

Data Integration for 2-D Areal Proportions of Facies

Sahyun Hong and Clayton V. Deutsch

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

Secondary information is important for facies modeling. In this work, we propose a new technique that allows obtaining likelihood distribution with multiple secondary variables. The proposed method calls for probability decomposition with reasonable data redundancy weights among data sources. After obtaining likelihood distribution the updated posteriori distribution that combines prior and likelihood information is derived. The work flow is applied to 2-D areal facies modeling.

Introduction

Bayesian updating is a fairly robust method with secondary data. Likelihood distribution, $P(y_1 | facies)$ can be simply calibrated, however, it becomes tedious and difficult to obtain $P(y_1, \dots, y_m | facies)$ when multiple secondary variables are considered (m secondary variables). Several methodologies consider the integration of several secondary information with assumption of conditional independence or full data independence. This kind of independence assumption enables ones to simplify the data integration process. As the number of correlated secondary variables increase, however, any independence assumption is untenable. Thus, data source inter-dependence, that is called data redundancy, should be considered very carefully. This paper aims to develop data integration method for 2-D facies modeling accounting for data correlation. 3-D facies modeling using the considered method is introduced in another report [3]. The benefits of the proposed work include: (1) improvements in likelihood calibration (2) combining prior and likelihood to build updated posterior probability. 2-D areal facies modeling was tested for the evaluation of the proposed method and implementations.

The proposed integration methodology is made up of two integration steps: (1) integrate first all secondary variables to model likelihood distribution of facies and (2) combine the likelihood into prior model constructed only with primary variable. Decoupling the influence of primary and secondary data in the early stages of reservoir modeling simplifies calculation. Any secondary data calibration technique such as non-parametric or parametric multivariate data classification and data clustering can be used for generating secondary likelihood distribution. These techniques, however, fairly depend on user-defined parameter settings, and so produce unstable results. Prior facies model can be acquired through indicator kriging and/or multipoint geostatistics. Figure-1 illustrates generalized data integration process for facies modeling. The first step explains for gathering secondary data information. In step 1, secondary data is used for generating likelihood distribution of facies through plausible method. We can reasonably assume no scale difference between secondary variables. The second step involves combining prior and likelihood distribution in order to finally build posterior distribution. Main difficulty in step 2 is to account for scale difference between prior model and likelihood model constructed using fine scaled primary data and coarse scaled secondary data, respectively. Scale inconsistency problem gets significant when building facies model in 3-D because vertical resolution of secondary data is several times than resolution of primary data. Horizontal resolution can be reasonably assumed to be consistent for primary and secondary data. We developed data integration technique for 2-D modeling.

