GEOPHYSICS, VOL. 62, NO. 6 (NOVEMBER-DECEMBER 1997); P. 1921—1930, 10 FIGS., 2 TABLES.

Extraction of geological information from acoustic well-logging waveforms using time-frequency wavelets

Naoki Saito* and Ronald R. Coifman‡

ABSTRACT

Recently developed classification and regression methods are applied to extract geological information from acoustic well-logging waveforms. First, acoustic waveforms are classified into the ones propagated through sandstones and the ones propagated through shale using the local discriminant basis (LDB) method. Next, the volume fractions of minerals are estimated (e.g., quartz and gas) at each depth using the local regression basis (LRB) method. These methods first analyze the waveforms by decomposing them into a redundant set of time-frequency wavelets, i.e., the orthogonal wiggle traces localized in both time and frequency. Then, they automatically extract the local waveform features useful for such classification and estimation or regression. Finally, these features are fed into conventional classifiers or predictors. Because these extracted features are localized in time and frequency, they allow intuitive interpretation. Using the field data set, we found that it was possible to classify the waveforms without error into sandstone and shale classes using the LDB method. It was more difficult, however, to estimate the volume fractions, in particular, that of gas, from the extracted waveform features. We also compared the performance of the LRB method with the prediction based on the commonly used ratio of compressional and shear-wave velocities, V_P/V_S, and found that our method performed better than the V_P/V_S method.

INTRODUCTION

Acoustic measurements have long been used in geophysical well logging to infer petrophysical properties or the lithology of subsurface formations (Paillet and Cheng, 1991). In sonic logging, velocities of *P*-, *S*-, and Stoneley-wave components (with or without their amplitudes) have been used to infer petrophysical or lithologic properties of the surrounding formations, such as porosity, mineralogy, grain contacts, fluid saturation, or rock types, such as sandstone, shale, and limestone (White, 1983; Paillet and Cheng, 1991; Murphy et al., 1993; Winkler and Murphy, 1995).

Extracting velocity information for each wave component, however, is not necessarily an easy task. There are two popular approaches for estimating velocities of wave components. One is a semiautomatic tracking of the first zero-crossing of the wave component, and the other is based on the semblance and coherency of the wave components among multiple waveforms (Kimball and Marzetta, 1984). Both have drawbacks. The former method often requires manual editing, since the positions of zero-crossings vary (sometimes wildly) from trace to trace. The latter method is computationally expensive and requires a data set recorded by tools equipped with multiple receivers.

The velocity and amplitude information for a particular wave component are just a portion of the information contained in the entire waveform shape. Thus, it is expected that the entire waveform shape contains more information about the lithology of the formation. In fact, the empirical relationship between the shapes of the waveforms and the lithology has been recognized for some time. Several attempts have been made to infer this information using the entire waveform shape (Hoard, 1983; Hsu, 1990). Most of these attempts have been based on statistical pattern recognition techniques because building an exact mathematical or physical model is complicated and difficult.

The first method to use waveform shape information systematically for this inference problem was that of Hoard (1983). His method is based on clustering the envelopes of the input waveforms. A few years later, Hsu (1990) recognized the importance of the relationships between individual wave components and lithologic information and proposed a different approach using the Karhunen-Loeve transform. Unfortunately, both methods suffer from the following: (1) These methods can extract only

Manuscript received by the Editor August 30, 1995; revised manuscript received January 7, 1997.
*Formerly Schlumberger-Doll Research, Old Quarry Road, Ridgefield, Connecticut 06877-4108; presently Dept. of Mathematics, University of California, Davis, CA 95616-8633. E-mail: saito@math.ucdavis.edu.
‡Department of Mathematics, Yale University, New Haven, Connecticut 06520. E-mail: coifman@math.yale.edu.

global waveform information; i.e., they cannot automatically extract the wave components localized in both time and frequency unless such localized features are explicitly supplied. (2) Their computational costs are high; the graph-theoretical clustering (with the Hilbert transform for envelope computation) costs at least $O(n^2 \log n)$, whereas the Karhunen-Loeve transform requires $O(n^3)$, where n is the number of time samples of each waveform. Moreover, both methods are "unsupervised" learning methods. In other words, these methods first extract waveform features without using any available lithologic or petrophysical information and then try to correlate the features with such information.

In this paper, we view the problem of inferring lithologic information from the entire waveform shape as a classification or regression problem, i.e., classification of the rock types or estimation of the volume fractions of minerals (e.g., quartz or gas) on the basis of the waveform shape information. Classification and regression are "supervised" learning methods. They use available lithologic or petrophysical information to construct algorithms to predict rock classes or volume fractions. The resulting algorithms are expected to predict such information when they are given a new set of sonic waveforms recorded in a geological environment similar to that of the training waveforms (which were used for constructing the algorithms). Then, we apply recently developed methods, the so-called local discriminant basis (LDB) and local regression basis (LRB) methods (Saito, 1994; Coifman and Saito, 1994; Saito and Coifman, 1994, 1995, 1996) to this inference problem. Both methods have automatic feature extraction capability. Given a training data set (i.e., waveforms and their associated lithologic information at specific depth levels), the LDB and LRB methods automatically extract useful waveform features for this inference task. For these features, we use a specific set of elementary waveforms known as time-frequency wavelets, that are well localized in both time and frequency. This local property of wavelets makes interpretation of classification and regression results far easier than with the conventional approaches directly applied to signals represented as either time samples or frequency components, since the time-domain representation is too local in time and too global in frequency, and the frequency-domain representation is too local in frequency and too global in time.

Our objective in this paper is to examine (1) how accurately we can classify lithology (in particular, sandstones or shale) on the basis of the sonic waveform features automatically extracted by the LDB method, (2) how accurately we can estimate the petrophysical quantities (in particular, volume fractions of quartz and gas) from the sonic waveform features automatically extracted by the LRB method, and (3) whether those features provide results that are easy to interpret. Our methodology is generic enough so that it should be useful for other geological settings, such as carbonate reservoirs.

The organization of the paper is as follows. In the next section, we formulate the problem and review the LDB and LRB methods. In the Data Description section, we give some background information for the field data set on which we test our methods. Then, we present the results and their interpretation in the Results section. In the Discussion section, we compare the results of the LRB method with the prediction based on the conventional physical quantity V_P/V_S. Finally, we present our conclusions.

SIGNAL CLASSIFICATION AND REGRESSION USING TIME-FREQUENCY WAVELETS

Signal classification and regression problems

Before proceeding further, we set up some notations and clarify our strategy. Let $\mathcal{X} \subseteq \mathbf{R}^n$ be an input signal space that consists of all possible input signals under consideration (i.e., sonic waveforms of n time samples). Let $\mathcal{Y} \subset \mathbf{R}$ be an output response space. For classification problems, $\mathcal{Y} = \{1, \ldots, K\}$ is a set of possible class names (e.g., 1 corresponds to sandstone, 2 corresponds to shale, etc.). For regression problems, $\mathcal{Y}$ is simply an interval in the real axis, e.g., $\mathcal{Y} = [0, 1]$ for the volume fraction of quartz or gas. Let $(\mathbf{X}, Y) \in \mathcal{X} \times \mathcal{Y}$ be a pair of an input signal and the corresponding response, which can be viewed as a random sample from a probability distribution P over $\mathcal{X} \times \mathcal{Y}$. A predictor (also often called a classifier in classification problems) is a function $d: \mathcal{X} \to \mathcal{Y}$, which predicts the response to each input signal $\mathbf{X} \in \mathcal{X}$. A learning (or training) data set $\mathcal{T}$ consists of N pairs of input signals and the corresponding responses; $\mathcal{T} = \{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_N, Y_N)\}$. We also assume the availability of data set $\mathcal{T}' = \{(\mathbf{X}_{N+1}, Y_{N+1}), \ldots, (\mathbf{X}_{N+M}, Y_{N+M})\}$, which is independent of $\mathcal{T}$ and still obeys the same probability law that generated $\mathcal{T}$. This is called a test data set and is used for the evaluation of predictors.

