PGE 337 Data Analytics and Geostatistics

Lecture 14: Indicator Simulation

Lecture outline . . .

- Indicator Methods
- Indicator Estimation and Simulation

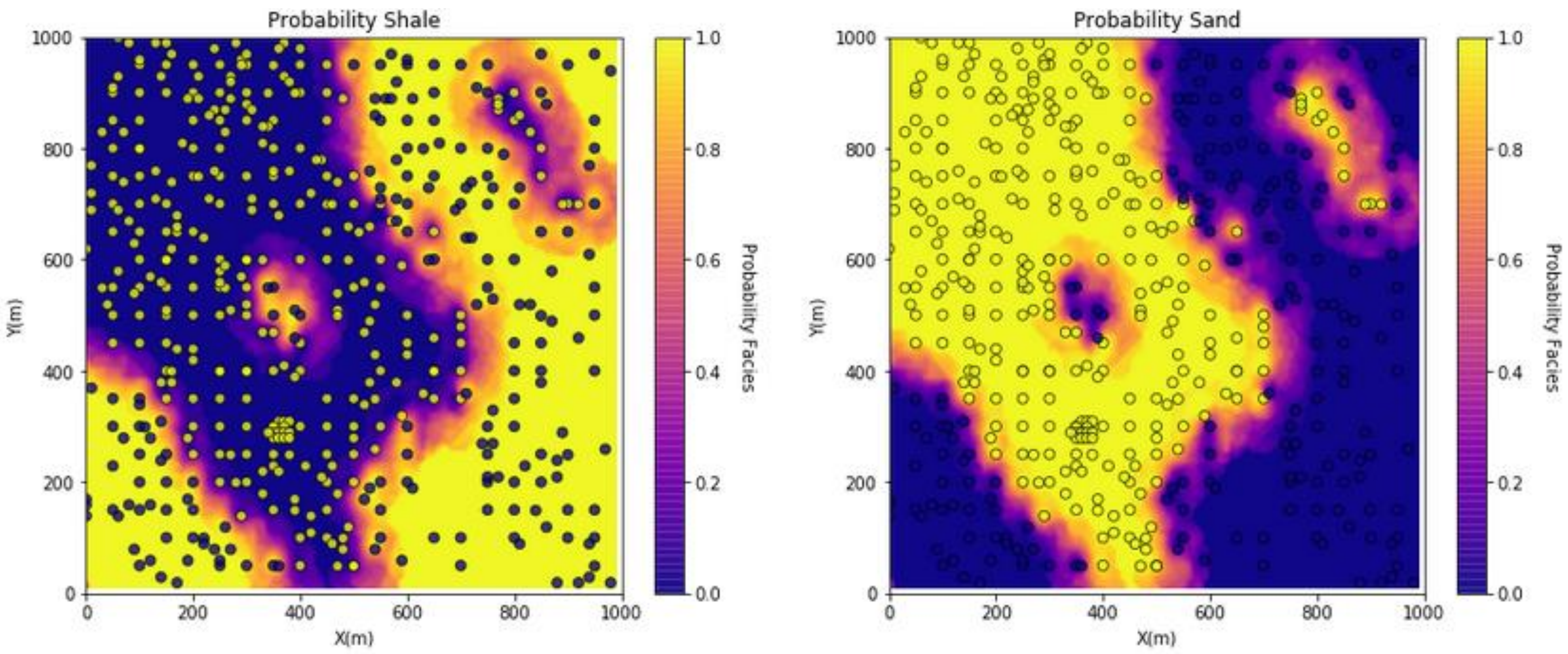
Michael Pyrcz, The University of Texas at Austin

Some Indicator Slides Modified from Prof. Clayton Deutsch's MinE612 Class.

Motivation

We need spatial estimation and simulation methods for categorical data and soft data.

- We can use a probability encoding known as the indicator transform.



PGE 337 Data Analytics and Geostatistics

Lecture 14: Indicator Simulation

Lecture outline . . .

- Indicator Methods

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

Some Indicator Slides Modified from Prof. Clayton Deutsch's MinE612 Class.

The Indicator Formalism and Indictor Kriging (IK)

- Indicator Methods:
  - Estimation and Simulation with categorical variables with explicit control of spatial continuity of each category
  - Estimation and simulation with continuous variables with explicit control of the spatial continuity of different magnitudes
  - Requires indicator coding of data, a probability coding based on category or threshold
  - Requires indicator variograms to describe the spatial continuity.

The Indicator Formalism and Indictor Kriging (IK)

Indicator coding is transforming a random variable / function to a probability relative to a category or a threshold.

- If $I\{\mathbf{u}: z_k\}$ is an indicator for a categorical variable,
  - What is the probability of a realization equal to a category?

$$I(\mathbf{u};\boldsymbol{z_k}) = \begin{cases} 1, & \text{if } Z(\mathbf{u}) = \boldsymbol{z_k} \\ 0, & \text{otherwise} \end{cases}$$

  - e.g. given threshold, $z_2 = 2$, and data at $\mathbf{u}_1, z(\mathbf{u}_1) = 2$, then $I\{\mathbf{u}_1; z_2\} = 1$
  - e.g. given threshold, $z_1 = 1$, and a RV away from data, $\mathrm{Z}(\mathbf{u}_2)$ then $I\{\mathbf{u}_2; z_1\} = 0.25$

- If $I\{\mathbf{u}: z_k\}$ is an indicator for a continuous variable,
  - What is the probability of a realization less than or equal to a threshold?

$$I(\mathbf{u};\boldsymbol{z_k}) = \begin{cases} 1, & \text{if } Z(\mathbf{u}) \leq \boldsymbol{z_k} \\ 0, & \text{otherwise} \end{cases}$$

  - e.g. given threshold, $z_1 = 6\%$, and data at $\mathbf{u}_1, z(\mathbf{u}_1) = 8\%$, then $I\{\mathbf{u}_1; z_1\} = 0$
  - e.g. given threshold, $z_4 = 18\%$, and a RV $\mathrm{Z}(\mathbf{u}_2) = N[16\%, 3\%]$ then $I\{\mathbf{u}_2; z_4\} = 0.75$

The Indicator Formalism and Indictor Kriging (IK)

Example of indicator transforms for a categorical variable.

Original Data	$I\{\mathbf{u}_\alpha; z_1 = 1\}$	$I\{\mathbf{u}_\alpha; z_2 = 2\}$	$I\{\mathbf{u}_\alpha; z_3 = 3\}$
$z(\mathbf{u}_1) = 3$	0	0	1
$z(\mathbf{u}_2) = 1$	1	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$z(\mathbf{u}_n) = 2$	0	1	0

Our $\mathrm{z}(\mathbf{u}_\alpha)$, $\alpha = 1, \ldots, n$, data become $k = 1, \ldots, K$ sets of n data, a new variable that indicates the probability of being each category.

The Indicator Formalism and Indictor Kriging (IK)

Example of indicator transforms for a continuous variable.

Original Data	$I\{\mathbf{u}_\alpha; z_1 \le 5\%\}$	$I\{\mathbf{u}_\alpha; z_2 \le 10\%\}$	$I\{\mathbf{u}_\alpha; z_3 \le 15\%\}$
$z(\mathbf{u}_1) = 12\%$	0	0	1
$z(\mathbf{u}_2) = 4\%$	1	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$z(\mathbf{u}_n) = 17\%$	0	0	0

Our $\mathrm{z}(\mathbf{u}_\alpha)$, $\alpha = 1, \ldots, n$, data become $k = 1, \ldots, K$ sets of n data, a new variable that indicates the probability of being less than or equal to each threshold.

