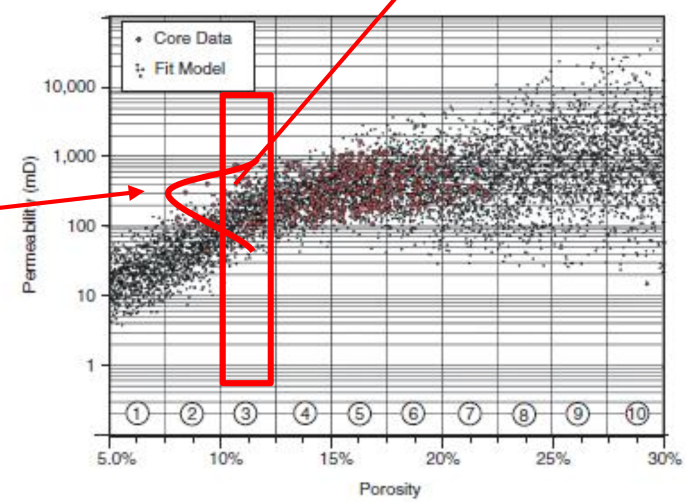
Cloud Transform Simulation

3. Exhaustive Primary Cosimulation

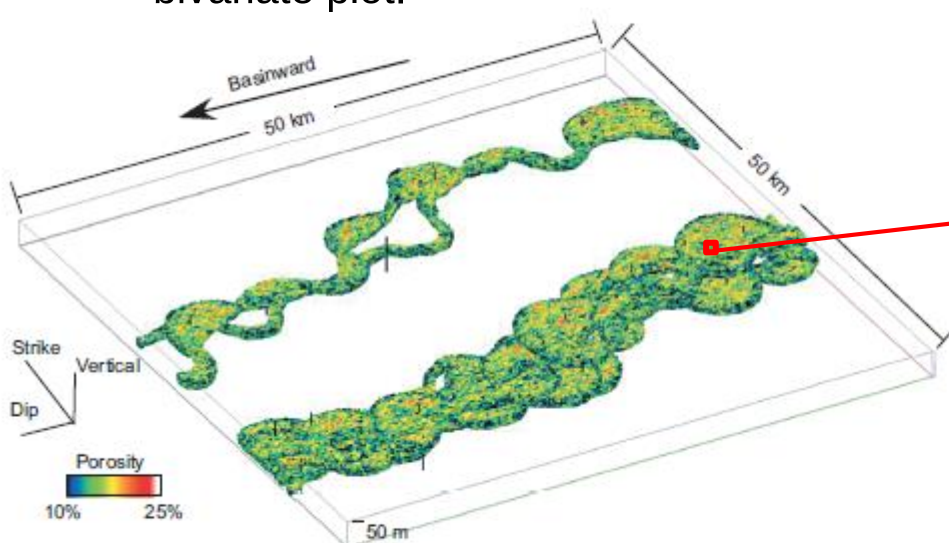


Workflow

1. Simulate realization of secondary variable (e.g. porosity)
2. For each location draw from the conditional distribution given the secondary realization with spatially correlated p-values
3. Check model histogram, vairogram and bivariate plot.



2. Conditional Primary | Secondary



1. Exhaustive Secondary Simulation

Cosimulation Review

- Cosimulation includes methods that simulate a property realization (primary) conditional to a previously simulated property realization (secondary)
- Two simulated properties without cosimulations will only have correlations imposed by data and outside correlation range of data will be uncorrelated.
  - This leads to unrealistic values and too high spatial uncertainty.
- Two methods are commonly applied to do this:
  - Collocated Cokriging simplifies the full Cokriging system
    - » Markov assumption – only need collocated secondary value
    - » Bayesian updating – get the cross variogram by scaling the primary variogram with correlation coefficient
    - » May not reproduce the cloud well (relationship between the 2 variables)
  - Cloud transform
    - » Forces reproduction of the cloud
    - » May not get the spatial continuity and distribution well
  - Know their assumptions and steps

Multivariate Modeling:
Multivariate Workflows

Lecture outline . . .

- **Spatial Bayesian Updating for Multivariate, Spatial Modeling**
  - Pyrcz and Deutsch (2014) Section 4.1.4 Multivariate Mapping
  - Originally published in an SPE paper by Deutsch, Ren and Leuangthong (2005)

Spatial Bayesian Updating

- **Spatial Bayesian Updating**
  - Here's a workflows for multivariate integration with spatial data updating.
  - This should help you get started with building subsurface models that include multiple information sources.

