

Data Integration with Direct Multivariate Density Estimation

Sahyun Hong and Clayton V. Deutsch

A longstanding problem in geostatistics is the integration of multiple secondary data in the construction of high resolution facies and property models. Data integration requires the multivariate distribution among the secondary data. One way for estimating multivariate pdf is an indirect estimation through combining univariate distributions conditioned to each secondary data. A challenge of indirect methods is to quantify and to account for data redundancy among secondary variables. Directly estimating the multivariate probability density between all secondary variables meets this challenge. A procedure for direct multivariate density estimation is presented with multivariate kernel functions.

Introduction

Reservoir characterization is an important aspect of reservoir management and the geologic model should be constructed using all relevant data. Reservoir properties observed at well locations are considered the most reliable; however, there are typically few wells with limited lateral coverage. This lack of resolution in spatial context enables us to consider different types of data that has high resolution in lateral. The most extensive data source results from seismic survey that provide multiple seismic attributes related to reservoir properties of interest.

Facies are often important in reservoir modeling because the petrophysical properties of interest are highly correlated with facies distribution. Knowledge of facies constrains the range of variability in porosity and permeability. In this paper, facies identification study based on seismic data or related attributes to be considered secondary have been investigated. Secondary data integration requires the modeling of multivariate distribution. One could build a multivariate pdf by combining univariate pdf that has been individually calibrated from secondary variable. Permanence of ratios, tau-model and lamda-model are examples of the indirect approach (ref PR,lamda,tau). These indirect methods simply approximate multivariate pdf. Challenges of indirect estimation methods include quantifying data redundancy between secondary variables and how to meet marginal conditions. The other way for building multivariate distribution is to use direct modeling. Saggaf et al. (2003) applied neural network algorithm to create facies distribution maps. Matos et al. (2007) used Kohonen self-organizing maps (SOMs) to cluster seismic attributes for facies labeling. In many applications, neural networks show good results. Their flexibility in algorithm design makes it be available in various different problems. These flexibilities require several model tuning parameters such as number of layers, neurons and connection links among neurons, and often need expert's experiences. It should be noted that final results are different depending on the status of algorithm tuning. Statistical analyses also have been applied using complete seismic information in seismic attributes; Fournier and Derain (1995) performed multivariate statistical analysis. In statistical methods, multivariate pdf for each facies is first established via either parametric or non-parametric way. Then, each modeling grids come to have probability values which are extracted from the established pdf.

In this paper, we proposed a statistical facies classification methodology in a supervised mode that requires well data to be calibrated with secondary variables. Well data is used as control data for the inferring the multivariate pdf for each facies. The multivariate pdf is inferred by non-parametric probability density estimation method. Inferred pdf is then updated to satisfy marginal constraints. Iterative updating process has been adopted for this.

Methodology

Multiple secondary data arise from seismic feature extraction or inversion. We reasonably assumed to have cleaned and extracted seismic information for the facies modeling. It is also reasonable to assume that those seismic attributes data are measured at all locations over the study area. The methodology developed in this paper is applicable with any number of attributes. The multivariate seismic data are denoted by:

$$\mathbf{Y}(\mathbf{u}) = [y_1(\mathbf{u}), \dots, y_{n_{\text{seis}}}(\mathbf{u})], \mathbf{u} \in A$$

where the random variable $\mathbf{Y}(\mathbf{u})$ represents the multiple seismic attributes at every location, n_{seis} is the number of seismic attributes to be considered, and A denotes the study area. In general, there are many attributes variables with $1 < n_{\text{seis}} < 10$. More than 10 variables would normally be combined together through dimension reduction preprocessing. Consider predicting a categorical variable called facies. In practice, there are a limited number of facies between 2 and 5. The facies become difficult to distinguish if they are subdivided into more than 5. The facies are represented by discrete random variable denoted as

$$S(\mathbf{u}), \mathbf{u} \in A$$

Random variable $S(\mathbf{u})$ can take one of $k = 1, \dots, K$ integer values over the study area. The facies are directly measured at a limited of well locations. The central problem addressed by this paper is the prediction of facies at location \mathbf{u} given seismic data \mathbf{Y} at the same location:

$$f_{S|\mathbf{Y}}(k; \mathbf{u}), \quad k = 1, \dots, K, \quad \mathbf{u} \in A$$

The uncertainty in facies at each location due to secondary data will be used for subsequent reservoir characterization. These probabilities are locally updated with the well data and used in facies simulation. Block co-kriging is a noble way to incorporate secondary and primary well data. The use of these probabilities is not discussed in this paper; the focus on the inference of these conditional probabilities. These are calculated with Bayes law:

$$f_{S|\mathbf{Y}}(k; \mathbf{u}) = \frac{f_{S\mathbf{Y}}(k, \mathbf{y}; \mathbf{u})}{f_{\mathbf{Y}}(\mathbf{y}; \mathbf{u})}, \quad \mathbf{u} \in A$$

The numerator is the joint probability distribution of \mathbf{Y} and k , and the denominator is the multivariate probability of the seismic variables. The challenge is in modeling or inferring the multivariate distribution $f_{S\mathbf{Y}}(k, \mathbf{y}; \mathbf{u})$.

Multiple seismic attributes are processed to provide one term indicating probability of facies given all seismic variables. Multivariate density estimation is applied to estimate probability. Our challenge is to derive local probability distribution of the facies variable given observed seismic data over the entire modeling locations. For the simple notation, location vector (\mathbf{u}) is dropped. The multivariate distribution of interest is $f_{S\mathbf{Y}}(k, \mathbf{y}; \mathbf{u})$ for all values of the \mathbf{y} vector and $k = 1, \dots, K$ facies. Building probability density distribution with continuous variables is more straightforward. Figure-1(a) illustrates bivariate density distribution in case of two continuous variables denoted as random variables Y and Z . RVs Y and Z are deemed single seismic variable and continuous variable of primary interest (for example porosity in this illustration), respectively. Univariate marginal distribution of the seismic variable Y is well understood because there are exhaustively sampled seismic data over all grids. Univariate marginal distribution of the primary variable Z is derived from well data samples. If few data is available statistical smoothing can be applied for marginal distribution. It must be ensured that integration of bivariate distribution over all y outcomes corresponds to marginal distribution of Z . Simultaneously, integration of bivariate distribution over all z outcomes must be equivalent to the marginal distribution of Y . Two marginal distributions, $f_Z(z)$ and $f_Y(y)$, are shown on each axis as solid line on Figure-1. Although variable of interest is not continuous but categorical, bivariate distribution can be constructed. The bivariate distribution with continuous Y and discrete variable S is shown in the bottom of Figure-1(b). The distribution is not continuous in the direction of random variable S but histogram bar chart denoted as $p_S(k)$. Summing up bivariate distribution over all y values must satisfy marginal probability of facies k , $p_S(k)$, $k = 1, 2$. The following equations relate the bivariate to the marginal distributions of Y and S :