PGE 337 Data Analytics and Geostatistics

Lecture 14: Indicator Simulation

Lecture outline . . .

- Indicator Estimation and 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

Some Indicator Slides Modified from Prof. Clayton Deutsch's MinE612 Class.

The Indicator Formalism and Indictor Kriging (IK)

Then we can use these indicator-based probabilities for spatial estimates for each category or threshold.

- The IK process consists of discretizing the interval of variability of the continuous attribute z with a series of *K* threshold values $z_k, k = 1, \ldots, K$.
- A conditional CDF is built by assembling the *K* indicator kriging estimates

$$z_I^*(\mathbf{u}_\alpha; z_k) = P^*(Z(\mathbf{u}_\alpha) \leq z_k | n)$$

  – where $z_I^*(\mathbf{u}_\alpha; z_k)$ is the indicator kriging estimate at location, $\mathbf{u}_\alpha$ for threshold z_k

or for the categorical case we work with each category.

$$z_I^*(\mathbf{u}_\alpha; z_k) = P^*(Z(\mathbf{u}_\alpha) = z_k | n)$$

  – where $z_I^*(\mathbf{u}_\alpha; z_k)$ is the indicator kriging estimate at location, $\mathbf{u}_\alpha$ for threshold z_k

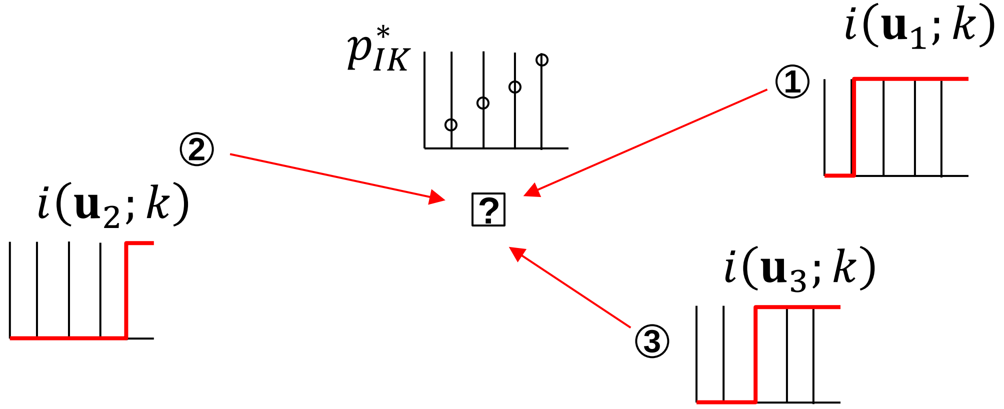
Conditional CDF / CCDF is the CDF at an unsampled locations estimated by local data.

The Indicator Formalism and Indictor Kriging (IK)

The indicator kriging estimator:

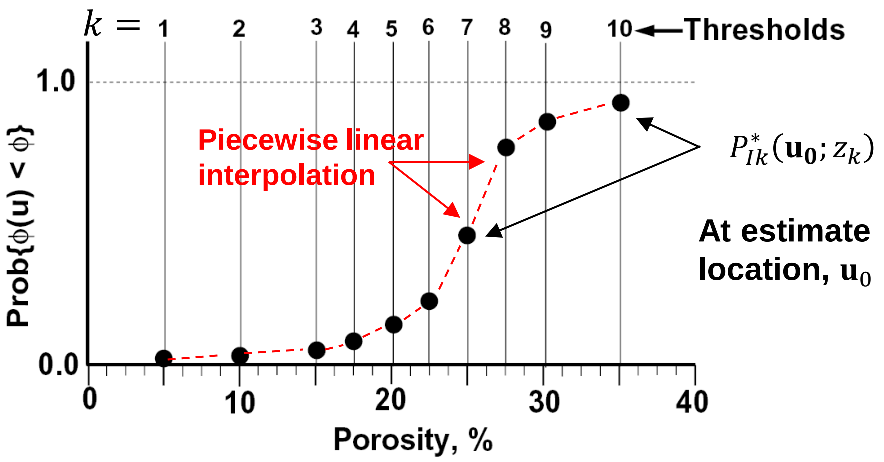
$$p_{IK}^*(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda_\alpha(k) \cdot i(\mathbf{u}_\alpha;k) + \left(1 - \sum_{\alpha=1}^{n} \lambda_\alpha(k)\right) \cdot p(k)$$

where $\lambda_\alpha(k)$ is the indicator kriging weight for data α and category / threshold k, $i(\mathbf{u}_\alpha;k)$ is the k category / threshold indicator transform of the data at location $\mathbf{u}_\alpha$ and $p(k)$ is the global or local mean categorical probability / continuous cumulative probability.

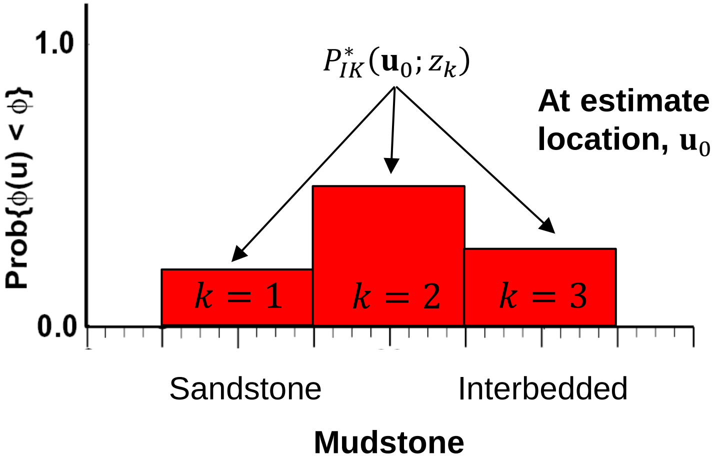


Result of Indicator Kriging (IK)

- **Indicator kriging estimates the conditional CDF at thresholds or categories at an unsampled location.**



Output for continuous indicator kriging at an estimate location, CDF estimated at thresholds.



Output for categorical indicator kriging at an estimate location, PDF.

- Establish a series of thresholds / categories:
  - May be related to critical thresholds and should enough thresholds to represent the local distributions of uncertainty
  - By estimating **probability ≤ for each threshold and interpolation** or **probabilities for each category** we are directly estimating the distribution of uncertainty at an unsampled location without distribution assumption.

Indictor Kriging (IK)
Unique Steps

Indicator kriging includes the following steps, for each threshold or category:

1. Application of indicator transform
2. Calculation of an indicator variogram
3. Indicator kriging to estimate the probability at an unsampled location

Indicator Variogram Models

How do we calculate spatial continuity for the indicator approach?

The indicator variogram:

$$\gamma_I(\mathbf{h}; z_k) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{u}_{\boldsymbol{\alpha}}; z_k) - i(\mathbf{u}_{\boldsymbol{\alpha}} + \mathbf{h}; z_k)]^2$$

- Note: for hard data the indicator transform $i(\mathbf{u}; z_k)$ is either 0 or 1, in which case the $[i(\mathbf{u}; z_k) - i(\mathbf{u} + \mathbf{h}; z_k)]^2$ is equal to 0 when the values at head and tail are $\leq \boldsymbol{z_k}$ (continuous) or $= \boldsymbol{z_k}$ (categorical) or 1 when they are different.