- **The Workflow**
  - Formulate a prior distribution with spatial estimation based on available well data
  - Combine all mapped secondary data into a likelihood distribution at all locations
  - Update prior from wells with combined secondary from all mapped information sources to calculate the posterior

Spatial Bayesian Updating

The Approach

1. Transform variables to standard normal (Gaussian mean = 0, standard deviation = 1.0)
2. **Well Data**: Calculate a prior distribution with spatial estimation based on available well data
3. **Secondary Data:** Combine all mapped secondary data into a likelihood distribution at all locations
4. **Bayesian Updating:** Update prior from wells with combined secondary from all mapped information sources to calculate the posterior
5. Back Transform results to original distribution for reporting uncertainty

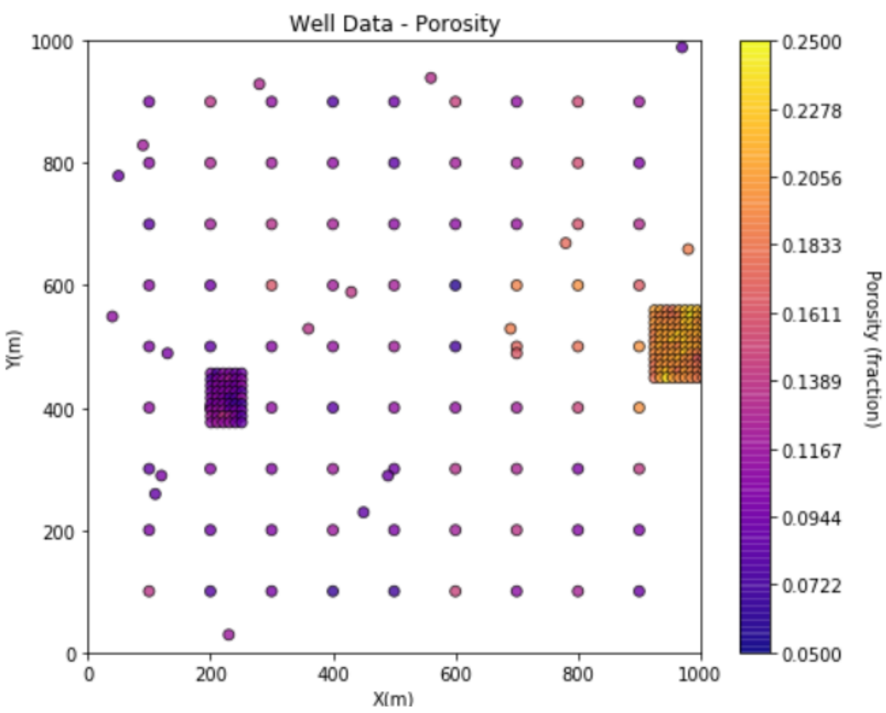
Spatial Bayesian Updating

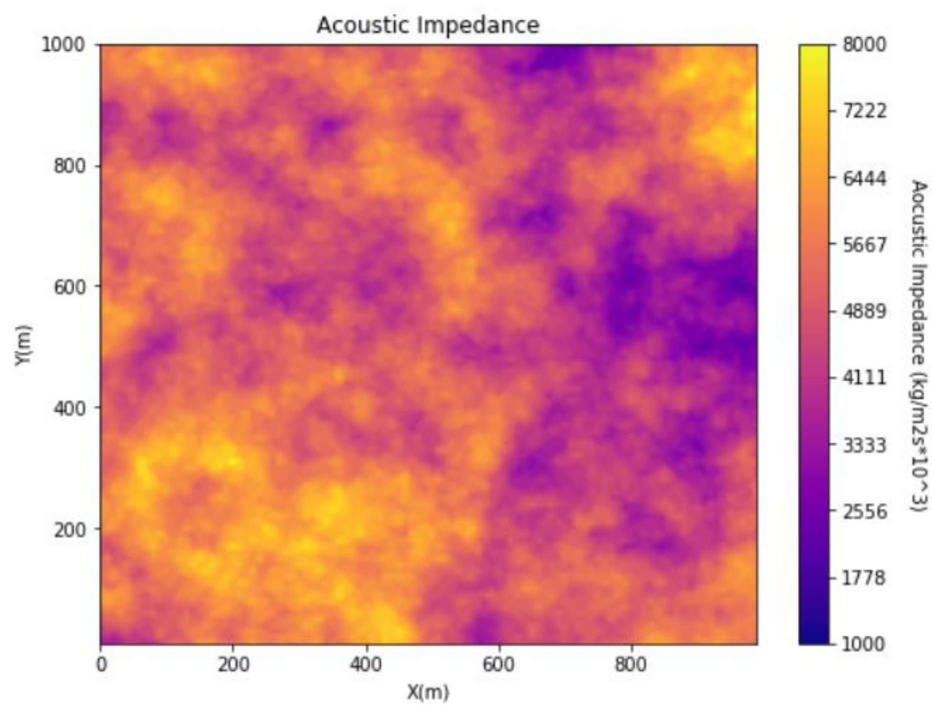
- **Advantages of this Workflow**
  - Quantify and visualize the contributions from wells and mapped secondary data
  - Account for spatial continuity in estimate and uncertainty for prior
  - Combine secondary sources accounting for redundancy and closeness
- Assumptions
  - Multivariate Gaussian after univariate normal score transform
  - Same volume support for primary and secondary data

