Mathematical Geology, Vol. 34, No. 2, February 2002 (© 2002)

Modeling Conditional Distributions of Facies From Seismic Using Neural Nets[1]

Jef Caers[2] and Xianlin Ma[2]

We present a general, flexible, and fast neural network approach to the modeling of a conditional distribution of a discrete random variable, given a continuous or discrete random vector. Although many more applications of the neural net technique could be envisioned, the aim is to apply the developed methodology to the integration of seismic data into reservoir models. Many geostatistical methods for integrating seismic data rely on a screening assumption of further away seismic events by the colocated seismic datum. Such assumption makes the task of modeling cross-covariances and local conditional distributions much easier. In many cases, however, the seismic data exhibit distinct and locally varying spatial patterns of continuity related to geological events such as channels, shale bodies, or fractures. The previous screening assumption prevents recognizing and hence utilizing these patterns of seismic data. In this paper we propose to relate seismic data to facies or petrophysical properties through a colocated window *of seismic information instead of the single colocated seismic datum. The variation of seismic data from one window to another is accounted for. Several examples demonstrate that using such a window improves the predictive power of seismic data.*

KEY WORDS: seismic inversion, neural networks, pattern recognition.

INTRODUCTION

Imaging of the subsurface and 3D modeling of reservoir properties represent a major challenge, given the sparsity of hard information. Wells provide quality information on the vertical variation, but do not provide much insight on how that vertical variation varies laterally. The most prevalent source of lateral variation is seismic data.

In geostatistics seismic data is typically viewed as a soft, indirect, or secondary information. To make proper use of such information, many alternative methods

[1]Received 10 July 2000; accepted 12 January 2001.

[2]Department of Petroleum Engineering, Stanford University, Stanford, California 94305-2220; e-mail: jef@pangea.stanford.edu

have been proposed:

- Use seismic data as a locally varying mean (Deutsch and Journel, 1998) informing the average porosity at any colocated location or as a prior probability for a certain facies to occur.
- A Bayesian approach where one models the likelihood of seismic data given hard facies or porosity data. The likelihood is used in either sequential simulation (Doyen and others, 1997) or Markov chain Monte Carlo simulation (Eide, Omre, and Ursin, 1996).
- Interpret a 2D seismic map (that has no vertical resolution) as secondary information about the vertically averaged porosity and perform colocated cokriging. The resulting estimated averaged porosity is then used to perform 3D stochastic simulation (Behrens, Macleod, and Tran, 1996; Journel, 1999; Xu, Tran, and Journel, 1992; Yao, in press).
- When a full 3D seismic cube is available, a full cokriging of soft and hard data could be considered. But such cokriging requires a full coregionalization model which typically calls for additional assumptions to alleviate the cross-covariance modeling effort:
  - **MM1** (Almeida and Journel, 1993): In the Markov Model 1, one assumes that the colocated primary data screens the influence of further away primary data on the central secondary data. This assumption is difficult to hold when the secondary variable is on a larger volume support than the primary, which is the case for seismic data.
  - **MM2** (Journel, 1999): In the Markov Model 2, it is assumed that the colocated secondary data screens the influence of further away secondary information on the primary central data. This assumption is more appropriate in presence of a secondary data with large volume support.

The increase in seismic resolution and accuracy of modern seismic acquisition and processing tools call for a much improved utilization of the seismic data in geostatistical modeling, beyond its use as mere colocated seismic information. Although the results obtained from using the above methods can be significantly different, the common approximation is the screening of further away seismic data by the colocated seismic datum; also the seismic attribute is treated as a smooth (non)-linear average representation of either petrophysical variable or lithofacies.

Looking at seismic images or cubes, one can often observe distinct local spatial patterns. These patterns are often related to certain depositional features such as faults, channels, or large shale bodies. In the above methods, such seismic spatial patterns are largely neglected when only the colocated seismic datum is retained. It therefore seems important to consider a whole colocated window of seismic data, as was originally proposed in Pairazian and Scheevel (1999). A window $W(\mathbf{u})$ is a series of neighbor locations including the central location $\mathbf{u}$. Such a window contains information about the local spatial pattern of

continuity of seismic which might be related to the lithofacies or petrophysical property at **u**.

In this paper, we propose to model the conditional distribution of a discrete variable given a random continuous or discrete vector. The developed methodology is general and can be applied to any field of the Earth Sciences where such modeling is convenient or needed. In this paper we apply the neural net methods to predict a facies indicator variable or a discretized petrophysical variable at location **u** given the seismic data observed, within the window $W(\mathbf{u})$. Such conditional distribution is the key ingredient of sequential simulation algorithms such as a p-field (Srivastava, 1992) or indicator simulation (Deutsch and Journel, 1998). Although the application of neural nets to the modeling of seismic properties is not new (e.g. Wong, Jian, and Taggart, 1995), most of the traditional neural net applications are limited to regression of lithofacies from seismic. We develop a neural network methodology for automatic modeling of *conditional distributions* of lithofacies.

METHODOLOGY

An exhaustive set of 2D or 3D seismic data is available, possibly consisting of multiple seismic attributes. Seismic is considered as a secondary variable $Z_2(\mathbf{u})$ informing either a continuous primary variable $Z_1(\mathbf{u})$ (petrophysical property) or a facies indicator variable $I(\mathbf{u}, s_k), k = 1, \ldots, K$ (facies). Many geostatistical estimation or simulation methods for secondary data integration call for modeling the prior conditional probabilities

$$\Pr\{Z_1(\mathbf{u}) \leq z_1 \mid z_2(\mathbf{u})\} \quad \text{or} \quad \Pr\{I(\mathbf{u}, s_k) = 1 \mid z_2(\mathbf{u})\},\ k = 1, \ldots, K \qquad (1)$$

Such modeling is based on the scatterplot between primary hard data $z_1(\mathbf{u}_\alpha), \alpha = 1, \ldots, n$, and the colocated secondary data $z_2(\mathbf{u}_\alpha), \alpha = 1, \ldots, n$, at well locations. Stochastic simulation algorithms such as p-field simulation and sequential indicator simulation make use of these conditional probabilities.

We propose to model the relationship between the hard variable $Z_1(\mathbf{u})$ at location **u** and a colocated *window* of seismic or secondary data. Such window is defined through the geometry of a template, that is the set of n_t locations

$$\mathbf{u} + \mathbf{h}_\beta, \quad \beta = 1, \ldots, n_t,\ \mathbf{h}_\beta \neq 0$$

At each hard datum location $\mathbf{u}_\alpha$, we therefore collect the following hard and soft data vector

$$\left\{z_1(\mathbf{u}_\alpha); z_2(\mathbf{u}_\alpha), z_2(\mathbf{u}_\alpha + \mathbf{h}_1), \ldots, z_2\left(\mathbf{u}_\alpha + \mathbf{h}_{n_t}\right)\right\}, \quad \alpha = 1, \ldots, n$$

Note that the central colocated seismic value $z_2(\mathbf{u})$ is included. Given this training database, one needs to model the following conditional distribution

$$\Pr\{Z_1(\mathbf{u}) \leq z_1 \mid z_2(\mathbf{u}), z_2(\mathbf{u}_\alpha + \mathbf{h}_1), \ldots, z_2(\mathbf{u}_\alpha + \mathbf{h}_{n_t})\} \tag{2}$$

or

$$\Pr\{I(\mathbf{u}, s_k) = 1 \mid z_2(\mathbf{u}), z_2(\mathbf{u}_\alpha + \mathbf{h}_1), \ldots, z_2(\mathbf{u}_\alpha + \mathbf{h}_{n_t})\}, \quad k = 1, \ldots, K \tag{3}$$

These more fully conditioned probabilities can be used in exactly the same fashion as the probabilities (1). A neural network methodology is developed for modeling the conditional distributions (2) and (3).

NEURAL NETWORKS

Discrete Case, Least-Square Method

Consider first the case of modeling the discrete distribution (3).