If we had complete knowledge of this probability model (i.e., an exact physical model that can explain the recorded phenomena with a valid stochastic noise model of the recording devices), it would be the end of the story. One could construct the so-called Bayes predictor, which gives the minimum error among all the predictors using the probability models. In practice, however, we do not have complete knowledge of the probability distribution P. For problems with a small dimensionality, e.g., $1 \leq n \leq 3$, it is feasible to estimate the probability distribution from the available training data set. For large-dimension problems, such as our sonic waveforms $n = 256$, however, it is essentially impossible to obtain a reliable estimate of the probability distribution because of the curse of dimensionality (Scott, 1992). We need a large number of training signals to estimate the probability distribution reliably. Also, feeding raw signals to conventional predictors, such as linear discriminant analysis, linear regression, or neural networks, is prohibitive. Noise in the signals and the large dimensionality of the signals overwhelm these methods. On the other hand, there often exists lower-dimensional important structures in the signals (Scott, 1992). In other words, signal classification and regression problems often have an intrinsic dimension $m \ll n$. Therefore, it would be much more efficient and effective to analyze the data and build predictors in this smaller-dimensional subspace $\mathcal{F}$ of $\mathcal{X}$. We call $\mathcal{F}$ a feature space and a map $f: \mathcal{X} \to \mathcal{F}$ a feature extractor. Then, the key is how to construct this "good" feature space $\mathcal{F}$ consisting of important features for our problems and design the corresponding feature extractor f. If we know precisely the underlying physical and mathematical models of the problem, we can design a mechanism to extract specific features relevant for that problem and may obtain its intrinsic dimension.

It is often difficult, however, to set up exact mathematical models for the problem we are interested in, such as estimating the volume fraction of gas from sonic waveforms. Therefore, in this paper, we adopt exploratory approaches; we want to find

important features for our problems by automatic procedures and examine their effectiveness. In turn, this may lead to our understanding or reconfirmation of the underlying physics of the problem.

Our strategy here for classification and regression problems can now be written as

$$d = g \circ f = g \circ \Theta_m \circ \underset{\sim}{\Psi}^T, \tag{1}$$

where $f: \mathcal{X} \to \mathcal{F}$ is a feature extractor, $g: \mathcal{F} \to \mathcal{Y}$ is any conventional predictor, and $\circ$ represents the composition operation, i.e., $d(\mathbf{X}) = g(f(\mathbf{X}))$. The feature extractor f here consists of two components. $\underset{\sim}{\Psi}$ is an n-dimensional orthogonal matrix (i.e., an orthonormal basis) selected from the so-called time-frequency dictionary (Wickerhauser, 1994; Saito, 1994) consisting of a large collection of the time-frequency wavelets briefly mentioned in the Introduction. Θ_m is a feature selector that selects the most important m ($<n$) coordinates (features) from n-dimensional coordinates. Most statistical literature focuses on the performance and statistical properties of various predictors g in equation (1). Some literature discusses the feature selector Θ_m on a given set of features. On the other hand, the LDB and LRB methods focus on f, in particular, how to select $\underset{\sim}{\Psi}$ from a finite collection of bases.

Time-frequency wavelets and dictionary

Most geophysical signals of interest are nonstationary, consisting of transients, edges, and/or local oscillations. To use them efficiently and have easily interpretable results for various tasks, one should have tools that can not only analyze the signals but also synthesize the signal components and features useful for those tasks. The traditional Fourier transform is not efficient enough to handle such nonstationary phenomena. It uses global oscillations to analyze local phenomena. Time-frequency wavelets are mathematical building blocks (basis functions or vectors) well localized in both time and frequency. Each such wavelet has specific time-frequency characteristics (i.e., location and duration of its support in the time and frequency domains). Wavelet packets and local trigonometric functions are examples of such wavelets and recently have drawn considerable attention from such diverse fields as signal and image processing, numerical analysis, and statistics. See Daubechies (1992), Meyer (1993), and Wickerhauser (1994) for the detailed properties of these wavelets and other applications.

In this paper, we use the so-called time-frequency dictionary extensively. This is a large collection of the time-frequency wavelets organized in a hierarchical fashion (a binary tree structure) (Wickerhauser, 1994; Saito, 1994). Each node in this binary tree represents a specific set of wavelets that spans the subspace occupying a specific region in the time-frequency plane. For signals of length n, this dictionary can contain up to $n(1 + \log_2 n)$ wavelets and more than 2^n different complete orthonormal bases (Wickerhauser, 1994). Let $\mathcal{D}$ be this time-frequency dictionary consisting of a collection of wavelets $\{\mathbf{w}_i\}_{i=1}^{N_w}$. This dictionary $\mathcal{D}$ also can be expressed as a list of all possible orthonormal bases (matrices) $\{\underset{\sim}{\mathbf{B}}_j\}_{j=1}^{N_B}$, where $\underset{\sim}{\mathbf{B}}_j = (\mathbf{w}_{j_1}, \ldots, \mathbf{w}_{j_n})$. This redundancy permits us to select the most suitable basis (coordinate system) $\underset{\sim}{\Psi}$ in equation (1) for given signals and a given objective. This implies that we can represent our signals as linear combinations of local features that are useful for our classification and regression tasks. This local property makes the interpretation of classification and regression results far easier than with the conventional approaches (i.e., various conventional predictors computed on signals represented entirely as either time samples or frequency components). Moreover, the selection of such a basis is still computationally efficient. Expanding a signal of length n into such tree-structured bases is fast, i.e., $O(n \log n)$ for a wavelet packet dictionary and $O(n(\log n)^2)$ for a local trigonometric dictionary. After expansion, the selection of a good basis $\underset{\sim}{\Psi}$, the so-called best-basis algorithm of Coifman and Wickerhauser (1992), is readily available. Let $\mathcal{M}^+(\underset{\sim}{\mathbf{B}}_j)$ be a measure of efficacy of the basis $\underset{\sim}{\mathbf{B}}_j$ for a certain task. Then, the best-basis algorithm selects

$$\underset{\sim}{\Psi} = \arg\max_{\underset{\sim}{\mathbf{B}}_j \in \mathcal{D}} \mathcal{M}^+(\underset{\sim}{\mathbf{B}}_j) \tag{2}$$

in $O(n)$. Clearly, if we use the deficiency of the basis $\mathcal{M}^-$ instead of efficacy, equation (2) should be written as

$$\underset{\sim}{\Psi} = \arg\min_{\underset{\sim}{\mathbf{B}}_j \in \mathcal{D}} \mathcal{M}^-(\underset{\sim}{\mathbf{B}}_j). \tag{3}$$

The heart of the matter is the measures of efficacy $\mathcal{M}^+$ or deficiency $\mathcal{M}^-$. The original best-basis algorithm was developed mainly for signal compression problems.