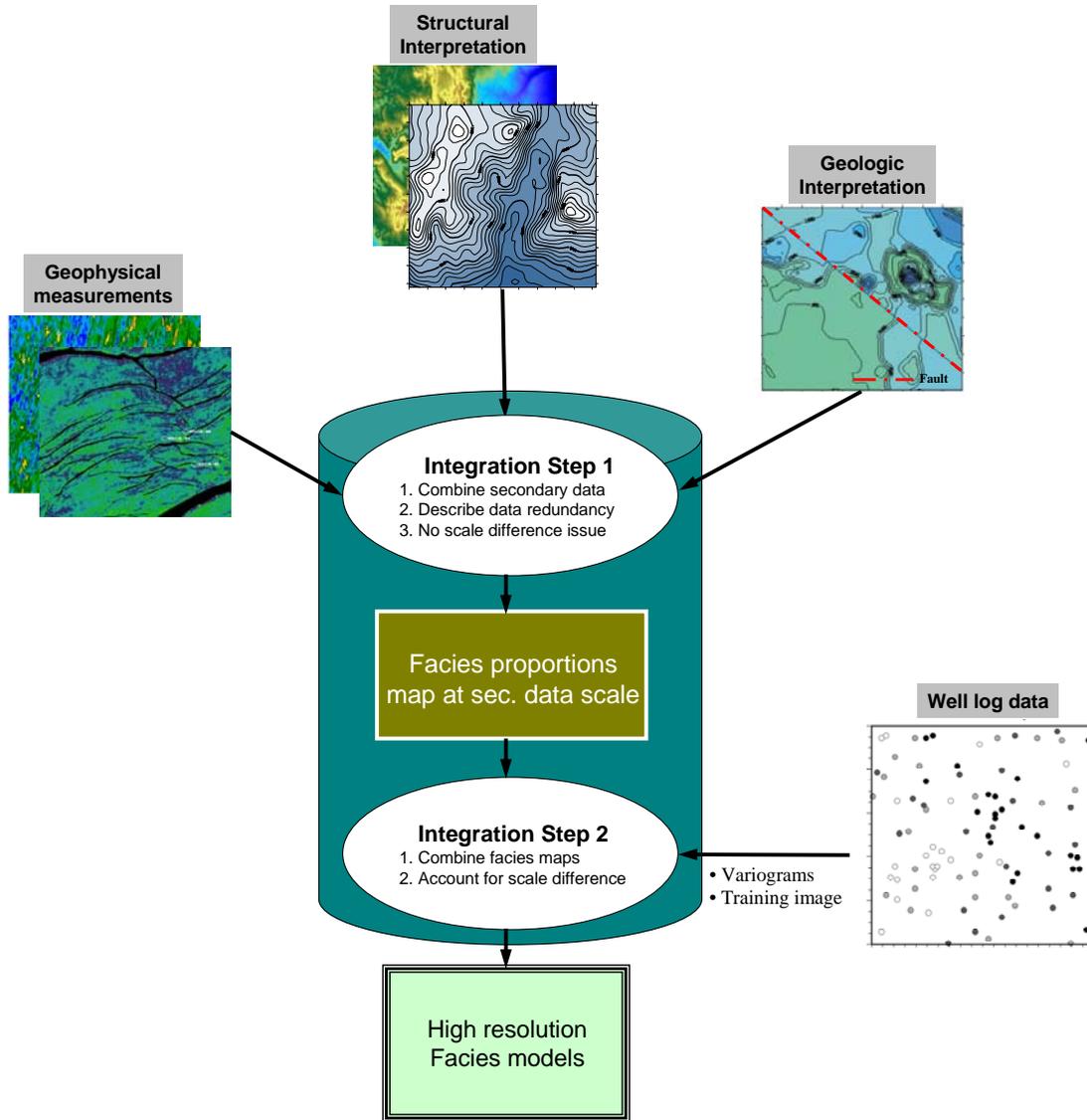


Figure-1: Data integration process for facies modeling.

Proposed Methodology

Categorical variable is considered and its updated probability conditioned to all relevant information is described. Updated probability of facies $k = 1, \dots, K$ at the location \mathbf{u} is noted as P^* and it is expanded as $P^* = P(k | \vec{i}, \vec{y}_1, \dots, \vec{y}_m)$ where \vec{i} is primary data (facies indicator at well sample location) and $(\vec{y}_1, \dots, \vec{y}_m)$ is secondary data and m is the number of secondary variables. One can reasonably assume that spatially related data to the estimation location \mathbf{u} have more impact on the estimated variable at the location. Primary data spatially close to the estimation location \mathbf{u} and collocated secondary variables (same as Markov-type hypothesis in collocated co-kriging) are retained and thus, the updated probability is re-expressed given relevant information,

$$P^* = P(k | \vec{i}(\mathbf{u}), y_1(\mathbf{u}), \dots, y_m(\mathbf{u}))$$

where, the vector $\vec{i}(\mathbf{u})$ means group of primary data surrounding the estimation location \mathbf{u} , i.e., $\vec{i}(\mathbf{u})$ summarizes $i(\mathbf{u}_\alpha)$, $\alpha \in$ local neighborhood of \mathbf{u} .

P^* can be expanded into the following expressions by Bayes law:

$$P^* = P(k | \vec{i}(\mathbf{u}), y_1(\mathbf{u}), \dots, y_m(\mathbf{u})) = P(k | \vec{i}(\mathbf{u}))P(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}) | k) \cdot C$$

$$\text{where, } C = P(\vec{i}(\mathbf{u})) / P(\vec{i}(\mathbf{u}), y_1(\mathbf{u}), \dots, y_m(\mathbf{u}))$$

For the derivation of above equation, the assumption $P(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}) | k, \vec{i}(\mathbf{u})) \approx P(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}) | k)$ is adopted, which means estimated facies at the location \mathbf{u} screens out the influence of surrounding primary indicator information, $\vec{i}(\mathbf{u})$. Finally updated probability is, therefore, decomposed at location \mathbf{u} into a product of two terms: priori probability, which is $P(k | \vec{i}(\mathbf{u}))$, and secondary likelihood, which is $P(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}) | k)$. The priori probability can be obtained from either indicator kriging or training image. Recently, estimation using training image is increasingly applied in reservoir modeling due to the ability to capture more realistically geologic features. This is one advantage of probability decomposition method since conventional co-kriging approach generates estimates using variograms.