- Therefore, the indicator variogram is ½ the proportion of pairs that change! The indicator variogram can be related to probability of change over a lag distance.

Indicator Variogram Models

	Cumulative Class %	Nugget	Exponential			Spherical		
			3a	sill	anis	range	sill	anis
First Cutoff	6.1	0.17	18.0	0.50	2.3	100.0	0.33	10.0
Second Cutoff	15.5	0.11	47.7	0.54	3.3	150.0	0.35	10.0
Third Cutoff	23.3	0.13	90.0	0.58	5.0	170.0	0.29	10.0
Fourth Cutoff	32.7	0.13	90.0	0.61	6.2	160.0	0.26	10.0
Fifth Cutoff	43.4	0.12	108.0	0.68	6.2	91.0	0.20	7.1
Sixth Cutoff	57.7	0.12	108.0	0.68	7.1	85.0	0.20	7.1
Seventh Cutoff	75.4	0.22	144.0	0.69	6.7	66.0	0.09	5.9

- Standardize all points and models to a unit variance from the original variance.
- Note $p(1 - p)$ is the variance of an indicator with proportion of 1’s as p.

- Model the variograms with smoothly changing parameters for a *consistent* description
  - Common dataset → imparts consistency
  - Consistency between indicator variograms reduces order relations in IK/SIS (more later)
  - Allows straightforward interpolation of models for new cutoffs

Table from Pyrcz and Deutsch (2014)

Indicator Estimation and Simulation

Some Comments on the Indicator Approach for Estimation and Simulation

- A variogram is needed for each threshold → more difficult inference problem, however, there is greater flexibility
- Resulting model of uncertainty is not Gaussian (avoid maximum entropy issue)
- More readily integrates data of different types (more later on soft data)
- May use Sequential Indicator Simulation (will demonstrate next as extension to indicator kriging).
  - Commonly used for categorical variable like facies
  - Sometimes used for continuous variables like porosity

Indicator Kriging (IK) Take 2

Indicator Kriging Workflow

For all locations:

- For all thresholds:
  - find all relevant data: *n*
  - code all data as indicator data at the current threshold:

$$i(\mathbf{u}_\alpha; z_k) = Prob^*\left\{Z(\mathbf{u})_\alpha \le z_k\right\}$$

  - estimate the indicator function at the current threshold at this location:

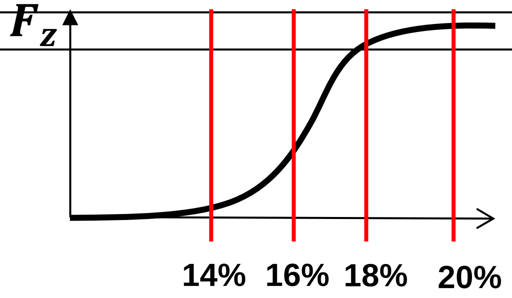
$$p^*_{IK}(\mathbf{u}; k) = \sum_{\alpha=1}^{n} \lambda_\alpha(k) \cdot i(\mathbf{u}_\alpha; k) + \left(1 - \sum_{\alpha=1}^{n} \lambda_\alpha(k)\right) \cdot p(k)$$

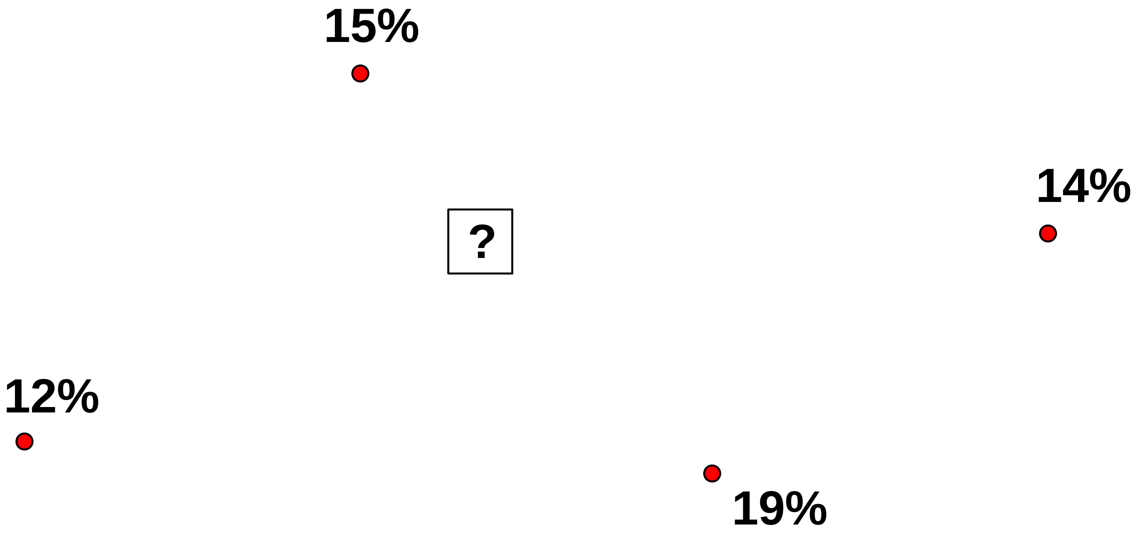
- Correct local distribution for order relations
- Use distribution for:
  - measure of uncertainty, probability intervals
  - probability to exceed given thresholds
  - E-type mean estimate, truncated statistics
  - stochastic simulation

Indicator Kriging (IK)
Take 2

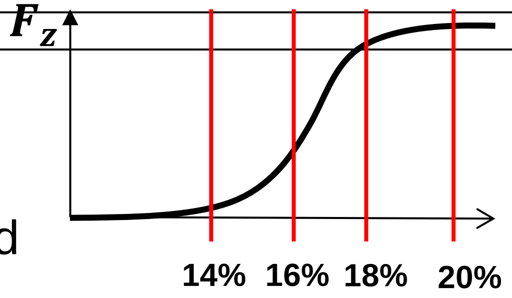
- Demonstration of Indicator Kriging

1. Assign thresholds





Indicator Kriging (IK)
Take 2



- Demonstration of Indicator Kriging

1. Assign thresholds
2. Calculate indicator variograms for each threshold
   a) indicator transform
   b) calculate variogram

$i(\mathbf{u}_1; z = z_1)$

0

?

$i(\mathbf{u}_4; z = z_1)$

1

$i(\mathbf{u}_2; z = z_1)$

1

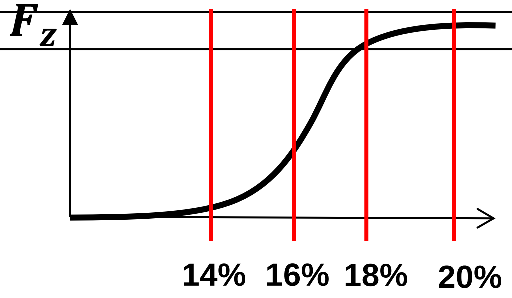
$i(\mathbf{u}_3; z = z_1)$

0

- Indicator Variogram:

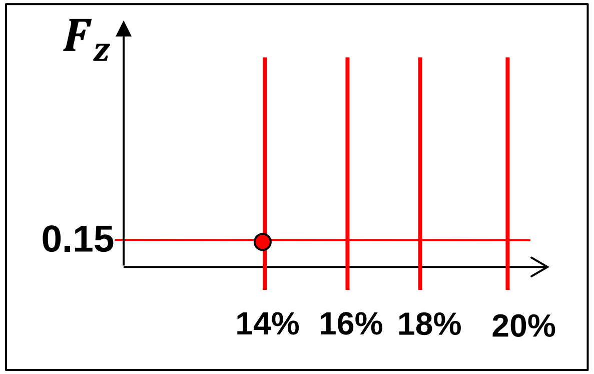
$$\gamma_I(\mathbf{h}; z_k) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [i(\mathbf{u}; z_k) - i(\mathbf{u}+\mathbf{h}; z_k)]^2$$

Indicator Kriging (IK)
Take 2



- Demonstration of Indicator Kriging

3. For each location:
4. For each threshold: $z = z_1 = 14\%$



$i(\mathbf{u}_1; z = z_1)$

0

?