Local discriminant basis (LDB)

A good basis for classification and discrimination should be the one through which we can "view" the classes as maximally separated point clouds in $\mathcal{X}$. Therefore, we need to measure a certain kind of "distances" among the signal classes under the coordinate system $\underset{\sim}{\mathbf{B}}_j$. To measure such distances, we need the probability density function (pdf) of each input signal class. This is, however, impossible to compute or estimate because of the high dimensionality of $\mathcal{X}$, as mentioned previously. Therefore, we evaluate each basis vector in $\underset{\sim}{\mathbf{B}}_j$ separately and sum up their efficacies. If we project an input signal $\mathbf{X} \in \mathcal{X}$ onto a unit vector $\mathbf{w}_i \in \mathcal{D}$, then its projection (or coordinate) $Z_i \triangleq \mathbf{w}_i \times \mathbf{X}$ also is a random variable, where $\times$ denotes the standard inner product in $\mathbf{R}^n$. We are interested in knowing how Z_i is distributed for each signal class so that we can quantify the efficacy of the direction $\mathbf{w}_i$ for classification. We refer to such a measure of efficacy as a discriminant measure. We also use the term discriminant power of $\mathbf{w}_i$, which is the actual value of such a measure evaluated at $\mathbf{w}_i$. Here, we must deal with two issues. One is to decide how to represent the distribution of Z_i; the other is how to define the differences or distances among those computed distributions among signal classes. In this paper, we consider only the following quantities to represent how Z_i is distributed:

1) Normalized energy of Z_i written as

$$V_i^{(y)} \triangleq \frac{E\left[Z_i^2 \mid Y = y\right]}{\sum_{i=1}^{n} E\left[Z_i^2 \mid Y = y\right]}, \tag{4}$$

where E is the mathematical expectation.

2) Probability density function of Z_i written as

$$q_i^{(y)}(z) \triangleq \int_{\mathbf{w}_i \cdot \mathbf{x} = z} p(\mathbf{x} \mid y)\, d\mathbf{x}. \tag{5}$$

In practice, however, the true conditional pdf $p(\mathbf{x}\,|\,y)$ is not available, as mentioned in the previous section. Therefore, we must estimate equations (4) and (5) using the available training data set. In other words, instead of equation (4) we use the sample estimate

$$\hat{V}_i^{(y)} = \frac{\sum_{k=1}^{N_y} \left| \mathbf{w}_i \cdot \mathbf{x}_k^{(y)} \right|^2}{\sum_{k=1}^{N_y} \left\| \mathbf{x}_k^{(y)} \right\|^2}, \tag{6}$$

where $\{\mathbf{x}_k^{(y)}\}_{k=1}^{N_y}$ is a set of available class y signals in the training data set $\mathcal{T}$. For equation (5), we estimate the pdf using commonly available methods, such as histograms or kernel density estimators (Scott, 1992). We denote the sample estimate of $q_i^{(y)}$ by $\hat{q}_i^{(y)}(z)$. We note here that the normalized energy [equation (6)] is simply a scalar for each coordinate, but the empirical estimation of the pdf per coordinate $\hat{q}_i^{(y)}(z)$ is a function (or a vector after discretization). Therefore, the estimated pdf is expected to carry more subtle information, such as phase information of input signals, than the normalized energy, whereas the normalized energy is definitely computationally much cheaper than the estimation of pdf.

Once we decide the representation of the distribution of the projections onto each vector in $\mathcal{D}$, the next question is how to define the efficacy of a basis $\underset{\sim}{\mathbf{B}}_j$ or how to measure "distances" among the signal classes under the coordinate system $\underset{\sim}{\mathbf{B}}_j$. There are many choices for such a distance measure (Basseville, 1989). In this paper, we consider only the relative entropy measure as $\mathcal{M}^+$ in equation (2). See Saito (1996) for some experiments using the other measures. The normalized energy for two-class problems can be written as

$$\mathcal{M}^+(\underset{\sim}{\mathbf{B}}_j) = \sum_{k=1}^{n} \hat{V}_{j_k}^{(1)} \log_2 \frac{\hat{V}_{j_k}^{(1)}}{\hat{V}_{j_k}^{(2)}}. \tag{7}$$

On the other hand, if we use the estimated pdf, we obtain

$$\mathcal{M}^+(\underset{\sim}{\mathbf{B}}_j) = \sum_{k=1}^{n} \int_{-\infty}^{\infty} \hat{q}_{j_k}^{(1)}(z) \log_2 \frac{\hat{q}_{j_k}^{(1)}(z)}{\hat{q}_{j_k}^{(2)}(z)}\, dz. \tag{8}$$

For K-class problems with $K > 2$, the simplest approach is to take $K(K+1)/2$ pair-wise combinations of the above measures. We first proposed to use equation (7) to obtain the LDB for a given classification problem (Coifman and Saito, 1994; Saito and Coifman, 1994). Therefore, we call the method using equation (7) the original LDB (OLDB) method. The LDB algorithm based on equation (8) is a relatively new strategy (Saito and Coifman, 1996), so we call this method the new LDB (NLDB) method.

After selecting the LDB, we still need to find the best $m(<n)$ features (or LDB coordinates) to supply to classifiers. This is an interesting but difficult problem. There is no theoretical answer to this problem [see, e.g., the counterintuitive example in Cover (1974)]. Therefore, we use the suboptimal but fast approach. We first sort the feature into descending order in its discriminant power [i.e., the summand of equation (7) or (8)]. We then take the first m features and construct the final classifier g in equation (1) on the basis of these features.

Local regression basis (LRB)

In regression problems, for a given input signal, one wants to predict a certain quantity of interest (e.g., volume fraction of gas) carried by the signal. Classification problems can be viewed as special cases of the regression problems by interpreting the class labels as the quantity of interest. The goal here is to extract a good prediction rule (predictor) from a given training data set that is applicable to a test data set and that provides better insight and understanding of the input-output relationships. An important factor in the efficacy of regression is the accuracy of the prediction. Thus, we adopt prediction errors such as the residual sum of squares as the measure of the deficiency $\mathcal{M}^-$ in equation (3), written as

$$\mathcal{M}^-(\underset{\sim}{\mathbf{B}}_j) = \sum_{i=1}^{N} \left[Y_i - \tilde{g}\left(\underset{\sim}{\mathbf{B}}_j^T \mathbf{X}_i; \mathcal{T}\right) \right]^2, \tag{9}$$

where $\tilde{g}$ is a certain regression function (e.g., linear regression) that clearly is dependent on the available training data set $\mathcal{T}$. Note that the measure (9) depends on the choice of the regression function $\tilde{g}$. We also note that $\tilde{g}$ here is not necessarily the same regression method as the final predictor g in the overall scheme (1). A major difference from the distance measures (7) and (8) in the classification problems is that we do not evaluate the individual coordinate for regression problems. Instead, we compute the regression error for each node (subspace) in the dictionary $\mathcal{D}$. This is because the regression error is generally not additive, i.e., $\mathcal{M}^-(\{\mathbf{w}_j, \mathbf{w}_k\}) \neq \mathcal{M}^-(\{\mathbf{w}_j\}) + \mathcal{M}^-(\{\mathbf{w}_k\})$. We call the basis obtained via equation (9) a LRB relative to $\tilde{g}$. After computing the LRB, we again need to select the best m LRB coordinates to supply to the final predictor g in equation (1). The situation is exactly the same as in the LDB case. There is no theory to obtain the best m features from n features for regression either. In this paper, we rely on the regression method g, which has a built-in feature selection mechanism. Examples include the regression tree (Breiman et al., 1993) and the linear regression model with stepwise variable selection (Draper and Smith, 1981).