Markov-type assumption leads to decomposition of the posteriori probability into likelihood and priori distribution which can be estimated regardless of likelihood term.

$$P^* = \boxed{P(k | \vec{i}(\mathbf{u}))} \boxed{P(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}) | k)} \cdot C$$

↓ *Unknown*
↓ *Priori distribution*
↓ *Likelihood distribution*

Obtaining likelihood distribution is the main process in the integration model 1 and then combining likelihood distribution into priori distribution is the main step in the integration model 2 (see Figure-1). Provided that $(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}))$ follows multivariate Gaussian distribution, $P(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}) | k)$, where $k = 1, \dots, K$, is also multivariate Gaussian distribution with different mean and covariance matrix corresponding to the facies, k . In such multivariate Gaussian case, probability of secondary variables given facies k is estimated easily based on Gaussian density function and data redundancy is implicitly considered by covariance matrix in the Gaussian function. However, in case of non-multivariate Gaussian, the likelihood distribution cannot be estimated based on Gaussian function. Non-parametric estimation is a solution to obtain joint distribution of $(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}))$ which is non-multivariate Gaussian, but non-parametric density estimation has potential problems related to the design of estimation algorithm. In this paper, we used Lamda-model to obtain likelihood distribution. Details of the Lamda-model and how to calibrate λ s are introduced in another report in this volume [2].

Experimental Results

Categorical variable estimation is performed with Amoco data. Amoco data has 62 indicator primary samples (# of facies1 = 36 and # of facies2 = 26), porosity, permeability and seismic amplitude at the same location. To get exhaustive secondary data sets, sequential Gaussian simulation was performed for each porosity, permeability and seismic amplitude at the desired modeling cell size (100 × 100). Thus, secondary data sets are treated as being sampled at the scale of primary data, which indicates there is no scale issue in integrating all variables (primary and secondary). In practice, however, scale difference between primary and secondary data must be carefully considered in data integration. Figure-2 illustrates the maps of primary and three secondary data. As a visual interpretation, facies 1 is distributed over the area where all secondary data have large values.

To utilize Lamda-model, conditional probability given secondary variable is required. Histogram smoothing technique is useful to provide conditional probability. Figure-3 illustrates sampled histogram and smoothed histogram based on sample histogram for facies 1 and 2. Thus, probability of facies 1 given secondary variable at certain location can be obtained using smoothed line. Figure-4 draws the calibrated probabilities of facies 1 from secondary data which indicate $P(k/y_1)$, $P(k/y_2)$ and $P(k/y_3)$. Figure-5 represents facies classification based on the probability map shown in Figure-4. Facies type (1 or 2) is assigned to every pixel with maximum probability. Facies distribution pattern using secondary 1 and 2 are similar each other except center and low-right area. Facies classification using secondary 3 looks noisy. These three probability maps are equivalently important, therefore, plausible integration model should maximize information contents.

Each probability maps are combined to generate the likelihood of each facies through the integration models. Lamda-model is applied to create likelihood maps. Method 1 and method 2 indicates two different schemes for estimating λ redundancy weights. Reference [2] introduces an algorithm for the Lamda-model method 1 and method 2 integration model.

Integrated likelihood of facies from the considered integration model is cross-validated in terms of quantitative measure of goodness. Classical entropy defined by Shannon in information theory and closeness to true facies are chosen for the evaluation criterion. How to define entropy and closeness is introduced in the reference [1]. As a summary, entropy is a measure of uncertainty of the integrated likelihood and closeness to true facies is a measure of how the integrated likelihood is close to true facies. Thus, lower entropy and higher closeness means better integrated results. Entropy and closeness are calculated over primary sample location. Table-1 summarizes goodness measures averaged over facies 1 and facies 2.

Quantitative goodness measure gives us there is no significant difference between method 1 and method 2. Estimated λ s are similar for method 1 and method 2. Figure-6 represents integrated likelihood maps from the considered method (method 1 and 2). Facies is classified with maximum likelihood as shown in the middle part in the Figure-6. From the facies classification, global proportions of facies are reproduced and they are very close to target proportions.

Table-1: Quantitative goodness measure of Lamda-model with method 1 and method 2.

		Closeness	Entropy
Lamda-model	Method1 for $k = 1$ $\lambda_1=0.63, \lambda_2=0.51, \lambda_3=0.54$ for $k = 2$ $\lambda_1=0.55, \lambda_2=0.53, \lambda_3=0.56$	0.728	0.088
	Method 2 For $k=1$ $\lambda_1=0.55, \lambda_2=0.71, \lambda_3=0.47$ For $k=2$ $\lambda_1=0.65, \lambda_2=0.41, \lambda_3=0.57$	0.733	0.09

Estimated likelihood distribution is merged to build posterior distribution as illustrated in Figure-7. The posterior distribution is obtained by

$$P^* = P(k | \vec{i}(\mathbf{u}), y_1(\mathbf{u}), \dots, y_m(\mathbf{u})) = P(k | \vec{i}(\mathbf{u}))P(y_1(\mathbf{u}), \dots, y_m(\mathbf{u}) | k) \cdot C$$

where, $C = P(\vec{i}(\mathbf{u})) / P(\vec{i}(\mathbf{u}), y_1(\mathbf{u}), \dots, y_m(\mathbf{u}))$