$i(\mathbf{u}_4; z = z_1)$

0

$i(\mathbf{u}_2; z = z_1)$

1

$i(\mathbf{u}_3; z = z_1)$

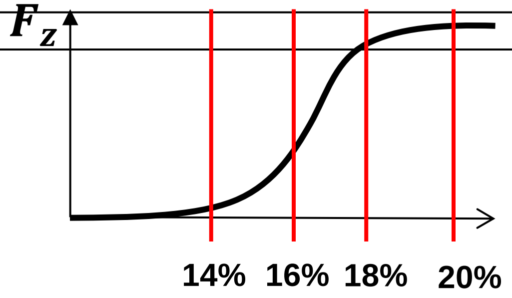
0

- Solve for indicator kriging estimate of probability porosity at location ? is < z_1.

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

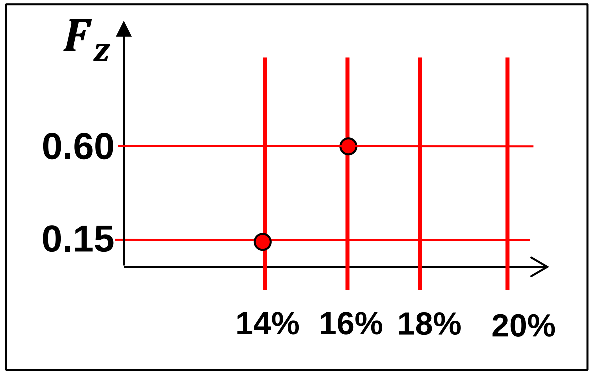
$$z^*(\mathbf{u}_1; z_1) = 0.15$$

Indicator Kriging (IK)
Take 2



- Demonstration of Indicator Kriging

3. For each location:
4. For each threshold: $z = z_2 = 16\%$



$i(\mathbf{u}_1; z = z_2)$
1

?

$i(\mathbf{u}_4; z = z_2)$
1

$i(\mathbf{u}_2; z = z_2)$
1

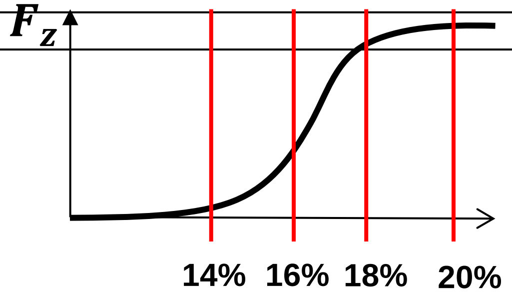
$i(\mathbf{u}_3; z = z_2)$
0

- Solve for indicator kriging estimate of probability porosity at location ? is < z_1.

$$\begin{bmatrix} C_{2_{1,1}} & \\ & C_{2_{n,n}} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{2_{0,1}} \\ C_{2_{0,n}} \end{bmatrix}$$

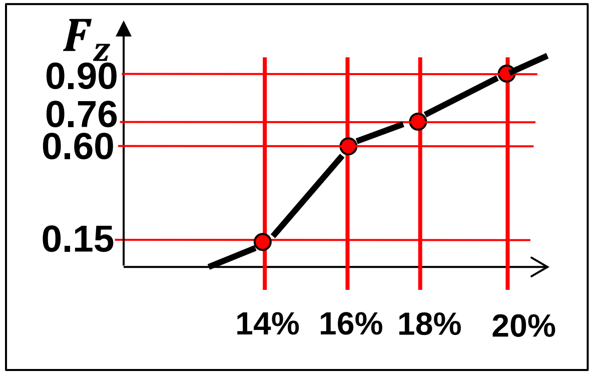
$$z^*(\mathbf{u}_0; z_2) = 0.60$$

Indicator Kriging (IK)
Take 2



- Demonstration of Indicator Kriging

3. For each location:
4. For each threshold: $z = z_4 = 20\%$



$i(\mathbf{u}_1; z = z_4)$

1

?

$i(\mathbf{u}_4; z = z_4)$

1

$i(\mathbf{u}_2; z = z_4)$

1

$i(\mathbf{u}_3; z = z_4)$

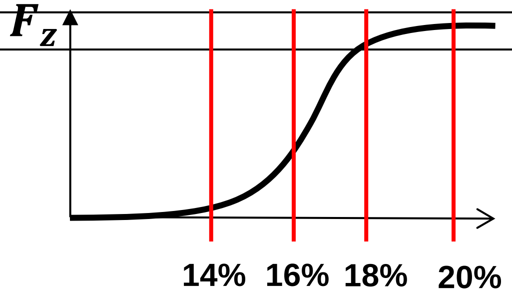
1

$$\begin{bmatrix} C_{2_{1,1}} & & \\ & & C_{2_{n,n}} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{2_{0,1}} \\ C_{2_{0,n}} \end{bmatrix}$$

$$z^*(\mathbf{u}_0; z_2) = 0.60$$

- Solve for indicator kriging estimate of probability porosity at location ? is < z_1.
- **Indicator kriging directly solves for the local distribution of CDF!**

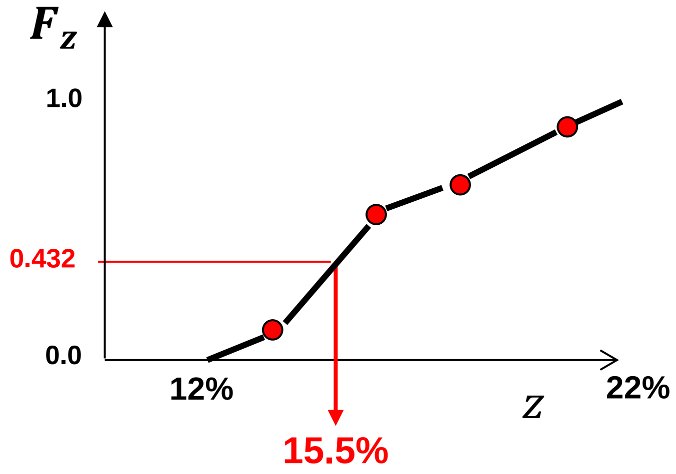
Sequential Indicator Simulation (SIS) Take 2



- Demonstration of Indicator Kriging

3. For each location:
4. For each threshold: $z = z_4 = 20\%$

Monte Carlo Simulation

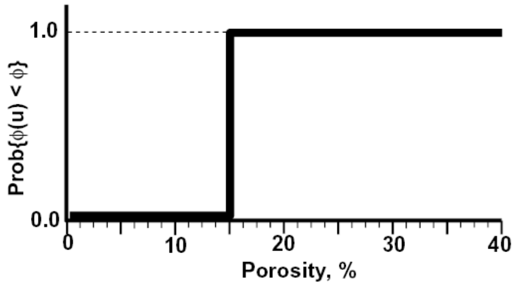


- Solve for indicator kriging estimate of probability porosity at location ? is < z_1.
- **Monte Carlo Simulation and treat the simulated value as data, indicator transform it and move to the next location on the random path!**