DATA DESCRIPTION

In this study, we use 3012 acoustic waveforms recorded in a certain well at every 0.5 ft (0.15 m) depth. These are the single-receiver waveforms, i.e., the common-offset gathers [the distance between the source and the receiver is 9 ft (2.74 m)]. Each waveform consists of 512 time samples with a sampling rate of 10 μs. No core samples from this well were available to us. Thus, we do not know the "ground truth" of the lithology here. Fortunately, however, the volume fractions of minerals at each depth level were available. These were computed by the volumetric analysis method from a set of geophysical well logs (Quirein et al., 1986; Cannon and Coates, 1990). No acoustic or elastic information was used to compute these volume fractions. In this study, we use the volume fractions of quartz and gas as the "lithologic" information. Because of the 9-ft distance between the source and receiver locations, we assume that each waveform carries the average lithologic information of a 9-ft interval. Therefore, we computed the average of these quartz and gas volume fractions over a 9-ft interval immediately below the receiver locations and used these averaged quantities as the lithologic information in this paper.

The region in which the well is located consists mainly of sandstone-shale sequences. Most sandstone layers contain either gas or water. Figure 1 shows the data set under study. From the volume fraction curves in Figure 1, we make the following observations about the geological layers in this data set.

1) L1—There is a thick sandstone layer containing gas around the depth indices ranging from 1600 to 2100.
2) L2—There is a shale layer around the depth indices ranging from 700 to 1100.
3) L3—There are alternating sandstone-shale sequences above the thick sandstone layer and below the shale layer described above.

We refer to the waveforms propagated through sandstone layers as "sand waveforms" and those propagated through shale layers as "shale waveforms." We observe by visual inspection the following waveform features from Figure 1.

1) W1—The *S*-wave components in the sand waveforms have much stronger energy and faster speed than those in the shale waveforms.
2) W2—The velocities of the *P*-wave components in the sand waveforms are higher than those in the shale waveforms.
3) W3—The velocities of the Stoneley-wave components in the shale waveforms (depth index, 700–1100) are lower than those in the sand waveforms, except for those at the bottom of the well.

The physics of wave propagation suggest that the *P*- and *S*-wave velocities are sensitive to the fluid content and the mineralogy and that the Stoneley-wave velocity is sensitive to the permeability of the formations as well as borehole conditions, such as rugosity and borehole diameter (White, 1983; Paillet and Cheng, 1991; Murphy et al., 1993; Winkler and Murphy, 1995). Because of the sensitivity to borehole conditions, we smoothly taper the Stoneley-wave component from each waveform and consider only the earlier part of the waveforms (i.e., the number of time samples is now 256 for each waveform).

Because the waveforms from neighboring wells were not available, we had to select both training and test data sets from this well. Therefore, we adopt the so-called ten-fold cross-validation method (Weiss and Kulikowski, 1991). We first divide the entire data set randomly into ten groups (i.e., each group consists of 301 waveforms and the corresponding lithologic information). We then repeat ten experiments by taking one group as a test data set and the remaining nine groups as a training data set at each time. Finally, we take an average over these ten results to obtain the mean prediction error estimate of our methods.

For waveform classification experiments, we do not use the entire data set. Instead, we select from the entire data set representative sand waveforms and shale waveforms as our working data set and simply ignore the rest. This is because the data set contains ambiguous layers, such as "shaly-sand" or "sandy-shale," that make our classification task difficult. This two-class assumption, however, is a good starting point to examine what features in the waveforms carry the discriminatory information between sandstone and shale and whether our methods extract such relevant waveform features automatically and classify the waveforms correctly. The selected subset of data consists of 201 contiguous depth levels from the main shale layer and 201 depth levels from three different sandstone layers (Figure 2). We use the ten-fold cross-validation method (each group now consists of 40 waveforms and the corresponding class names) to compute the average misclassification rate.

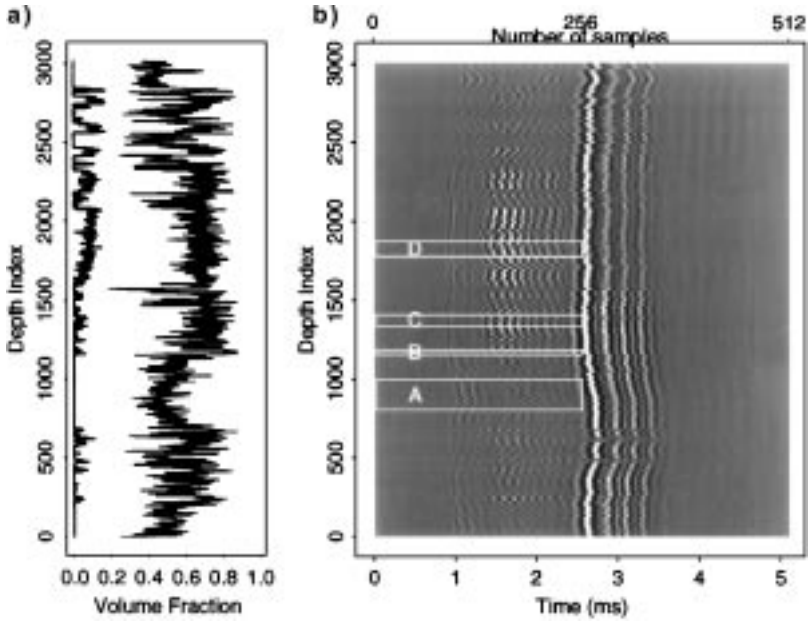

FIG. 1. Data set used in this study. (a) Two curves representing volume fractions of quartz (solid line) and gas (thick solid line having smaller volume fractions). (b) Acoustic waveforms recorded at the corresponding depth levels, shown as a gray-scale image. The depth index 0 corresponds to the deepest level. The windows indicated by A, B, C, and D, were used for the classification study in this paper.

RESULTS

Since velocity information is important in this study, a natural choice for the time-frequency decomposition is the local trigonometric transforms rather than the wavelet packets. It is easier to manipulate the time (or velocity) information in the local trigonometric transforms than in the wavelet packets. Hence, we use a dictionary of local sine bases in this study.