In general, neural networks are most commonly used for building nonlinear regression models between a set of random variables $\mathbf{X} = \{X_1, \ldots, X_N\}$ (input variables) and a set of (target) variables $\mathbf{Y} = \{Y_1, \ldots, Y_M\}$. The vector $\mathbf{X}$ is input into the neural network to produce an output $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}, \boldsymbol{\theta})$$

$\mathbf{g} = (g_1, \ldots, g_m)$ represents a multivariate nonlinear function which depends on the architecture of the network, as defined by the various connections between the network nodes. The simplest neural network is a feed-forward network that connects the input to a middle layer and then to the output layer (see Fig. 1(A)). The middle layer is called the *hidden layer*. The following "regression" function g was adopted:

$$g_m(\mathbf{x}, \boldsymbol{\theta}) = \sum_{l=1}^{L} o_l T\left(\sum_{n=1}^{N} w_{l,n} x_n\right), \quad m = 1, \; n = 1, \ldots, N \tag{4}$$

where o_l are L output weights and $w_{l,n}$ are $L \times N$ input weights linking the input to the hidden layer, and L is the number of nodes of the hidden layer. T is a nonlinear nondecreasing transfer function. We take for T the following expression

$$T(a) = \frac{1}{1 + \exp(-a)}$$

The parameters o_l and $w_{l,n}$ are represented by a single vector $\boldsymbol{\theta}$, the parameters of the neural network.

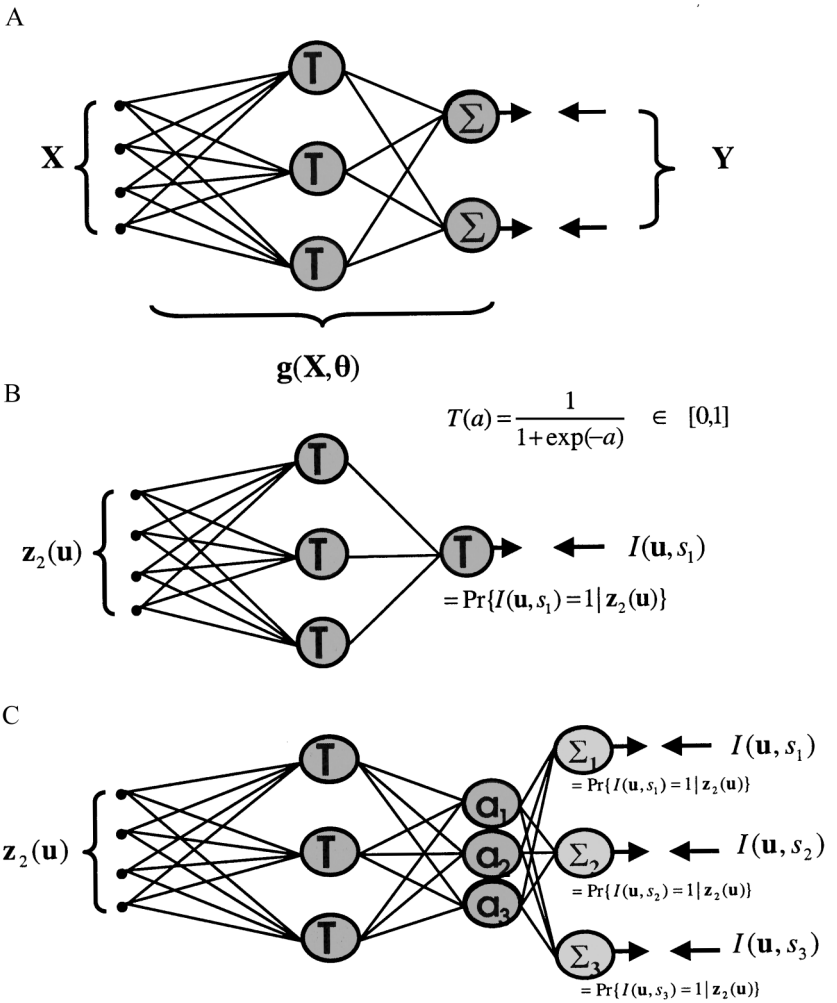

Figure 1. Neural nets. (A) Least-squares case, (B) likelihood case, two facies, and (C) likelihood case, multiple facies.

The training of the neural networks consists of fitting the parameter vector $\boldsymbol{\theta}$ to the training data. Training data consists of observed pairs of the vectors $\{\mathbf{x}_i, \mathbf{y}_i\}, i = 1, \ldots, n$. The parameters $\boldsymbol{\theta}$ are determined by minimizing a least-square criterion

$$E(\boldsymbol{\theta}) = \frac{1}{2n} \sum_{i=1}^{n} \| \mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}) - \mathbf{y}_i \|^2$$

The parameter values at the minimum of $E(\boldsymbol{\theta})$ are denoted as $\hat{\boldsymbol{\theta}}$. To determine the number of hidden layer nodes and in order to avoid overfitting of the training data, one splits the training set in two parts: one for training, and the other for testing and validation purposes (Bishop, 1995).

For a single output neural network, it can be shown (see Bishop, 1995, p. 201) that the neural network output can be interpreted as a conditional expectation of the variable Y, given the input vector $\mathbf{X}$

$$g(\mathbf{x}, \boldsymbol{\theta}) \cong \mathrm{E}[Y \mid \mathbf{X} = \mathbf{x}]$$

This interpretation calls for the following conditions:

- A large data set is available, i.e. n is large.
- The neural network accurately models the training data set.
- The neural network has good generalization capacity, in that it will predict accurately an output y from any given input vector $\mathbf{x}$ not in the training data set.

In this paper, however, we are interested in modeling conditional probabilities not conditional expectations. For the discrete case, with two facies categories s_1, s_2, the training data set consists of a set of n data vectors of the type:

$$\left\{i(\mathbf{u}_\alpha, s_1); z_2(\mathbf{u}_\alpha), z_2(\mathbf{u}_\alpha + \mathbf{h}_1), \ldots, z_2\left(\mathbf{u}_\alpha + \mathbf{h}_{n_t}\right)\right\}, \quad \alpha = 1, \ldots, n \qquad (5)$$

Next, consider the following vector of soft data

$$\mathbf{z}_2(\mathbf{u}) = \left\{z_2(\mathbf{u}), z_2(\mathbf{u} + \mathbf{h}_1), \ldots, z_2\left(\mathbf{u} + \mathbf{h}_{n_t}\right)\right\}$$

For the two facies case, there is a single binary indicator output (Fig. 1(B)). The target output of the neural network is the indicator variable $I(\mathbf{u}, s_1)$. Since the output is interpreted as a conditional expectation and since

$$\mathrm{E}\{I(\mathbf{u}, s_1) \mid \mathbf{z}_2(\mathbf{u})\} = \Pr\{I(\mathbf{u}, s_1) = 1 \mid \mathbf{z}_2(\mathbf{u})\}$$

the output of the neural network, written as $g(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta})$ with input vector $\mathbf{z}_2(\mathbf{u})$, can be interpreted as a conditional probability:

$$\Pr\{I(\mathbf{u}, s_1) = 1 \mid \mathbf{z}_2(\mathbf{u})\} = g(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta})$$

However, with least-squares estimation methods, there is no guarantee that the probability $\Pr\{I(\mathbf{u}, s_1) = 1 \mid \mathbf{z}_2(\mathbf{u})\}$ is in the interval [0, 1]. Least-square methods

are known to work well when the target output variable Y valued in $[-\infty, \infty]$ is a Gaussian variable. In this particular case the variable is discrete and thus far from being Gaussian; hence, a different minimization criterion should be considered to avoid such order relation.