Thus, posterior distribution P^* has spatial interdependence information between facies variable as well as indirect information from secondary data variables. Sequential simulation based on the posterior

distribution is performed over all location. Figure-8 represents E-type facies out of 100 simulations both for Lamda method 1 and method 2. E-type facies map is uncleaned. Table-2 summarizes the global proportions estimated from E-type facies map. For the comparative study, we applied indicator simulation with only primary, Permanence of ratios and Tau integration model and results are compared (See the details of Permanence of ratios and Tau model in reference [1]). Lamda-model shows the best reproduced global proportions of each facies and Tau-model reproduced global proportions as well.

Table-2: Reproduced global proportions for different integration models. 100 realizations are generated and E-type facies is extracted from realizations.

	Global proportions from E-type facies					Target proportions
	Only primary	PR	Tau-model	Lamda Method 1	Lamda Method 2	
Facies 1	0.616	0.621	0.548	0.57	0.599	0.581
Facies 2	0.384	0.379	0.452	0.43	0.401	0.419

Reproduction of spatial structures is shown in Figure-9 in terms of reproduced variograms. Variograms of 100 realizations in N-S direction are drawn and the modeled N-S variogram is shown as thick line in the figure. Even though Tau-model shows good reproduced global proportions, variograms of realizations depart from the modeled variogram. Lamda-model still has a good accordance with the model variogram.

Conclusions

A secondary data integration technique for facies modeling is introduced. The priori and secondary likelihood distributions are generated separately and then combined to build an updated distribution. Estimation of likelihood distribution, $P(k|y_1, \dots, y_m)$, is tedious and difficult in practice. Instead we propose a probability decomposition approach that decouples the likelihood into the product of each conditional probability. The considered Lamda-model calls for estimating redundancy among secondary data and two different ways of measuring redundancy weight (Lamda-model with method 1 and method 2) were applied to 2-D facies modeling. Updated posteriori distribution was obtained by Bayesian derivation which combines prior and likelihood and final 2-D facies distribution was made by sequential simulation based on the updated distribution. Lamda-model showed good performance in reproduction of global proportions of facies and in reproduction of variograms compared to other integration models.

References

- C. V. Deutsch. A Short Note on Cross Validation of Facies Simulation Methods. In *Report 1, Centre for Computational Geostatistics*, Edmonton, AB, Canada, 1998
- S. H. Hong and C. V. Deutsch. Methods for Integrating Conditional Probabilities for Geostatistical Modeling. In *Report 9, Centre for Computational Geostatistics*, Edmonton, AB, Canada, 2007
- S. H. Hong and C. V. Deutsch. High Resolution Facies Modeling in Presence of Large-Scale Probabilistic Data. In *Report 9, Centre for Computational Geostatistics*, Edmonton, AB, Canada, 2007

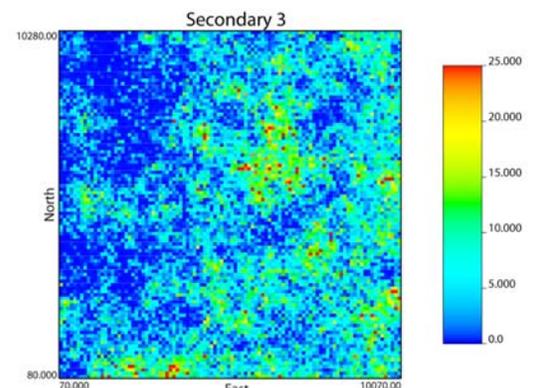
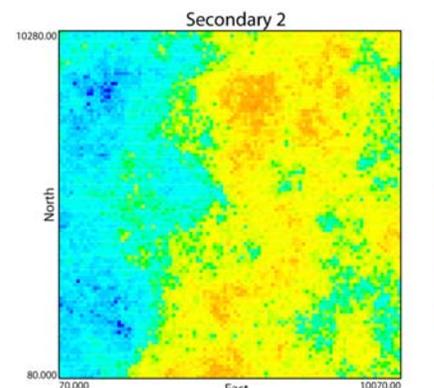
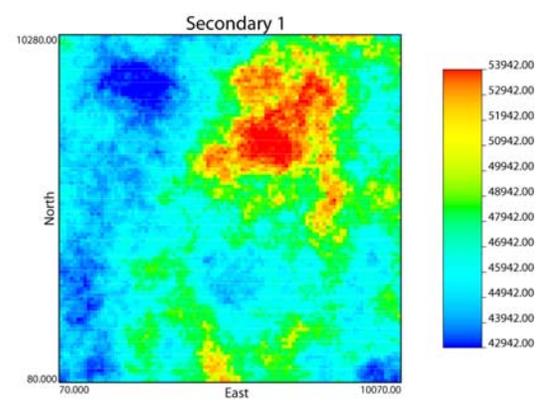
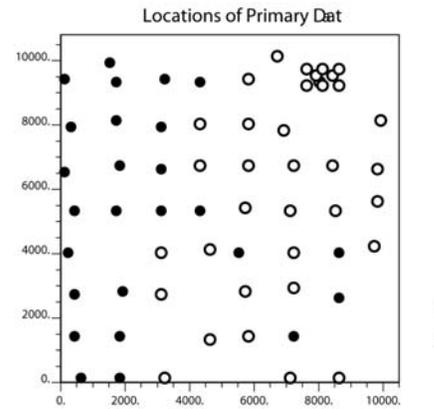
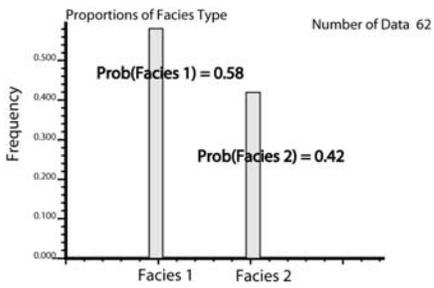