$$\sum_{k=1}^2 f_{S\mathbf{Y}}(k, y) = f_{\mathbf{Y}}(y), \quad \forall y$$

$$\int_{-\infty}^{\infty} f_{S\mathbf{Y}}(k, y) dy = p_S(k), \quad k = 1, 2$$

$p_s(k)$ is a probability of facies k and it is calculated from overall well data. We referred to it as global probability of facies k . Bivariate distributions of Y and S are 1-D lines as shown in the Figure-1(b). Consideration of one seismic variable is simple but more than one variable is challenge. Provided that two seismic variables, $\mathbf{Y}=[y_1, y_2]$ then trivariate distribution of RVs (S, \mathbf{Y}) should be modeled. Figure-2 illustrates the trivariate distribution. As same as in bivariate pdf modeling shown in Figure-1, marginal constraints mentioned above must be met in trivariate pdf modeling. Integration of $f_{SY}(k=1, y_1, y_2)$ and $f_{SY}(k=2, y_1, y_2)$ should amount to $f_Y(y_1, y_2)$, which can be interpreted as summation in vertical direction in Figure-2. Simultaneously, sum of $f_{SY}(k=1, y_1, y_2)$ over all y_1, y_2 amounts to $p_s(k=1)$, and sum of $f_{SY}(k=2, y_1, y_2)$ over all y_1, y_2 amounts to $p_s(k=2)$. This second marginal constraints are interpreted as horizontal summation in the Figure-2. For general high dimensional problems, two marginal constraints specified above are adopted in the following iterative procedure:

Step 1. Initialize the distributions $f_{SY}^{(0)}(k, \mathbf{y}), \forall \mathbf{y}, \forall k$

Step 2. Reset the distributions $f_{SY}^{(0)}(k, \mathbf{y}), \forall \mathbf{y}, \forall k$ to ensure that the marginal of the global facies probabilities is reproduced:

$$f_{SY}^{(1)}(k, \mathbf{y}) \leftarrow f_{SY}^{(0)}(k, \mathbf{y}) \frac{f_S(k)}{\int_{-\infty}^{\infty} f_{SY}^{(0)}(k, \mathbf{y}) d\mathbf{y}}, \forall k$$

Step 3. Reset the distributions $f_{SY}^{(1)}(k, \mathbf{y}), \forall \mathbf{y}, \forall k$ to ensure that the marginal distribution of seismic attributes data vector \mathbf{y} is reproduced:

$$f_{SY}^{(2)}(k, \mathbf{y}) \leftarrow f_{SY}^{(1)}(k, \mathbf{y}) \frac{f_Y(\mathbf{y})}{\sum_{k=1}^K f_{SY}^{(1)}(k, \mathbf{y})}, \forall \mathbf{y}$$

Step 4. Set $f_{SY}^{(2)}(k, \mathbf{y})$ to $f_{SY}^{(0)}(k, \mathbf{y})$ and return to step 2 until there are no changes in the multivariate distribution.

Global probabilities of $k = 1, \dots, K$, $p_s(k)$ is the first requirement to be used as an input in the iterative procedure. Well data is the primary information source and it records petrophysical properties including facies type at well locations. Thus, occurrences of each facies could be calculated using well data. Multivariate pdf $f_{SY}(k, \mathbf{y}), k = 1, \dots, K$ is the second requirement for the iterative procedure. In practice, multivariate normal model after univariate normal score transformation of $(y_1, \dots, y_{n_{seis}})$ is commonly adopted for multivariate pdf inference, $f_{SY}(k, \mathbf{y}), k = 1, \dots, K$ since its simplicity and reasonable efficiency. Inferred distributions become facies dependent pdf with n_{seis} dimension by:

$$f_{SY}(k = 1, \mathbf{y}) = N_{n_{seis}}(\boldsymbol{\mu}_{k=1}, \boldsymbol{\Sigma}_{k=1})$$

⋮

$$f_{SY}(k = K, \mathbf{y}) = N_{n_{seis}}(\boldsymbol{\mu}_{k=K}, \boldsymbol{\Sigma}_{k=K})$$

For the applicability in general data distribution, we used kernel density estimation techniques as a nonparametric method (Scott, 1992 and Silverman, 1993). Probability density function is constructed from the observed data so that it has good reflection of natural data distributions. However, there is no guarantee that the experimental initial pdf $f_{SY}(k, \mathbf{y})$ that is obtained by non-parametric modeling would satisfy the two constraints. In the following, we theoretically derived that the inferred multivariate pdf $f_{SY}(k, \mathbf{y})$ does not satisfy the marginal constraint. In case of binary facies, the joint pdf is non-parametrically obtained using kernel function $\mathbf{w}(\bullet)$:

$$f_{SY}(k = 1, \mathbf{y}) = \frac{1}{nh_{k=1}^{n_{seis}}} \sum_{i=1}^n \mathbf{w} \left(\frac{1}{h_{k=1}} (\mathbf{y} - \mathbf{Y}(\mathbf{u}_i)) \right) \times \mathbf{I}(k(\mathbf{u}_i) = 1)$$

$$f_{SY}(k=2, \mathbf{y}) = \frac{1}{nh_{k=2}^{n_{seis}}} \sum_{i=1}^n \mathbf{w} \left(\frac{1}{h_{k=2}} (\mathbf{y} - \mathbf{Y}(\mathbf{u}_i)) \right) \times \mathbf{I}(k(\mathbf{u}_i)=2)$$

