PETROPHYSICS, VOL. 47, NO. 2 (APRIL 2006); P. 93–111; 14 FIGURES, 2 TABLES

New Approach for Solving Inverse Problems Encountered in Well-Logging and Geophysical Applications

R. Freedman[1]

ABSTRACT

This paper proposes a new method for solving inverse problems for which a large calibration database exists consisting of pairs of inputs and corresponding outputs. Distinct pairs of input and output data in the database correspond to different states of the underlying physical system whose properties are the subject of the inversion. Either empirical measurements or numerical computations, e.g., using a forward model, can be used to construct the database.

The method uses radial basis function (RBF) interpolation, a method for approximating smooth and continuous multivariate functions of many variables. RBF interpolation is used to derive a non-linear mapping function from which the properties of physical systems (e.g., crude oils, sandstones, etc.) are predicted from input measurements that are not in the database.

An advantage of this method over the traditional approach of fitting data to a forward model is that it can be used to solve well-logging and geophysical inverse problems associated with unknown forward models. In other cases where accurate forward models are known, the method can be used to solve inverse problems in real time, where these might otherwise be computationally too expensive or, as in many practical cases, lead to ill-behaved non-linear minimization. Construction of a robust database that spans the physical range of input and output data encountered in practice is essential for accurate predictions. Database construction is the most challenging part of applying the method.

The method is intuitive and does not require iterative training. It is easier to implement than inversion methods based on artificial neural networks, which use non-intuitive multilayered networks and require lengthy iterative training.

Testing and benchmarking of the inversion method is performed on three challenging example problems of current interest in well logging. Example 1 shows that viscosities of dead crude oils can be predicted from nuclear magnetic resonance (NMR) relaxation time measurements. The predicted viscosities are in good agreement with those measured by a Couette-type viscometer and are more accurate than viscosities predicted using published empirical correlations. The first example also shows that molecular compositions of dead crude oils can be predicted from NMR relaxation time measurements. The predicted compositions are in good overall agreement with those obtained from gas chromatography. The method used in this paper can also be applied to databases constructed from measurements made on live crude oils. Example 2 shows that borehole-corrected signals can be predicted from raw measurements made by a 3D induction tool. The predicted borehole-corrected signals are in excellent overall agreement with the target tool responses (i.e., for homogeneous media). Example 3 shows that mass density and molecular composition of dead crude oils can be predicted from near-infrared (NIR) spectra. These examples demonstrate that the methodology proposed in this paper is applicable to a large variety of problems encountered in well logging and geophysical applications.

Keywords: petrophysical interpretation; inversion methods; radial basis function

Manuscript received by the Editor June 1, 2005; revised manuscript received February 3, 2006.

[1]Schlumberger Oilfield Services, 125 Industrial Blvd., Sugar Land, TX 77478; E-mail: freedman1@slb.com

INTRODUCTION

The interpretation of well logging measurements often involves formulating and solving a mathematical inverse problem. A typical challenge is to predict the physical properties of some underlying physical system from a suite of measurements. Indeed, this is the sole reason for performing the measurements in the first place.

For example, measurements could be from a borehole logging tool or a suite of tools for which the underlying physical system is the porous, fluid-filled rock formation surrounding the borehole. In this case the physical properties predicted from the measurements might include porosity, permeability, fluid type and saturation, and bed thickness. For geophysical exploration, the measurements could be surface measurements of reflected seismic wave energy, as a function of wavelength, made at different receiver locations. In this case, the underlying physical system is the subsurface, consisting of layers of porous sediments. The subsurface properties of interest are the hydrocarbon-bearing layers, if any, and their subsurface depths and thicknesses.

The traditional approach used to solve inverse problems is to fit a theoretical or empirically derived forward model to the measurements (Tarantola, 2005). The forward model is a function of a set of model parameters that are either identical to or related to the physical properties of the underlying physical system. Selecting the values of the model parameters that minimize the difference between the actual measurements and those predicted by the forward model is assumed to solve the inverse problem. This basic assumption is itself riddled with difficulties and can lead to erroneous solutions because most well logging and geophysical inverse problems are ill-posed, i.e., the solutions are not unique.