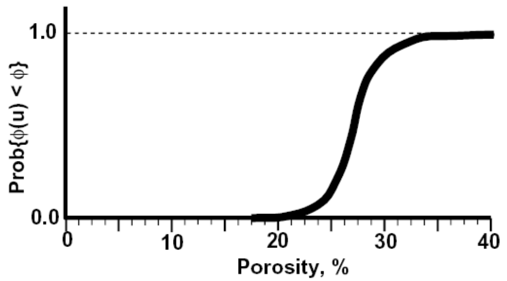
Data For Indicator Estimation and Simulation

Hard and Soft Data with the Indicator Approach

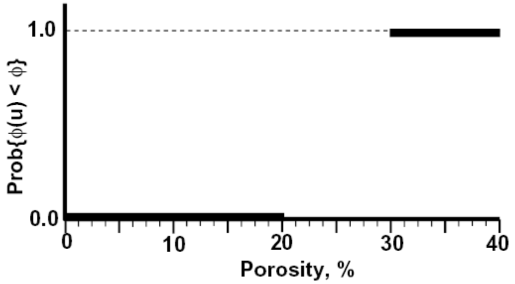
- Various constraints that may be applied to indicator coding



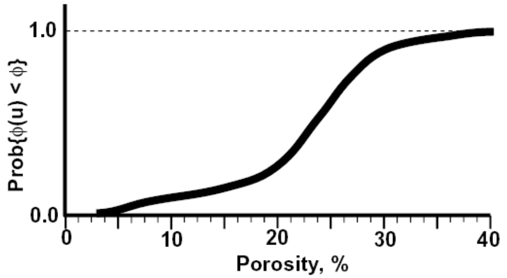
Local Hard Indicator Data



Prior Distribution From
Soft Secondary Data



Inequality Constraint From
Soft Secondary Data



Global Distribution From
Stationary Histogram

Recall:

$$I(\mathbf{u};z_k) = \begin{cases} 1, & \text{if } Z(\mathbf{u}) \le z_k \\ 0, & \text{otherwise} \end{cases}$$

Order Relations Correction

What is Order Relations? Why is it an Issue with Indicator Methods?

- Nonmonotonic behavior in continuous CDF (negative slope) or sum of categorical probabilities not equal to one.
- Cumulative probability at each threshold was solved with potentially a difference indicator variogram and with a separate kriging.
- There is no direct constraint to impose slope ≥ 0.0.
- Correction:
  - Continuous - take average of the forward and reverse constrained slope ≥ 0
  - Categorical - normalize sum of probabiiltes $= 0$

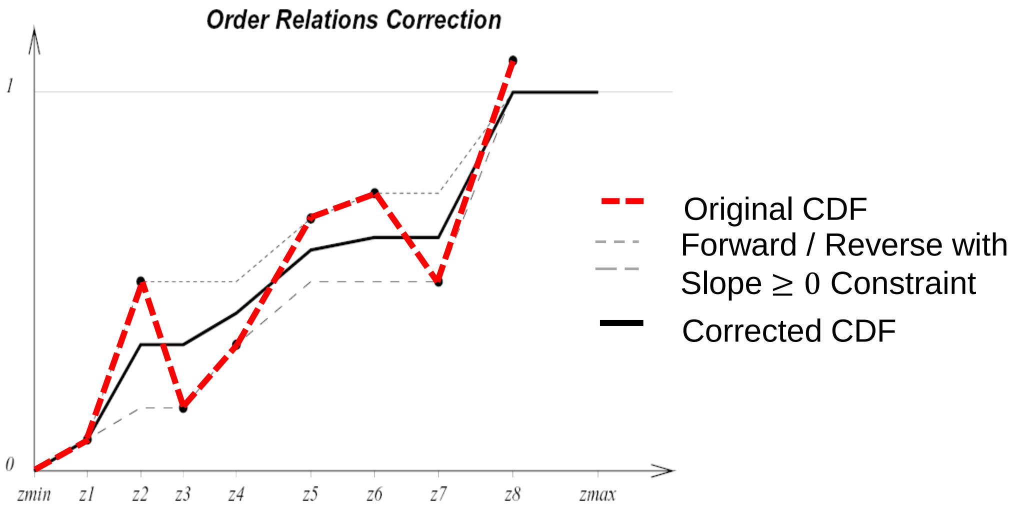
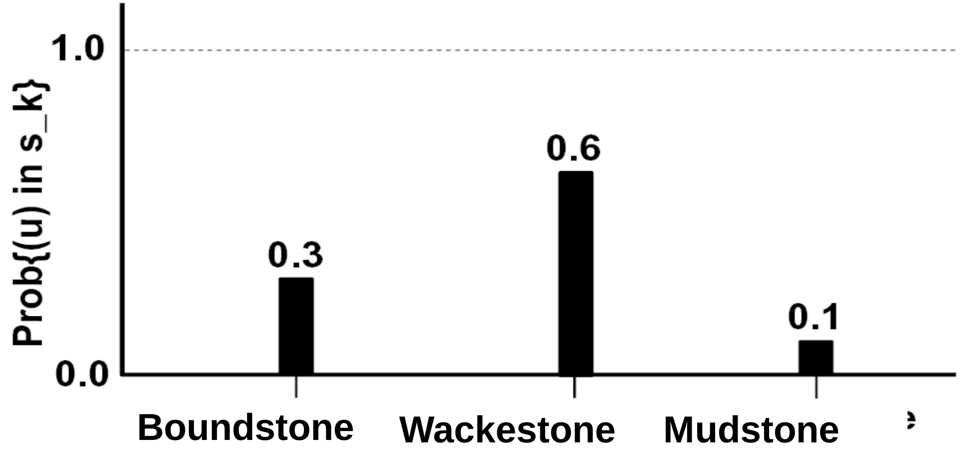


image from Pyrcz and Deutsch (2014)

Indicator Methods with Categorical Variables



- Consider a set of *K* categories: $z_k, k = 1, \ldots, K$
- Indicator variable for each location and each category:

$$i(\mathbf{u}_\alpha; z_k) = \begin{cases} 1 \text{ if category } z_k \text{ prevails at location } \mathbf{u}_\alpha \\ 0 \text{ if not} \end{cases}$$