Waveform classification by LDB

We computed both OLDB and NLDB with the ten-fold cross-validation method. Let $\mathcal{T}_i$ be the ith training data set and $\mathcal{T}_i'$ be the corresponding test data set ($i = 1, \ldots, 10$) in the cross-validation experiments. Note that $\mathcal{T}_i \cup \mathcal{T}_i'$ is the entire working data set shown in Figure 2 for each i. Our experiments are described as follows.

begin
  for $i := 1$ **to** 10 **step** 1 **do**
    Compute an LDB from the training data set $\mathcal{T}_i$;
    for $m := 5$ **to** 100 **step** 5 **do**
      Select the most discriminant m coordinates from the LDB coordinates;
      Build a classifier using these m features of the training signals in $\mathcal{T}_i$;
      Compute the misclassification rate $\epsilon_{m,i}$ by supplying the corresponding m features of the test signals in $\mathcal{T}_i'$;
    enddo
  enddo
  Compute the average misclassification rate for each m as $\epsilon_m = (1/10)\sum_{i=1}^{10} \epsilon_{m,i}$;
end

We used the averaged shifted histograms method (Scott, 1992) as the pdf estimator in the NLDB algorithm. For the classifier, we used the linear discriminant analysis (LDA) and the classification tree (CT). For the details of these standard classification methods, see Fukunaga (1990) for LDA and Breiman et al. (1993) for CT. We examined the dependence of classification performance on the number of selected features m. The results for $m = 5, \ldots, 100$ in steps of five are summarized in Figures 3 and 4. From these plots, we make the following observations.

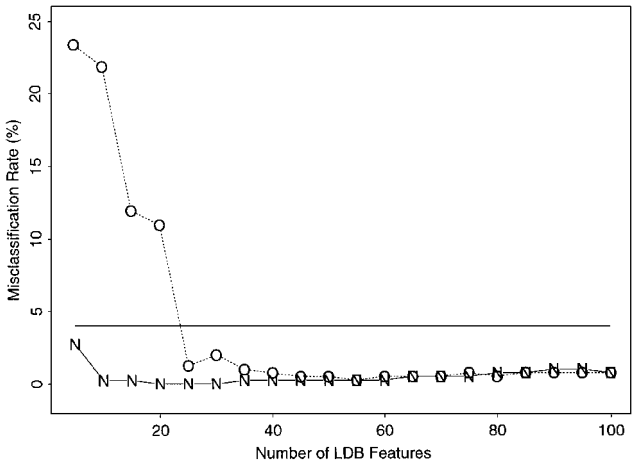

FIG. 3. Misclassification rates using LDA as a classifier versus the number of top LDB features retained. The plots with symbols O and N correspond to the results obtained using the OLDB and NLDB algorithms, respectively. The constant level line at about 4% indicates the performance of LDA applied directly to the signals represented in the standard coordinate system (of 256 time samples).

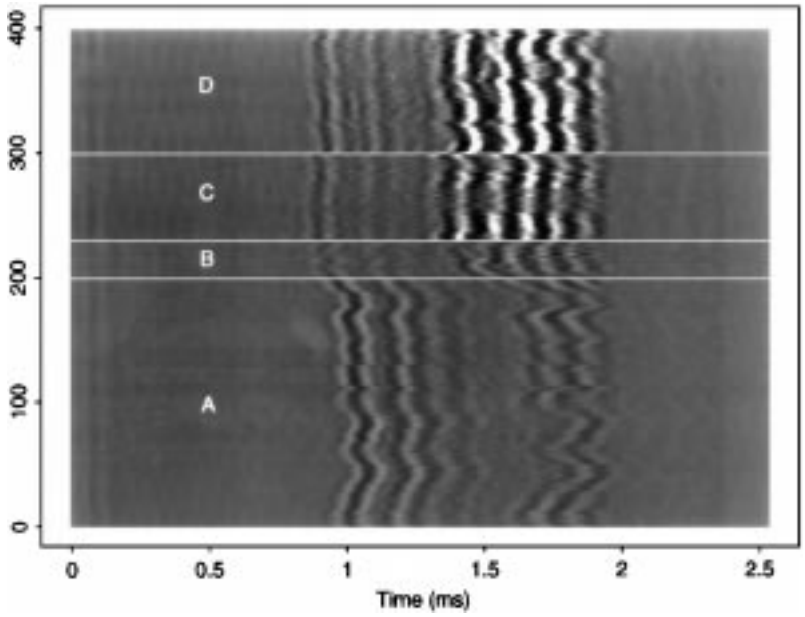

FIG. 2. Subset of Figure 1 selected for the classification study. The bottom 201 recordings correspond to the shale region (window A in Figure 1). The top 201 recordings correspond to the sandstone regions [windows B, C (water-sand), and D (gas-sand)]. The waveforms have been smoothly tapered to eliminate the Stoneley wave components.

1) No misclassification occurs with LDA in the top 20, 25, and 30 NLDB features.
2) Using LDA with less than 40 features, NLDB outperforms OLDB. The difference is small for more than 45 features.
3) Using CT, OLDB performs better than NLDB, but the result in the standard basis (STD) is even better.

These top 20 to 30 NLDB vectors or wavelets, with which the perfect classification was obtained, are essentially the same throughout the 10 cross-validation experiments; these features are quite stable. Figure 5 shows the top 20 NLDB wavelets constructed from the $\mathcal{T}_1$ data set. From this figure, we observe that the NLDB algorithm selected the wavelets supported in the first half-time interval, where the P-wave components exist. In other words, we can conclude that it is possible to use only P-wave components to discriminate sand waveforms from shale waveforms in this data set. We would like to emphasize that these features were selected completely automatically once the data set was given. Since LDA produces the best linear combination of given features, we can synthesize the most discriminant feature. Figure 6 compares the feature synthesized from the top 20 NLDB wavelets with the feature obtained by applying LDA directly to the original signals represented as time samples (i.e., on the standard basis). The difference is very clear. LDA applied directly to the time samples is very noisy and too sensitive to extremely local events in time, since this synthesized feature is a linear combination of unit impulses over the whole time interval. For example, the big peaks around $t = 2$ ms must be an effect of the tapering used to remove the Stoneley-wave components. On the other hand, the feature synthesized from the top 20 NLDB wavelets is very similar to the P-wave components, because this feature is a linear combination of sinusoids supported in the first half-time interval. Our classification scheme is equivalent to the following intuitive procedure: (1) correlate the discriminant feature in Figure 6 with each recorded waveform (i.e., inner product of this feature and each waveform) and (2) classify a waveform as the sand waveform class if the correlation is larger than some threshold but as the shale waveform class otherwise. The reason for success is that this synthesized feature, in particular, in the time segment from 0.8 ms to 1.2 ms, is "in phase" with the P-wave components of the sand waveforms but "out of phase" with those of the shale waveforms. The NLDB algorithm successfully extracted this discriminant feature, i.e., the difference in P-wave velocity, automatically.

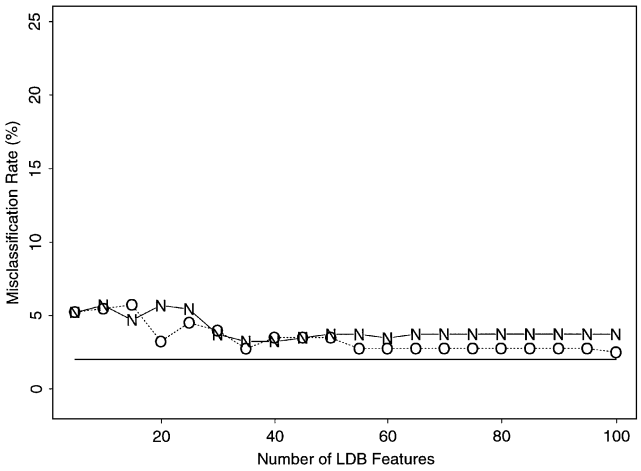

FIG. 4. Misclassification rates using CT as a classifier versus the number of top LDB features retained. The constant level line at about 2% indicates the performance of CT applied directly to the signals represented in the standard coordinate system. O = OLDB results; N = NLDB results.

