Incorporation of Spatial Characteristics Into Volcanic Facies and Favorable Reservoir Prediction

Hong Tang, SPE, Louisiana State U. and **Hancheng Ji,** U. of Petroleum, Beijing

Summary

Compared to clastic reservoirs, volcanic reservoirs exhibit higher heterogeneity. Lithological facies type is one of the most important indicators of favorable volcanic reservoirs. Traditionally, facies are identified by core observation or log classification. However, spatial-distribution characteristics and geological conceptual models, which are important in the early stages of exploration, are seldom incorporated quantitatively in facies prediction. Based on previous work, a new method has been developed to incorporate volcanic spatial information with limited well data (three wells) to improve facies prediction. This method was applied to a volcanic clastic reservoir of the Cretaceous Yingchen member of the Xinshan fault depression, northeastern China. For better well control, an artificial neural network (ANN), a beta-Bayesian method (BBM), and a discriminant analysis (DA) algorithm, were used to predict log-based facies. Confidence analysis was applied to evaluate the log facies prediction. Analysis of variance (ANOVA) verifies that the overall prediction accuracy is above 82%.

Indicator kriging was used to estimate the conditional probabilities of facies occurrence given residual thickness. This is based on the assumption that the residual thickness of the volcanic formation is controlled by distance from the eruption center, a major factor defining the geological facies. The geological conceptual models (areal sedimentary facies maps and diagenetic facies maps) were converted into the conditional probability of facies occurrence in given geological settings using multinomial logistic regression. These conditional probabilities were combined with well-log facies data within a Bayesian framework. Three favorable reservoirs were predicted based on the method above, and the predictions were proved by the subsequent drilling.

Introduction

Volcanic reservoir quality is controlled by both lithofacies and diagenetic effects. Traditionally, these effects are qualified by core observation and well-log (especially image-log) interpretation. The spatial distribution of reservoirs is characterized by high-resolution seismic interpretation. Even though the importance of these features is well known among the geological community, it is difficult to quantify and integrate these data into reservoir modeling and flow simulation.

The area of study is in the early Cretaceous (Yingchen) volcanic formation, Xinshan fault depression, Songliao basin, northeastern China. Several factors have made the traditional approach less practical. First, the available well data are limited (three wells were drilled in this area). Second, the volcanic formation is deeply buried (3000 to 6000 m) in the Xinshan area. It has a high seismic amplitude contrast with bounding sedimentary formations but low intralayer reflection. Seismic properties appear homogeneous for most volcanic formations (Zhao 1999); furthermore, compared to a clastic reservoir, the spatial distribution of volcanic reservoirs is less continuous. Reservoir quality is highly heterogeneous because of complex lithology, facies and diagenetic overprints. Detailed characterization is, thus, more difficult to conduct.

This paper (SPE 90847) was first presented at the 2004 SPE Annual Technical Conference and Exhibition, Houston, 26–29 September, and revised for publication. Original manuscript received for review 4 June 2004. Revised manuscript received 4 April 2006. Paper peer approved 26 July 2006.

There are two major challenges: (1) to quantify conceptual geological models and integrate them with other types of data (e.g., seismic and log data) and (2) to accurately identify volcanic lithofacies with limited well-log data.

To incorporate diverse data into lithofacies prediction, reservoir-quality assessment, and uncertainty reduction, two types of methods are frequently used: statistical (geostatistics) methods and ANN methods. Geostatistics methods such as cokriging or indicator cokriging (Goovaerts 1998; Deutsch and Journel 1998; Yarus and Chambers 1994) are versatile and widely used. However, these methods require cross-variogram models for all indicators, which are tedious and time-consuming to construct. In addition, this approach does not guarantee better results (Deutsch and Journel 1998). Alternatively, object-based simulation, such as Boolean simulation (Yarus and Chambers 1994), can be used to integrate geological models into reservoir modeling and reproduce the exact geometry of the facies model. It requires good geometric parameters, which can be estimated from outcrop analogs or detailed seismic interpretation. However, those parameters are seldom available and are less representative in this fault-complicated area. ANN methods are powerful for integrating high-dimensional data and expressing complex, nonlinear relationships between input and output. Several successful applications using neural networks for data incorporation and facies prediction have been reported (Wong et al. 1995; Siripitayananon et al. 2001; Bhatt and Helle 2002). However, the computation cost of ANNs is high, and they do not provide any estimate of uncertainties

A new method is developed in this paper to integrate geological deterministic features (lithofacies and diagenesis models) with log and seismic residual thickness data. Seismic residual thickness is used because it is more reliable than other interpreted seismic attributes. Log lithofacies was identified using statistical, neural-network, and hybrid methods. The performance of these methods is compared using ANOVA.

Basically, the method includes four connected steps: classification of log lithofacies, transformation of seismic residual thickness to give the probability of reservoir quality, quantification of reservoir geological models into reservoir quality, and the combination of all three types of data into the classification of reservoir quality. This mixed method preserves both petrophysical information close to the well and seismic-geological information distant from the well. It is systematic, straightforward, and easy to compute.

This paper briefly describes the main methods used for facies classification and data integration, followed by an introduction to the geological model. Finally, the application of this method is discussed.

For brevity, in this paper, reservoir-quality level is referred as "reservoir quality"; the geological prior probability of reservoir quality is referred as "prior." Reservoir quality is defined as an association of lithofacies and diagenesis facies. This definition is illustrated in the reservoir geological model section later in this paper.

Description of Methods

Because volcanic lithofacies consist of complex mineral components and are vulnerable to diagenetic transformation, traditional lithology-interpretation methods using a single-log curve such as spontaneous potential (SP) or gamma ray (GR) are not adequate. Multivariate statistics and ANNs can use multiple logs to quantify lithofacies and are used widely for complex facies classification. One purpose of this paper is to compare the performances of three methods: DA, probabilistic neural networks (PNNs), and a BBM. Furthermore, multinomial logistic regression is a powerful classification tool used widely in production risk mitigation (Wisnie and Zhu 1994; Shultz and Fischbeck 1999) but rarely applied in reservoir facies prediction.