- Same procedure for indicator kriging with continuos variable
- Order relations:

$$\text{all } i(\mathbf{u}; z_k)^*, k = 1, \ldots, K \text{ must be } \geq 0$$

$$\sum_{k=1}^{K} i(\mathbf{u}; z_k)^* = 1.0$$

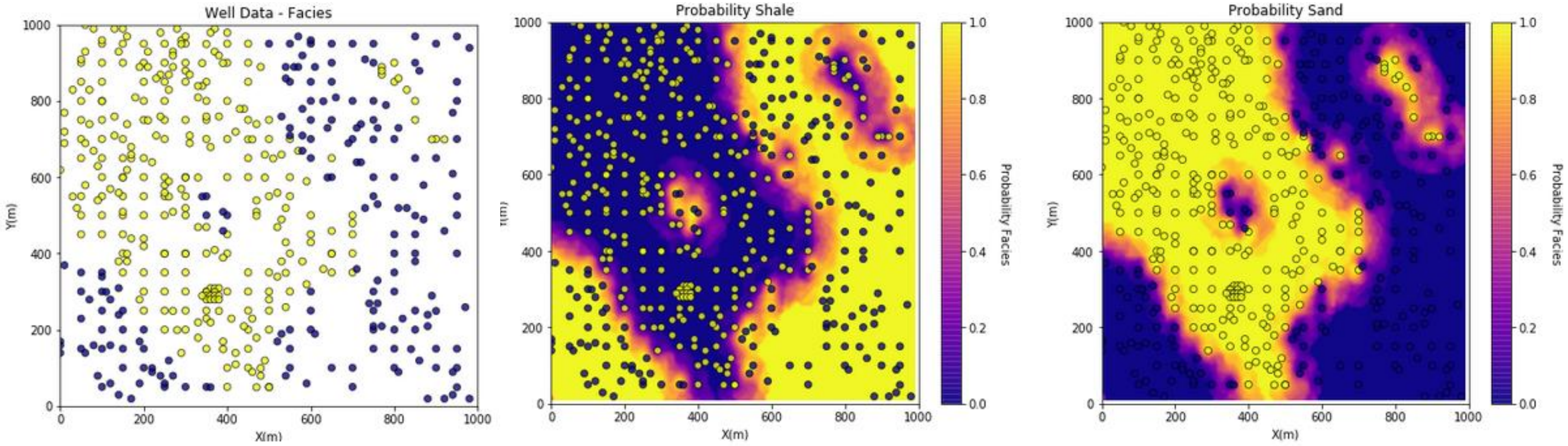
- Results are the probabilities that categories $z_k, k = 1, \ldots, K$ at location $\mathbf{u}$

Example of Categorical Indicator Kriging

Categorical indicator kriging

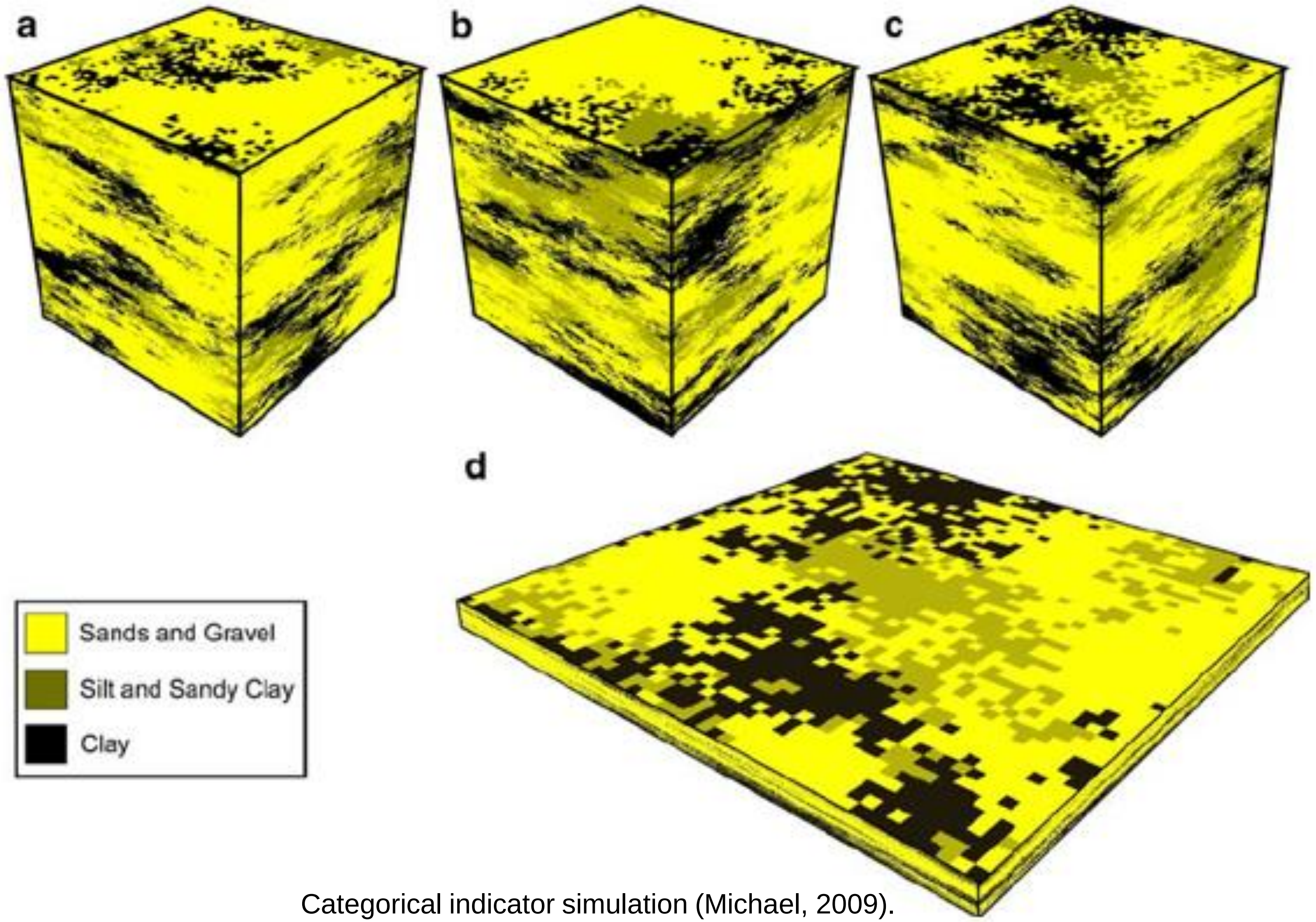
- Estimate the probability of sand and shale at locations between the wells
- Order relations correction for probability closure

$$P(sand; \mathbf{u}_\alpha) + P(shale; \mathbf{u}_\alpha) = 1.0, \forall\, \mathbf{u}_\alpha$$



Categorical indicator kriging.

Example of Categorical Indicator Simulation

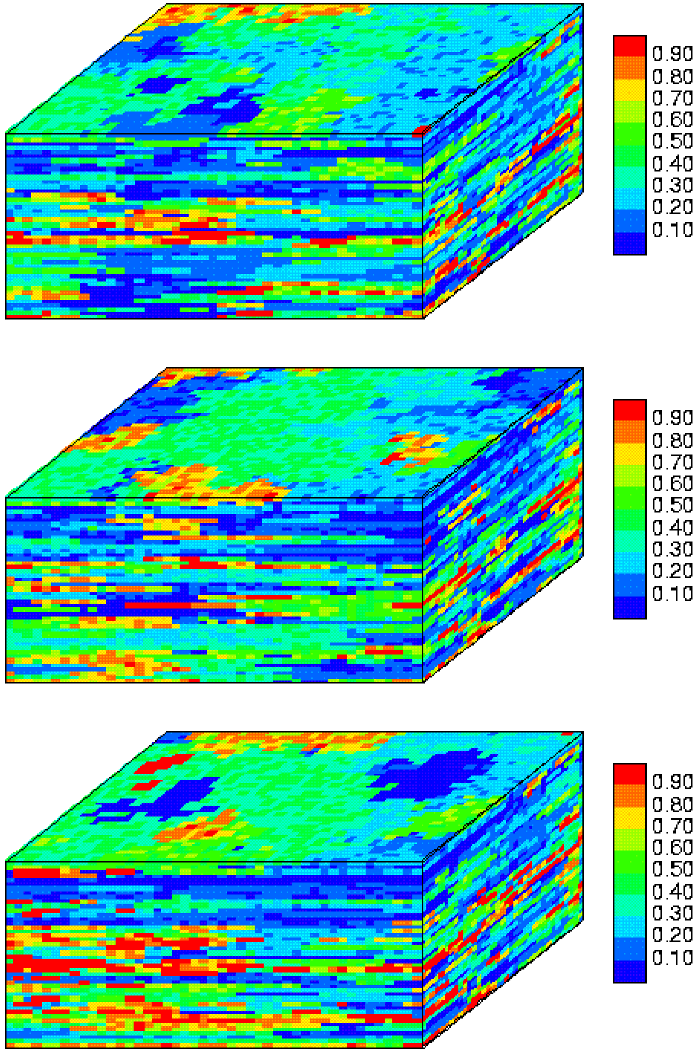


Categorical indicator simulation (Michael, 2009).