Indicator $\mathbf{I}(k(\mathbf{u}_i)=1)$ is defined as 1 if facies at well sample location \mathbf{u}_i is code 1 otherwise 0. $\mathbf{I}(k(\mathbf{u}_i)=2)$ is defined as 1 if facies at well sample location \mathbf{u}_i is code 2 otherwise 0. n is the total number of well sample data, that is $n = n_{k=1} + n_{k=2}$ where n_k is the number of well sample denoting facies k . $h_{k=1}$ and $h_{k=2}$ are kernel smoothing window that is normally depending on the number of data points $n_{k=1}$ and $n_{k=2}$: large bandwidth if small number of data points and small bandwidth if large number of data points. According to the marginality, the below equation should be met.

$$\sum_{k=1}^2 f_{SY}(k, \mathbf{y}) = f_Y(\mathbf{y})$$

Left-hand side of above relation is expanded as,

$$\begin{aligned} \sum_{k=1}^2 f_{SY}(k, \mathbf{y}) &= f_{SY}(k=1, \mathbf{y}) + f_{SY}(k=2, \mathbf{y}) \\ &= \frac{1}{nh_{k=1}^{n_{seis}}} \sum_{i=1}^{n_{k=1}} \mathbf{w} \left(\frac{1}{h_{k=1}} (\mathbf{y} - \mathbf{Y}(\mathbf{u}_i)) \right) + \frac{1}{nh_{k=2}^{n_{seis}}} \sum_{i=1}^{n_{k=2}} \mathbf{w} \left(\frac{1}{h_{k=2}} (\mathbf{y} - \mathbf{Y}(\mathbf{u}_i)) \right) \end{aligned}$$

Right-hand side is a known marginal pdf consisting of exhaustively sampled secondary data. It can be obtained via non-parametric way,

$$f_Y(\mathbf{y}) = \frac{1}{Nh^{n_{seis}}} \sum_{i=1}^N \mathbf{w} \left(\frac{1}{h} (\mathbf{y} - \mathbf{Y}(\mathbf{u}_i)) \right)$$

where N is total number of secondary data values. Left-hand side and right-hand side are not equivalent as long as number of data points to be used for inferring pdf is not same. In all cases, N is different from n_k : N is much larger than n_k , $k = 1, \dots, K$. Thus, we proved marginal constraint is not satisfied.

Iterative scheme enables the initial distributions to be updated distributions under marginal constraints. This methodology is illustrated with a simple univariate example. Figure-3 shows distribution of single seismic data for a small area with 62 well data. The solid continuous line represents the marginal distribution of seismic variable. The distribution of the seismic data at the 62 wells is shown by the histogram bar chart. The light gray portion of the histogram bars corresponds to facies 2 and the black portion corresponds to facies 1. The initial pdf to the seismic data distributions of facies 1, $f_{SY}(k=1, \mathbf{y})$ and facies 2, $f_{SY}(k=2, \mathbf{y})$ are shown in the Figure-4(a). The summation of $f_{SY}(k=1, \mathbf{y})$ and $f_{SY}(k=2, \mathbf{y})$ is shown in the densely dashed red line. It should match the known marginal distribution represented by black solid line. The correction procedure described above was applied to the initial distributions. Figure-4(b) shows the corrected distributions of facies 1 and facies 2 using same line color scheme.

Multiple Secondary Data Integration

Synthetic Data

To describe quantitative improvements of the proposed seismic data integration method, we first performed facies modeling using the synthetic reference data. A synthetic data generated by physical depositional modeling was chosen for the test, which contain binary facies, lobe (coded as 1) and shale (coded as 0) facies. The reference image has a dimension of $256\text{m} \times 256\text{m} \times 1\text{m}$. Well data is extracted from the reference image at every $40\text{m} \times 40\text{m}$. These extracted well data are to be used as control data for seismic-facies calibration. Reference and extracted well data are shown in Figure-6. Global proportions of facies are calculated with well samples as $p(\text{shale})=17/49=0.35$ and $p(\text{lobe})=32/49=0.65$. Although sample data is extracted with regular basis from reference image, facies proportions are somewhat different from the reference data, which produces $p(\text{shale})=0.45$ and $p(\text{lobe})=0.55$. Debiasing technique with exhaustive seismic data could correct the biased global proportions calculated using sparse well data (see the following Debiasing section). Two seismic attributes were simulated with $256 \times 256 \times 1$ grid locations. Gaussian

variables were first simulated then they were correlated with non-linear fashion. Simulated seismic variables are standardized with zero mean and unit variance. Exhaustive data plot of simulated variables (y_1, y_2) is shown in Figure-7(a) and their relation is non-linear with linear correlation $\rho = 0.65$. Collocated seismic attributes data of shale and lobe is also plotted on the exhaustive scatter plot. Data scatter plots represent simulated variables y_1 and y_2 have good discriminant capabilities between shale and lobe: higher y_1 and y_2 values tend to classify as lobe, and lower y_1 and y_2 values tend to classify as shale. Bivariate pdf of (Y_1, Y_2) is drawn in Figure-7(b).

The initial multivariate distributions are first inferred by Gaussian kernel method using well data. Smoothing window size h is set as 0.5 for y_1 and y_2 . Choice of too small window size could reflect the data point, but it risks at oversmoothing. Left column of Figure-8 illustrates the initial distributions $f_{SY}^{(0)}(k, \mathbf{y})$, $k = 1, 2$ obtained from the Gaussian kernel estimation method. They reflect reasonably data scatter plots shown in Figure-7(a); however, the sum of those initial distributions (bottom left of Figure-8) does not correspond to the marginal bivariate distribution $f_Y(\mathbf{y})$ shown in the bottom of Figure-7. The proposed method was applied and the updated multivariate distributions denoted as f^* are shown in the right column of Figure-8. After applying the proposed method, the updated multivariate distributions have detailed variations that cannot be reliably observed in initial estimation. Summation of $f_{SY}^{(*)}(k = 1, \mathbf{y})$ and $f_{SY}^{(*)}(k = 2, \mathbf{y})$ exactly amounts to the marginal bivariate distribution $f_Y(\mathbf{y})$. To evaluate the performance of the method, the mean absolute deviation (mAD) statistics is calculated based on the underlying known reference distributions estimated using the entire reference image. Table-1 summarizes the efficacy of the proposed updating method. Numeric values in the parenthesis indicates the percentage in improvements. The resulting probability of facies 0 (shale) is shown in Figure-9.

Table 1. The mean absolute deviation of the initial and updated multivariate distributions are calculated with reference distributions.