Discrete Case, Likelihood Method

Consider again the case of a discrete variable with only two categories. We now use a likelihood principle for determining the parameters $\boldsymbol{\theta}$ of the neural network. Recall that in maximum likelihood estimation, as opposed to least-squares estimation, one maximizes the probability of observing a data set, *given* a probability model for that data set. In our case, the data set consists of the information jointly scanned from the wells and seismic, namely the training data set (5). We would like the output of the neural network $g(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta})$ to represent the conditional probability $\Pr\{I(\mathbf{u}, s_2) = 1 \mid \mathbf{z}_2(\mathbf{u})\}$. The probability of observing the *single* data vector $\{i(\mathbf{u}_\alpha, s_1); \mathbf{z}_2(\mathbf{u}_\alpha)\}$ follows a Bernoulli distribution with

$$p(i(\mathbf{u}_\alpha, s_1) \mid \mathbf{z}_2(\mathbf{u}_\alpha)) = g(\mathbf{z}_2(\mathbf{u}_\alpha), \boldsymbol{\theta})^{i(\mathbf{u}_\alpha, s_1)}[1 - g(\mathbf{z}_2(\mathbf{u}_\alpha), \boldsymbol{\theta})]^{1-i(\mathbf{u}_\alpha, s_1)}$$

The likehood $L(\boldsymbol{\theta})$ of observing *all* the data (5), assuming that the data are drawn *independently* from the Bernoulli distribution, is then given by

$$L(\boldsymbol{\theta}) = \prod_{\alpha=1}^{n} g(\mathbf{z}_2(\mathbf{u}_\alpha), \boldsymbol{\theta})^{i(\mathbf{u}_\alpha, s_1)}[1 - g(\mathbf{z}_2(\mathbf{u}_\alpha), \boldsymbol{\theta})]^{1-i(\mathbf{u}_\alpha, s_1)} \tag{6}$$

The assumption of independence is clearly not a correct one and its impact on the final estimated parameters will have to be evaluated. This likelihood function needs to be maximized in the parameters $\boldsymbol{\theta}$ or, equivalently, the negative log-likelihood is minimized

$$\begin{aligned} E = -\log L(\boldsymbol{\theta}) = -\sum_{\alpha=1}^{n} & i(\mathbf{u}_\alpha, s_1) \log g(\mathbf{z}_2(\mathbf{u}_\alpha), \boldsymbol{\theta}) \\ & + (1 - i(\mathbf{u}_\alpha, s_1)) \log(1 - g(\mathbf{z}_2(\mathbf{u}_\alpha), \boldsymbol{\theta})) \end{aligned} \tag{7}$$

In the least-squares method, the neural network output could be interpreted as a conditional probability due to the use of a least-square criterion to find the optimal parameters set $\boldsymbol{\theta}$. Since the training criterion is now a maximum likelihood criterion, the neural network output cannot be directly interpreted as a conditional probability. This problem is resolved by changing the linear output node of the

neural network in Figure 1(A) to be a logistic function: we therefore propose the following neural network architecture, see Figure 1(B):

$$g(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}) = T\left(\sum_{l=1}^{L} o_l T\left(\sum_{\beta=1}^{n_t} w_{l,\beta} z_2(\mathbf{u}+\mathbf{h}_\beta)\right)\right) \tag{8}$$

This mode of training network is also termed *logistic regression* in the statistical literature (Christensen, 1997). Note that the function F is bounded by the interval [0, 1]; hence, the conditional probability modelled this way is always permissible.

Discrete Case, Multiple Categories

Consider now the case of a discrete variable with multiple categories, i.e. more than two. We extend the neural network (8) to have multiple outputs, each output representing a single category s_k.

$$g_k(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}) = T\left(\sum_{l=1}^{L} o_{k,l}\, T\left(\sum_{\beta=1}^{n_t} w_{l,\beta} z_2(\mathbf{u}+\mathbf{h}_\beta)\right)\right), \quad k = 1, \ldots, K$$

and extend the likelihood function (6) to multiple categories as follows:

$$L(\boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{\alpha=1}^{n} g_k(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta})^{i(\mathbf{u}_\alpha, s_k)}$$

The challenge remains now in estimating the parameters of the neural network by maximizing this likelihood function such that all probabilities are permissible, that is:

$$0 \le g_k(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}) \le 1 \;\forall k \quad \text{and} \quad \sum_{k=1}^{K} g_k(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}) = 1$$

The first conditions will hold because the function F is properly bounded; however, the last condition is not guaranteed. Therefore, we generalize the logistic function F to make the outputs sum to unity by performing the following transformation on the neural network output activations

$$g_k(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}) = \frac{\exp(a_k(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}))}{\sum_{k'=1}^{K} \exp(a_{k'}(\mathbf{z}_2(\mathbf{u}), \boldsymbol{\theta}))}$$

where $a_k(\cdot, \theta)$ are the so-called output activations given by the expression

$$a_k(\mathbf{z}_2(\mathbf{u}), \theta) = \sum_{l=1}^{L} o_{k,l} T \left(\sum_{\beta=1}^{n_t} w_{l,\beta} z_2(\mathbf{u} + \mathbf{h}_\beta) \right)$$

The latter generalization of the logistic function is also known as the softmax condition in the neural network literature (Bridle, 1990). The softmax condition ensures that the multiple neural network outputs are conditional probabilities.

Continuous Case

When modeling continuous variables one can rely on a discretization of the variable into multiple classes, and then using the above proposed methods for modeling the discrete conditional probabilities. Alternatively, one can rely on a mixture of densities network as proposed in Caers and Journel (1998).

PRINCIPAL COMPONENT ANALYSIS

The proposed method models the conditional distribution of a single facies value at $\mathbf{u}$ given a colocated window of n_t seismic data. The number n_t depends on the size of the particular window retained, and may be large on 3D grids.

To reduce the dimensionality n_t a principal component analysis is often proposed to preprocess seismic data (Fournier and Derain, 1995; Pairazian and Scheevel, 1999). Principal component analysis is a dimension-reduction method that builds upon the redundancy present in most multicomponent data sets. The scanning of the seismic data produce multiple realizations (one per window) of a random vector $\mathbf{Z}_2(\mathbf{u})$. The variables $Z_2(\mathbf{u} + \mathbf{h}_\beta)$ that make up this random vector are not independent of each other, hence inducing considerable redundancy in the data set $\{\mathbf{Z}_2(\mathbf{u}_\alpha), \alpha = 1, \ldots, n\}$. To remove that redundancy and henceforth reduce the dimension of $\mathbf{Z}_2(\mathbf{u})$ a principal component analysis of the random vector $\mathbf{Z}_2(\mathbf{u})$ is performed. Because of the reduction of dimension, the resulting vector will have less variance than the original random vector $\mathbf{Z}_2(\mathbf{u})$. The percentage of variance reproduction, depending on the number of retained components d, is usually taken to be above 85%.

It is often useful to substract from each sampled vector $\mathbf{z}_2(\mathbf{u}_\alpha)$ its local mean $m(\mathbf{u})$ as defined from a local window; this allows focusing on the local variability of the seismic z_2-data.

EXAMPLES

Several examples are now given to illustrate the methodology. The examples are synthetic and intend to provide an understanding of why the window approach should be preferred over the traditional single value colocated approach.

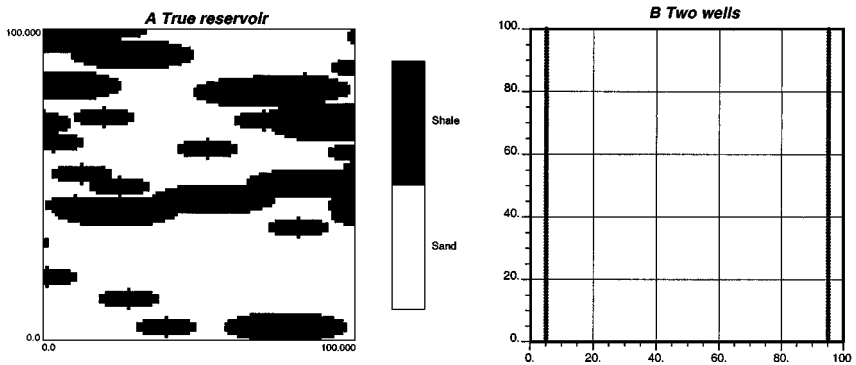

Figure 2. Shale reservoir. (A) Actual reservoir and (B) two well locations.

Shale Reservoir

Using the GSLIB program ellipsim (Deutsch and Journel, 1998), a vertical section of a shale deposit is simulated (Fig. 2(A)). The overall shale proportion is 36%. Two vertical wells are drilled through that section at the locations shown in Figure 2(B). On the basis of this exhaustive data set a seismic amplitude data set in constructed as follows: seismic waves reflect in the subsurface because of changes in the rock impedance values Imp(**u**). For this model, we consider a constant shale impedance of 7000 (gr/cm^3) (m/s) and a constant background impedance of 5000 (gr/cm^3) (m/s). From the rock-impedance values one can calculate the reflection coefficients

$$C(\mathbf{u}) = \frac{\text{Imp}(x, d+1) - \text{Imp}(x, d)}{\text{Imp}(x, d+1) + \text{Imp}(x, d)}$$

where x is the horizontal coordinate, d is the depth coordinate, and $\mathbf{u} = (x, d)$.