Example of Continuous Indicator Simulation

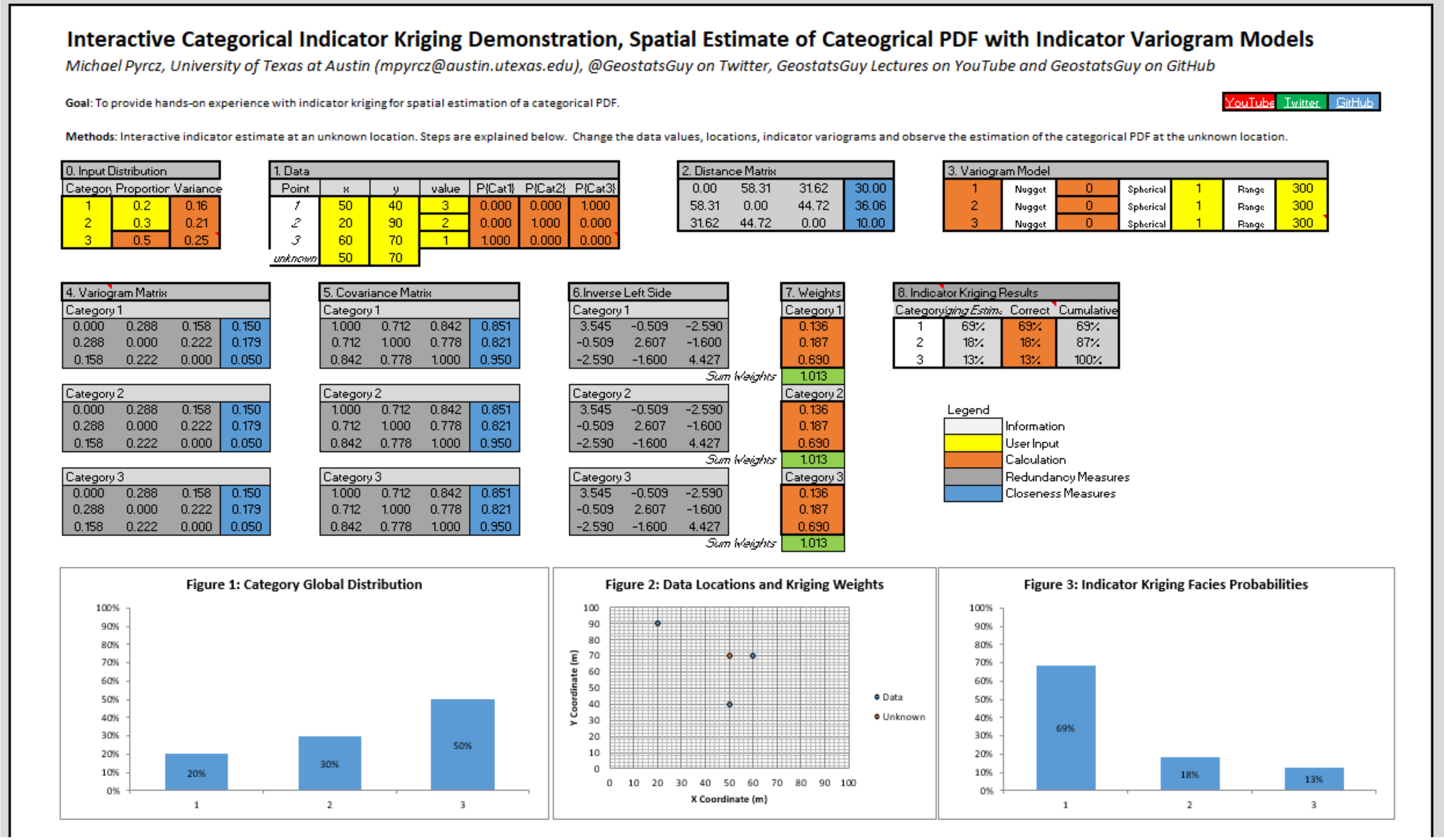
Continuous indicator simulation

- See the discontinuity across the continuous thresholds?
- My estimate is 0.7, 0.5, 0.3 and 0.1 were used.
- This is a known artifact with the continuous indicator simulation.



Three Vsh realizations for a shale group (Meehan and Verma, 1994)

Categorical Indicator Kriging Hands-on


Interactive Categorical Indicator Kriging Demonstration, Spatial Estimate of Cateogrical PDF with Indicator Variogram Models

Michael Pyrcz, University of Texas at Austin (mpyrcz@austin.utexas.edu), @GeostatsGuy on Twitter, GeostatsGuy Lectures on YouTube and GeostatsGuy on GitHub

YouTube Twitter GitHub

Goal: To provide hands-on experience with indicator kriging for spatial estimation of a categorical PDF.

Methods: Interactive indicator estimate at an unknown location. Steps are explained below. Change the data values, locations, indicator variograms and observe the estimation of the categorical PDF at the unknown location.

0. Input Distribution

Category	Proportion	Variance
1	0.2	0.16
2	0.3	0.21
3	0.5	0.25

1. Data

Point	x	y	value	P{Cat1}	P{Cat2}	P{Cat3}
1	50	40	3	0.000	0.000	1.000
2	20	90	2	0.000	1.000	0.000
3	60	70	1	1.000	0.000	0.000
unknown	50	70				

2. Distance Matrix

0.00	58.31	31.62	30.00
58.31	0.00	44.72	36.06
31.62	44.72	0.00	10.00

3. Variogram Model

1	Nugget	0	Spherical	1	Range	300
2	Nugget	0	Spherical	1	Range	300
3	Nugget	0	Spherical	1	Range	300

4. Variogram Matrix

Category 1			
0.000	0.288	0.158	0.150
0.288	0.000	0.222	0.179
0.158	0.222	0.000	0.050

Category 2			
0.000	0.288	0.158	0.150
0.288	0.000	0.222	0.179
0.158	0.222	0.000	0.050

Category 3			
0.000	0.288	0.158	0.150
0.288	0.000	0.222	0.179
0.158	0.222	0.000	0.050

5. Covariance Matrix

Category 1			
1.000	0.712	0.842	0.851
0.712	1.000	0.778	0.821
0.842	0.778	1.000	0.950

Category 2			
1.000	0.712	0.842	0.851
0.712	1.000	0.778	0.821
0.842	0.778	1.000	0.950

Category 3			
1.000	0.712	0.842	0.851
0.712	1.000	0.778	0.821
0.842	0.778	1.000	0.950

6. Inverse Left Side / 7. Weights

Category 1			Category 1
3.545	-0.509	-2.590	0.136
-0.509	2.607	-1.600	0.187
-2.590	-1.600	4.427	0.690
		Sum Weights	1.013

Category 2			Category 2
3.545	-0.509	-2.590	0.136
-0.509	2.607	-1.600	0.187
-2.590	-1.600	4.427	0.690
		Sum Weights	1.013

Category 3			Category 3
3.545	-0.509	-2.590	0.136
-0.509	2.607	-1.600	0.187
-2.590	-1.600	4.427	0.690
		Sum Weights	1.013

8. Indicator Kriging Results

Category	Kriging Estim.	Correct	Cumulative
1	69%	69%	69%
2	18%	18%	87%
3	13%	13%	100%

Legend: Information; User Input; Calculation; Redundancy Measures; Closeness Measures

Exercises:

1. Set the variogram range shorter than data spacing.
2. Place a datum on the unknown location.
3. Set all the data to the same category.