Discriminant Analysis. DA uses a discriminant criterion to classify each observation into one of the mutually exclusive groups. The discriminant criterion is determined by generalized squared distance, which is a function of a covariance matrix of existing sample data or cores. Depending on sample-data availability, such a criterion can be computed from either the individual within-group covariance matrices or a pooled covariance matrix. The discriminant criterion is expressed as a function of covariance matrices. With adequate data, the individual within-group covariance matrices yield a quadratic function; the pooled covariance matrix yields a linear function, which is useful when sample data are rare (*SAS/SAT User Guide* 2002).

DA can also take into account priors, termed as vectors of lithofacies fractions, the sum of which is equal to 1. Priors can be estimated in several ways. First, with no geological knowledge, priors simply can be assumed to be equally probable; hence, the prior probability is $1/n$ for all groups, with n representing the number of groups. Alternatively, the fraction of training sets can be treated as a prior by assuming that the training-set fraction reflects the real fraction of lithofacies. Practically, the prior can also be adjusted on the basis of geological knowledge.

Probabilistic Neural Network. PNN uses a radial basis function (RBF), which is a symmetric Gaussian type function. The RBF is similar to the variogram model in kriging, but it does not require the fitting of experimental variogram models. The estimator is a linear combination of different RBFs. The weights of linear combinations are estimated by a gradient-descend method. The design of PNN is guaranteed to converge to a Bayesian classifier, provided that enough training data are available. However, it involves more computation steps and thus is slower than other kinds of neural networks (Neural Network Toolbox 1998).

PNN has two layers: the radius basis layer and the competitive layer. The radius basis layer computes distances from the input vector to training input vectors. The competitive layer sums these contributions for each class of inputs to produce a vector of probability. Then, the most probable group is assigned as the output group. MATLAB6.1 (Neural Network Toolbox 1998) is used to design a PNN.

Bayes Theorem and BBM. The Bayes method is a classic statistical method that can be expressed as:

$$P(Q_i|L) = \frac{P(Q_i)f(L|Q_i)}{\sum_{i=1}^{n} P(Q_i)f(L|Q_i)} \quad i = 1 \ldots n, \quad \ldots\ldots\ldots\ldots\ldots (1)$$

where $P(Q_i|L)$, the posterior probability, is a conditional probability such as reservoir quality from given log data. $P(Q_i)$ is the prior, and $f(L|Q_i)$ is a conditional probability density function (PDF) such as log reading for each reservoir-quality type. This is often referred to as "likelihood." The denominator is a normalization term.

As with the estimation of priors, there are several ways to compute PDF $f(L|Q_i)$. One direct method is the bin-dependent method (Karpur et al. 2000). Conditional data are subdivided into several intervals or bins (Q_i, $i = 1 \ldots n$); then, the proportion within all intervals can be calculated as $f(L|Q_i) = n_i/N$, where n_i = counts within each interval, and N = total counts for all intervals. However, such a bin-dependent method requires a large number of cores, which is often impractical. Furthermore, selection of bin size is subjective and less accurate. As noted in Karpur et al. (2000), "If too few bins are selected, the facies occurrence probability (FOP, referred to in this paper as posterior probability) lacks the ability to discriminate between adjacent log readings; if there are too many bins, the FOP will not be estimated precisely."

Alternatively, the BBM (Tang et al. 2004; Armenta et al. 2003) can overcome the above dilemma. This method uses an empirical beta function to fit the distribution of $f(L|Q_i)$ instead of selecting discrete bins. The beta function **(Fig. 1)** is chosen because it is simple to understand and flexible to manipulate. The shape of the beta function is controlled by only two factors: shape factor a_β and position factor b_β. For example, when a_β and b_β are equal, the PDF curves are symmetric. When a_β and b_β both equal 1, the beta function becomes a uniform function. Combined with the Bayes theorem, BBM can be used to predict lithofacies.

Multinomial Logistic Regression (MLR). Logistic regression is a regression technique for analysis of categorical data. It is often used to investigate the relationship between categorical data and a set of explanatory variables. MLR uses the logarithm transformation of odds ratio, termed as Logit. The Logit is defined as the natural log of fraction of one facies to the reference facies. The categorical variables can be specified as linear combinations of explanatory variables. Thus, interaction terms of these categorical variables can be specified in the same way as the general linear model. Unlike ordinary linear regression, MLR uses the maximum likelihood (ML) method to estimate regression coefficients. In this study, Logit is used to compute the likelihood of a certain quality type given binary indicators such as the existence of erosional facies. Posterior probability for each quality type is inferred, which can be combined with other types of data.

Modeling Steps

The following four steps are applied in reservoir-quality classification **(Fig. 2).**

1. Well-log lithofacies are classified and verified with statistical and ANN methods. The resulting lithofacies proportions are assumed to approximate the real reservoir lithologic proportions that are taken as "prior" for the following steps.
2. By assuming that the volcanic residual thickness is controlled by the distance from the eruption center, interpreted seismic residual thickness is applied to estimate the conditional probability of the reservoir quality using an ordinary indicator kriging method.
3. MLR translates geological deterministic models (lithofacies and diagenesis model) into probability of reservoir quality. Training sets or calibration data come from both drilled areas and areas far from the eruption center, which can be safely assumed to have poor reservoir quality. The trained model is used to predict the whole study area. The likelihood derived from posterior probability can be combined with indicator kriging output within the Bayes framework.
4. Bayes' theorem was used to combine well-log data and seismic thickness with a quantified geological model to predict the reservoir quality.

Reservoir Geological Model

The geological background is discussed extensively in the literature (Feng et al. 2003; Yang et al. 2003; Guo 2002; Tang et al. 2001; Chen 2002).

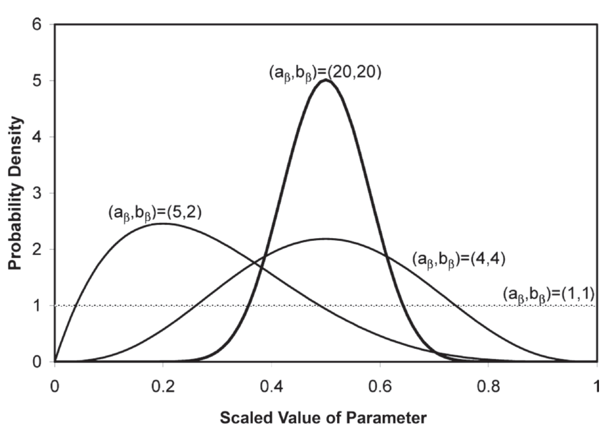

Fig. 1—Beta function PDF and related parameters (*SAS/SAT User Guide* 2002).