The main cause of non-unique solutions to inverse problems is that the resulting system of equations is often underdetermined because there are more unknowns than independent measurements. Therefore, there are an infinite number of solutions that will fit the data. Moreover, for noisy data many inverse problems are unstable, i.e., small changes in the data caused by noise fluctuations can result in large changes in the solution. Imposing conditions to constrain the solution space is used to circumvent these problems. One commonly used approach is to add a regularization (i.e., smoothing) term to the objective function that is minimized when fitting the measurements to the forward model. Other conditions imposed on the solutions often include constraints or bounds on the permitted range of solutions. The regularization approach in conjunction with the constraints selects a particular smooth solution out of the infinite number of solutions that fit the data.

The traditional approach has other limitations and caveats that render it unsuitable for providing accurate solutions to many problems of interest. For example, petrophysical systems are typically so complex that accurate forward models are not available. A well-known example is Archie's equation, used to predict water saturations from electrical conductivity or resistivity measurements made on porous rocks. Archie's equation contains exponents and parameters that have been observed to have a wide range of values for reservoir rocks. The wide range of values is caused by the complexity of porous rocks and can result in water saturation predictions with large errors.

There are other well-logging problems for which reasonably accurate forward models can be derived from the laws of physics, but fitting the forward model to the measurements is often computationally too expensive for real-time applications. They include, for example, forward models for the responses of logging tools (electromagnetic, sonic, nuclear) deployed in boreholes that penetrate modeled earth formations of various descriptions, including formations that have dipping beds with different thicknesses and physical properties.

The purpose of this paper is to introduce a model-independent inversion method that overcomes many of the aforementioned problems. The new approach is based on data and does not require a forward model. Instead of non-linear optimization it uses radial basis function (RBF) interpolation, a numerical analysis method for approximating smooth and continuous multivariate functions that has been developed by applied mathematicians over the past 30 years. The book by M.D. Buhmann (2003) provides a thorough description of RBF interpolation. The mathematical foundation underlying RBF interpolation is part of modern approximation theory in the field of numerical analysis (Powell, 2001).

Another model-independent database approach for solving inverse problems uses artificial neural networks (ANN). However, the ANN approach has drawbacks. Its implementation is not intuitive and it requires lengthy, iterative training that is not guaranteed to converge to a solution. In the ANN literature, RBF neural networks are discussed. These are three-layer neural networks that use RBFs in the hidden layer (Haykin, 1999). They should not be confused with the RBF interpolation method discussed in this paper, i.e., the method discussed here is totally independent of neural networks.

This paper introduces RBF interpolation and discusses how it can be applied to the solution of well logging and other geophysical inverse problems. Then it goes on to demonstrate the application of the method to three important and challenging well-logging problems. The first example is the prediction of viscosity and the molecular

composition of crude oils from nuclear magnetic resonance (NMR) relaxation time measurements. The second example is the prediction of borehole-corrected signals from raw signals acquired with a 3D tri-axial induction tool. The third example is the prediction of the density and molecular composition of crude oils from near-infrared absorption measurements similar to those made by a formation-sample testing tool. It will be clear from the examples and from the theory that the method can be applied to the inversion of well-logging and geophysical measurements from any type of sensor (e.g., electromagnetic, acoustic, nuclear, mechanical, electromechanical, or thermal.)

