High Resolution Facies Modeling in Presence of Large-Scale Probabilistic Data

Sahyun Hong and Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil and Environmental Engineering
University of Alberta

Facies modeling precedes petrophysical property modeling since porosity and permeability are highly related to facies type. Integration of secondary data with direct facies observations provides more realistic models. Cokriging is a traditional geostatistical way to incorporate secondary variable. However, generalized linear regression model between primary and secondary variable is inadequate especially when mixing indicator primary and continuous secondary variable. The scale inconsistency between primary and secondary data is also an issue. Secondary data have limited vertical resolution, but relatively high areal resolution. Primary data has high vertical resolution, but limited areal coverage. In this work, we propose a new method based on Bayesian updating rule to combine primary with secondary data. Large-scaled probability cubes are generated first and fine-scaled probability cube is updated with large-scaled probabilistic data. Sequential simulation is performed to generate multiple realizations based on the updated distribution.

Introduction

The identification and mapping of facies is important to reliable reservoir characterization because the porosity and permeability are highly correlated with facies type. Knowledge of facies constraints the range of variability in porosity and permeability and moreover facies type constraints fluid saturation. In this work, our concern is to construct realistic 3-D distributions of the facies that may be used in subsequent reservoir decision making.

Facies identification is primarily based on the inspection of well-log data that provide exact facies type at only well location. Well-log data has limited horizontal resolution although high resolution in vertical. Seismic data usually of great value in constraining facies models; it is a really extensive over the reservoir and can be sensitive to facies variations. Seismic-derived attribute such as acoustic impedance of course show variations according to facies type. Integration of well-data and seismic attributes could give more reliable construction of 3-D facies distribution.

The use of co-kriging is a traditional way to integrate primary and secondary variable. In the framework of co-kriging, integration of secondary data is based on the auto-covariance and the cross-covariance between primary and secondary variables. Co-kriging approach, however, requires tedious joint modeling of primary and secondary variables. Besides, the joint modeling of variogram is inadequate especially when modeling mixture of discrete primary and continuous secondary data.

In this work, we propose a new technique to integrate hard primary and soft secondary data. A new method does not depend on co-kriging and does not require joint modeling of variogram. Instead, we build a local posteriori ccdf of each facies with Bayesian updating rule over all grids. Finally, sequential simulation is performed based on the posteriori ccdf.

Methodology

This paper aims to stochastic model 3-D facies by combining well hard data and soft secondary data. The posteriori ccdf of each facies is built and sequential simulation is made based on the ccdf over all modeling grids. To simplify the discussion, we consider two facies coded as integer 0 and 1. Multiple attributes of seismic data are used as secondary variable. Seismic amplitude is a representative attribute derived from raw seismic data. Acoustic impedance also can be obtained by seismic data inversion and it is a good measurement of facies. We specify secondary variables as $y_1(\mathbf{u})$ and $y_2(\mathbf{u})$, $\mathbf{u} \in$ entire domain. Facies type is completely specified by the indicator i being either 0 or 1. Associated with each modeling grid, secondary variables exist which is assumed to provide indirect measurement of facies. Figure 1 shows the schematic diagram of scale difference between primary and secondary in vertical direction. Scale in horizontal direction is assumed to be same for the variables.

The target probability to be estimated at the location **u** is expressed by

$$P^*_{i=0}(\mathbf{u}) = P(i=0 \mid i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u}))$$

$$P^*_{i=1}(\mathbf{u}) = P(i=1 \mid i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u}))$$

We call a target probability $P^*_{i=0}(\mathbf{u})$ or $P^*_{i=1}(\mathbf{u})$ as the posteriori probability or updated probability conditioned to all information. $i_1,\ldots,i_n$ indicate surrounding hard primary data of the location **u**. $y_1(\mathbf{u})$ and $y_2(\mathbf{u})$ indicate secondary variables at co-location **u**. We assumed the co-located secondary variables have greater impact on the estimation of facies at the location than nearby secondary variables. This assumption is reasonable since the secondary variables have usually larger volume support than modeling cell size hence co-located secondary variables have maximum information about the facies estimated at modeling grid **u**.

The inference of the posteriori probability is divided into several steps as following Bayesian decomposition,

$$P^*_{i=0}(\mathbf{u}) = P(i=0, i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u})) \frac{1}{P(i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u}))}$$

$$= P(i=0 \mid i_1,...,i_n) \frac{P(i_1,...,i_n)}{P(i=0, i_1,...,i_n)} \frac{P(i=0, i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u}))}{P(i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u}))}$$

$$= P(i=0 \mid i_1,...,i_n) P(y_1(\mathbf{u}), y_2(\mathbf{u}) \mid i=0, i_1,...,i_n) \frac{P(i_1,...,i_n)}{P(i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u}))}$$

The second term, $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 0,i_1,\ldots,i_n)$, is approximated as $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 0)$ because the estimated facies type at the location **u** is only correlated the secondary variables at that location. Thus, we have

$$P^*_{i=0}(\mathbf{u}) = P(i=0 \mid i_1,...,i_n) P(y_1(\mathbf{u}), y_2(\mathbf{u}) \mid i=0) \bullet C(\mathbf{u}) \quad \text{for facies } i=0 \tag{1}$$

$$P^*_{i=1}(\mathbf{u}) = P(i=1 \mid i_1,...,i_n) P(y_1(\mathbf{u}), y_2(\mathbf{u}) \mid i=1) \bullet C(\mathbf{u}) \quad \text{for facies } i=1$$