The recorded seismic amplitude $A(x, d)$ is modeled as a convolution of a vertical series of reflection coefficients and a seismic wavelet $w(d, \psi)$

$$A(x, d) = \sum_{j=-J}^{J} c(x, d+j) w(j, \psi)$$

A Ricker wavelet is considered (Yilmaz, 1987):

$$w(d, \psi) = (1 - 2\pi^2 d^2 \psi^2)\ \exp(-\pi^2 d^2 \psi^2)$$

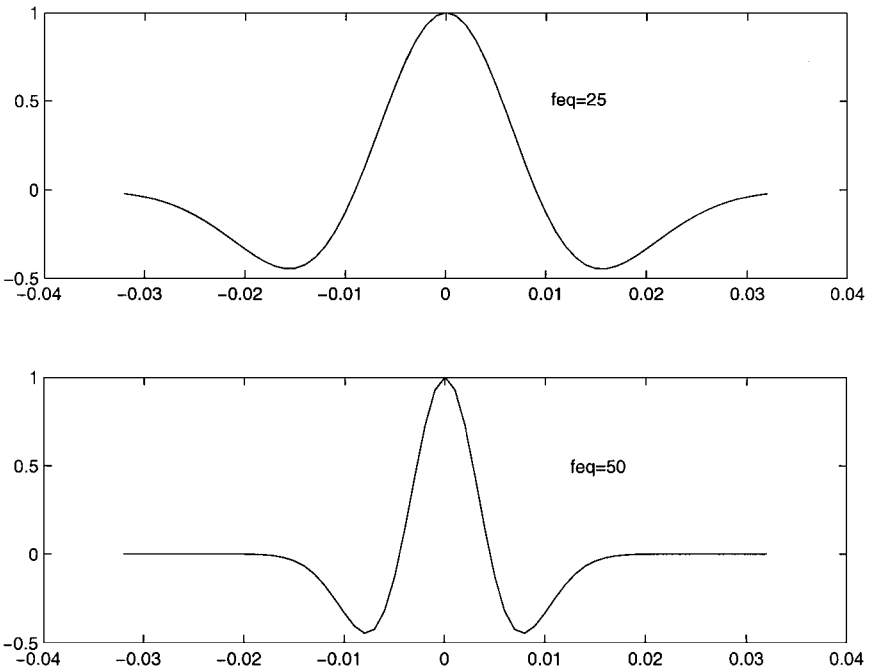

Figure 3. Wavelets with different peak frequencies (top) $\psi = 25$, (bottom) $\psi = 50$.

The coefficient ψ defines the peak frequency content of the wavelet: the higher the ψ, the higher the frequency content, and the more narrow the wavelet. We use the two wavelets shown in Figure 3 (top, bottom) with $\psi = 25$ and $\psi = 50$ to construct two seismic data sets. Figure 4 shows the two resulting seismic amplitude data sets.

A general problem with geophysical data processing is the deconvolution operation, i.e. starting from a seismic amplitude image $A(x, d)$ the task is to find the seismic impedance Imp(**u**). To perform such an operation, one needs to determine the wavelet w, which is not an easy task if only a few wells are available, the seismic data is noisy, or if the wavelet is location-dependent. In the following, no knowledge of the wavelet is assumed.

Consider first the case of constructing a seismic amplitude data set with $\psi = 25$. The histograms of seismic amplitude per facies is shown in Figure 5(A) and (B). Given only the per facies seismic amplitude information, it would be impossible to distinguish shale from sand since the two histograms overlap. To prove this conjecture, a neural network with 10 nodes in the hidden layer is used to model the conditional probability $\Pr\{I(\mathbf{u}, s_1) = 1 \mid A(\mathbf{u})\}$ of sand given the single colocated amplitude datum.

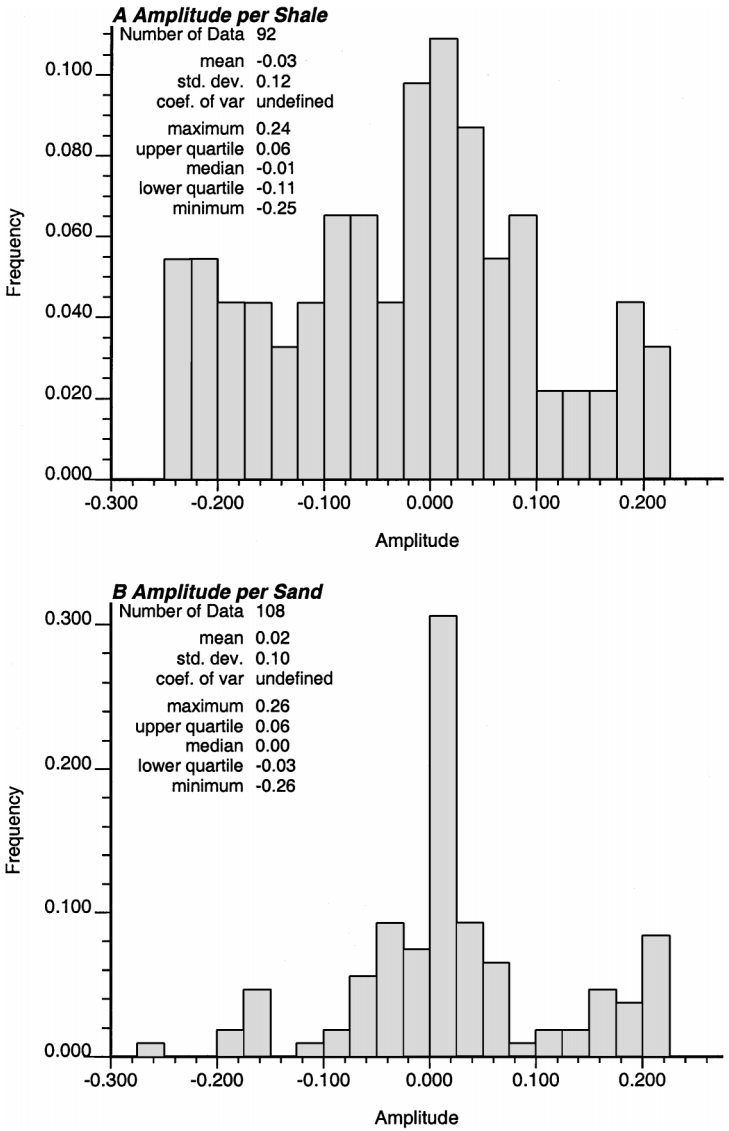

Figure 4. Amplitude (A) for shale and (B) for sand at the two wells for frequency $\psi = 50$.

To evaluate the resulting conditional probability, a so-called Bayesian confusion matrix is constructed as follows: given a seismic datum at an unsampled location $\mathbf{u}$, the neural-net–determined probability $\Pr\{I(\mathbf{u}, s_k) = 1 \mid z_2(\mathbf{u})\}$ is used to perform classification, i.e. location $\mathbf{u}$ is classified into

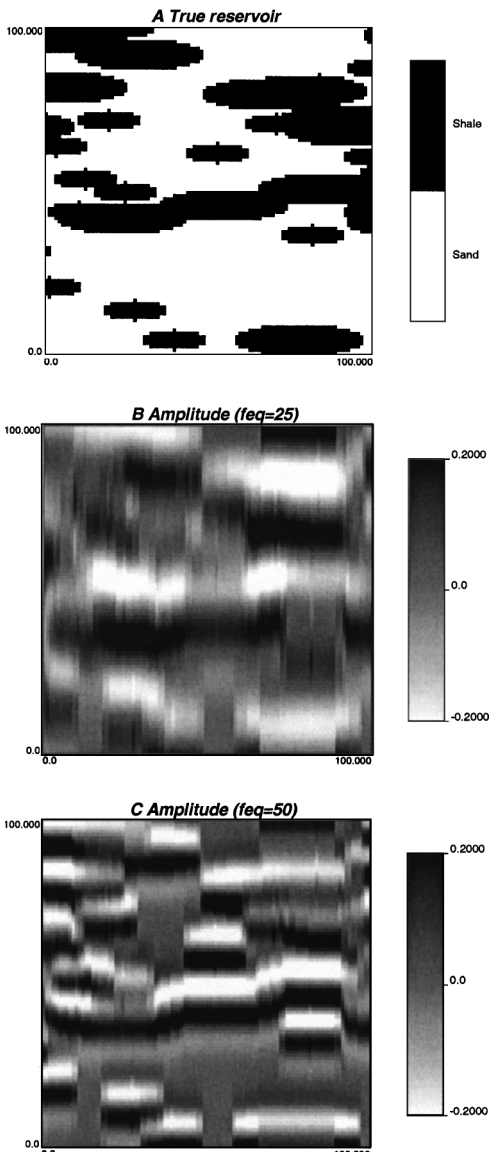

Figure 5. Seismic obtained from convolution case study. (A) Actual reservoir, (B) amplitude for freq. $\psi = 25$, and (C) amplitude for freq. $\psi = 50$.