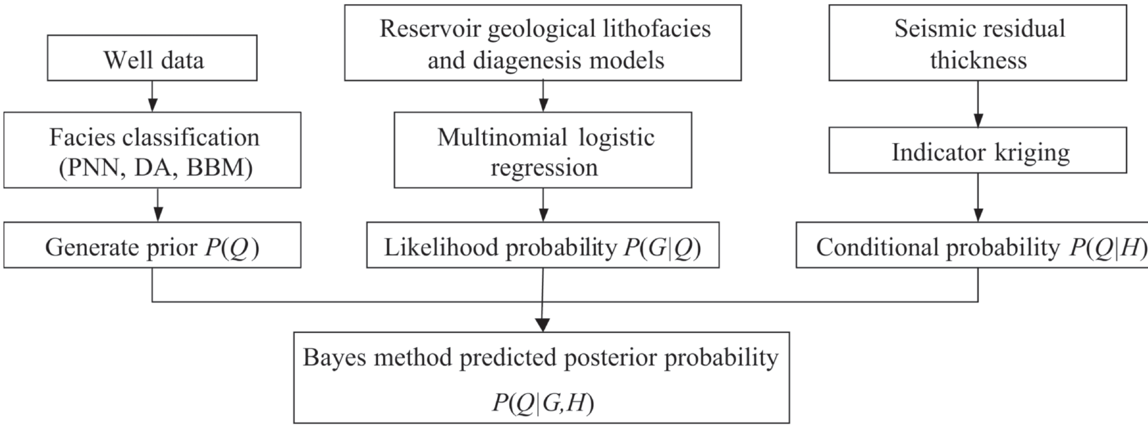

Fig. 2—Flow chart of volcanic reservoir-quality prediction. Using different statistical and neural-network methods, three types of data are integrated to predict reservoir quality within the Bayesian framework.

Structural and Sedimentary Model. In the study area, the early Cretaceous [locally called early Yingchen (Y1)] is subdivided into two parts. Because during Y1, the active basin boundary Xuxi fault, as well as the local fault series, controlled the volcanic eruption centers, volcanic sediments were widely distributed. For the later Yingchen (Y2), the basin-boundary fault activity decreased. The basin was changed from a rift basin to a constantly subsiding depression. As a result, volcanic sedimentation was reduced and deposition changed to normal fluvial, lacustrine sediments.

During the main rifting period Y1, the basin-boundary fault controlled the sedimentary basin. The volcanic structures were distributed along the local fault series, which were divided into a west main fault zone, a middle normal fault zone, and an east reverse fault zone. The volcanic structures in the study area are characterized as stratovolcano **(Fig. 3).** These are formed by repetitive volcanic eruptions. They are composed of volcanic clastic sediments and small lava flows. Strombolian and vulcanian-style weak eruptions are predominant; they include multiplayer, multigenetic volcanic and clastic sediments (Zhao 1999; Chen 2002). The volcanic lithology includes rhyolite, volcanic tuff, andesine, and basalt. Lithofacies include air-fall facies, base-gush facies, lava-flood facies, and pyroclastic facies. Compared with the juxtaposing clastic deposits, the distribution of these facies changes rapidly, both laterally and vertically. During the Y2 period, basin sediments changed from volcanic sediments and volcanic sedimentary sediments to normal clastic sediments. Braided fluvial and fluvial-fan delta sediments are the main sedimentary facies. To combine well lithofacies with the geological conceptual model, they are simplified into four major quality types (I–IV) based on petrophysical, seismic, and geological characteristics and spatial distributions **(Fig. 4).** The volcanic lithofacies and their related reservoir qualities are tabulated in **Table 1.**

Diagenesis Model. Because minerals in volcanic deposits are less stable than those in clastic deposits, they are more vulnerable to diagenetic effects. According to a previous study, structural fractures, diagenetic (including hypergenetic) fluid flow (downward meteorological flow and upward acid flow), burial processes, and paleothermal fields are the main diagenetic effects that change reservoir quality. Thus, three types of diagenetic facies are included in our model (Fig. 4):

- Erosion-enhanced facies (A).
- Structural fracture-enhanced facies (B).
- Compressed and cemented facies (C).

Due to the lack of drilling and log information, the diagenetic model (as well as the sedimentary facies model) is generated on the basis of the geological conceptual model and our hypothesis.

Reservoir-Quality Geological Classification. On the basis of drilling observations and geological analogs from the surrounding area, reservoir-quality types are classified on the basis of porosity, permeability, residual thickness, and lithologic facies (Table 1). Reservoir quality increases with an increase in residual thickness. Continuous thickness is then classified into four categories (Types I, II, III, and IV). Indicator kriging uses semivariogram models estimated for each indicator to predict the probability of occurrence for reservoir quality. The geological model is coded as binary data (indicators) in this study. For example, pixels of erosion facies are coded 1, otherwise 0. Multinomial logistic regression uses these geological indicators to quantify reservoir quality.

Application

Log Lithofacies Classification. All lithofacies were classified using DA, BBM, and PNN. ANOVA was applied to analyze their classification performances. In Well ZS6, six lithofacies were identified: compact gray felsites, compact slate, shale, siltstone, silt sand, and conglomerate. For all methods, training data were randomly selected from half the core samples. The other half were used for validation. At uncored normal sediment intervals, training data were derived from mud-log and well-log interpretations. To test the variations between each method, these procedures were repeated four times by randomly choosing four training sets.

Probabilistic Neural Network. The PNN has two layers: the radius basis layer and the competitive layer. The RBF is the distance between weight vector W and input vector V multiplied by the bias b. The RBF is expressed as:

$$A = \text{Radbas}\,(\|W.V\|b), \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

where $\|W.V\|$ is the Euclidian distance between input vector V and weight vector W, which is analog to the distance matrix in DA, and Radbas represents the RBF. Bias b is similar to the prior in DA.