Spatial Bayesian Updating

An Example of this Workflow in Python

Predicting porosity from acoustic impedance map.



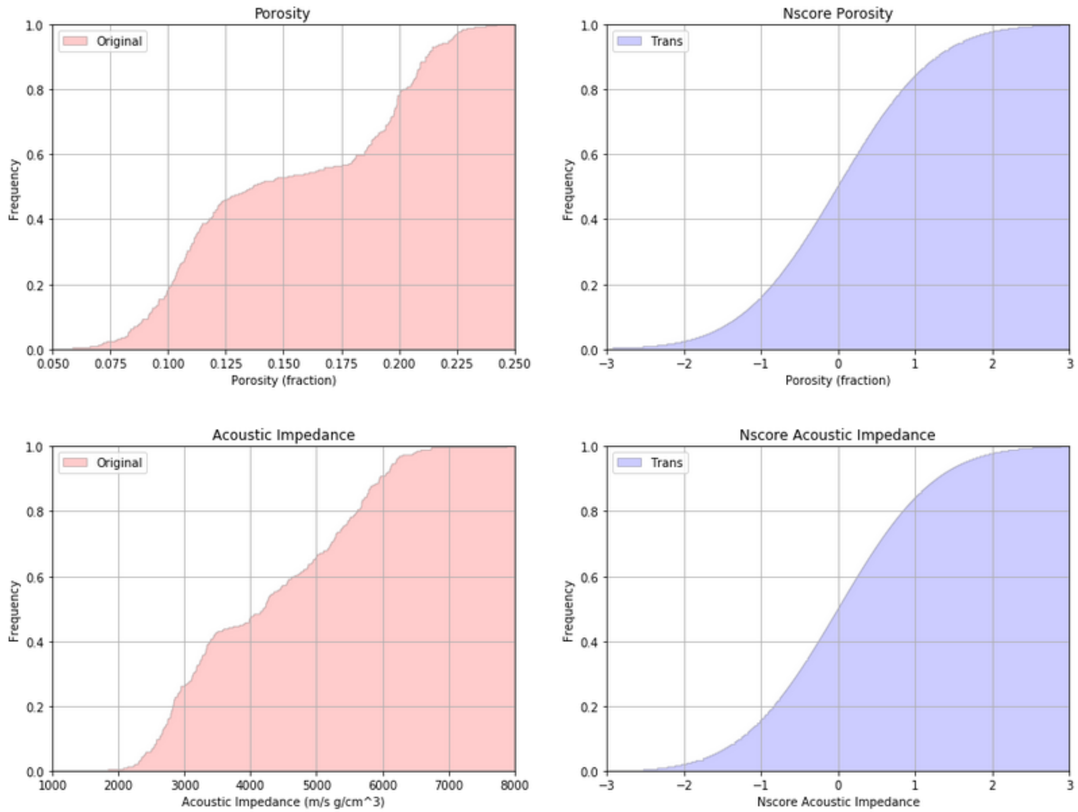


We will only use a single mapped secondary variable for simplicity.