Figure-2: Primary indicator data and three secondary data sets.

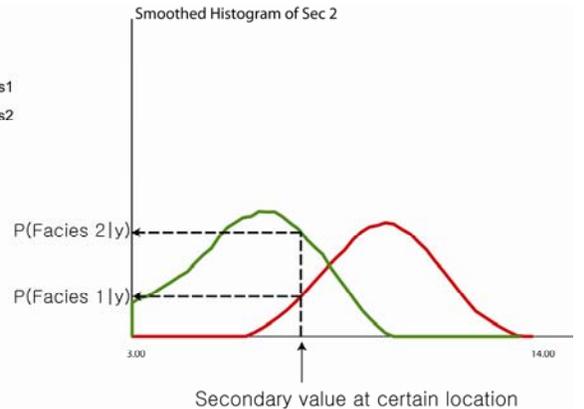
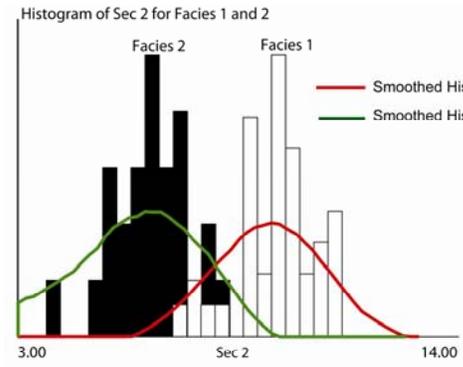


Figure-3: Estimation of probability given secondary variables, $P(k|y_2)$, based on smoothed histogram.