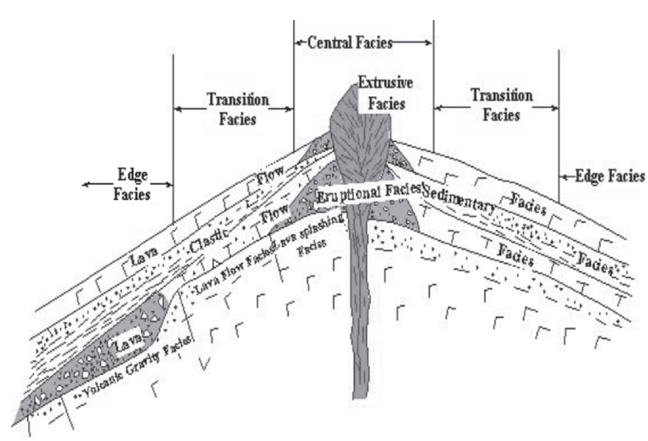

Fig. 3—Volcanogenetic lithofacies model (Zhao 1998).

Fig. 4—Geological lithofacies model and diagenesis model. (Left) The lithofacies model indicates that the volcanic eruption center controls the lithofacies distribution. F1: central facies; F2: transitional facies; F3: edge facies; F4: sedimentary or pyrosedimentary facies. (Right) Diagenesis model: (A) fracture-enhanced facies, (B) erosion-enhanced facies, (C) compressed and cemented facies.

The RBF has a symmetric bell shape; when the input of RBF is 0, it has a maximum value of 1. As the distance between W and V decreases, the output increases, and vice versa. Thus, the bias b allows the sensitivity of the neurons to be adjusted. As noted in Neural Network Toolbox (1998), "When the bias or spread = 0, the network will act as a nearest-neighbor classifier. As the bias becomes larger, the designed network will take into account several nearby vectors." By adjusting the spread b, PNN can reach the optimum value for the training set **(Fig. 5).**

Discriminant Analysis. The discriminant criterion was measured by the generalized squared distance of training sets. Because of the unbalanced sample size, some lithofacies groups had fewer data than others. The classification criterion based on the pooled covariance matrix yielded better results (84.7%) than that based on the individual within-group covariance matrices (75.2%). Priors computed from the sample proportion yielded better results than uniform priors. Cross validation proved useful to check the performance of classification **(Table 2).** In one case in which pooled variance was used, the overall accuracy was very good (91%). This may be because of the strong log-response contrast between volcanic and clastic lithology. The prediction accuracy for slate and felsite was even better (at 97% and 100%, respectively).

The BBM also generated an 82.5% prediction accuracy **(Fig. 6).** ANOVA indicates that within a 95% confidence interval, there are no significant differences between all the methods (P value = 0.754) **(Table 3).**

Quantification of Seismic Residual Thickness

The quality type was defined in terms of binary indicators.

$$i(x, y) = \begin{cases} 1, h(x, y) \text{ within certain thickness range} \\ 0, h(x, y) \text{ not within that range} \end{cases} \quad \ldots\ldots\ldots (3)$$

Indicator semivariograms were computed using these quality indicators **(Fig. 7).**

The reservoir quality prior derived from the above log classification was treated as a global cumulative distribution function (CDF) to guide indicator kriging. Euclidean distances between observation pairs were computed in the major and minor axis directions. This represents the major and minor continuous directions of a given reservoir type. Nested spherical and exponential models were used to fit the parabolic trends of these variograms. Ranges or spatial continuity of certain indicators varied from 6500 to 23 000 m. Indicators I and IV show zonal heterogeneity. Longer ranges and lower variances in the major direction indicated that the major axis direction (northwest/southeast) was more continuous and has less variation. Zonal anisotropy was modeled by adding a geometrically anisotropic variogram. This variogram was designed to have no effect along the major axis by providing it with a very large range. There seems to be a hole effect in Indicator II. The smooth semivariograms are partly caused by large and regular grids. Small nugget effects were added to semivariograms to avoid numerical instability (Goovaerts 2002). These variogram models

TABLE 1—VOLCANIC RESERVOIR QUALITY DIVISION (ZHAO 1998): RESERVOIR TYPE

	Porosity (%)	Thickness (m)	Permeability ($\times 10^{-2} \mu m^2$)	Fracture Prevalence	Volcanic Lithofacies
I	> 10	120–180	> 10	High	Air fall facies, lava flow facies, pyrodebris flow facies
II	5–10	80–120	1–10	Medium	Lava flow facies, base gush facies, pyrodebris flow facies, pyroclastic facies
III	1–5	60–80	0.1–1	Medium-low	Pyrodebris flow facies, pyroclastic facies, pyroclastic sedimentary facies
IV	< 1	< 60	< 0.1	Low	Intrusive facies, hypovolcanic facies, pyroclastic sedimentary facies

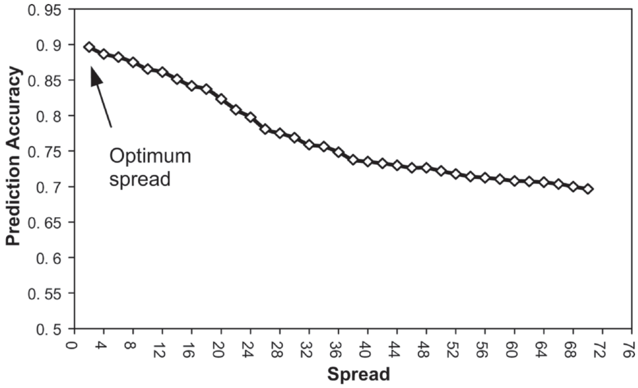

Fig. 5—Relationship of PNN prediction accuracy and spread (bias).

were used in indicator kriging. For each pixel, the posterior probabilities of occurrence for four reservoir qualities were computed using the IK3D procedure (Deutsch and Journel 1998). As expected, the prediction results presented in **Fig. 8** illustrate that the residual thickness is spatially elongated in the northwest/southeast direction. However, because kriging smooths the real spatial variation, the prediction result is too continuous and too smooth. Geological-model information should be included to update the reservoir-quality prediction.