where, the unknown term $C(\mathbf{u}) = P(i_1,...,i_n) / P(i_1,...,i_n, y_1(\mathbf{u}), y_2(\mathbf{u}))$ is location dependent, but is independent of facies. The posteriori probability equation (1) can be interpreted as the product of two conditional probabilities. The first term is a probability of facies given primary hard information and the second term is joint probability of y_1 and y_2 at location **u** given specific facies type. The first term is referred to as *a priori* distribution that is obtained from only primary well data. Indicator kriging is a one way to build a priori distribution. The second term is referred to as a *secondary likelihood* distribution. Likelihood distribution needs to model joint facies-conditional probability, $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 0)$ and $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 1)$. We estimated the secondary likelihood using Lamda-model discussed in paper 105 and 205 in this report. The Lamda-model decompose the joint conditional probability $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 0)$ and $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 1)$ into the product of each conditional probability with redundancy weights λ_1 and λ_2,

$$P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 0)= P(y_1(\mathbf{u})|i = 0)^{\lambda 1} \cdot P(y_2(\mathbf{u})|i = 0)^{\lambda 2} \quad \text{for facies } i = 0$$

$$P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 1)= P(y_1(\mathbf{u})|i = 1)^{\lambda 1} \cdot P(y_2(\mathbf{u})|i = 1)^{\lambda 2} \quad \text{for facies } i = 1$$

How to optimize the redundancy weights are discussed in paper 105 and 205 in this report so that we do not want to introduce details here. The constant C(**u**) term is obtained by the probability property $P^*_{i=0}(\mathbf{u}) + P^*_{i=1}(\mathbf{u}) = 1$.

Overall, the posteriori probability is the combination of the influence of primary and secondary variables. Advantage of this approach is that we can adopt the most appropriate method to build a priori and secondary likelihood for the considered primary and secondary data. For instances, a priori distribution can be obtained using training image provided that the underlying geology is too complex to model using indicator kriging. To obtain secondary likelihood is similar to the process of supervised image classification so that many image processing techniques or statistical methods can be applied even though those techniques are demanding computational efforts.

Sequential Simulation

Locally updated probability using primary and secondary information is obtained by equation (1). Now we advanced the new approach to be applicable in the sequential simulation context. Let us consider the first two simulation node be **u’** and **u’’**. The local posteriori distribution of facies at **u**’ is built,

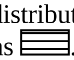
$$P^*_{i=0}(\mathbf{u}') = P(i=0 \mid i_1,...,i_n) P(y_1(\mathbf{u}'), y_2(\mathbf{u}') \mid i=0) \bullet C(\mathbf{u}')$$

$$P^*_{i=1}(\mathbf{u}') = P(i=1 \mid i_1,...,i_n) P(y_1(\mathbf{u}'), y_2(\mathbf{u}') \mid i=1) \bullet C(\mathbf{u}')$$

$i_1,\ldots,i_n$ indicate the surrounding primary data nearby the simulation node **u**’. Random number in [0,1] is generated and assigned to the posteriori ccdf. Facies realization is created at node **u**’. For a next node **u**’’, the local posteriori distribution at **u**’’ given all conditioning information should be built. The posteriori probabilities are expressed by equation (1) ,

$$P^*_{i=0}(\mathbf{u}'') = P(i=0 \mid i_1,...,i_n, i_{u'}) P(y_1(\mathbf{u}''), y_2(\mathbf{u}'') \mid i=0) \bullet C(\mathbf{u}'')$$

$$P^*_{i=1}(\mathbf{u}'') = P(i=1 \mid i_1,...,i_n, i_{u'}) P(y_1(\mathbf{u}''), y_2(\mathbf{u}'') \mid i=1) \bullet C(\mathbf{u}'')$$

A priori distribution at node **u**’’ is constructed based on $i_1,\ldots,i_n$ data and the previously simulated facies type $i_{u'}$. However, the tricky case is that the simulation node **u**’ and **u**’’ are inside within the same large secondary block. Let’s see the Figure 2 in detail. Scale of primary and secondary in lateral is same as modeling grid size. Vertical scale of secondary variables is four times than that of primary variable so that four small cells are included within one secondary large cell. The local posteriori distribution of each facies at first node **u**’ is built using equation (1). Facies realization at node **u’** is drawn as . In the next simulation node **u**’’, the conditioning information comprehend original primary data $i_1,..,i_n$ and the previously simulated $i_{u'}$ and co-located secondary variables. In this example, y_1(**u’’**) and y_2(**u**’’) are equal to y_1(**u**’) and y_2(**u**’), respectively since **u**’ and **u**’’ are included in the same secondary variable block. In simulation node **u**’’, the approximation is not valid anymore. The co-located secondary variables are related to co-located facies as well as the previously simulated facies (see the valid approximation equation of Figure 2 (B)).

The secondary likelihood distribution that we derived in equation (1) is different from the new secondary likelihood noted as valid approximation in Figure 2 (B). New form of likelihood distribution is difficult to be estimated. It requires modeling of joint probability given facies type at multiple locations, such as $P(y_1(\mathbf{u}'''), y_2(\mathbf{u}''') \mid i=0, i_{\mathbf{u}'}, i_{\mathbf{u}''}) \bullet C(\mathbf{u}''')$ where $\mathbf{u}'''$ is the current simulation node, $i_{\mathbf{u}'}$ and $i_{\mathbf{u}''}$ are previously simulated facies, and cell locations $\mathbf{u}', \mathbf{u}'', \mathbf{u}'''$ are included within same large secondary cell. We have proposed a sequential indicator simulation algorithm that can still use the equation (1) accounting for a priori and likelihood distribution. The posteriori distribution is estimated over the entire modeling cells first. In estimation mode, there is no scale inconsistency related problem discussed in Figure 2. After that, spatially correlated probability values are generated and these values are assigned to the inverse ccdf to create facies realization. This simulation process is called *p*-filed simulation which has the following procedures:

1. Calculate conditional probabilities using secondary variables.

e.g., $P(y_1(\mathbf{u})|i = 0)$, $P(y_2(u)|i = 0)$

2. Integrate conditional probability to generate secondary likelihood. This integration process could use Lamda-model, simple PR-model and Tau-model(See the reference paper)

   e.g. $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 0) = \Phi[\ P(y_1(\mathbf{u})|i = 0), P(y_2(u)|i = 0)\ ]$, Φ is an integration model

3. Apply indicator kriging or training image to primary data in order to estimate a priori probability

   e.g. $P(i = 0|i_1,\ldots,i_n)$

4. Build the posteriori probability using equation (1) over all modeling cells.

5. Generate spatially correlated probability values [0,1] over cells, $p(\mathbf{u})$

6. Apply correlated random p-value to the inverse cumulative posteriori distribution built in step 4 and assign facies realization to each simulation node.

   e.g. $i_u = F^{*-1}(p(\mathbf{u}))$, i_u is a facies realization at **u**, F^{*-1} is an inverse ccdf of the posteriori distribution

7. Finish the first facies realization

8. Go to step 5 and 7 for the next realization

This sequential simulation process involves two main steps; one is to build the posteriori distribution once regardless of the number of realization. The other is to generate correlated probability field according to realization number and this p-value accounts for spatial inter-dependency.

Examples

A 50 × 50 × 20 3-D synthetic test examples are considered for the evaluation of a new method. One Gaussian variable was simulated first with 50 × 50 × 20 grids. Other two Gaussian variables were simulated at 50 × 50 × 5 with retaining correlation($\rho = 0.65$) to the first simulation. The first simulation is used to make primary hard data and the other simulations are used as soft secondary data. Over the entire grids, we assigned integer code 0(noted as sand) if the simulated value is less than 0 and assigned code 1(noted as shale) if the simulated value is greater than 0. And 20 wells were randomly selected as primary sampling location. Figure 3 describes primary data. Extent of modeling area is 50m × 50m × 20m and modeling cell size is 1m × 1m × 1m. Each well samples facies type at every 1m depth hence total primary indicator data is 400(=20 well locations × 20 samples at each well). Figure 4 represents the variogram of facies 0 in horizontal and in vertical direction. Figure 5 and 6 illustrates the simulated secondary 1 and secondary 2 data. Horizontal resolution is same as primary data(or modeling grid resolution), but vertical resolution of secondary data is coarser than primary data. Total sample number is 12500(= 50 × 50 × 5).

As a first step, we built facies-conditional probability such as $P(y_1(\mathbf{u})|i = 0)$, $P(y_1(\mathbf{u})|i = 1)$ and $P(y_2(\mathbf{u})|i = 0)$, $P(y_2(\mathbf{u})|i = 1)$. Secondary data values are extracted at facies 0 sampled locations and histogram of those values is built. Smoothed histogram is produced based on the data histogram. Figure 7 shows the calibration of facies proportion using secondary variables. Solid line is a smoothed histogram estimated based on the data histogram corresponding to facies type(0 or 1). From the smoothed histogram, we extracted a probability $P(i = 0\ |y_1(\mathbf{u}'))$ or $P(i = 0\ |y_2(\mathbf{u}'))$ given secondary variable y_1 or y_2 at location **u'**. Facies-conditional probability at location **u'** is obtained by

$$P(y_1(\mathbf{u}')\,|\,i=0)=\frac{P(i=0\,|\,y_1(\mathbf{u}'))P(y_1(\mathbf{u}'))}{P(i=0)} \text{ and } P(y_1(\mathbf{u}')\,|\,i=1)=\frac{P(i=1\,|\,y_1(\mathbf{u}'))P(y_1(\mathbf{u}'))}{P(i=1)}$$

$$P(y_2(\mathbf{u}')\,|\,i=0)=\frac{P(i=0\,|\,y_2(\mathbf{u}'))P(y_2(\mathbf{u}'))}{P(i=0)} \text{ and } P(y_2(\mathbf{u}')\,|\,i=1)=\frac{P(i=1\,|\,y_2(\mathbf{u}'))P(y_2(\mathbf{u}'))}{P(i=1)}$$

These calibrated probability is then integrated to produce the secondary likelihood probability $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 0)$ and $P(y_1(\mathbf{u}),y_2(\mathbf{u})|i = 1)$ through data integration model. We used the Lamda-model to integrate and Figure 8 represents the integrated facies distribution in 3-D. Estimated λ_1 and λ_2 weights through the Lamda-model are 0.298 and 0.495, respectively.

The posteriori distribution is the product of a priori and likelihood distribution. Figure 9 illustrates the posteriori probability of facies 0 in 3-D. A priori distribution is obtained by indicator kriging and integrated with secondary likelihood to produce the updated probability. One realization of correlated probability value is shown in the Figure 9.

Two facies realizations using the considered simulation method are shown in the Figure 10. To check the update after considering secondary variable, we plot realizations from SISIM in the left column. As a way to evaluate the simulation method, spatial structures and global proportions of each facies should be reproduced. Figure 11 shows the reproduction of variograms. Solid line is the modeled variogram and dashed lines are calculated variogram from 10 realization. Dots indicate experimental variogram calculated using the original sparse data. Variogram reproduction in horizontal and in vertical direction is checked. 10 realizations obtained from the proposed method show low uncertainty with retaining good accordance to the modeled variogram. Reproduction of global proportion is also checked with 10 realizations.

Table 1: Global proportions of facies are estimated from 10 realizations

	Original data	10 realization with SISIM	10 realization with the proposed method
Facies 0	0.4325	0.470	0.419
Facies 1	0.5675	0.531	0.581

Discussions and Conclusions

3-D facies model was built using primary data and large-scaled secondary data. The considered method is based on Bayesian updating process. In Bayesian updating process, a priori distribution and secondary likelihood distribution were obtained separately. The updated distribution is the product of the two distributions. Sequential facies simulations were generated based on the posteriori distribution that resulted from the proposed method. Scale inconsistency between primary and secondary caused a problem in sequential simulation context. Thus, we adopted a p-filed simulation approach. Synthetic 3-D examples showed that the proposed method combines primary and secondary data and realizations based on the updated distribution preserved spatial structures and reproduced global proportions of facies.