Proportions of Facies 1

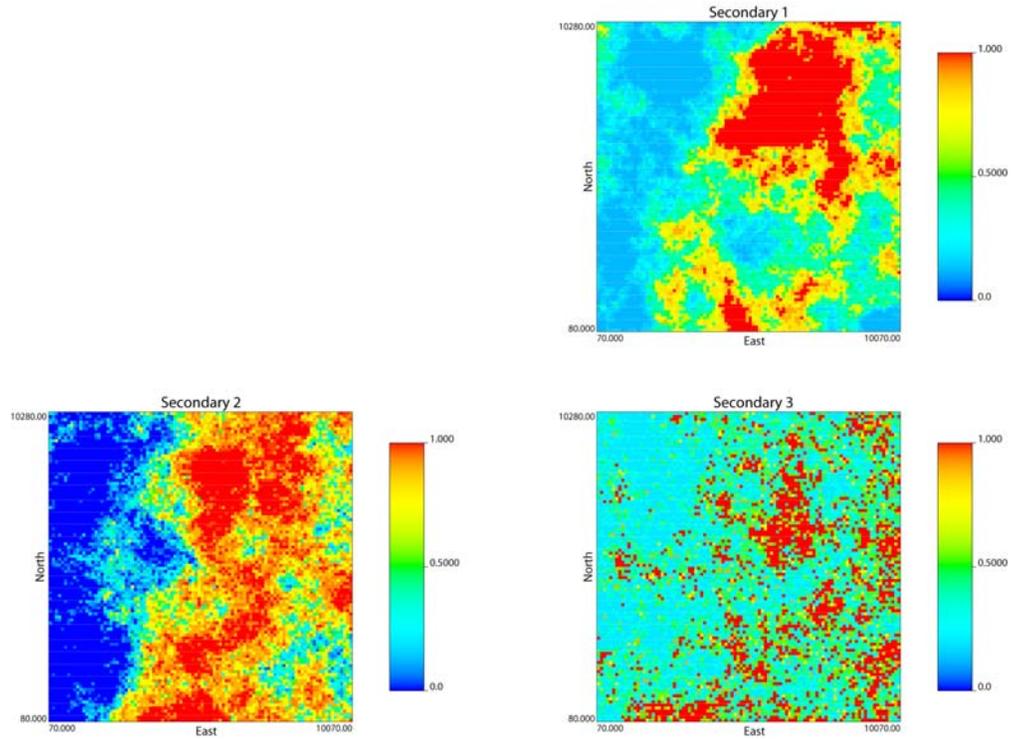


Figure-4: Proportions of facies 1 using secondary variables. Each map indicates $P(k|y_1)$, $P(k|y_2)$, and $P(k|y_3)$ over all locations.

Proportions of Facies 1

Facies Classification(Max. prob)

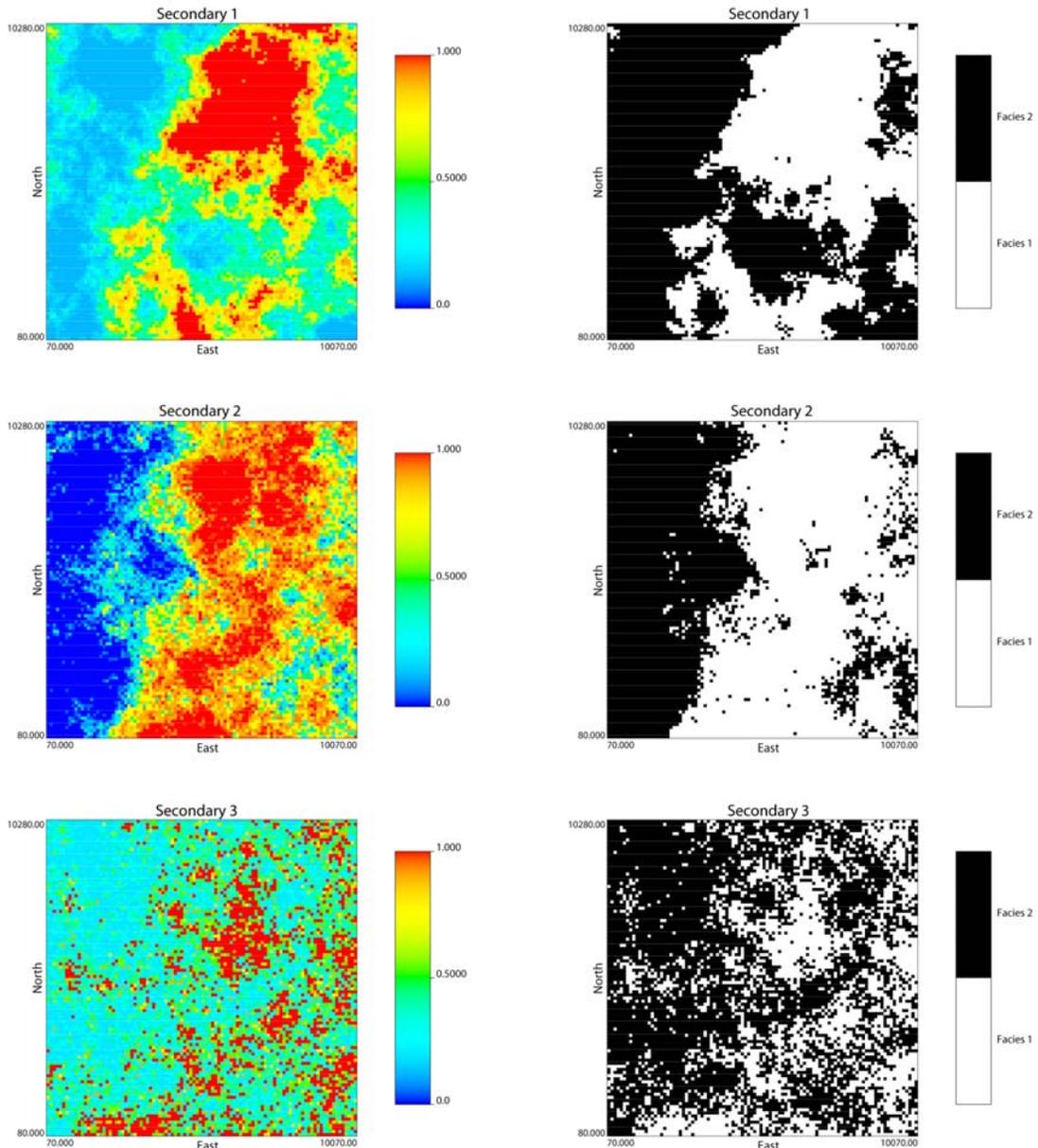


Figure-5: Facies classification is obtained with maximum probability from each secondary variable. Lambda-model combines three information maps to provide final integrated result.