Table 1. Bayesian Confusion Matrix (Amplitude Data Set, Freq. $\psi = 25$; Single Colocated Datum)

	Shale	Sand
Shale	17	83
Sand	28	72

class s_k if

$$\Pr\{I(\mathbf{u}, s_k) = 1 \mid z_2(\mathbf{u})\} > \Pr\{I(\mathbf{u}, s_j) = 1 \mid z_2(\mathbf{u})\} \quad \forall j \neq k \tag{9}$$

The Bayesian confusion matrix collects the results of this classification: the entry in the kth row and lth column reports the percentage of classification into category s_k when the facies at location $\mathbf{u}$ in the true reservoir is actually of category s_l. A classification with good performance should have large entries on the diagonal and small off-diagonal entries. The resulting Bayesian confusion matrix, given in Table 1, shows a reasonably correct classification of sand at 72%, but the shale bodies are severely misclassified (only 17% correct classification). This misclassification is clearly depicted in Figure 6(D).

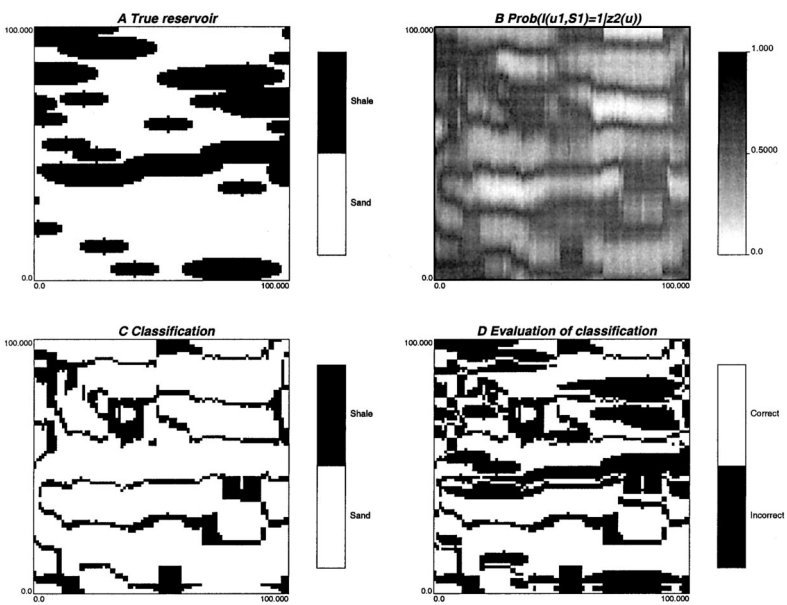

Figure 6. Amplitude data set for freq. $\psi = 25$. (A) Actual reservoir, (B) probability of shale using the single colocated seismic datum, (C) Bayesian classification results, and (D) correctly versus incorrectly classified pixels.

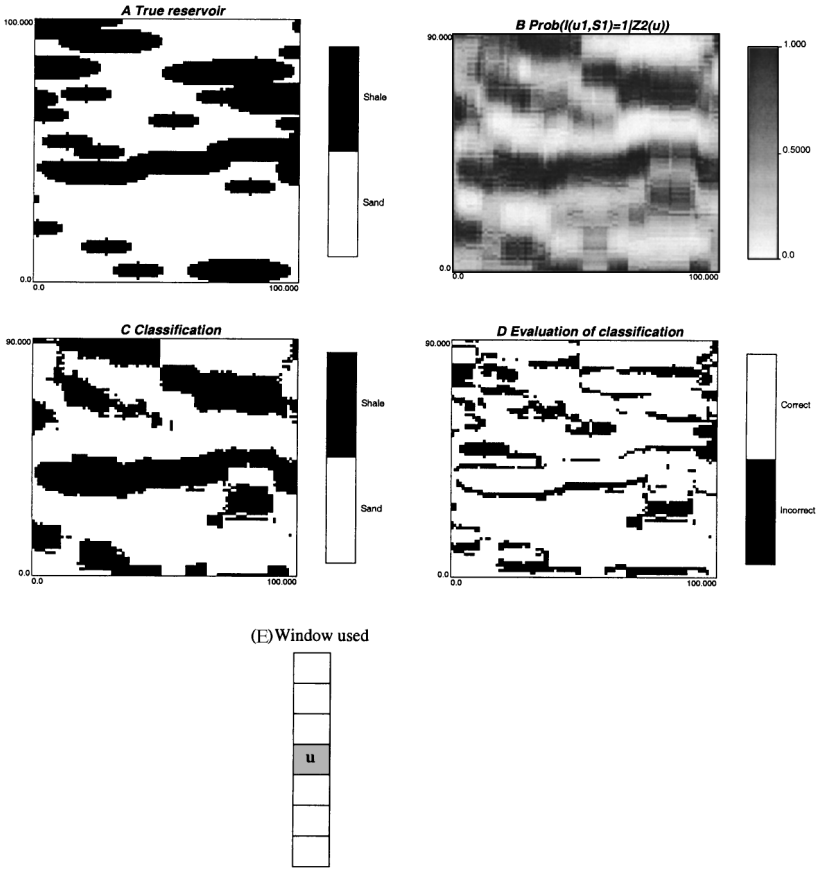

Figure 7. Amplitude data set for freq. $\psi = 25$. (A) Actual reservoir, (B) probability of shale using a window of seismic data, (C) Bayesian classification results, and (D) correctly versus incorrectly classified pixels.

Next, we model the conditional probability at each location using the colocated window approach. A neural network with 10 nodes in one hidden layer is used to model the conditional probability $\Pr\{I(\mathbf{u}, s_1) = 1 \mid \mathbf{z}_2(\mathbf{u})\}$. A window with geometry shown in Figure 7(E) is used. Figure 7(B) shows the resulting probability map. Under Bayes rule (9), the Bayesian confusion matrix is shown in Table 2. The result of the classification is given in Figure 7(C). Compared to the actual reservoir, the large shale bodies are mostly correctly classified, and the proportion of correctly classified sand is also high. This is a strong result since the histograms of seismic amplitude per facies gave almost no discrimination between facies.

A similar study was conducted with amplitude defined at the peak frequency $\psi = 50$. A larger value for ψ entails a wavelet with higher-frequency content,

Table 2. Bayesian Confusion Matrix (Amplitude Data Set, Freq. $\psi = 25$: Window Approach)

	Shale	Sand
Shale	72	28
Sand	19	81

hence, the resulting seismic amplitude data set has more resolution, as shown in Figure 5(C). First the traditional colocated approach is used to model the conditional probability $\Pr\{I(\mathbf{u}, s_1) = 1 \mid A(\mathbf{u})\}$. The resulting Bayesian confusion matrix, given in Table 3, shows a correct classification rate of sand at 87%, but the shale bodies remain largely misclassified, with only 28% correct classification. This misclassification is clearly shown in Figure 8(D).

Next the window approach with the window geometry shown in Figure 9(E) is used. The window size is taken smaller than for the frequency $\psi = 25$, because the wavelet dispersion in Figure 3 is smaller. A neural network with 10 nodes in one hidden layer is used to model the conditional probability $\Pr\{I(\mathbf{u}, s_1) = 1 \mid \mathbf{z}_2(\mathbf{u})\}$. Figure 9(B) shows the probabilities map. The result of the classification is given in Figure 9(C). Figure 9(D) gives the evaluation of that classification. The corresponding Bayesian confusion matrix is shown in Table 4. Compared to Table 2, the sand classification is near perfect with only 4% misclassified. Figure 9(C) shows that even small shale bodies are classified correctly: a result due to the higher-frequency content of the wavelet, allowing to pick up small-scale variabilities in amplitude.