Spatial Bayesian Updating

Transform All Features to Standard Normal

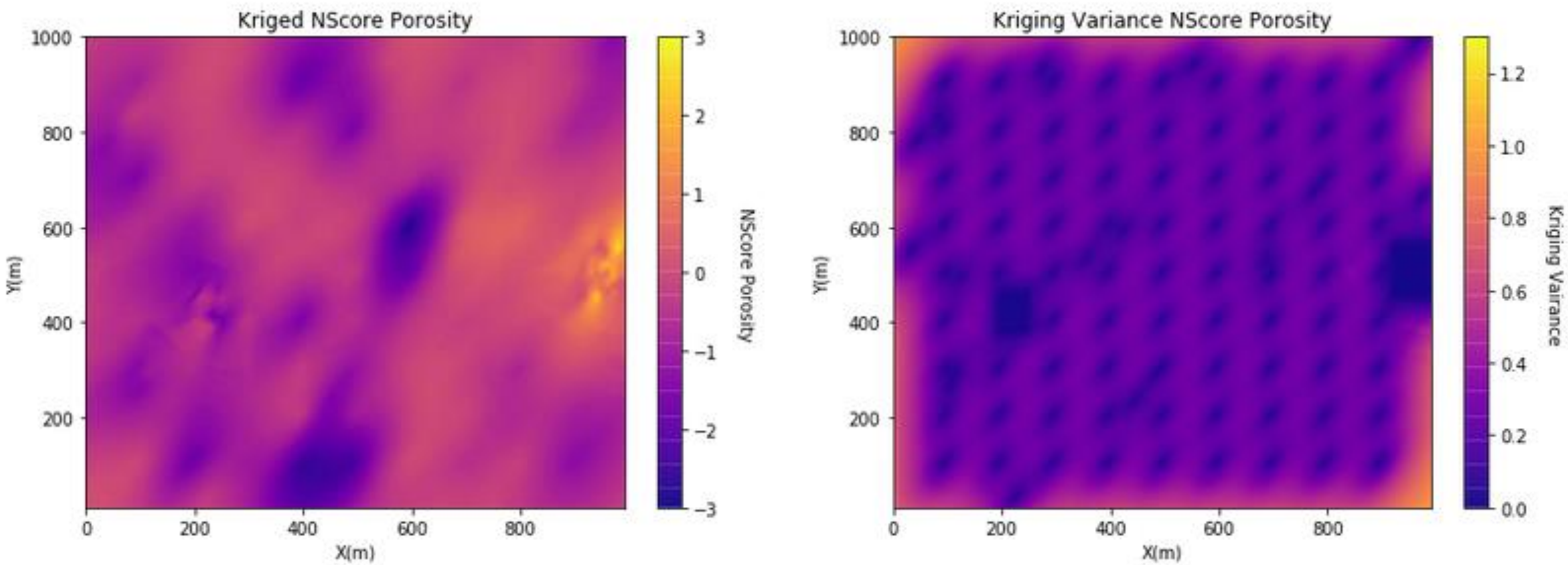
Transform the distributions of porosity and AI to Gaussian with mean of zero and standard deviation of 1.0.



Spatial Bayesian Updating

Simple Kriging to Estimate Well-based Prior

Based on stationary mean (could include trend) and variogram model



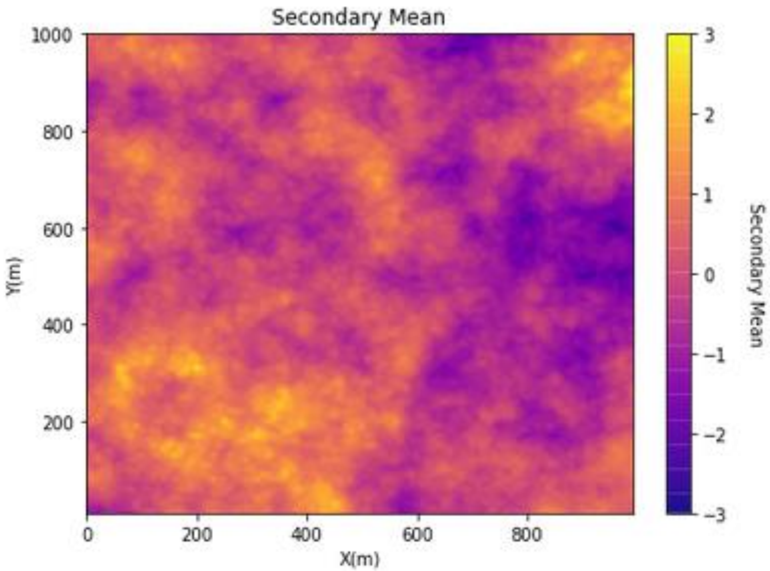
- We have a prior Gaussian mean and standard deviation at all locations
- With the Gaussian assumption we have a complete distribution at all locations.