References

Deutsch, C.V. and Journel, A.G., 1998, *GSLIB: Geostatistical Software Library: and User's Guide*. Oxford University Press, New York, 2nd Ed.

Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation, Oxford University Press, New York.

Hong, S. and Deutsch, C.V., 2007 Methods for Integrating Conditional Probabilities for Geostatistical Modeling, *Centre for Computatinal Geostatistics Annual Report 9 (paper-105)*.

Doyen, P.M. and Psaila, D.E., 1994, Bayesian Sequential Indicator Simulation of Channel Sands From 3-D Seismic Data in The Oseberg Field, Norwegian North Sea, SPE 28382 *SPE* 69th *Annual Technical Conference and Exhibition SPE*.

Hong, S. and Deutsch, C.V., 2007 Data Integration for 2-D Areal Proportions of Facies, *Centre for Computatinal Geostatistics Annual Report 9 (paper-205)*.

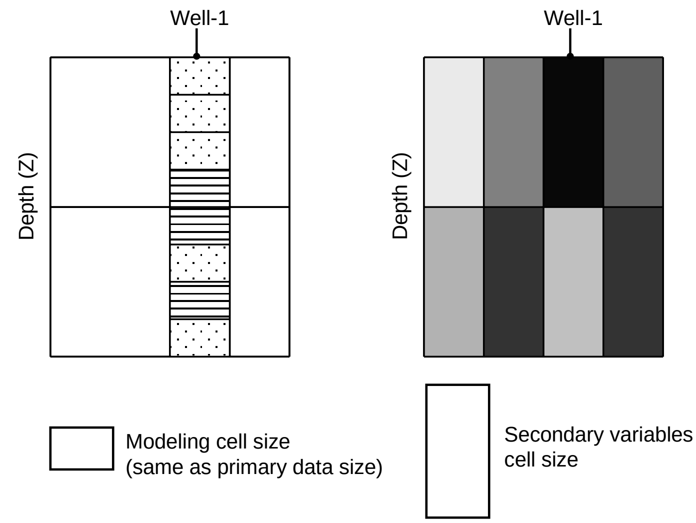


Figure 1: Schematic diagram of scale difference in vertical direction between primary and secondary variables. Different gray color represents different secondary values.

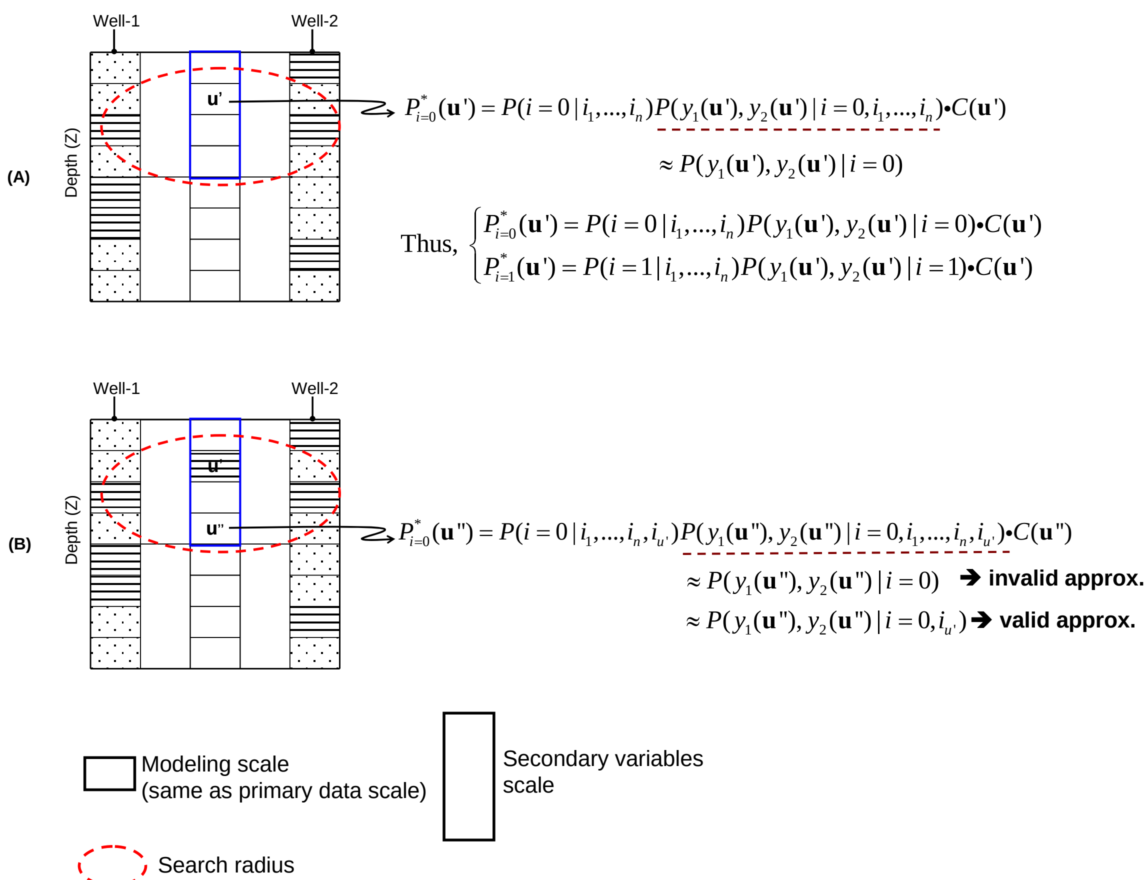
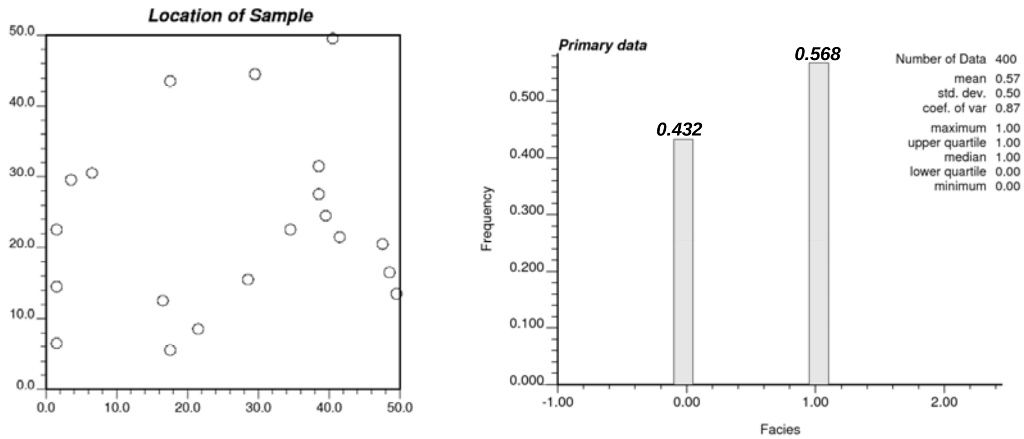