Deep Fluvial Reservoir

Consider next a deep fluvial reservoir such as present in the North Sea. A prototype for such reservoir is the Stanford V reservoir (Mao and Journel, 1999; available at http://ekofisk.stanford.edu/SCRFweb.html). The Stanford V synthetic reservoir depicts a channel formation with known lithofacies, various petrophysical properties as well as forward simulated seismic data, in this case a low resolution seismic impedance. To obtain the latter a frequency-domain Born filter was applied to the true rock impedance values (Mavko, Mukerji, and Dvorkin, 1998).

Table 3. Bayesian Confusion Matrix (Amplitude Data Set, Freq. $\psi = 50$, Single Colocated Datum)

	Shale	Sand
Shale	28	72
Sand	13	87

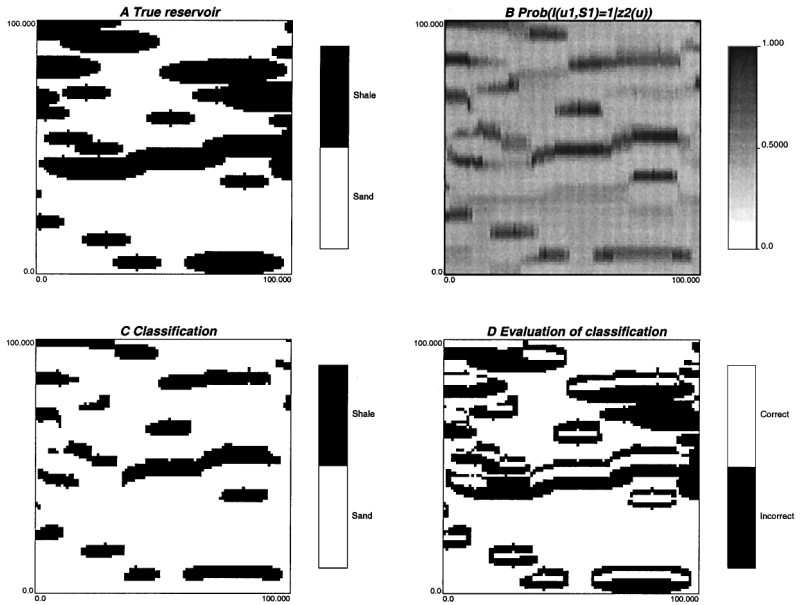

Figure 8. Amplitude data set for freq. $\psi = 50$. (A) Actual reservoir, (B) probability of shale using the single colocated seismic datum, (C) Bayesian classification results, and (D) correctly versus incorrectly classified pixels.

The five top layers are retained (out of 30) to perform this exercise. The grid size is $100 \times 120 \times 5$. The total sand proportion in this portion of the reservoir equals 56%. Figure 10(A) show a single horizontal slice of the reservoir together with the true rock impedance in Figure 10(B) and the measured seismic impedance in Figure 10(C). The true impedance reflects clearly the channels, while the seismic impedance provides a much blurred image of the channel structure. Thirty vertical wells are extracted from the true reservoir as shown in Figure 10(D). The well sample sand proportion is 60%, slightly greater than the reference population proportion of 56%.

Figure 11(A) and (B) show the histograms of seismic impedance within channel and within mud. The mean seismic impedance in channel is 8059 (gr/cm^3) (m/s) with a standard deviation of 495 (gr/cm^3) (m/s), which is close to the

Table 4. Bayesian Confusion Matrix (Amplitude Data Set, Freq. $\psi = 50$, Window Approach)

	Shale	Sand
Shale	69	31
Sand	4	96

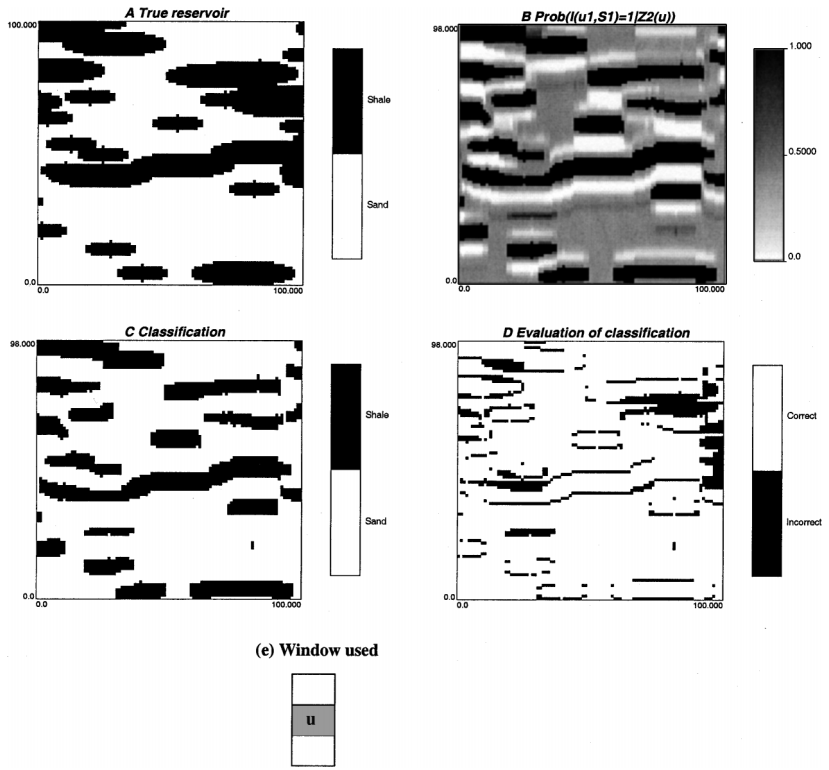

Figure 9. Amplitude data set for freq. $\psi = 50$. (A) Actual reservoir, (B) probability of shale given a window of seismic data, (C) Bayesian classification results, and (D) correctly versus incorrectly classified pixels.

mean mud seismic impedance of 8489 (gr/cm^3) (m/s) with a standard deviation of 592 (gr/cm^3) (m/s). Hence the single colocated seismic impedance is not a good discriminator of channel versus mud.

To quantify the discriminatory power of the colocated seismic impedance, we develop an approach different from that used for the previous case studies. This is merely to illustrate the existence of alternatives to using neural networks for modeling conditional distributions. The conditional distribution $\Pr\{I(\mathbf{u}, s_1) = 1 \mid z_2(\mathbf{u})\}$ is modelled as follows: the histograms of impedance per facies are fairly symmetric and short-tailed, as shown in Figure 11(A) and (B), and hence they could be modeled using a normal distribution. Denote by $\mathcal{N}(m, \sigma)$ a normal distribution with mean m and standard deviation σ. The distribution (likelihood) of seismic impedance related to sand facies s_1 is modelled as

$$f_1(z_2 \mid i(\mathbf{u}; s_1) = 1) = \mathcal{N}(8059, 495) \tag{10}$$

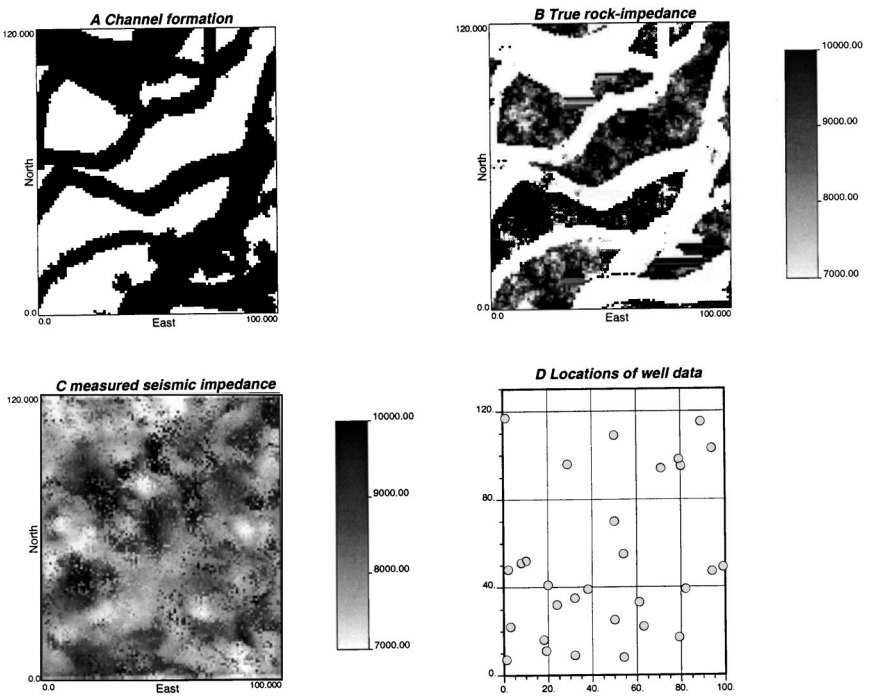

Figure 10. Fluvial channel reservoir case study. (A) Actual reservoir, (B) rock impedance, (C) seismic impedance, and (D) well locations.