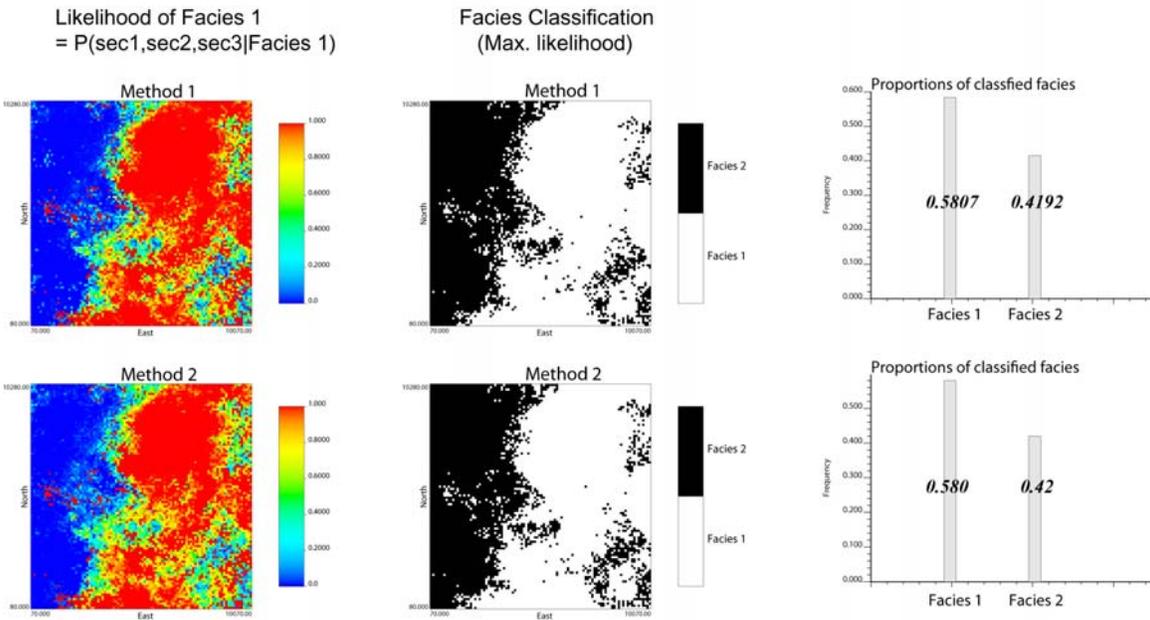
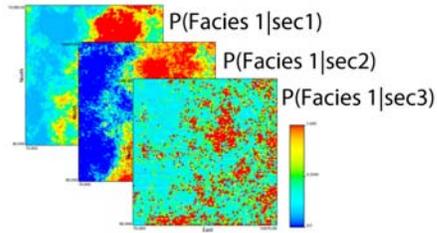


Figure-6: Integrated result from Lamda-model with method 1 and method 2. Integrated facies classification is obtained with max. likelihood. Global proportions of each facies are calculated based on facies classification.

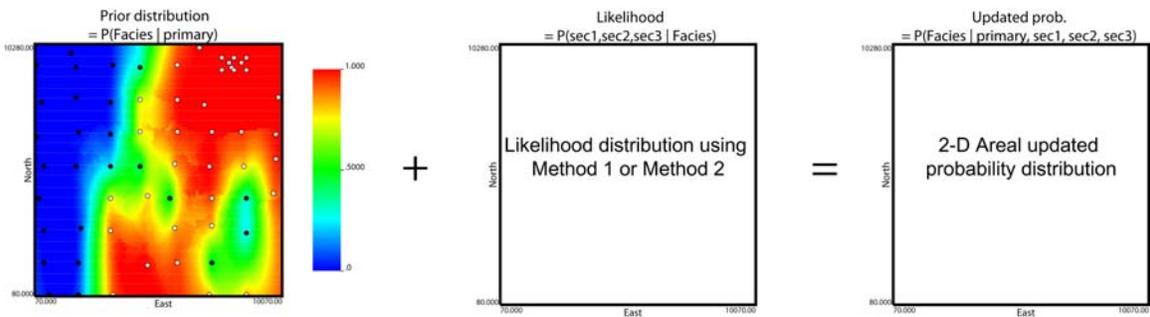


Figure-7: Likelihood distribution obtained by Lamda-model should be merged to prior distribution. Prior distribution is acquired by sequential indicator simulation. Final result is an updated posterior probability over modeling grids.