Figure 2: Scale inconsistency between primary and secondary variables makes it difficult to perform sequential indicator simulation using the equation (1)



Primary Indicator Data		
Spatial extent	X-axis	0 – 50 m
	Y-axis	0 – 50 m
	Z-axis	0 – 20 m
# of sampling well location		20
# of samples at each well		20
Data type		Sparsely sampled integer Sand (coded as 0) Shale (coded as 1)
Total # of samples		400
Sampling cell size		$1 \times 1 \times 1$ (in meter)

Figure 3: Description of primary data.

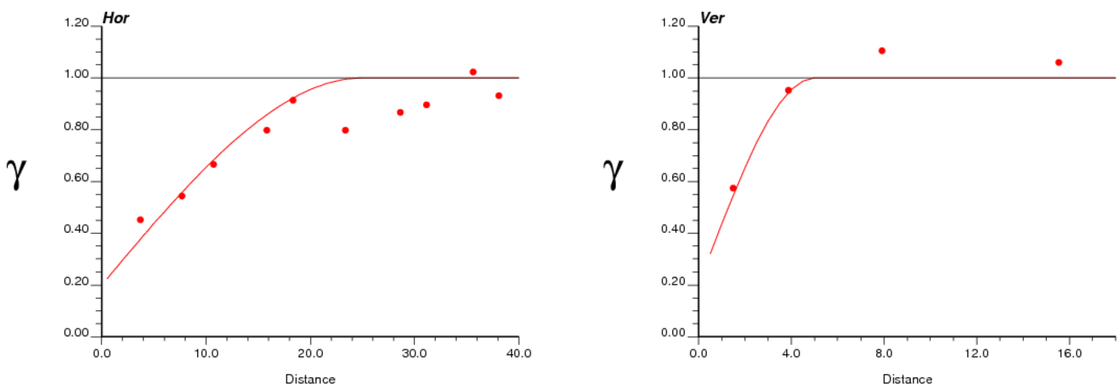
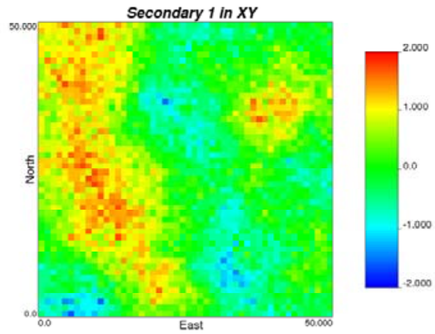
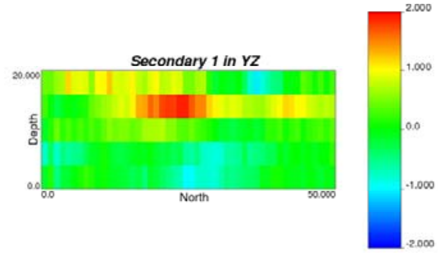
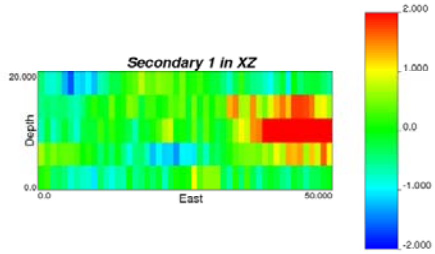


Figure 4: Standardized variogram of facies 0 in horizontal and in vertical direction. Variogram of facies 1 is identical to the variogram of facies 0.

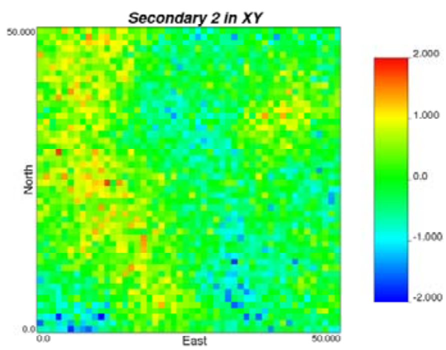


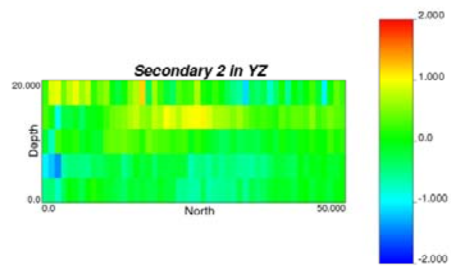


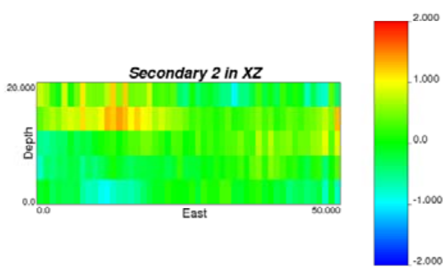


Secondary 1 Data		
Spatial extent	X-axis	0 – 50 m
	Y-axis	0 – 50 m
	Z-axis	0 – 20 m
Data Types		Exhaustively sampled continuous values
Sampling cell size		$1 \times 1 \times 4$ (in meter)
Total # of samples		12500

Figure 5: Simulated secondary 1 data.