RBF INTERPOLATION

Mathematical description

Let $\vec{f}(\vec{x})$, $\vec{x} \in R^n$, and $\vec{f} \in R^m$, be a real-valued vector function of n variables, and let values of $\vec{f}(\vec{x}_i) = \vec{y}_i$ be given at N distinct points, $\vec{x}_i$. The interpolation problem is to construct the function $\vec{F}(\vec{x})$ that approximates $\vec{f}(\vec{x})$ and satisfies the interpolation equations:

$$\vec{F}(\vec{x}_i) = \vec{y}_i\,, \quad i = 1,2,\ldots,N\,. \tag{1}$$

RBF interpolation chooses an approximating or mapping function of the form

$$\vec{F}(\vec{x}) = \sum_{i=1}^{N} \vec{c}_i \varphi(\|\vec{x} - \vec{x}_i\|)\,. \tag{2}$$

The non-linear function, φ, is known as an RBF. The argument of the RBF depends on the Euclidean norm in n-dimensional space; i.e.,

$$\|\vec{x} - \vec{x}_i\| = \sqrt{\sum_{m=1}^{n} (x_m - x_{i,m})^2}\,. \tag{3}$$

The functions are called "radial" because they depend only on the distance, not the direction, of $\vec{x}_i$ from an arbitrary input vector, $\vec{x}$, at which the function is to be evaluated.

The real coefficients, $\vec{c}_i$ in equation (2) are determined by requiring that the interpolation equations (equation (1)) be satisfied exactly. Therefore, the coefficients are a linear combination of the given function values,

$$\vec{c}_i = \sum_{j=1}^{N} \Phi^{-1}_{i,j} \vec{y}_j\,, \tag{4a}$$

where $\Phi_{i,j} \equiv \varphi(\|\vec{x}_i - \vec{x}_j\|)$ is the $N \times N$ interpolation matrix. If the condition number of the interpolation matrix is large and if the given function values are noisy, the coefficients determined from equation (4a) can be unstable and one should replace equation (4a) by the regularized solution (Haykin, 1999),

$$\vec{c}_i = \sum_{j=1}^{N} (\Phi + \gamma I)^{-1}_{i,j} \vec{y}_j\,, \tag{4b}$$

where I is the $N \times N$ identity matrix and γ is a real and positive regularization parameter.

One of the nice features of RBFs is that for certain functional forms, including Gaussian, multiquadric, and inverse multiquadric forms, mathematicians have proved that the interpolation matrix is nonsingular (e.g., Micchelli, 1986). This means that for these functions, the mapping function in equation (2) is uniquely determined. RBF interpolation has other attractive properties not possessed by classical interpolation schemes such as polynomial splines or finite difference approximations. First, RBF interpolation does not require the data to be on a uniform lattice and has been shown to provide good results with scattered datasets (Buhmann, 2003). Second, numerical experiments have shown the somewhat surprising result that for a given number of data points, N, the accuracy of the interpolation is independent of the number of independent variables, n, even for very large values of n (Powell, 2001).

Using RBF interpolation to solve inverse problems

The previous section introduced RBF interpolation from a purely mathematical point of view. In this section I will discuss how this modern method for approximating functions of many variables can be used to solve inverse problems.

Inverse problems encountered in well logging and geophysical applications involve predicting the physical properties of some underlying system given a set of measurements. Consider a database having a set of distinct input data $\vec{x}_i \in R^n$ (i.e., the inputs are n-dimensional vectors) and a set of corresponding real outputs, $\vec{y}_i \in R^m$, for $i = 1,\ldots,N$, where N is the number of cases in the database. The different cases in the database represent different states of the underlying physical system. In the mathematical language of RBF interpolation, $\vec{y}_i$ values represent samples of the function that we want to approximate, and $\vec{x}_i$ values are the distinct points at which the function is given. The database inputs, $\vec{x}_i$, represent the measurements from which we would like to predict the physical properties of the underlying system. The database outputs, $\vec{y}_i$, are the physical properties we want to predict from the measurements. The database is used to construct a mapping function such that,

given measurements $\vec{x}$ that are not in the database, one can predict the properties $\vec{F}(\vec{x})$ of the physical system that is consistent with the measurements. The mapping function solves the inverse problem by predicting the physical properties of the system from the measurements. Geometrically the mapping function is the hypersurface that defines the database outputs in terms of the inputs in multi-dimensional space.