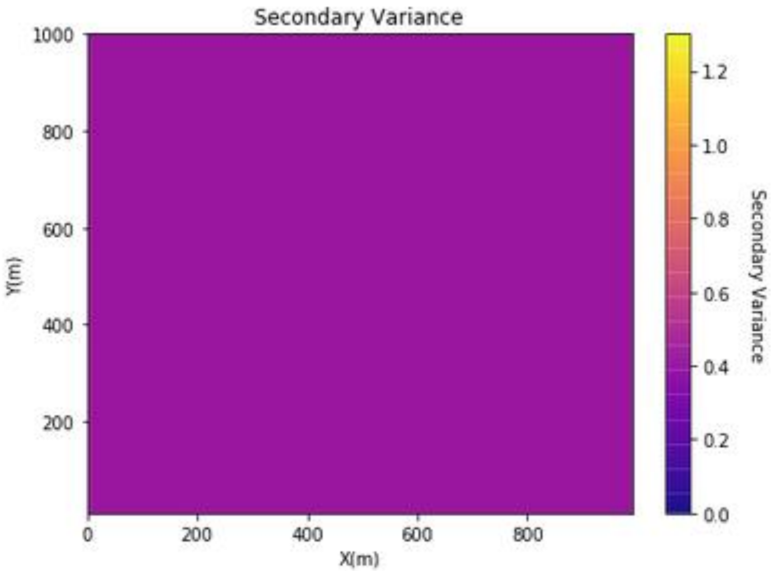
Spatial Bayesian Updating

Combine Secondary Variables to Estimate Map-based Likelihood

with multivariate kriging at each location. We can do this with the single

Note the variance is constant, since the secondary data is available at all locations and the model is homoscedastic!