Secondary 2		
Spatial extent	X-axis	0 – 50 m
	Y-axis	0 – 50 m
	Z-axis	0 – 20 m
Data Types		Exhaustively sampled continuous values
Sampling cell size		$1 \times 1 \times 4$ (in meter)
Total # of samples		12500

Figure 6: Simulated secondary 2 data.

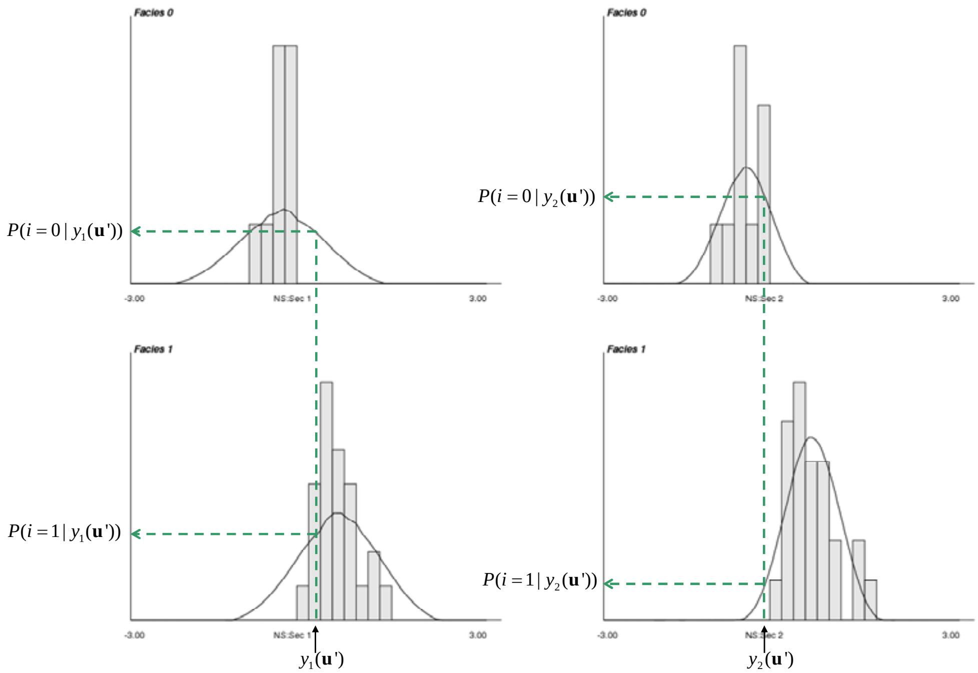


Figure 7: Calibration of facies proportion using secondary data.

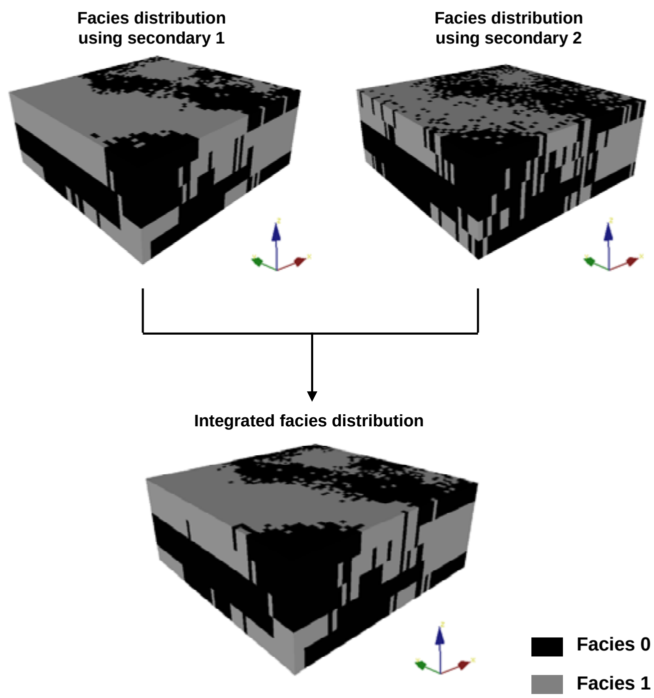


Figure 8: Calibrated facies distribution from secondary 1 and secondary 2 data. The Lamda-model is used to integrate two facies distribution.