It is useful to explain the construction of a database by giving an example that will be discussed in more detail in a subsequent section. The proposed problem is to predict the viscosity of dead crude oils from NMR measurements. NMR theory and experiments both show that viscosity can be predicted from NMR measurements of relaxation times and diffusion coefficients. However, the existing empirical correlations relating NMR measurements to viscosity can have large errors. RBF interpolation can be used to make more accurate predictions. The database is constructed by acquiring NMR measurements on a suite of N oil samples over a representative range of viscosities, temperatures, and pressures. The inputs to the database would be, for each oil sample, a vector whose elements are the NMR data (e.g., the amplitudes in a transverse-magnetization, relaxation-time distribution or in a diffusion-coefficient distribution) for that sample and the measurement temperature and pressure. The database outputs are the properties we wish to predict. In this example, we want to predict viscosity. Therefore, the database outputs are obtained by measurement, using a viscometer, of the viscosity for each oil sample at the same temperature and pressure as the NMR measurements. Using this database, a radial basis mapping function can be constructed so that oil viscosity can be predicted from NMR, temperature, and pressure measurements.

Clearly for this approach to work there must exist a deterministic, unique, and functional relationship between the input measurements and the properties one would like to predict from those measurements. For each $\vec{x}_i$ in the database there is a unique output, $\vec{y}_i$, as required for a single-valued function. This is obvious from the method of database construction. There can, however, be different input measurements in the database that correspond to the same database output. For example, two different crude oils in the database could have identical viscosities but different NMR T_2 distributions. The fact that this is still a functional relationship follows from the definition of a single-valued function, as learned in algebra, for which $\vec{x}$ is the independent variable and $\vec{F}(\vec{x})$ is the dependent variable or function. For each $\vec{x}$ there should be only one value of the function $\vec{F}(\vec{x})$, but different values of $\vec{x}$ can correspond to the same value of $\vec{F}(\vec{x})$. The RBF mapping function in equation (2) is a single-valued function of the inputs, $\vec{x}$, and, therefore, it predicts unique values for the outputs, $\vec{y}$. The method is capable of providing accurate results for smooth continuous multivariate mapping functions provided that the calibration database is sufficiently well populated. Unfortunately, there are no general rules or theorems for quantifying the accuracy of the predictions. The results are very much problem dependent, and one must rely on numerical or empirical testing to assess the accuracy of the predictions for a specific application and calibration database.

Normalized Gaussian RBFs

The RBFs used in this paper are normalized Gaussian RBFs defined by the equation:

$$\varphi(\|\vec{x}-\vec{x}_i\|)=\frac{\exp\left(-\dfrac{\|\vec{x}-\vec{x}_i\|^2}{2s_i^2}\right)}{\displaystyle\sum_{j=1}^{N}\exp\left(-\frac{\|\vec{x}-\vec{x}_j\|^2}{2s_j^2}\right)}. \tag{5}$$

As will become apparent when the example problems are discussed, one of the attractive features of using multivariate Gaussian functions is that they can be written as products of univariate functions. These functions are normalized in the sense that the summation over the database inputs, $\vec{x}_i$, is equal to unity for all $\vec{x}$, i.e.,

$$\sum_{i=1}^{N}\varphi(\|\vec{x}-\vec{x}_i\|)=1. \tag{6}$$

It is also easily seen from equation (5) that

$$0\le\varphi(\|\vec{x}-\vec{x}_i\|)\le 1. \tag{7}$$

By combining equations (2) and (5), one can write the mapping function for Gaussian RBFs as

$$\vec{F}(\vec{x})=\frac{\displaystyle\sum_{i=1}^{N}\vec{c}_i\exp\left(-\frac{\|\vec{x}-\vec{x}_i\|^2}{2s_i^2}\right)}{\displaystyle\sum_{i=1}^{N}\exp\left(-\frac{\|\vec{x}-\vec{x}_i\|^2}{2s_i^2}\right)} \tag{8}$$