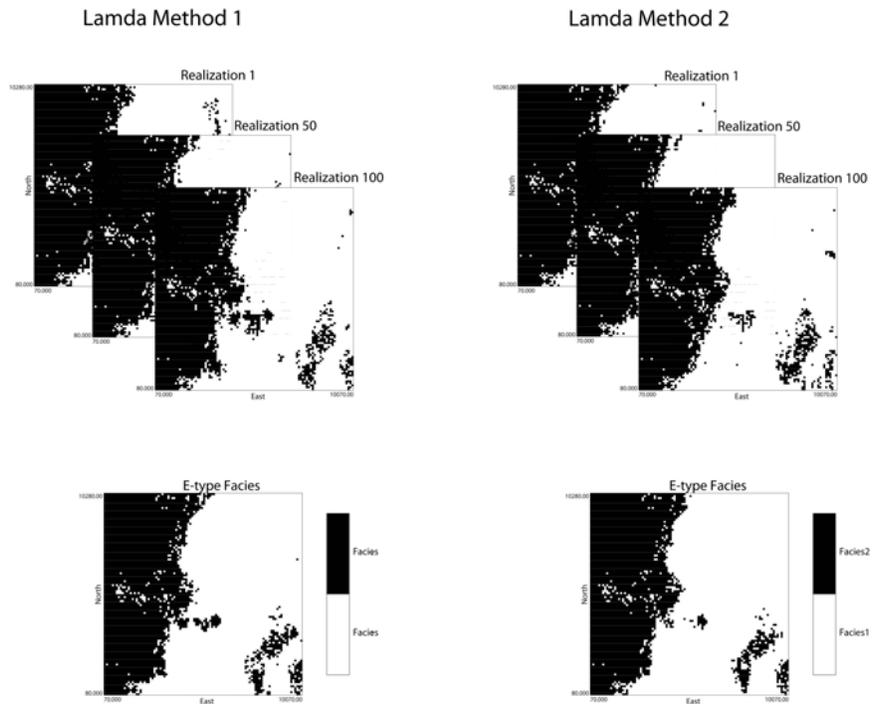


Figure-8: E-type facies (uncleaning) from 100 sequential realizations. Sequential simulation is performed based on posterior distributions that result from Lamda integration model.

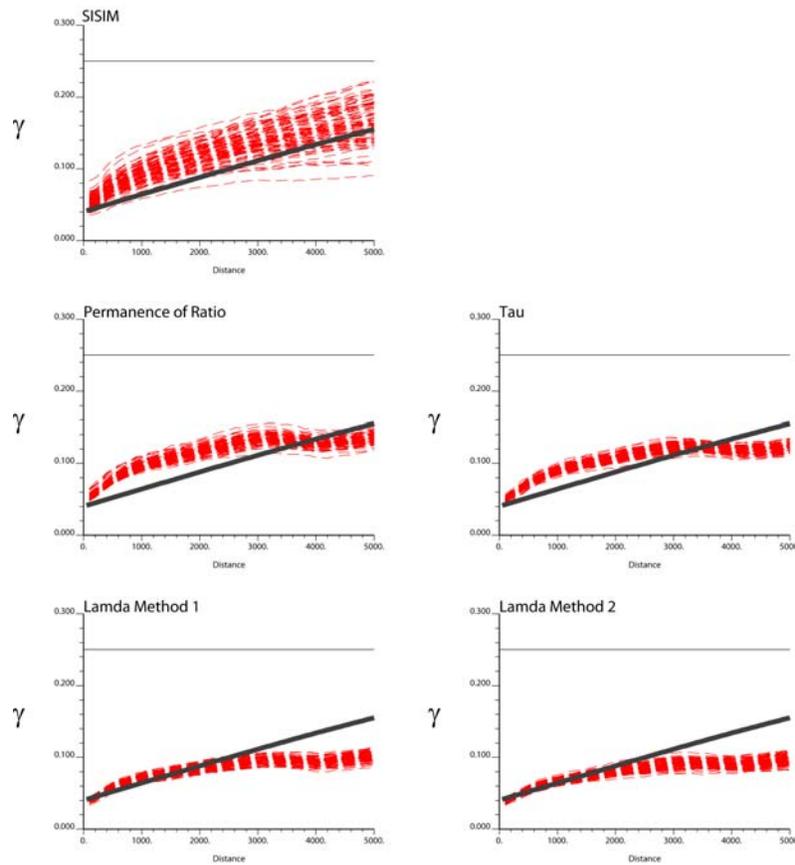


Figure-9: Variogram reproduction of facies realizations.