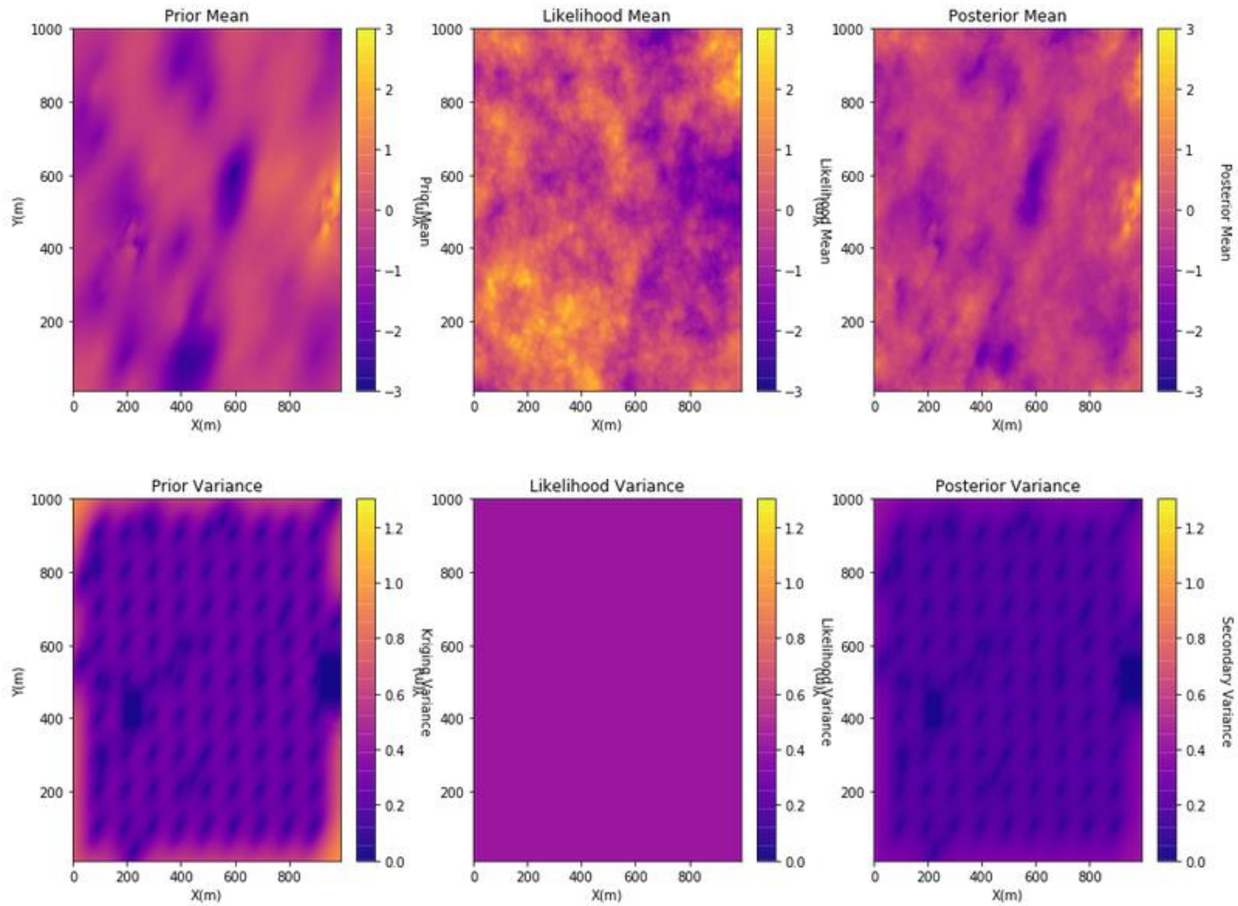


- We have a secondary likelihood Gaussian mean and standard deviation at all locations
- With the Gaussian assumption we have a complete distribution at all locations.

Spatial Bayesian Updating

Bayesian Updating

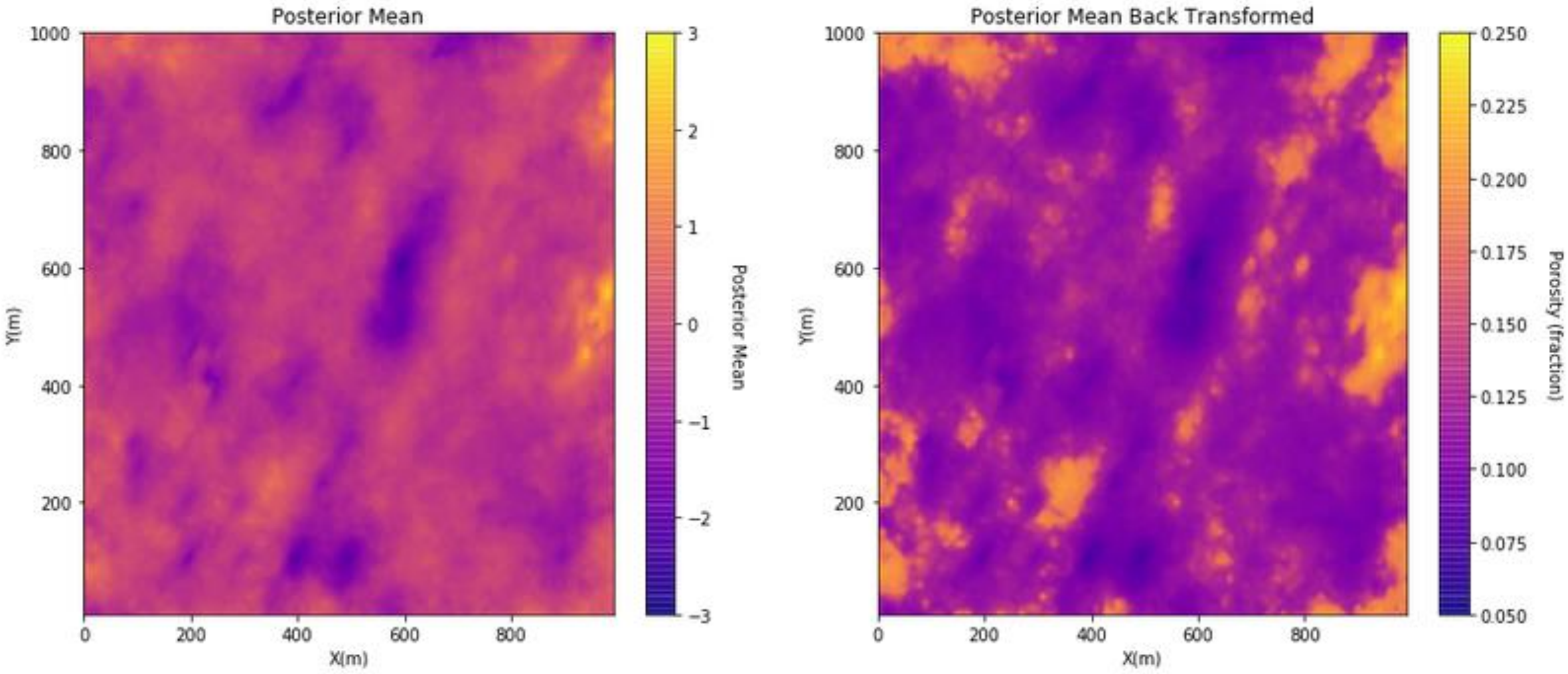
Calculate the updated posterior from calculate prior and likelihood distributions at all locations.