The width, s_i, of the Gaussian RBF centered at $\vec{x}_i$ is representative of the range or spread of the function in the input space. The optimal widths, for accurate approximations, should be of the order of the nearest-neighbor (NN) distances in the input space. The idea is to pave the input space with basis functions that have some overlap with nearest neighbors but negligible overlap with more distant neigh-

bors. This ensures that for an input $\vec{x}$ that is not in the database, the output $\vec{F}(\vec{x})$ is computed as a weighted average of contributions from those database inputs $\vec{x}_i$ that are nearest to $\vec{x}$. The interpolation points in the database that are far from $\vec{x}$ make a negligible contribution to $\vec{F}(\vec{x})$. These remarks suggest a protocol for dealing with very large databases where N is of the order of 10^6 or larger. An acceptable estimate of $\vec{F}(\vec{x})$ can be obtained by considering a subset of the database values of $\vec{x}_i$ (e.g., a few hundred points) that are nearest to $\vec{x}$. This works because RBF interpolation is a local approximation. Thus, except for identifying the few hundred points, the cost of estimating $\vec{F}(\vec{x})$ is independent of N. The variance in $\vec{F}(\vec{x})$ caused by random uncorrelated measurement errors in $\vec{x}$ can also easily be computed from equation (8) (e.g., see Appendix A of Freedman et al., 1998). One finds that increasing the widths of the Gaussian functions reduces the variance in $\vec{F}(\vec{x})$. This is intuitive because increased widths mean that more interpolation points are used, and this leads to a more robust estimate of $\vec{F}(\vec{x})$.

Ideally the calibration database of input and output measurements should be acquired with negligible noise. In this paper the widths for the RBFs are determined heuristically by selecting widths comparable to NN Euclidean distances separating the input vectors. Once the widths are selected, the determination of optimal coefficient vectors in equation (8) is a linear problem (e.g., see equation (4)). The next section discusses an approximation for the coefficient vectors and provides some intuitive insight into how the radial basis mapping function interpolates in the database to predict the output vectors.

Nadaraya-Watson Regression Estimator

An intuitive understanding of how the mapping function in equation (8) predicts an output vector for an input vector not in the database can be gained by considering the Nadaraya-Watson Regression Estimator (NWRE). The NWRE is based on a simple approximation for the coefficient vectors (Haykin, 1999). The interpolation equations for the mapping function in equation (8) can be written in the form:

$$\vec{F}(\vec{x}_j) = \frac{\vec{c}_j + \sum_{\substack{i=1 \\ i\neq j}}^{N} \vec{c}_i \exp\left(-\frac{\left\|\vec{x}_j - \vec{x}_i\right\|^2}{2s_i^2}\right)}{1 + \sum_{\substack{i=1 \\ i\neq j}}^{N} \exp\left(-\frac{\left\|\vec{x}_j - \vec{x}_i\right\|^2}{2s_i^2}\right)}. \qquad (9)$$

The summations in equation (9) can be neglected if one neglects the overlap of the database RBFs. The NWRE approximation assumes that the interpolation matrix in equation (4) is diagonal and leads to a simple approximation for the coefficient vectors, such that

$$\vec{F}(x_j) \equiv \vec{y}_j = \vec{c}_j \ . \qquad (10)$$

This simple approximation replaces the coefficient vectors in equation (8) with the database output vectors $\vec{y}_j$. For many practical problems, the NWRE approximation works well and is always a good starting point for testing RBF predictions. Computing the coefficients using equation (4) provides a refinement to the approximation.