and for the mud facies

$$f_2(z_2 \mid i(\mathbf{u}; s_1) = 0) = \mathcal{N}(8489, 592) \tag{11}$$

The conditional distribution of sand *given* seismic impedance can then be calculated using Bayes' relation as follows

$$\Pr\{I(\mathbf{u}, s_1) = 1 \mid z_2(\mathbf{u})\} = \frac{f_1(z_2(\mathbf{u}) \mid i(\mathbf{u}; s_1) = 1) \Pr\{I(\mathbf{u}, s_1) = 1\}}{f(z_2)}$$

and for the mud facies

$$\Pr\{I(\mathbf{u}, s_1) = 0 \mid z_2(\mathbf{u})\} = \frac{f_2(z_2 \mid i(\mathbf{u}; s_1) = 0) \Pr\{I(\mathbf{u}, s_1) = 0\}}{f(z_2)}$$

To avoid having to model the mud marginal distribution $f(z_2)$, we consider the ratio of these two probabilities

$$r(\mathbf{u}) = \frac{\Pr\{I(\mathbf{u}, s_1) = 1 \mid z_2(\mathbf{u})\}}{\Pr\{I(\mathbf{u}, s_1) = 0 \mid z_2(\mathbf{u})\}}$$

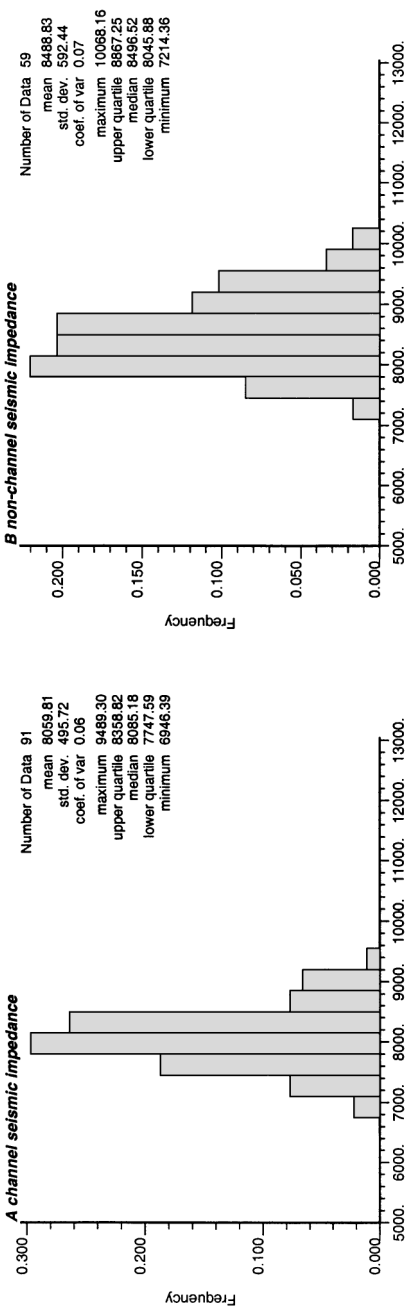

Figure 11. Histograms of impedance values in channel versus nonchannel, as obtained from 30 wells.

Table 5. Bayesian Confusion Matrix (Fluvial Reservoir, Single Colocated Datum)

	Shale	Sand
Shale	60	40
Sand	81	19

The ratio $r(\mathbf{u})$ provides a measure of the power of discriminating sand versus nonsand based on the colocated seismic impedance value. If $r(\mathbf{u})$ is close to one, then that discriminatory power is low. Using the well data the analytical expression for the ratio $r(\mathbf{u})$ is calculated based on the normal distribution assumptions (10), and then applied to each seismic datum to obtain a 3D volume of $r(\mathbf{u})$ values. Figure 13(A) shows a horizontal slice of that volume. Using $r(\mathbf{u})$, one can perform a classification of seismic impedance as follows:

$$\text{if } r(\mathbf{u}) > 1, \quad \text{then } i(\mathbf{u}; s_1) = 1, \quad \text{else } i(\mathbf{u}; s_1) = 0$$

This rule is exactly the same as the Bayes' rule (9). The classification results can be compared with the actual true facies occurrence at location $\mathbf{u}$. Table 5 shows the resulting Bayes' confusion matrix which, as expected, indicates poor discrimination scores.

Next, a 3D window is considered for scanning the seismic data at the well locations. The window used has 32 pixels and its geometry is displayed in Figure 12. A principal-components analysis is performed on the input seismic data to retain the first eight principal components which explain 90% of the total variability. Because the neural network output models the conditional probability $\Pr\{I(\mathbf{u}, s_1) = 1 \mid \mathbf{z}_2(\mathbf{u})\}$, the ratio $r(\mathbf{u})$ can be calculated from that output directly. Figure 13(B) shows a horizontal slice of the ratio $r(\mathbf{u})$ using the neural network approach. When compared to Figure 13(A), Figure 13(B) shows much more white and black areas,

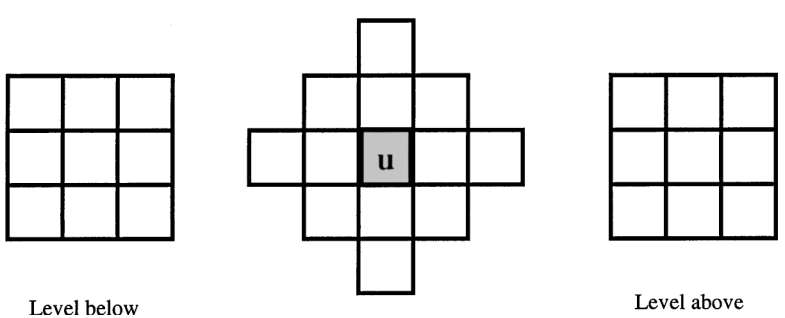

Figure 12. 3D template used for the fluvial reservoir case study.