	Initial pdf	Updated pdf
$f_{SY}(k=0, y_1, y_2)$.000028	.000015 (46%)
$f_{SY}(k=1, y_1, y_2)$.0000251	.0000129 (49%)
$f_{SY}(k=0, y_1, y_2) + f_{SY}(k=1, y_1, y_2)$.000058	.000009 (84%)

Debiasing with Secondary data

There are few well data and they are potentially located in areas of the reservoir with indications of high production, thus preferentially located with respect to the distribution of the Y . The marginal distribution of facies proportions, $p_S(k)$, $k = 1, \dots, K$ is a required input to the iterative updating approach; however, the marginal proportions are themselves uncertain. It is common that the well data are preferentially located in areas of high quality. These areas of high quality are likely identified by the exhaustive secondary data. It may be necessary to preprocess the available data to infer global representative proportions. A proposed debiasing methodology would be to (1) infer multivariate pdf with kernel smoothing, but do not standardize according to facies proportions $p_S(k)$, $k=1, \dots, K$ (described in Step 2 in the methodology section), (2) estimate the facies proportions for each particular outcome of the seismic data, and (3) calculate the representative proportions by integrating over all possible outcomes of the secondary data.

$$p_S^{Debiased}(k) = \int_{-\infty}^{\infty} f_{SY}^{(0)}(k, \mathbf{y}) \frac{f_Y(\mathbf{y})}{\sum_{k=1}^K f_{SY}^{(0)}(k, \mathbf{y})} d\mathbf{y}, \quad \forall k$$

The debiased facies proportion, $p_S(k)$, is used for the marginal constraint in the updating procedure described in the above section. Table-2 shows the facies proportions computed using the sampled well data and using debiasing technique for the first synthetic example. Reference global proportion of each facies obtained from reference image is shown as well.

Table 2. Comparison of global proportions of each facies with well data only, with debiasing technique, and with the whole reference image

	$p_s(k=0)$	$p_s(k=1)$
47 well sample	0.35	0.65
Debiasing with simulated seismic data	0.42	0.58
Reference data	0.45	0.55

West Angola Reservoir

Deepwater clastic reservoir has been considered for the applicability and effectiveness of the proposed methodology. Turbidite channel typed reservoir located on the west coast of Angola was chosen for the study. The reservoir field is located in an average water depth of 1200ft, approximately 50 miles offshore. The reservoir model is about 5850m × 4425m × 700ft consisting of 78 × 59 × 116 grid.

Total 533,832(=78×59×116) grid blocks are defined for the modeling and approximately 100,000 are active blocks. Average number of active grids in vertical direction is 40 that results in 40 × 6ft=240ft average thickness. However, reservoir thickness varies from 6ft to 492ft. Figure-10 shows the 28 wells and active blocks color-coded by one of seismic attribute. 28 wells including 20 production and 8 injection wells and 3-D seismic survey explored the reservoir. Four distinct facies have been identified in core analysis: three sand facies with varying quality (coded facies 2 through 4) and shale facies (coded facies 1). Sand facies consists of high net-to-gross channel, poor and good net-to-gross sand facies. Each facies is recognized based on quantitative net-to-gross level. Table 3 describes the varying NTG ratio depending on facies types.

Table 3. Defined facies with varying Net-to-Gross ratio. Four facies categories are identified according to the NTG ratio: highest quality channel, two sand and shale.

	NTG (%)
Facies 1 (shale)	3.8
Facies 2 (poor quality sand)	47.3
Facies 3 (good quality sand)	63
Facies 4 (channel)	90.7

Seismic inversion generated five seismic attributes: acoustic impedance (S1), Vsh using three different inversion technique (S2 ~ S4) and amplitude (S5). Inverted attributes were obtained with the facies modeling scale (6ft vertical interval). Three seismic attributes were selected as seismic information for building seismic-derived facies model based on the correlation between seismic and facies. At small scale, seismic attributes have low correlation coefficient with maximum correlation of 0.3. Facies 2 and 3 are nearly uncorrelated to the selected three seismic data. Table-4 describes the correlation coefficient.

Table 4. Correlation coefficient between seismic data and facies

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>
*S1	0.34	0.00	-0.08	-0.31
S2	0.07	-0.06	-0.03	-0.02
S3	0.09	-0.03	-0.04	-0.05
*S4	0.20	0.02	-0.03	-0.21
*S5	0.25	-0.09	-0.02	-0.20

The average proportions of each facies are calculated with 28 well data (Table-5). The amount of sand 50% and non-pay facies is 46% of the reservoir volume. The highest quality facies (channel) is about 30% of the reservoir. These facies proportions are computed with limited 28 well data and they might be biased statistics due to preferential drilling or spatial bias in data locations. Facies proportions are re-calculated using the debiasing technique discussed above then we obtained the debiased facies proportions shown in Table-5. Debiased facies proportions do not much change due to relatively low correlation among seismic and facies variables.

Table 5. Global proportions of four facies.

	Proportions with 28 wells	Re-calculated proportions with seismic data
Facies 1 (shale)	0.46	0.49
Facies 2 (poor quality sand)	0.12	0.113
Facies 3 (good quality sand)	0.11	0.112
Facies 4 (channel)	0.31	0.289

Seismic attributes are first calibrated with 28 well data through non-parametric multivariate modeling. Multivariate Gaussian kernel function was applied to the considered four variables (3 seismic variables and 1 categorical variable of primary interest) resulting in quadivariate probability distributions. Figure-11 demonstrates the resulting probability cubes of each facies in 3-D. Figure 12 shows the histogram of the obtained facies probabilities. Shale and channel facies has relatively good separation and two sand facies (facies 2 and 3) are highly overlapped. Meandering pattern of facies 4 can be observed.

The use of seismic data at small scale could not reflect the real vertical resolution of seismic data even though seismic inversion was performed at fine scale (modeling scale). In utilizing seismic data, realistic vertical averaging should be required because “real” seismic resolution is much coarser, and short scale noise masks the value of seismic information. Reasonably high correlation between averaged seismic and facies proportions could be obtained at a single vertical block ($58 \times 59 \times 1$). Table 6 summarizes the correlation coefficient between vertically averaged seismic variables and facies proportions.