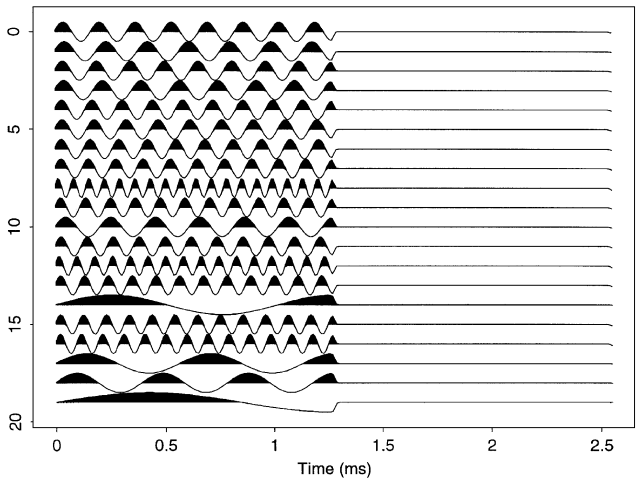

FIG. 5. Top 20 NLDB vectors that allowed perfect classification. All of them are supported in the first half-time interval, which captures P-wave components.

Estimation of volume fraction of minerals by LRB

We now describe our results for a more challenging problem, the estimation of volume fractions of quartz and gas using the sonic waveforms. We used the entire data set shown in Figure 1 and split it randomly into ten groups for the cross-validation experiments, which are described as follows.

begin
for $i := 1$ **to** 10 **step** 1 **do**
Compute an LRB relative to $\tilde{g}$ from the training data set $\mathcal{T}_i$;

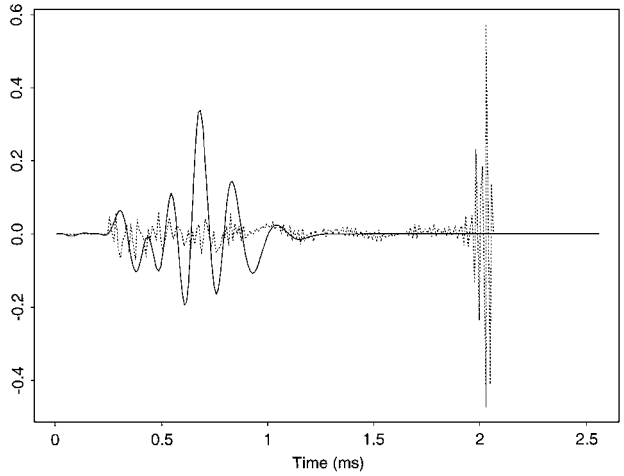

FIG. 6. Local feature synthesized by the linear combination of the top 20 NLDB wavelets (solid line) and that synthesized by the linear combination of the unit impulses (dashed line). The weights were supplied by the LDA method in both cases.

Supply the entire LRB coordinates to the predictor g, which does the automatic feature selection;
Predict the responses in $\mathcal{T}_i$ and compute the relative ℓ^2 error $R(\mathcal{T}_i)$;
Compute the LRB coordinates of the test signals in $\mathcal{T}_i'$, predict the responses, and compute the relative ℓ^2 error $R(\mathcal{T}_i')$;

enddo

Compute the average relative ℓ^2 errors as $R = (1/10)\sum_{i=1}^{10} R(\mathcal{T}_i)$, $R' = (1/10)\sum_{i=1}^{10} R(\mathcal{T}_i')$;

end

As a final predictor g in equation (1) and a basis evaluator $\tilde{g}$ in objective function (9), we adopted the linear regression model with step-wise variable selection (LMSTEP) and the regression tree (RT). For the details of these regression methods, see Draper and Smith (1981) for the LMSTEP and Breiman et al. (1993) for the RT. The relative ℓ^2 error R for a data set $\mathcal{S} = \{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_L, Y_L)\}$ is defined as

$$R(\mathcal{S}) \triangleq \left(\frac{\sum_{j=1}^{L} (Y_j - \hat{Y}_j)^2}{\sum_{j=1}^{L} Y_j^2} \right)^{1/2}, \qquad (10)$$

where $\hat{Y}_j$ is the prediction of Y_j using the regression method.

We summarize the average prediction errors in Table 1. As a comparison, we also estimated these volume fractions from the original waveforms represented in the STD, i.e., as time samples. From Table 1, we make the following observations.

Table 1. Average prediction (relative ℓ^2) errors for the volume fractions of minerals using the LRB; STD denotes the waveforms represented in the standard coordinates (i.e., as time samples).

	Error for quartz		Error for gas	
Method	Training	Test	Training	Test
RT on STD	0.0630	0.0747	0.3456	0.4783
LMSTEP on STD	0.1073	0.1125	0.6281	0.6754
RT on LRB RT	0.0629	0.0732	0.3293	0.4486
LMSTEP on LRB RT	0.1067	0.1117	0.6060	0.6747
LMSTEP on LRB LMSTEP	0.1047	0.1135	0.6041	0.6730
RT on LRB LMSTEP	0.0666	0.0833	0.3352	0.4575

1) For both quartz and gas volume estimates, the best results were obtained with RT applied to the LRB-RT coordinates [i.e., RT is used for both g in equation (1) and $\tilde{g}$ in equation (9)].
2) The use of LMSTEP as the final predictor g generated larger errors in both the quartz and the gas volume fraction estimates than did the use of RT.
3) The errors in the gas volume were much larger than those in the quartz volume.

The first two observations suggest that the linear combinations of LRB wavelets or impulses are not suitable features for the estimation of volume fractions of minerals. In other words, there must be a nonlinear relationship between these volume fractions and the LRB and STD coordinates. This nonlinearity may be the reason for the good performance of the RT, which is a nonlinear predictor. Even the RT on STD produced results better than any obtained by setting $g =$ LMSTEP.

As for the third observation (the larger errors for gas than for quartz), there may be two reasons. First, the acoustic velocities contain little information on hydrocarbon saturation (Williams, 1990). (See also the Discussion section.) Second, only a small number of gas waveforms exist in our data set. In fact, only about 20% of the entire depth levels have a gas volume fraction of more than 0.01 (i.e., the distribution of the gas volume fractions is very skewed toward the lower values), whereas the distribution of the quartz volume fraction is well balanced at about 0.6.

In Figure 7, we show cross plots of the quartz volume fractions given by the volumetric analysis method of Quirein et al. (1986) and Cannon and Coates (1990) versus our best estimates by RT on LRB-RT coordinates. In this case, the RT method selected 27 wavelets (LRB-RT vectors), on average, in these 10 cross-validation experiments. Throughout these ten experiments, four wavelets were consistently selected and significantly reduced the prediction errors in comparison with the other wavelets. In Figure 8, we show these four "common" or "survived" wavelets in our ten cross-validation experiments.