Quantification of the Geological Model

Even though it is widely accepted in the geology community that the conceptual geological model should act as a guide in reservoir characterization (particularly with sparse well control), few studies have been reported that quantify the conceptual models. The multivariate statistical method provides a powerful vehicle to quantify the subjective geological model. In this study, the lithofacies model and the diagenesis model are quantified as binary indicators:

$$i(x, y) = \begin{cases} 1, \text{ lithofacies or diagenesis facies exists.} \\ 0, \text{ lithofacies or diagenesis facies does not exist.} \end{cases} \quad \ldots\ldots\ldots (4)$$

The logistic regression method classifies these indicators into categorical quality types. The linear model is given by

$$\log\left(\frac{\boldsymbol{P}_i}{\boldsymbol{P}_{\text{ref}}}\right) = \beta_0 + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \ldots + \beta_m\mathbf{X}_m + \varepsilon \quad i = 1, \ldots n-1, \quad \ldots\ldots\ldots (5)$$

where $\log\left(\frac{\boldsymbol{P}_i}{\boldsymbol{P}_{\text{ref}}}\right)$ is the Logit of the odds ratio, $\boldsymbol{P}_i$ is the probability of group I, and $\boldsymbol{P}_{\text{ref}}$ is the reference-group probability. Any group can be a reference group (Reservoir Type IV is chosen in this study). $\mathbf{X}_i$ is a vector of binary lithofacies and diagenetic facies indicators. β represents the regression coefficients; n is the number of groups, such as quality types; m is the number of binary indicators, and ε is the error term. Similar to log facies classification, where core data are used as training data, the training data here come from either a confirmed drilling area or an edge area. The edge area has less volcanic residual thickness and can be assumed as a low-quality volcanic reservoir.

The ML method was used to estimate regression coefficients. A probability vector for all reservoir types was computed in each location. The reservoir type with the highest probability was assigned to that location **(Fig. 9).** As expected, this visually reproduces the predicted lithofacies and diagenesis models. However, more distant from the drilled areas, the model tends to lack detail. This is because such a model only takes into account the point geological data and lacks spatial information.

Fortunately, logistic regression provides a conditional PDF, which can be integrated for spatial reservoir-type classification within a Bayesian framework.

Integrated Reservoir-Quality Prediction

A mixed method was introduced to combine well-log data, seismic data, and geological data to predict quality type. Instead of geological priors, the probability of reservoir quality $P(Q_i|H)$ was computed from seismic residual thickness and treated as a prior in Eq. 1.

Although there was no existing software to compute the PDF, or likelihood, a simple two-step method proposed by Goovaerts (2002) was used in this study. First, by assuming a constant prior $P(Q_i) = 1/n$, the conditional PDF $f(G|Q_i)$ is proportional to the posterior Bayesian probability from Eq. 1. This was termed "pseudoposterior" probability $P_p(Q_i|G)$, which yields the same results as the theoretical expression. Then, instead of $f(G|Q_i)$ itself, the pseudoposterior probability $P_p(Q_i|G)$ was used in the Bayes equation:

$$P(Q_i|G, H) = \frac{P(Q_i|H)P_p(Q_i|G)}{\sum_{i=1}^{n} P(Q_i|H)P_p(Q_i|G)} \quad i = 1 \ldots n, \quad \ldots\ldots\ldots (6)$$

where $P(Q_i|G,H)$, the prediction, is the posterior probability of a certain quality type given the geological data G and the residual thickness H. The mixed method generates a more reasonable map **(Fig. 10)** and integrates all three types of data. Close to the well, the drilling data are honored; the geological model and seismic data are dominant at places farther from the wells. Meanwhile, the spatial trend computed from seismic geostatistics interpretation is preserved.

Discussion

Log-Prediction Evaluation. One advantage of the statistical classification method is that predictions can be evaluated statistically **(Fig. 11).** Two probability curves have been introduced to evaluate the prediction. First, the probability of the most probable facies represents the overall confidence (OC) of the prediction. If the OC

TABLE 2—DA PREDICTION ACCURACY CROSS VALIDATION (WELL ZS6)

Observed Lithofacies	Observation Classified Into Lithofacies 1	2	3	4	5	6	Total
1	66	0	2	0	0	0	68
2	0	380	0	0	0	0	380
3	2	1	363	1	29	1	397
4	3	0	1	25	1	0	30
5	0	0	47	2	82	0	131
6	0	1	4	1	2	123	131
Total	71	382	417	29	114	124	1137
Prediction Accuracy	0.97	1.00	0.91	0.83	0.63	0.94	0.91

Facies No.: 1. Condensed gray felsites; 2. condensed slate; 3. shale; 4. siltstone; 5. silt sand; 6. conglomerate.

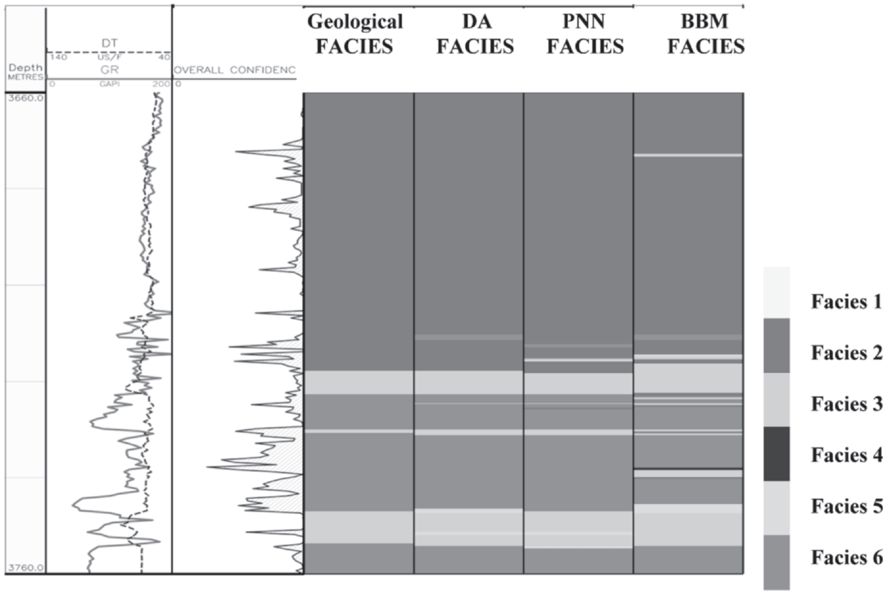

Fig. 6—DA, PNN, and BBM illustrate good prediction results. The prediction accuracy is evaluated by statistics—overall confidence (from BBM).

is high, the prediction is more certain and vice versa. High OC also means that the log responses of a facies are very similar to the log responses of the identified facies. Second, the probability difference between the two most probable facies illustrates the distinction ability (DIA) of the log prediction. A high DIA means that the log response at a depth is highly different from the responses of other facies. In other words, the facies prediction is conclusive, and vice versa. When the two most probable facies are equally probable, DIA equals zero. It means that at least one alternative facies is as equally possible as the prediction. Thus, the alternative facies should be generated for reference.