Table 6. Correlation coefficient between vertically averaged seismic variables and facies proportions

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>
*S1	0.62	-0.32	-0.04	-0.55
S2	0.18	-0.05	-0.03	-0.22
S3	0.10	-0.08	-0.08	-0.23
*S4	0.47	-0.15	-0.05	-0.52
*S5	0.23	-0.04	-0.07	-0.31

Vertically averaged seismic variables, S1, S4 and S5 were used for facies modeling due to their reasonably high correlation. It should be noted that vertical averaging of seismic data provides better correlation coefficient with facies. Estimated areal facies probabilities are shown in Figure-13. Shale and channel facies are better recognized rather than poor and good sand facies.

Conclusions

The main goal of this work is to account for all available geologic information in facies modeling. One way to integrate multiple secondary data is probability combination approaches that calibrate each data source and then combine with considering data redundancy. Probability combination schemes, however, have a few disadvantages: (1) they require a data redundancy weight estimation process, and (2) there is no guarantee of satisfying marginal pdf. In this study, we develop the direct density estimation technique. The target joint probability was derived directly using known marginal density distributions. The method is applied to integrate multiple seismic attributes for facies modeling. The proposed method effectively reproduces all data sources.

References

- Schwedersky-Beto, G., Marcella Maria de Melo Cortez, Marcos Fetter Lopes and Deutsch, C. V., 1999, Facies modeling accounting for the precision and scale of seismic data: application to Albacora Field, Centre for Computational Geostatistics Report 1.
- Besag, J., 1986, On the statistical analysis of dirty pictures: *Journal of Royal Statistical Society B*, **43**, 259-302.
- Chen, Q. and Sidney, S., 1997, Seismic attribute technology for reservoir forecasting and monitoring: *The Leading Edge*, **16**, 445.
- Coulon, J. P., Lafet, Y., Deschizeaux, B., Doyen, P. M. and Duboz, P., 2006, Stratigraphic elastic inversion for seismic lithology discrimination in a turbiditic reservoir: 76 th Annual International Meeting, SEG, Expanded Abstract, 2092–2096.
- Deutsch, C. V., 1998, Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection: *Computers and Geosciences*, **24**, pp. 551–562.
- Deutsch, C. V., 2002, *Geostatistical reservoir modeling*: Oxford University Press, Inc.
- Dumay, J. and Fournier, F., 1988, Multivariate statistical analyses applied to seismic facies recognition: *Geophysics*, **53**, 1151–1159.
- Fournier, F. and Derain, J. F., 1995, A statistical methodology for deriving reservoir properties from seismic data: *Geophysics*, **60**, 1437–1450.
- Johnson, R. A. and Wichern, D. W., 2007, *Applied multivariate statistical analysis*: Pearson Education, Inc.
- Matos, M. C., Osorio, P. L. M. and Johann, P. R. S., 2007, Unsupervised seismic facies analysis using wavelet transform and self-organizing maps: *Geophysics*, **72**, no. 1, P9–P21.
- Pyrz, M. J., 2004, Integration of geologic information into geostatistical models, Ph.D. thesis, University of Alberta.
- Saggaf, M. M., Toksoz, M. N., and Marhoon, M. I., 2003, Seismic facies classification and identification by competitive neural networks: *Geophysics*, **68**, 1984–1999.
- Scheevel, J. R. and Payrazyan, K., 1999, Principal component analysis applied to 3D seismic data for reservoir property estimation: Annual Technical Conference, SPE, Expanded Abstract.
- Silverman, B. W., 1993, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Scott, D. W., 1992, *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley and Sons, Inc., New York.
- Thodoridis, S. and Koutroumbas, K., 1999, *Pattern recognition*: Academic Press.
- Valet, L., Mauris, G. Bolon, P. and Keskes, N., 2001, Fuzzy seismic facies detection with interactive defuzzification method: IFSA World Congress and NAFIPS International Conference.

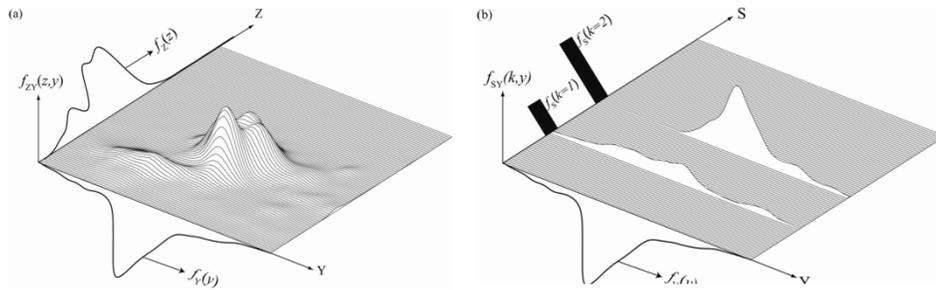


Figure 1: Bivariate relation between two variables. (a) represents bivariate distribution of continuous random variable Y and Z and (b) represents bivariate distribution of continuous variable Y and categorical variable S. Marginal distribution and marginal probability are shown on each axis.

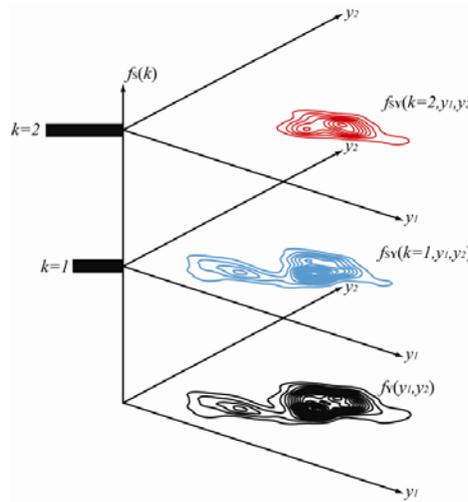


Figure 2: Trivariate modeling of RVs (S,Y). Two continuous variables and one discrete variable are considered.

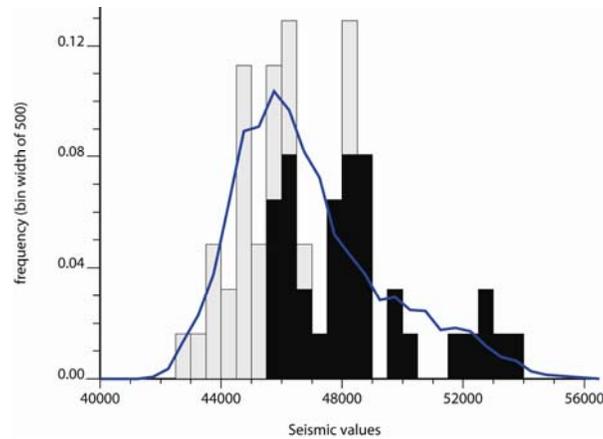


Figure 3: Distribution of a single seismic data. Gray and black bar charts correspond to two different facies, 2 and 1. The solid line represents the marginal distribution for the grid of 4225 exhaustive seismic values.