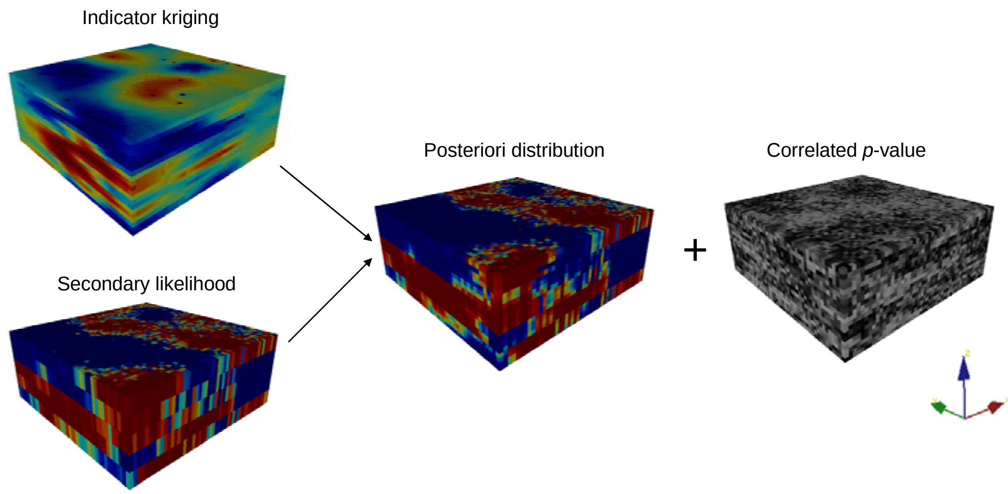
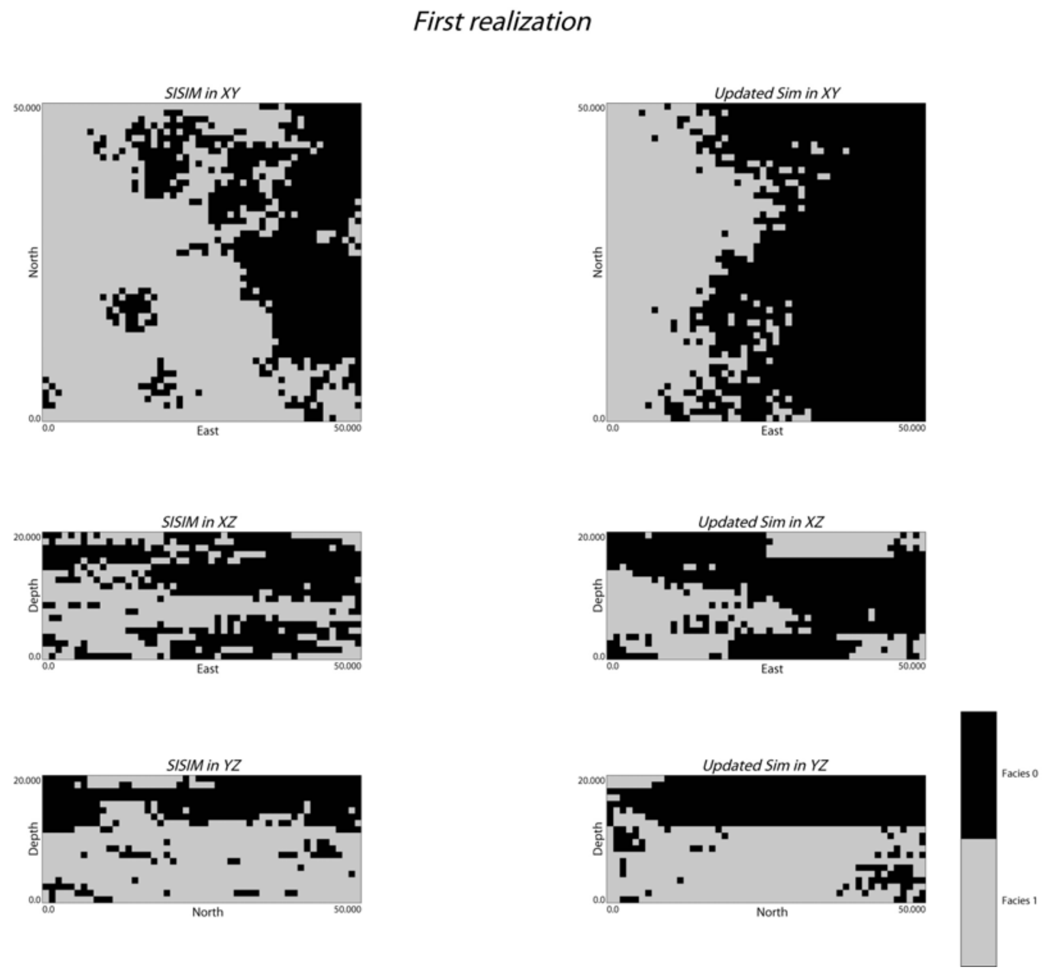


Figure 9: 3-D cubes represent probabilities of facies 0. A priori distribution obtained by indicator kriging is integrated with secondary likelihood to result in posteriori distribution. One realization of correlated *p*-values is shown. Multiple sets of *p*-values are generated and used to generate multiple facies realizations.



(continued in the next page)