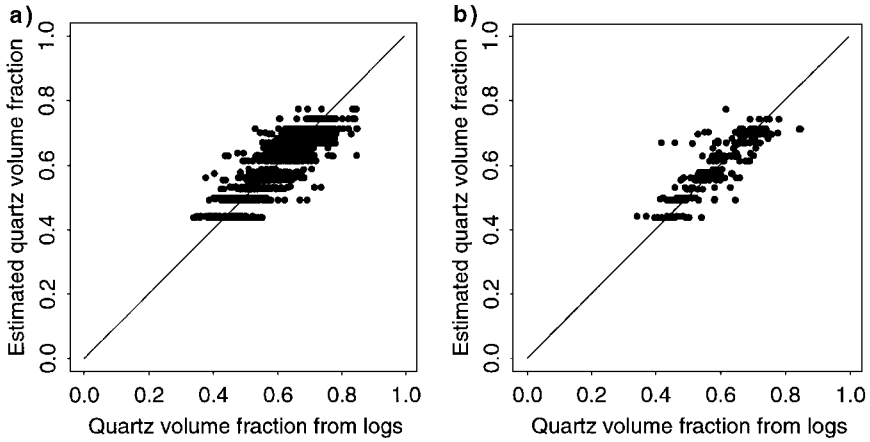

FIG. 7. Cross plots of the quartz volume fractions estimated by the volumetric analysis method (Quirein et al., 1986; Cannon and Coates, 1990) versus those estimated by the proposed method (with RT on LRB-RT coordinates). (a) Results for the training data set $\mathcal{T}_2$. (b) Results for the corresponding test data set $\mathcal{T}_2'$.

From Figure 8, we observe that the *P*-wave components play a more important role for the quartz volume fraction estimates, since the algorithm selected three wavelets in the time interval in which the *P*-waves reside. This is consistent with the classification problem discussed above, in which all the discriminant information (whether sand or shale) is contained in the *P*-wave components. In fact, a further examination of the plots in Figures 1, 2, and 8 shows that the three wavelets of the LRB indices, 81, 85, and 89, are located at the beginning of the *P*-wave components of the sand waveforms. On the other hand, the shale waveforms in the time interval in which these three wavelets are located have very little activity. The *P*-waves have not yet arrived in the shale region.

In Figure 9, we show cross plots of the gas volume fractions. RT on LRB-RT coordinates again produced the best results and selected 34 wavelets, on average. However, the selected features were more variable among the ten experiments than in the quartz case. Four of the ten experiments produced the three global sinusoids as the "survived" wavelets, whereas the other six experiments produced two wavelets located in the *S*-wave time interval and one wavelet located in the *P*-wave time interval. These three wavelets are shown in Figure 10. The frequency contents of the three global sinusoids are essentially the same as those of these three wavelets. This implies that the particular frequency components of the *S*-waves are important for the prediction of the gas volume fractions in this data set. This can be understood by examining Figures 1 and 2. The gas waveforms have distinctive frequency components with high amplitudes. LRB automatically found these specific features (wavelets 151 and 191) for gas volume estimation. The physical reason for this high-amplitude effect of the gas waveforms may be the difference in the shear velocity of the gas sand layers and the velocity in the borehole mud ($V_S > V_{\text{mud}}$) in this well. Wavelet 94 is located in the middle of the *P*-wave components of the sand waveforms, and it also works as a "sand detector."

DISCUSSION

In this study, we automatically extracted the waveform features, predicted the quantity of interest, and provided some

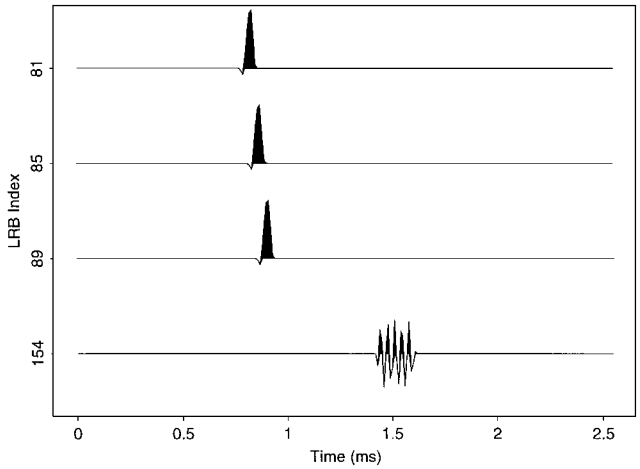

FIG. 8. LRB-RT wavelets that contributed most to the quartz volume fraction estimation.

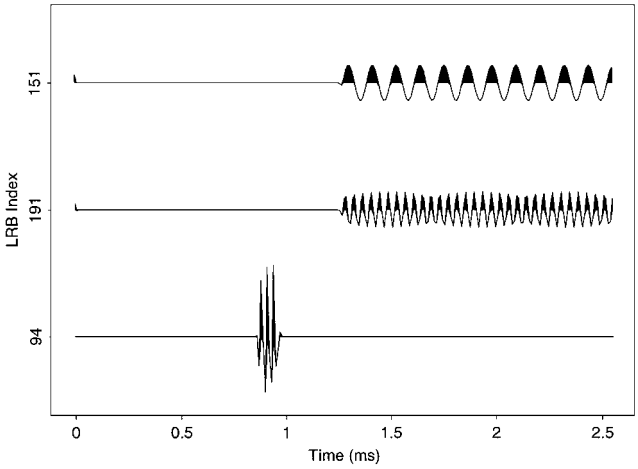

FIG. 10. LRB-RT wavelets that contributed most to the gas volume fraction estimation.

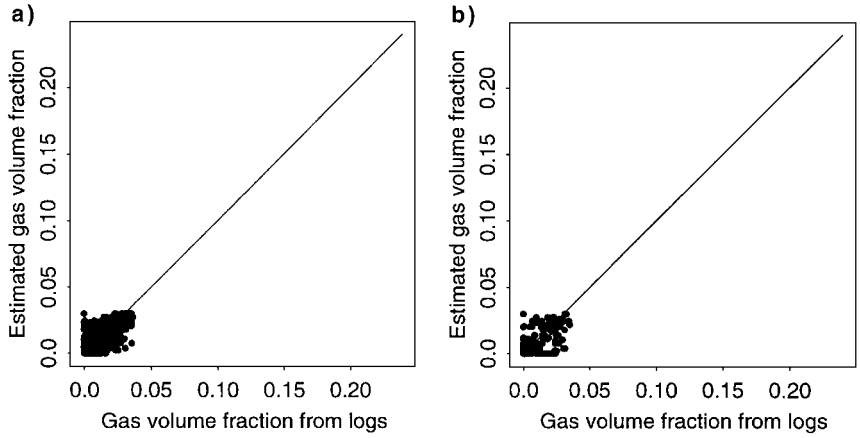

FIG. 9. Cross plots of the gas volume fractions estimated by the volumetric analysis method (Quirein et al., 1986; Cannon and Coates, 1990) versus those estimated by the proposed method (with RT on LRB-RT coordinates). (a) Results for the training data set $\mathcal{T}_2$. (b) Results for the corresponding test data set $\mathcal{T}_2'$. Note that the axis range just covers the porosity range in this well (0 to 0.24).