Drilling Validation. On the basis of seismic interpretation, the volcanic formation is seen to be widely distributed (the area is approximately 255 km^2, with an average thickness of 150 m); however, favorable volcanic reservoirs are quite limited. Three Type I reservoir targets are predicted; these reservoirs are 60 km^2 in total area based on the analysis of sedimentary facies and diagenetic effects.

A dry well, ZS6, was drilled in the late 1990s (Fig. 9). It turned out to contain condensing felsites and base slate with no production. The FS9 well on Target II encountered a fractured reservoir and achieved a gas production rate of 50 938 m^3/d. The ZS8 well, located at Target I, was drilled later; it found a fractured and erosion-enhanced reservoir. A production test proved the presence of a natural gas-cap reservoir and achieved a production rate of 11 221 m^3/d.

Based on the prediction, a new well, ZS10, was drilled. As expected, it encountered a fractured reservoir. After fracture enhancement, a natural-gas production rate of 40 000 to 60 000 m^3/d was achieved (Yang et al. 2003; Guo 2002; Tang et al. 2001; Chen 2002). This validated the reservoir-quality prediction.

Conclusions

DA, PNN, and BBM are useful in the well-log classification of complex volcanic lithofacies. By adjusting the parameters in the statistical models and neural-network, a prediction accuracy of 82% can be achieved. An ANOVA test indicates these independent classifications to be comparable in quality.

The BBM using the beta function to fit a conditional PDF avoids the error caused by the traditional bin-dependent method. It is easy to manipulate and has good prediction accuracy. The classification method can be applied in a spreadsheet.

Indicator kriging can be used to convert seismic-interpreted residual thickness into reservoir-quality types. Conditional probability can be generated for all indicators to integrate with other data.

Geological lithofacies and diagenesis models are important and are used qualitatively in geological communities. These now can be quantified into indicator variables. With these quantified indicators, the multinomial logistic regression method can be used to predict reservoir quality.

Assuming independence between the geological model and seismic data, the Bayes method can be used to generate an integrated prediction of reservoir quality from seismic data, geological

TABLE 3—METHOD COMPARISON OF ANOVA

	Groups	Count	Sum	Average	Variance	
	BBM	4	3.30	0.825	0.0008333	
	DA	4	3.32	0.830	0.0008667	
	PNN	4	3.39	0.848	0.0035583	
Source of Variation	SS	df	MS	F	P-Value	F crit
Between groups	0.001117	2	0.0005583	0.319	0.735	4.256
Within groups	0.015775	9	0.0017528			
Total	0.016892	11				

Note: SS = sum of squares; df = degree of freedom; MS = mean square; F = F-value; P = P-value; F crit = critical value of F.

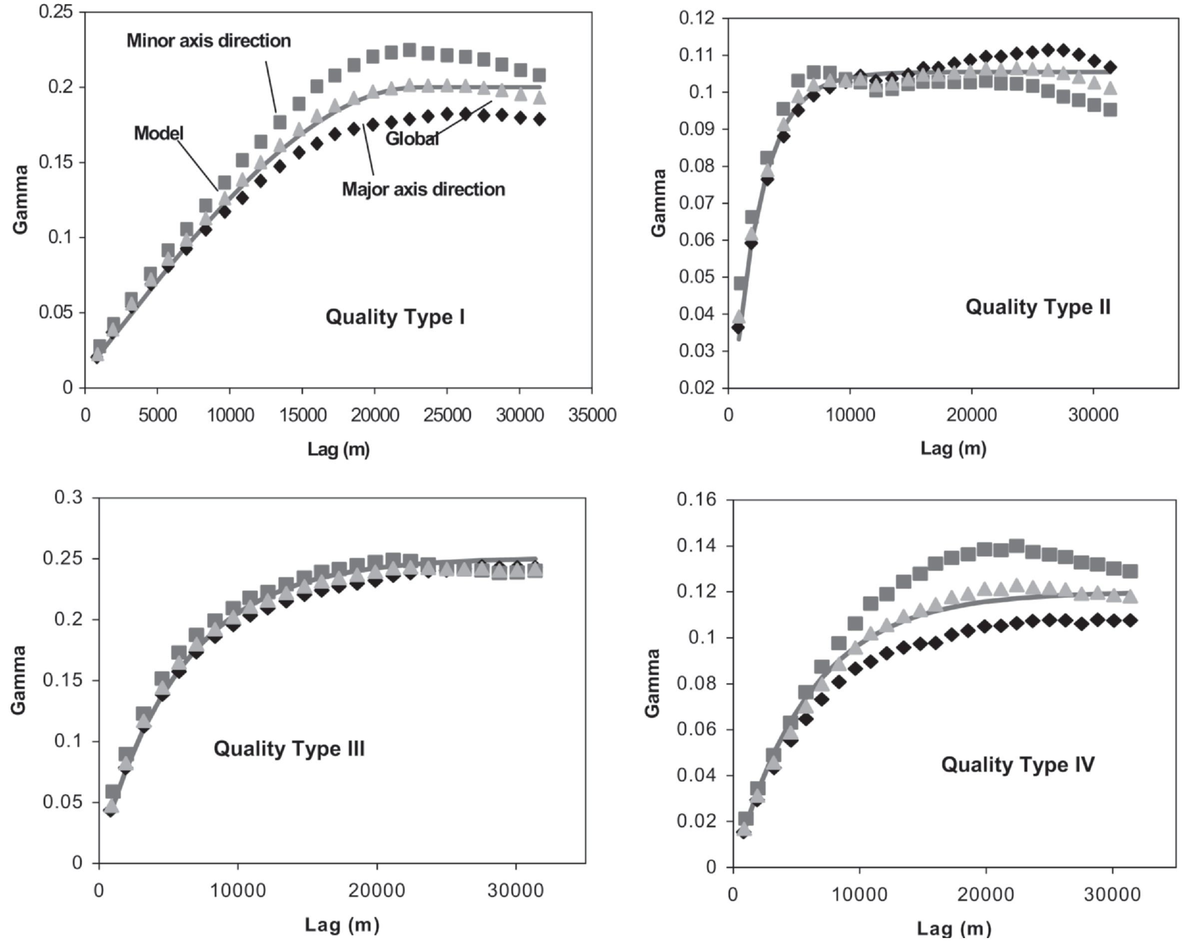

Fig. 7—Semivariogram models for four indicators. These semivariograms are computed from normal score transformed residual thickness (a); (d) Quality Types I and IV have zonal heterogeneity; the major axis direction is more continuous than the minor axis direction.

models, and well data. This method achieved good prediction results, which honored drilling information and predicted three targets. Subsequent drilling verified the prediction.