Files/Examples/Indicator_Kriging_Demo.xlsx

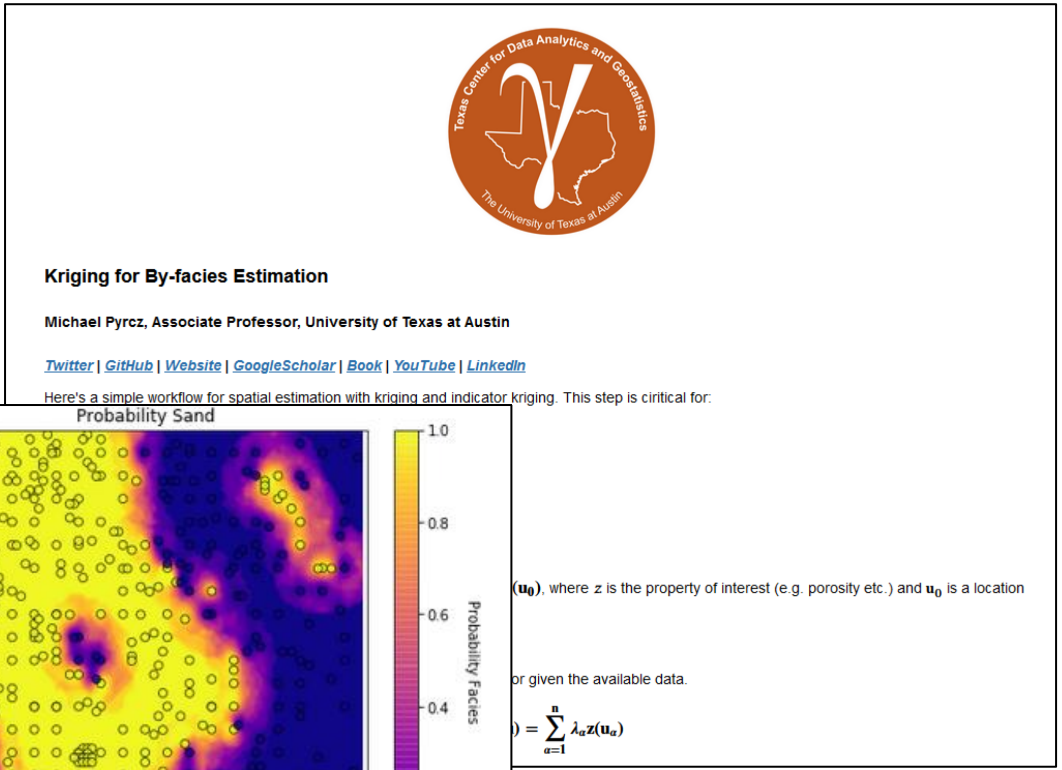
Indicator Kriging in Python

Variogram Calculation Workflow in Python

Walkthrough and try to:

- Change the variogram and search parameters.

- File is:
- GeostatsPy_kriging.ipynb



Kriging for By-facies Estimation

Michael Pyrcz, Associate Professor, University of Texas at Austin

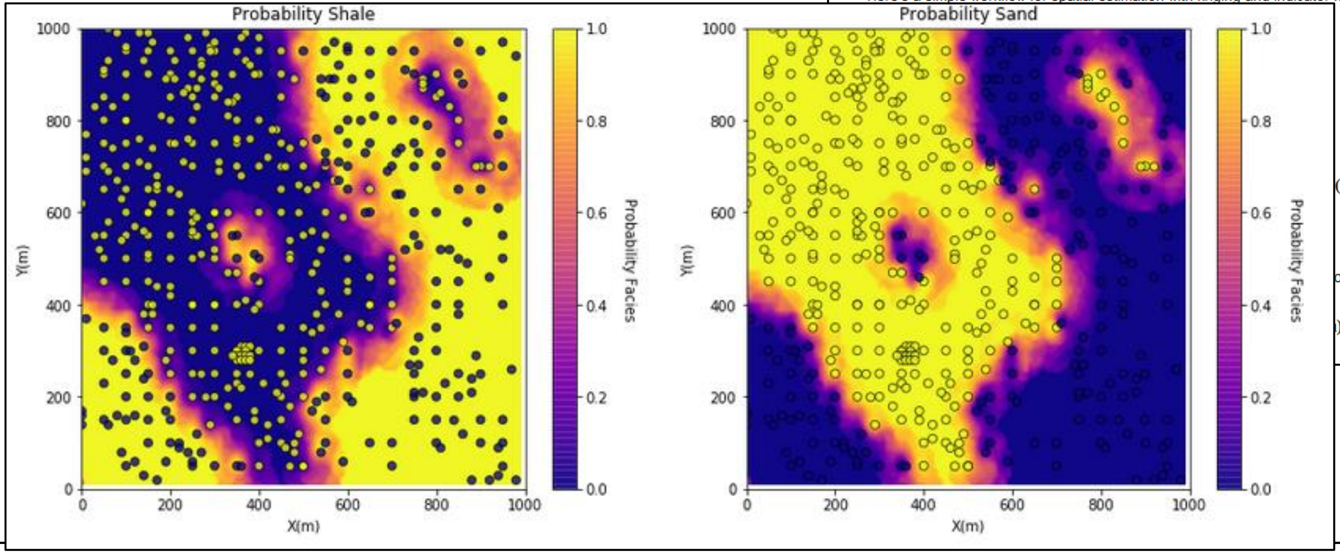
Twitter | GitHub | Website | GoogleScholar | Book | YouTube | LinkedIn

Here's a simple workflow for spatial estimation with kriging and indicator kriging. This step is cirical for:

$(\mathbf{u_0})$, where z is the property of interest (e.g. porosity etc.) and $\mathbf{u_0}$ is a location

or given the available data.

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



Summary of Indicators

Indicator Transforms

- Probability coding of the data and estimates
- Indicator transform of continuous variable with thresholds
- Indicator transform of categorical variable, by-category

Indicator for Spatial Estimation

- Applied to estimate local CDF without assuming Gaussian distribution

Indicator for Spatial Simulation

- Applied more often for categorical simulation
  - Replace simple / ordinary kriging with indicator kriging in the sequential context
  - i.e. Monte Carlo from indicator estimated CDF and transform simulated value for each threshold and use as data (sequential approach).

PGE 337 Data Analytics and Geostatistics

Lecture 14: Indicator Simulation

Lecture outline . . .

- Indicator Methods
- Indicator Estimation and Simulation

Michael Pyrcz, The University of Texas at Austin

Some Indicator Slides Modified from Prof. Clayton Deutsch's MinE612 Class.