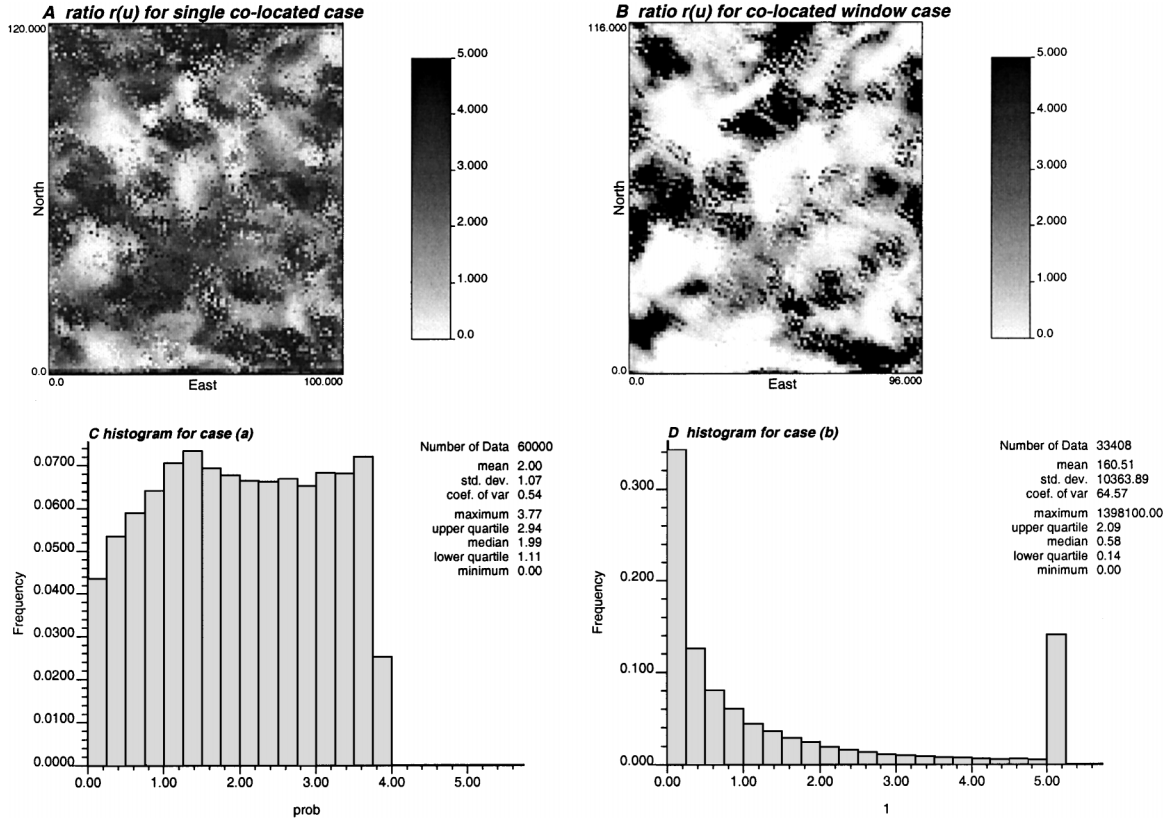

Figure 13. Fluvial reservoir case study. (A and B) Discrimination ratio $r(\mathbf{u})$ maps; (C and D) histograms of the ratios $r(\mathbf{u})$.

Table 6. Bayesian Confusion Matrix (Fluvial Reservoir, Window Approach)

	Shale	Sand
Shale	58	42
Sand	20	80

indicating that these areas are clearly distinguished as mud and sand respectively. Figure 13(A) has more gray areas, indicating a lower power of discrimination when using the single colocated seismic datum. The histograms of $r(\mathbf{u})$ corresponding to the two approaches confirm the higher discrimination ratio of the window approach.

Table 6 gives the Bayes confusion matrix based on the ratio $r(\mathbf{u})$ obtained with the window approach. The shale classification score of 58% is roughly the same as for the colocated case of 60%; however, the sand classification score has increased dramatically to 80%.

CONCLUSIONS

The neural net application to seismic data is not new, but has generally been limited to the use of the neural net as a nonlinear regression model. In this paper, we propose a simple, fast, and flexible modeling of the conditional distribution of facies, given multiple seismic data. Such modeling is comparable to the modeling of mixture distributions (e.g. Titterington, 1985). Indeed, the nodes in the hidden layer of the neural network can be viewed as single univariate distributions, as demonstrated in Caers and Journel (1998). Equation (4) is similar to the expression of a finite mixture distribution.

In an application of this general methodology to seismic data, we show that a whole colocated window of seismic data should be used for modeling the conditional distribution of facies. This allows taking into account local patterns in the seismic data, which may be associated to geological events leading to presence or absence of a specific facies at the center of the window.

The neural network approach automatically models the conditional distribution of several facies given such window of seismic data. A prior principal-components analysis of the seismic data values within the window allows a reduction of the window dimensions, and thus allows using larger windows while limiting the training effort for neural network.

Several synthetic case studies illustrated the approach proposed. When the local variability of the seismic continuity changes from one facies to another, the window approach is shown to be considerably more accurate. The approach seems to be particulatly useful for modeling seismic amplitude information without

the need for a deconvolution operation. Two practical case studies of this approach and how it can be integrated into geostatistical methodologies are presented in Caers and Haas (2001) and Caers, Avseth, and Mukerji (2001).

REFERENCES

Almeida, A., and Journel, A. G., 1993, Joint simulation of multiple variables with a Markov-type coregionalization model: Math Geol., v. 26 no. 5, p. 565–588.

Bishop, C. M., 1995, Neural networks for pattern recognition: Oxford University Press, Oxford. 482 p.

Behrens, R., Macleod, M. K., and Tran, T. T., 1996, Incorporating seismic attribute maps in 3D reservoir models, *in* Proceedings to the SPE Annual Conference and Technical Exhibition: Society of Petroleum Engineers, Denver, Oct. 4–7, 1996, SPE no. 36499.

Bridle, J. S., 1990, Probabilistic interpretation of feedforward classification network outputs, with relation to statistical pattern recognition, *in* Fogelman, F., and Herault, J., eds., Neurocomputing: Algorithms, architectures and applications: Springer, New York, p. 227–236.

Caers, J., Avseth, P., and Mukerji, T., 2001, Geostatistical integration of rock physics, seismic amplitudes and geological models in North-Sea turbidite systems: Lead. Edge, v. 20, no. 3, p. 308–312.

Caers, J., and Haas, A., 2001, Calibrating an automated seismic interpretation tools from human expert knowledge, *in* Third IMA Conference on modeling permeable rocks, Cambridge, March 27–29, 2001.

Caers, J., and Journel, A. G., 1998, Stochastic reservoir simulation using neural networks, *in* Proceedings to the SPE Annual Conference and Technical Exhibition on Formation evaluation and reservoir geology, Society of Petroleum Engineers, New Orleans, Sept. 28–30, 1998, SPE no. 49026, p. 321–336.

Christensen, R., 1997, Log-liner models and logistic regression: Springer, New York, 483 p.

Deutsch, C. V. M., and Journel, A. C., 1998, GSLIB—Geostatistical software library and user's guide, 2nd edn.: Oxford University Press, New York, 369 p.

Doyen, Ph., Psaila, D. E., den Boer, L. D., and Jans, D., 1997, Reconciling data at seismic and well-log scales in 3D earth modeling, *in* SPE Annual Technical Conference and Exhibition: Society of Petroleum Engineers, p. 465–474, SPE no. 38698.

Eide, A. L., Omre, H., and Ursin, B., 1996, Stochastic reservoir charaterization conditioned on seismic data, *in* Baafi, E., and Schofield, N., eds., International geostatistics congress: Kluwer Academic, Wollongong.

Fournier, F., and Derain, J. F., 1995, A statistical methodology for deriving reservoir properties from seismic data: Geophysics, v. 60, no. 5, 1437–1450.

Journel, A. G., 1999, Markov models for Cross-covariances: Math. Geol., v. 31, no. 8, 955–964.

Mao, S., and Journel, A. G., 1999, Generation of a reference petrophysical and seismic 3D data set: The Stanford V reservoir, *in* Stanford Center for Reservoir Forecasting Annual Meeting. Available at: http://ekofisk.stanford.edu/SCRF.html.

Mavko, G., Mukerji, T., and Dvorkin, J., 1998, The rock physics handbook: Academic Press, Cambridge, 329 p.

Pairazian, K., and Scheevel, J., 1999, Principal component analysis applied to 3D seismic data for reservoir property estimation, *in* Proceeding to the SPE Annual Meeting and Technical Exhibition: Society of Petroleum Engineers, Houston, Oct. 3–6, 1999, SPE no. 36499.

Srivastava, M., 1992, Reservoir characterization with probability field simulation, *in* Proceeding to the SPE Annual Meeting and Technical Exhibition: Society of Petroleum Engineers, Washington, DC, SPE no. 24753, p. 927–938.

Titterington, D. M., 1985, Statistical analysis of finite mixture distributions: Wiley, Chichester.

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. Petroleum Geol., v. 18, no. 2, 191–206.

Xu, W., Tran,T., and Journel, A. G., 1992, Integrating seismic data for reservoir modeling: The co-located cokriging alternative. SPE no. 24742.

Yao, T., in press, Porosity modeling in a W. Texas carbonate reservoir conditioned to seismic data: Solving the difference of scale and precision problem.

Yilmaz, O, 1987, Seismic data processing, Society of Exploration Geophysicists, Tulsa, OK, 526 p.