Spatial Bayesian Updating

Back-transform to Original Distribution

Move from Gaussian transformed variables and Gaussian assumption (Gaussian space) back to the original distribution for porosity (original space)

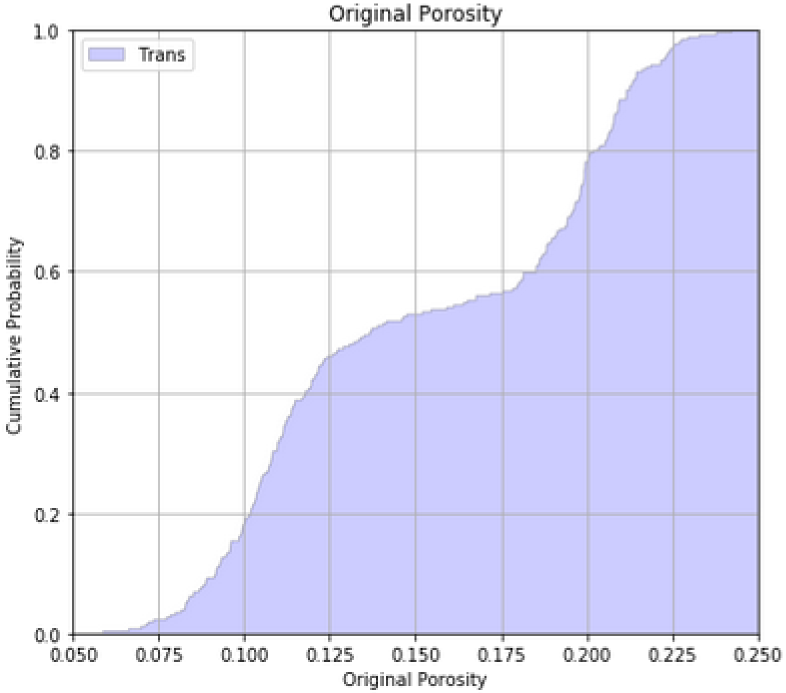


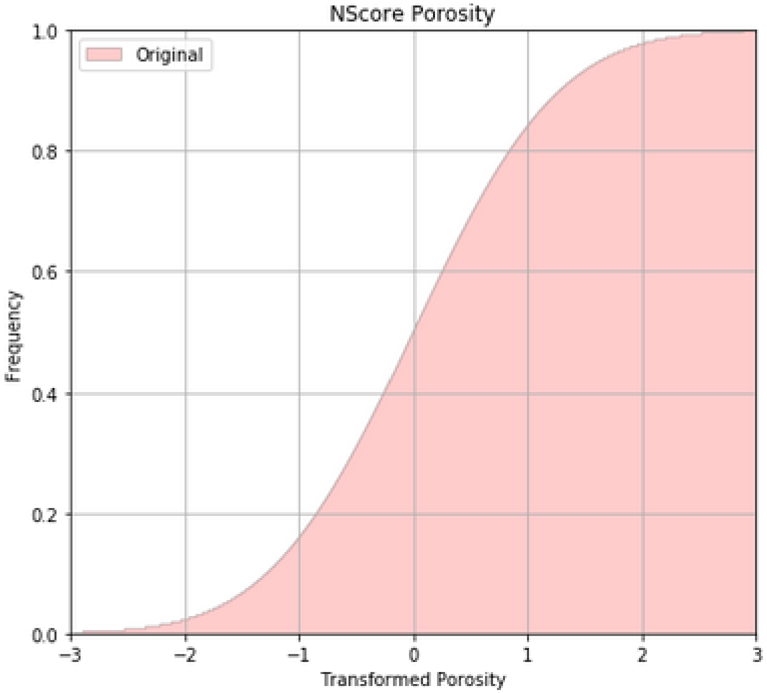
Spatial Bayesian Updating

Back-transform to Original Distribution

Note, the bimodal appearance in the original units.

Check the original distribution to confirm bimodal.



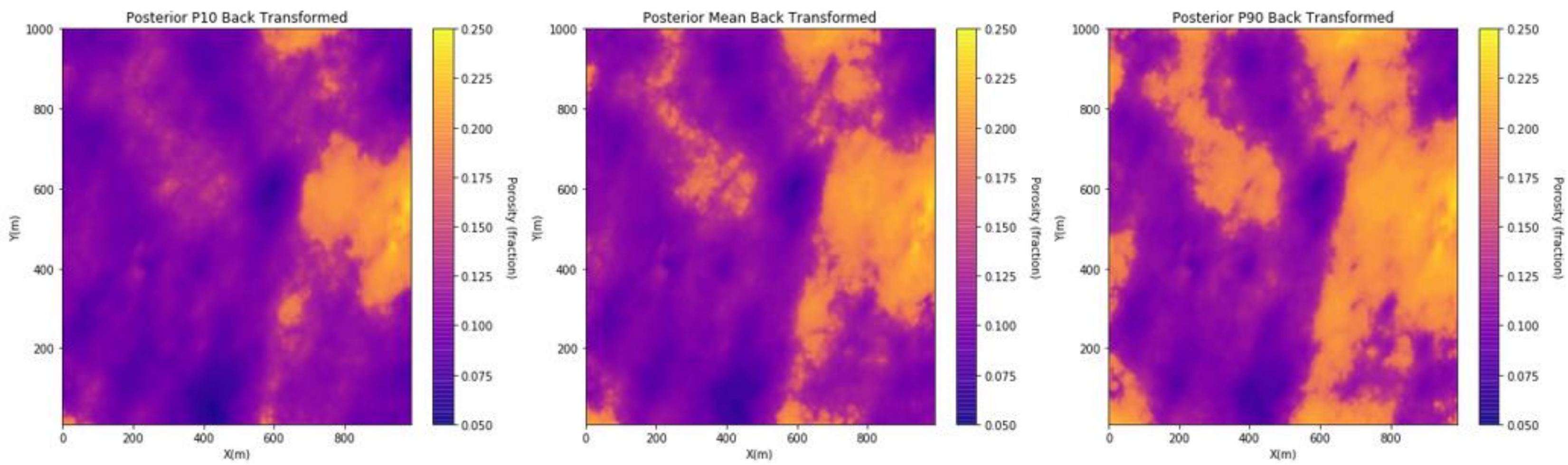


Spatial Bayesian Updating

Complete Uncertainty Model

Back-transform enough percentiles to calculate the entire non-parametric distribution at each location.

Here's the P10 and P90:



Spatial Bayesian Updating

Updated Simulated Realizations

Apply p-field simulation to calculate stochastic simulations of porosity

- integrating all secondary, mapped data
- integration local well data

Multivariate New Tools

Topic	Application to Subsurface Modeling
Cosimulation	Directly model simulated realizations that have a specified level of correlation. *Permeability is cosimulated conditioning to well data, trends and the previously simulated porosity realization.*
Spaital Prior	Build a spatial prior based on well data. *The well data results in a local porosity prior with a mean of 12% and standard deviation of 3%.*
Secondary Likelihood	*Build a mapped secondary data likelihood that combines all the the available secondary data.* *The mapped geological interpretations and acoustic impedance results in a local porosity likelihood with a mean of 15% and standard deviation of 2%.*
Updated Posterior	*Update the local prior with the local likelihood to calculate the local posterior.* *The updated porosity local posterior has a mean of 14% and a standard deviation of 1.6%.*

Multivariate Modeling: Multivariate Workflows

Lecture outline . . .

- Cosimulation
- Spatial Updating