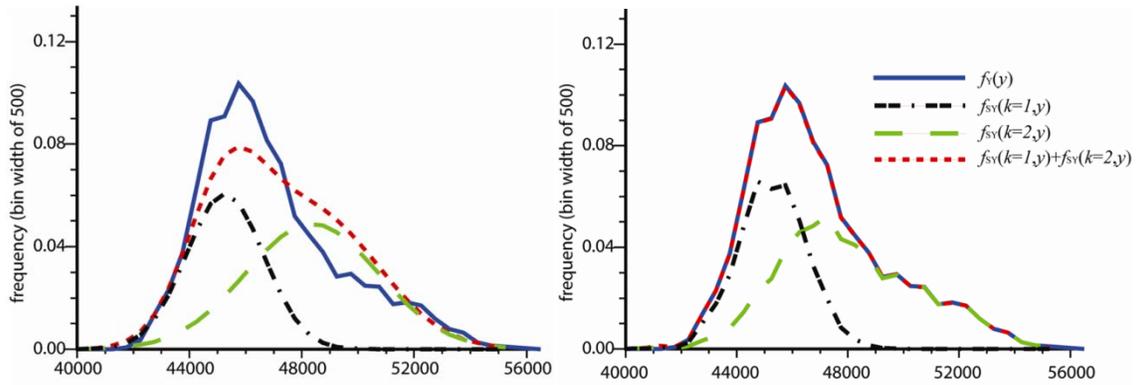


Figure 4: Initial fits to seismic data distributions of facies 1 and facies 2 (a) and the updated distributions are shown (b). The sum of the two distributions (classified as black and green dashed lines) is shown in red dashed line. It amounts to the well established marginal distribution after employing the iterative correction scheme.

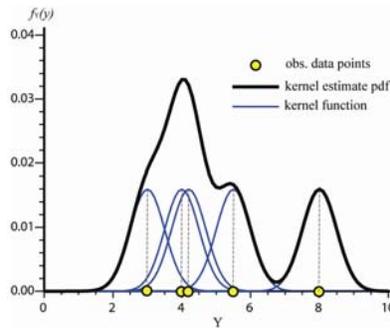


Figure 5: Kernel estimation method in univariate case. Gaussian kernel function is used for the demonstration.

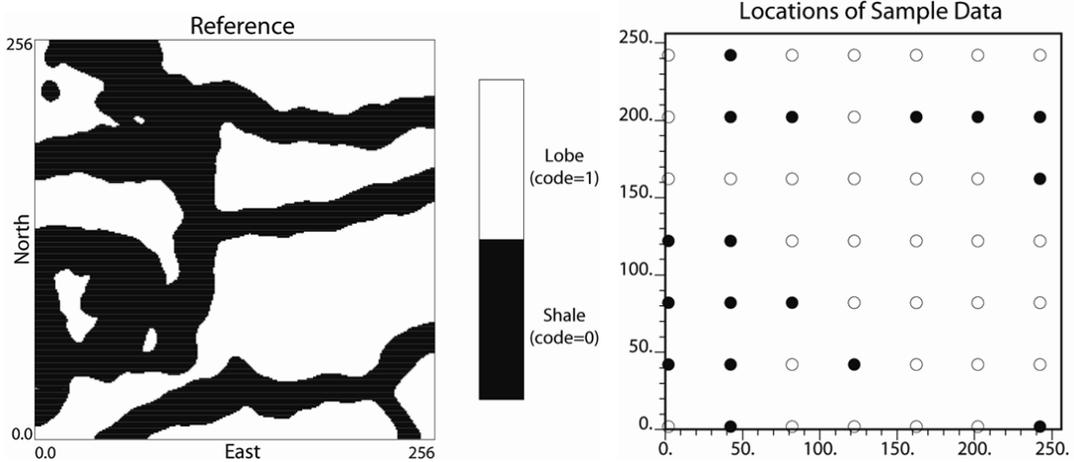


Figure 6: Reference synthetic image and well sample data extracted from the reference image.

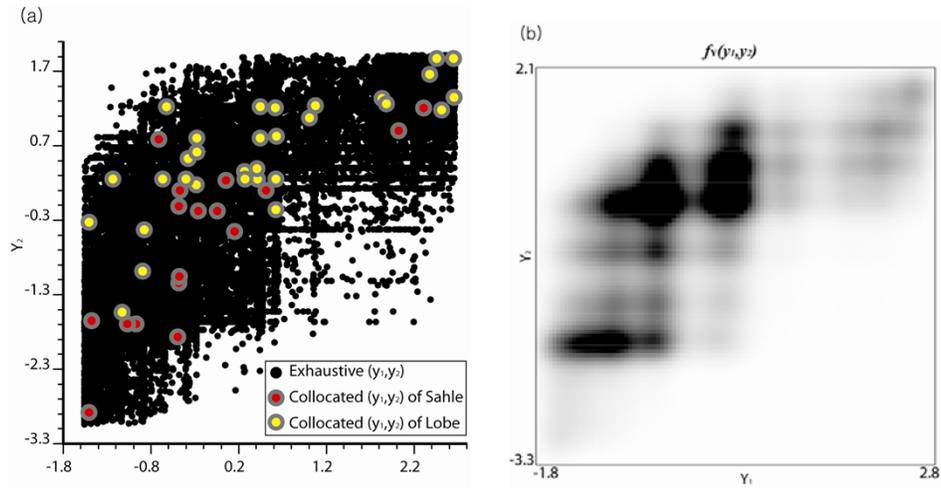


Figure 7: Data scatter plots of simulated seismic variables (a) and their estimated bivariate pdf (b). 62 seismic variable (y_1, y_2) measured at well locations are plotted in (a) as well.



Figure 8: Multivariate distributions of (S, Y) are inferred through kernel estimation method in the left column. The described methodology was applied to the initial distributions and their updated distributions are demonstrated in the right column.