Nomenclature

A = RBF output
b = bias
h = residual thickness range
H = residual thickness
L = wireline-log data
n = total number of groups
$\boldsymbol{P}$ = probability

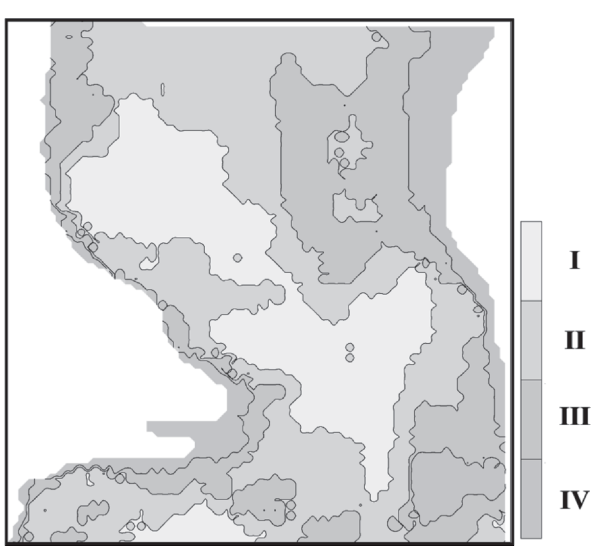

Fig. 8—Reservoir-quality geostatistical prediction. Seismic indicator kriging results are too smooth because of the kriging.

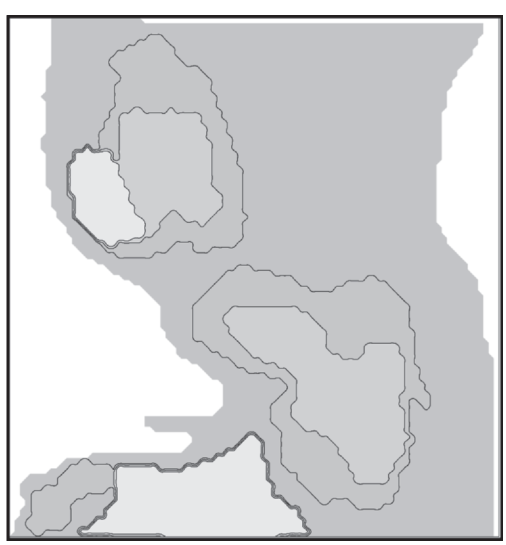

Fig. 9—Reservoir-quality statistical prediction. Using logistic regression, geologic models are transformed into reservoir quality. However, it lacks details distant from the well (same legend as Fig. 8).

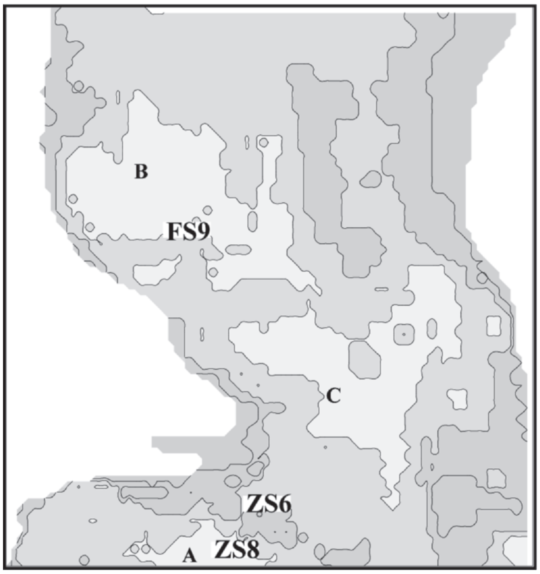

Fig. 10—Reservoir-quality geostatistical prediction results. Mixed method integrates well-log, seismic, and geological-model data, which is a more reasonable model. Three targets are predicted. Targets I and II honored three drilled wells; Target III is proved by later drill practice (same legend as Fig. 8).

Q = reservoir-quality types
V = input vector
W = weight vector
x = x-coordinate location
$\mathbf{X}$ = independent factors
y = y-coordinate location
β = regression coefficients

Subscripts

i = number of groups
ref = reference

Acknowledgments

The Geophysics and Well Logging Co., Daqing Oil Field, CNPC, and U. of Petroleum, China, supported this work. Cao Guanyin performed the seismic residual thickness interpretation. Zhao Chenlin provided many useful suggestions about the geological modeling. Christopher White provided the original idea of the BBM. Feng Wang and Qiang Xu have provided many useful suggestions for this paper.

References

Armenta, M. et al. 2003. Completion length optimization in gas wells. Petroleum Society's Canadian Intl. Petroleum Conference, Calgary, 10–12 June.

Bhatt, A. and Helle, H.B. 2002. Determination of facies from well logs using modular neural networks. *Petroleum Geoscience* **8** (3): 217.

Chen, J. 2002. Volcanic lithofacies and volcanic structure analysis, Yingchen formation, Xujiaweizi rift depression, Songliao Basin. PhD dissertation, Geosciences U. of China.

Deutsch, C.V. and Journel A.G. 1998. *GSLIB Geostatistical Software Library and User Guide,* second edition. New York City: Oxford U. Press Inc.

Feng, Z. et al. 2003. Study on inclusion of deep buried igneous reservoir and forming period of gas pool in Song Liao basin. *Natural Gas Geoscience* **14** (6): 436.

Goovaerts, P. 1998. Ordinary cokriging revisited. *Mathematical Geology* **30** (1): 21. DOI: http://dx.doi.org/10.1023/A:1021757104135.