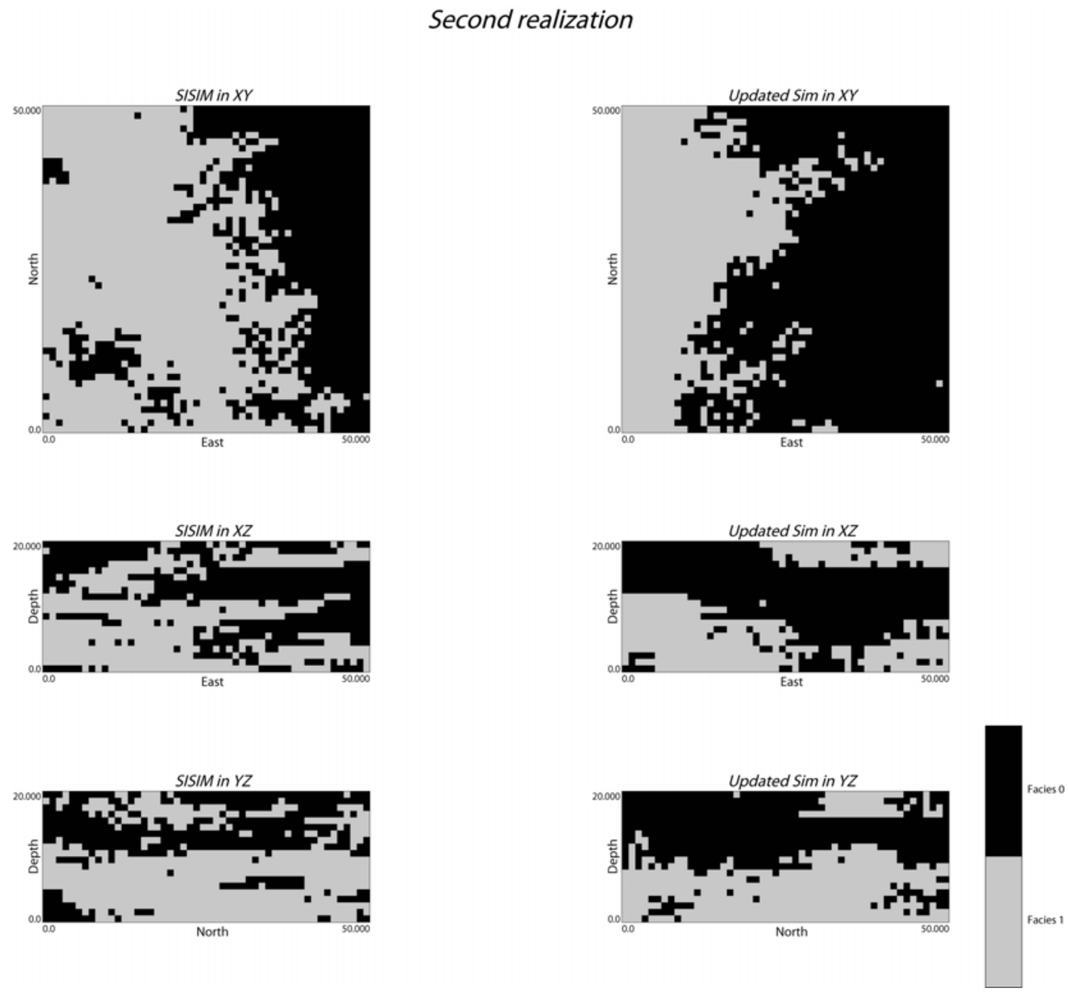


Figure 10: Two resulting realizations of facies are shown. For the comparison, SISIM realization is plotted as well.

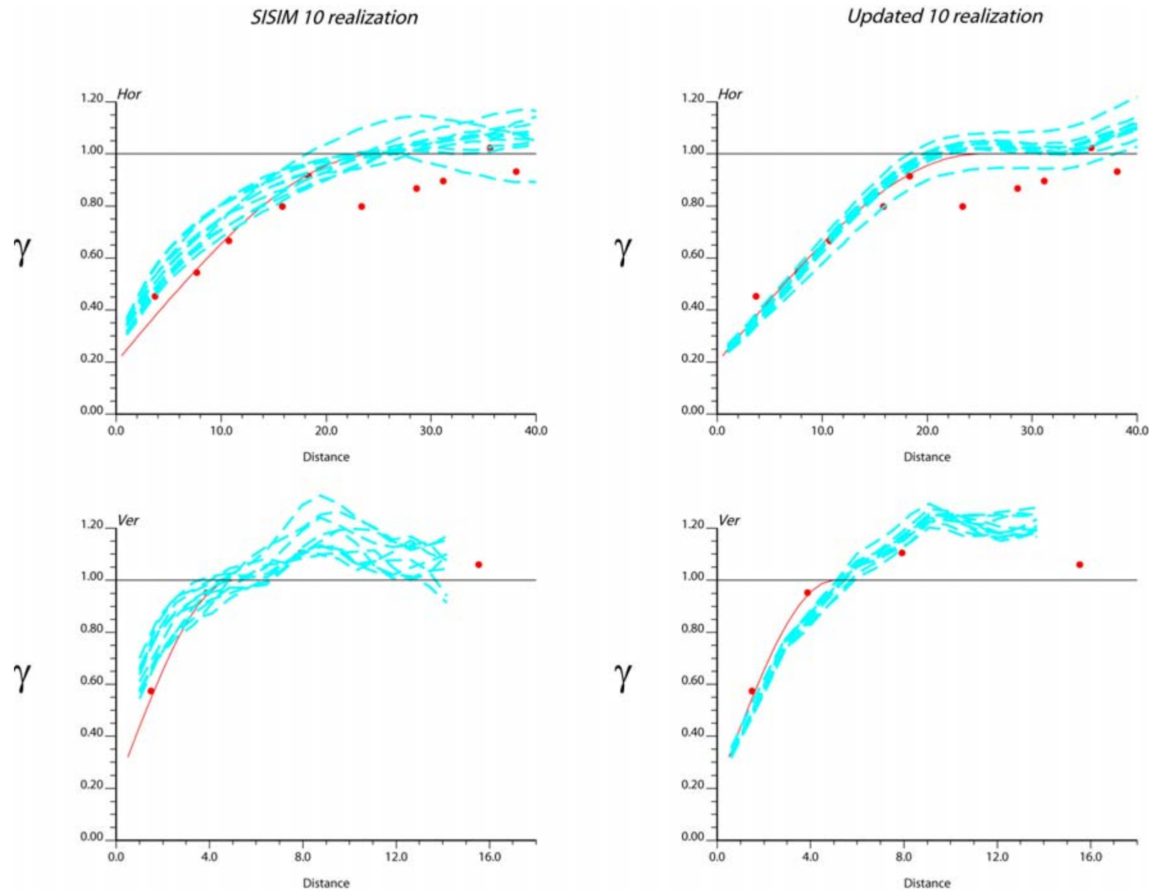


Figure 11: Checking the variogram reproduction with 10 realizations.