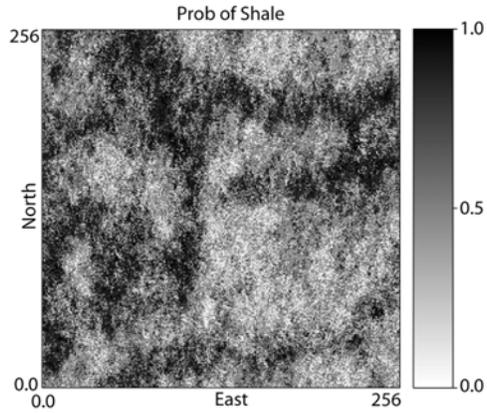


Figure 9: Probability map of shale conditioned to seismic data after employing the proposed method.

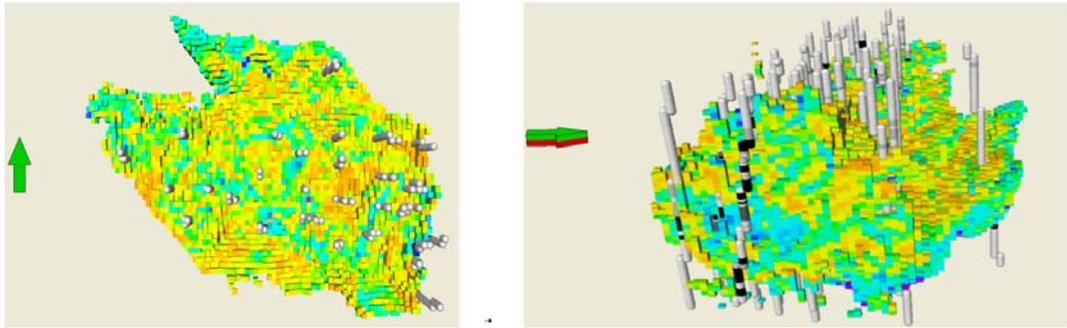


Figure 10: 3-D visualization of the reservoir grids. 28 vertical drilling wells are shown and one of seismic attribute is color-coded.

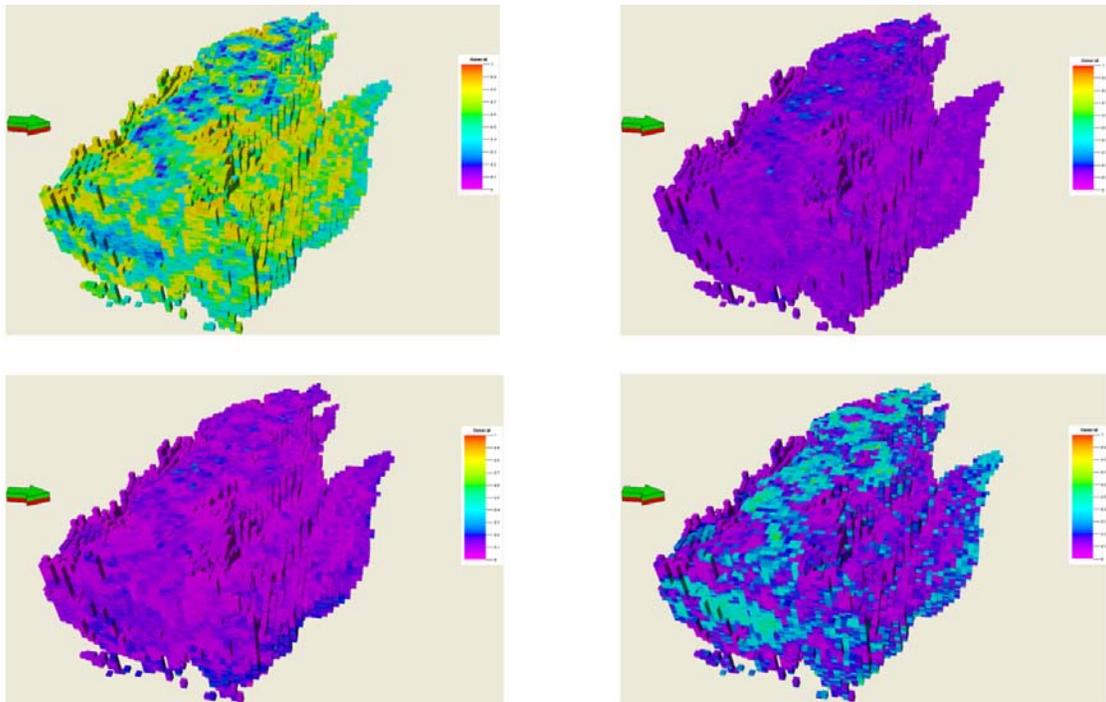


Figure 11: Estimated probability cubes for 4 facies (shale, poor sand, good sand and channel from left upper corner). Facies 2 and 3 are barely recognized, but shale and channel has good recognition.

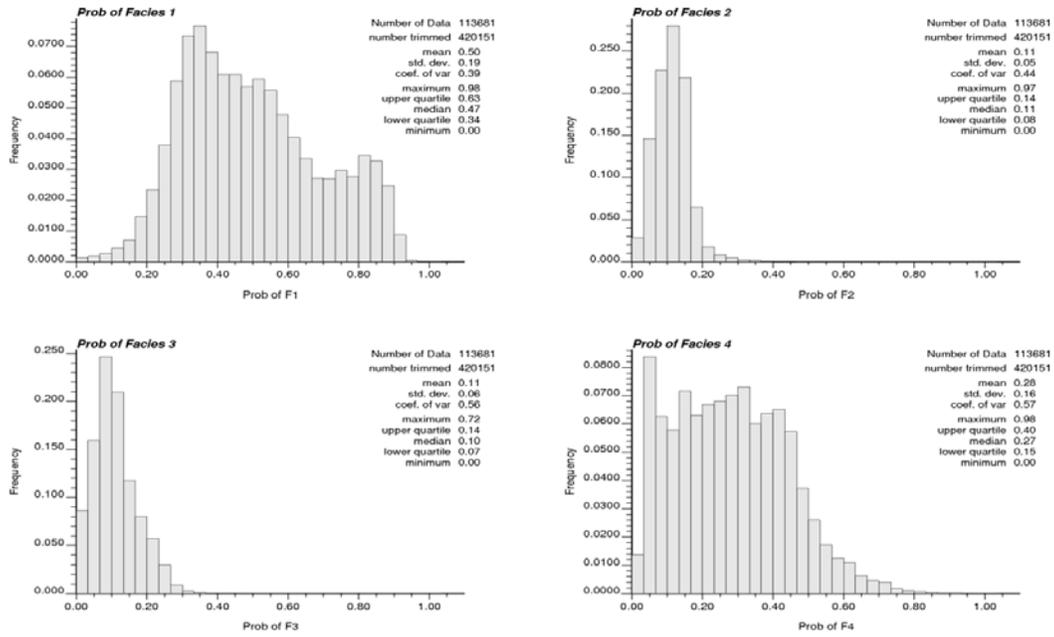


Figure 12: Histograms of the obtained facies probabilities. Facies 1 and 4 has a good separation, and facies 2 and 3 show highly overlapped distributions.