Table 2. Average prediction errors for the quartz and gas volume fractions using regressions on the V_P/V_S values with the 10-fold cross-validation method.

	Error for quartz		Error for gas	
Method	Training	Test	Training	Test
RT on V_P/V_S	0.0981	0.1127	0.4906	0.5754
LM on V_P/V_S	0.1095	0.1095	0.5980	0.6001

interpretations. Thus, it is important to compare the performance of the automatic procedures with that of the regression analysis for the quantities directly derived from the theory of wave propagation. The physics of wave propagation suggest that the key parameter for predicting lithology is the ratio of P- and S-wave velocities, V_P/V_S (Tatham, 1982; Williams, 1990). Although this quantity is still affected by other factors, such as cracks, pores, and geometry, it is related directly to Poisson's ratio, a characteristic property of elastic solids (Murphy et al., 1993; Tatham, 1982). In our data set, the velocities of the P- and S-wave components at each depth level were computed using the semblance-based algorithm (Kimball and Marzetta, 1984), which requires recording of multiple waveforms for each shot (in this case, eight receivers). This means that the average depth resolution of these velocities was 3.5 ft (1.07 m). Therefore, to compare performance with that in the experiments of the previous section, we averaged V_P and V_S values over a 9-ft window and then computed V_P/V_S values. For a fair comparison with our methods, we used RT and the simple linear regression model (LM) to predict the quartz and gas volume fractions from the V_P/V_S values. We again used the ten-fold cross-validation method to compute the average prediction errors as shown in Table 2. From Table 2, we can observe that the results of the LRB-RT method were superior to those of the V_P/V_S method, whereas the results of the LRB-LMSTEP method were inferior to the V_P/V_S method results. This is understandable, since V_P/V_S does not contain the volume information of the saturating fluid (Paillet and Cheng, 1991; Williams, 1990).

Williams (1990) proposed an algorithm to identify the type of saturating fluid based on V_P/V_S. His algorithm, however, is not a predictor of the volume of the saturating fluid, as he emphasizes. Therefore, we cannot use his algorithm for our volume estimation problem.

CONCLUSION

In this paper, we applied the LDB and LRB methods (Saito, 1994; Coifman and Saito, 1994; Saito and Coifman, 1994, 1995, 1996) to classify rock types and to predict the volume fractions of quartz and gas from acoustic well-logging waveforms. Using these methods, we successfully extracted the features useful for predicting such information automatically. Our results also allowed us to make interpretations in an intuitive manner and essentially agreed with explanations by the physics of wave propagation.

The LDB and LRB methods tested in this paper are generic in nature, and we believe that they can be used for other classification and regression tasks in geophysical problems. Their extension to images or multichannel waveforms also is quite straightforward.

ACKNOWLEDGMENTS

We thank the following people for useful discussion and help on data preparation: Kai Hsu of LWD-Schlumberger; Mike Kane, Hsiu-Lin Liu, Bill Murphy, and Bikash Sinha of Schlumberger-Doll Research; and Jim Marfice of GeoQuest. We also greatly appreciate various fruitful suggestions and criticisms from the anonymous reviewers, which improved this paper significantly.

REFERENCES

Basseville, M., 1989, Distance measures for signal processing and pattern recognition: Signal Processing, **18**, 349–369.
Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1993, Classification and regression trees: Chapman and Hall.
Cannon, D. E., and Coates, G. R., 1990, Applying mineral knowledge to standard log interpretation: 31st Ann. Logging Symp., Soc. Prof. Well Log Analysts, Paper V.
Coifman, R. R., and Saito, N., 1994, Constructions of local orthonormal bases for classification and regression: Comptes Rendus Acad. Sci. Sér. I, **319**, No. 2, 191–196.
Coifman, R. R., and Wickerhauser, M. V., 1992, Entropy-based algorithms for best basis selection: IEEE Trans. Inf. Theory, **38**, 713–719.
Cover, T. M., 1974, The best two independent measurements are not the two best: IEEE Trans. Syst. Man Cybern., **SMC-4**, 116–117.
Daubechies, I., 1992, Ten lectures on wavelets: Soc. Ind. Appl. Math.
Draper, N. R., and Smith, H., 1981, Applied regression analysis, 2nd ed.: John Wiley & Sons, Inc.
Fukunaga, K., 1990, Introduction to statistical pattern recognition, 2nd ed.: Academic Press Inc.
Hoard, R. E., 1983, Sonic waveform logging: A new way to obtain subsurface geologic information: 24th Ann. Logging Symp., Soc. Prof. Well Log Analysts, Paper XX.
Hsu, K., 1990, Wave separation and feature extraction of acoustic well-logging waveforms using Karhunen-Loeve transformation: Geophysics, **55**, 176–184.
Kimball, C. V., and Marzetta, T. L., 1984, Semblance processing of borehole acoustic array data: Geophysics, **49**, 274–281.
Meyer, Y., 1993, Wavelets: Algorithms and applications: Soc. Ind. Appl. Math.
Murphy, W., Reischer, A., and Hsu, K., 1993, Modulus decomposition of compressional and shear velocities in sand bodies: Geophysics, **58**, 227–239.
Paillet, F. L., and Cheng, C. H., 1991, Acoustic waves in boreholes: CRC Press, Inc.
Quirein, J., Kimminau, S., LaVigne, J., Singer, J., and Wendel, F., 1986, A coherent framework for developing and applying multiple formation evaluation models: 27th Ann. Logging Symp., Soc. Prof. Well Log Analysts, Paper DD.
Saito, N., 1994, Local feature extraction and its applications using a library of bases: Ph.D. thesis, Yale Univ.
———1996, Classification of geophysical acoustic waveforms using time-frequency atoms: Proc. Am. Statist. Assoc. Statistical Computing.
Saito, N., and Coifman, R. R., 1994, Local discriminant bases, *in* Laine, A. F., and Unser, M. A., Eds., Mathematical imaging: Wavelet applications in signal and image processing II: Proc. SPIE, **2303**, 2–14.
———1995, Local discriminant bases and their applications: J. Math. Imaging Vision, **5**, 337–358.
———1996, Improved local discriminant bases using empirical probability density estimation: Proc. Am. Statist. Assoc. Statistical Computing Section, 312–321.
Scott, D. W., 1992, Multivariate density estimation: theory, practice, and visualization: John Wiley & Sons, Inc.
Tatham, R. H., 1982, V_P/V_S and lithology: Geophysics, **47**, 336–344.
Weiss, S. M., and Kulikowski, C. A., 1991, Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems: Morgan Kaufman Publ.
White, J. E., 1983, Underground sound: Applications of seismic waves: Elsevier Science Publ. Co., Inc.
Wickerhauser, M. V., 1994, Adapted wavelet analysis from theory to software: A. K. Peters, Ltd.
Williams, D. M., 1990, The acoustic log hydrocarbon indicator: 31st Ann. Logging Symp., Soc. Prof. Well Log Analysts, Paper W.
Winkler, K. W., and Murphy, W. F., III, 1995, Acoustic velocity and attenuation in porous rocks, *in* Ahrens, T. J., Ed., Rock physics and phase relations: A handbook of physical constants: Am. Geophys. Union, 20–34.