Goovaerts, P. 2002. Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data. *J. of Geographical Systems* **4** (1): 99. DOI: http://dx.doi.org/10.1007/s101090100077.

Guo, Z. 2002. Petroleum Geology, reflection on how to make a break-

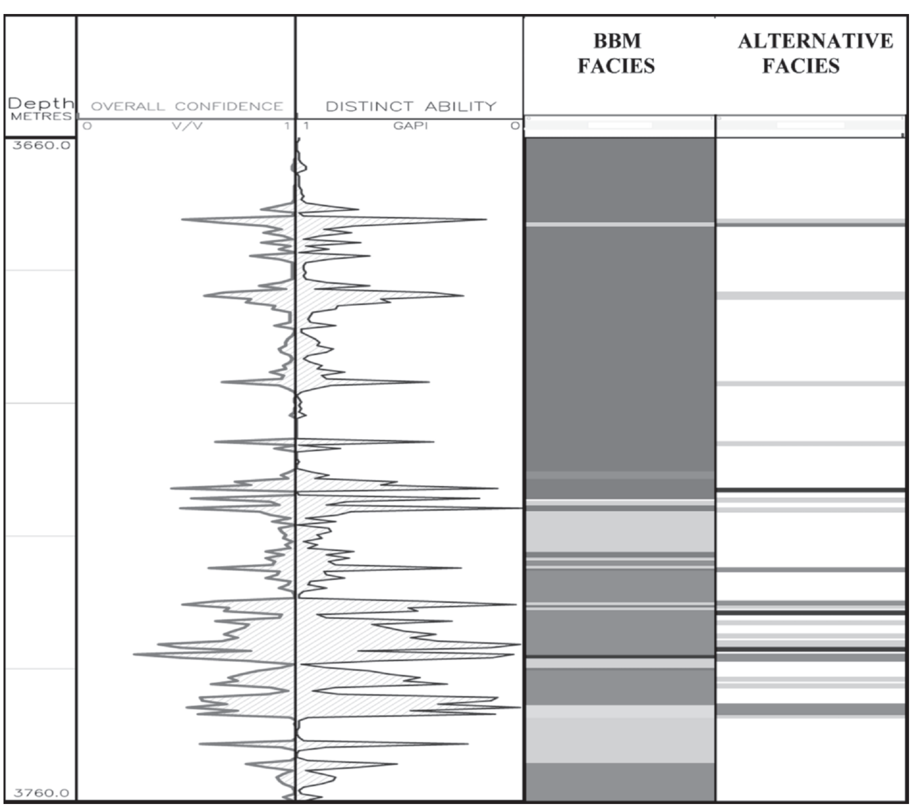

Fig. 11—Uncertainty evaluation for ZS6 volcanic BBM classification. The probability of the most probable facies represents the OC of the prediction. The probability difference of the two most probable facies is referred to as DIA. The enclosed area of the two parameters is an indicator of prediction uncertainty. The larger enclosed area means that the prediction is more uncertain. If the DIA value is less than 0.35, an alternative facies is shown for reference. The legend is the same as Fig. 6.

through in hydrocarbon exploration in north part of Songliao Basin. *Petroleum Geology and Development of Daqing* **17** (5): 2.

Karpur, L., Lake, L., and Sepehrnoori, K. 2000. Probability logs for facies classification. *In Situ* **24** (1): 57.

Neural Network Toolbox User's Guide for use with MATLAB. 1999. MathWorks Inc.

SAS/SAT User Guide, version 9. 2002. Cary, North Carolina: SAS Inst. Inc.

Shultz, J.R. and Fischbeck, P. 1999. Predicting Risks Associated With Offshore Production Facilities: Neural Network, Statistical, and Expert Opinion Models. Paper SPE 52677 presented at the SPE/EPA Exploration and Production Environmental Conference, Austin, Texas, 1–3 March. DOI: 10.2118/52667-MS.

Siripitayananon, P., Chen, H.C., and Hart, B.S. 2001. A new technique for lithofacies prediction: back-propagation neural network. *Proc.,* 39th Annual ACM-SE Conference, Athens, Georgia, 16–17 March.

Tang, H. et al. 2001. The application of artificial neural network technique in complex lithologic formation—An example of volcanic formation in Xinshan fault depression, Song Liao Basin. *Petroleum Science* **6** (3): 64.

Tang, H. et al. 2004. Comparison of Multivariate Statistical Algorithms for Wireline Log Facies Classification. Paper presented at the AAPG Annual Meeting, Dallas, 18–21 April.

Wisnie, A.P. and Zhu, Z. 1994. Quantifying Stuck Pipe Risk in Gulf of Mexico Oil and Gas Drilling. Paper SPE 28298 presented at the SPE Annual Technical Conference and Exhibition, New Orleans, 25–28 September. DOI: 10.2118/28298-MS.

Wong, P.M., Jian, F.X., and Taggart, I.J. 1995. A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and permeability predictions. *J. Petrol. Geol*. **18** (2): 191.

Yang, F. et al. 2003. Prediction of distribution of volcanic reservoir in Xinshan area of Song Liao basin through integrated seismic and geologic study. *Special Oil and Gas Reservoirs* **10** (1): 96.

Yarus, J.M. and Chambers, R.L. 1994. Stochastic Modeling and Geostatistics Principles, Methods, and Case Studies. *AAPG Computer Applications in Geology*, No. 3: 77.

Zhao, C. 1999. *Volcanic Reservoir and Oil/Gas in Liaohe Oil Field.* Beijing: Petroleum Industry Press.

SI Metric Conversion Factors

bbl × 1.589 873	E–01 = m^3
ft × 3.048*	E–01 = m
sq mile × 2.589 988*	E+00 = km^2

*Conversion factor is exact.

Hong Tang is a research scientist at Chevron Energy Technology Co. in Houston, where he works on reservoir modeling and simulation. e-mail: hongtang@chevron.com. He holds an MS degree in geoscience from the Petroleum U. of China and a PhD degree in petroleum engineering from Louisiana State U. **Hanchen Ji** is an associate professor at the Petroleum U. of China. e-mail: upchj@gmail.com. He holds MS and PhD degrees in sedimentology and diagenesis from the Petroleum U. of China.