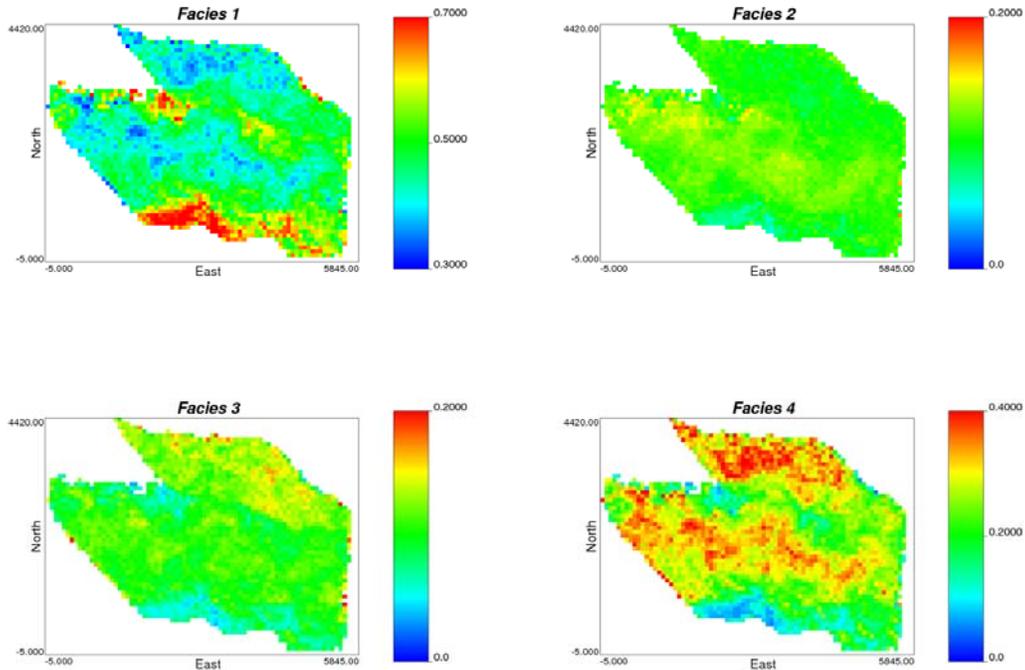


Figure 13: Estimated facies probability map in 2-D. Vertically averaged seismic data is integrated. Shale (facies 1) and channel (facies 4) has good recognition.

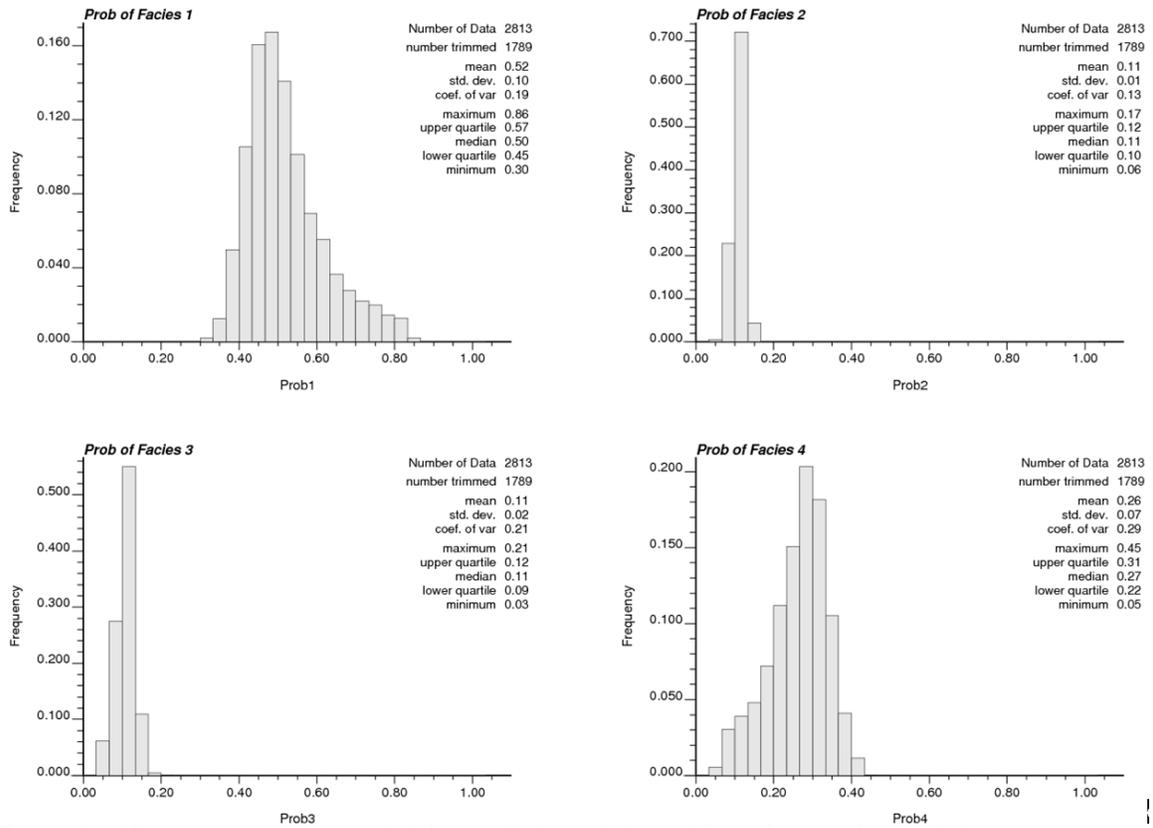


Figure 14: Histograms of facies probabilities. Shale and channel facies has good separability.