By combining equations (8) and (10), one obtains the NWRE mapping function:

$$\vec{F}(\vec{x}) = \frac{\sum_{i=1}^{N} \vec{y}_i \exp\left(-\frac{\left\|\vec{x} - \vec{x}_i\right\|^2}{2s_i^2}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{\left\|\vec{x} - \vec{x}_i\right\|^2}{2s_i^2}\right)}. \qquad (11)$$

Note that in the limit of very large s_i, $\vec{F}(\vec{x})$ approaches the sample mean of the database output vectors. In the limit of very small s_i, $\vec{F}(\vec{x})$ approaches the output vector $\vec{y}_j$ that corresponds to the database input vector $\vec{x}_j$ that is closest to $\vec{x}$. In general, $\vec{F}(\vec{x})$ is a weighted average of the database output vectors with RBF weighting factors determined by the closeness of $\vec{x}$ to the database input vectors. Note that the NWRE approximation in equation (11) does not satisfy the interpolation conditions of equation (1).

The NWRE approximation can be improved by determining optimal coefficient vectors such that the interpolation equations are satisfied. The problem is linear if the widths of the Gaussian RBFs are fixed. Interpolation conditions lead to a set of linear equations for the coefficient vectors, whose solution can be written in matrix form as,

$$C = \Phi^{-1} \cdot Y \ , \qquad (12)$$

where the $N \times m$ matrix, C, is given by

$$C = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,m} \\ c_{2,1} & c_{2,1} & \cdots & c_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ c_{N,1} & c_{N,2} & \cdots & c_{N,m} \end{bmatrix}, \qquad (13)$$

where the ith row of C is the transpose of the coefficient vector for the ith database case. That is, the first subscript on

each coefficient runs from 1 to N and denotes a particular database case and the second subscript denotes a particular element of the database output vectors and runs from 1 to m. The matrix , whose inverse appears in equation (12), is the $N \times N$ positive definite matrix of Gaussian RBFs, i.e.,

$$\Phi = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots & \varphi_{1,N} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots & \varphi_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_{N,1} & \varphi_{N,2} & \cdots & \varphi_{N,N} \end{bmatrix}, \tag{14}$$

where the matrix elements are the normalized Gaussian RBFs:

$$\varphi_{i,j} = \frac{\exp\left(-\dfrac{\left\|\vec{x}_i - \vec{x}_j\right\|^2}{2s_j^2}\right)}{\displaystyle\sum_{j=1}^{N} \exp\left(-\frac{\left\|\vec{x}_i - \vec{x}_j\right\|^2}{2s_j^2}\right)}. \tag{15}$$

The $N \times m$ matrix, Y, in equation (12) contains the database output vectors:

$$Y = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{N,1} & y_{N,2} & \cdots & y_{N,m} \end{bmatrix}. \tag{16}$$

Note that the ith row is the transpose of the database vector $\vec{y}_i$. The solution for the coefficients given in equations (12-16) improve the NWRE approximation by determining optimal coefficient vectors with the caveat of having fixed widths for the Gaussian RBFs.

APPLICATIONS TO WELL-LOGGING INVERSE PROBLEMS

This section provides examples of RBF interpolation applied to three important and challenging well-logging inversion problems. The first example addresses the prediction of crude oil viscosity and molecular composition from NMR measurements. It has obvious importance for downhole fluid analysis, which is a rapidly expanding area of well-logging technology. The second example addresses the prediction of borehole-corrected data values from raw signals acquired with a 3D tri-axial induction-logging tool. The third example addresses the prediction of fluid density and composition from near-infrared (NIR) absorption measurements like those made by NIR optical spectrometers used in fluid-sampling tools.

Predicting viscosity from NMR measurements

This is an example of an inversion problem for which an accurate forward model does not exist. It is instructive to first review the empirical correlations used by the existing methods (Freedman et al., 2001). The existing methods rely on the following empirical equations to estimate viscosity (η):

$$\eta = \frac{aT}{T_{2,LM} f(gor)}, \tag{17}$$

and

$$\eta = \frac{bT}{D_{LM}}. \tag{18}$$

Equations (17) and (18) provide estimates of viscosity from NMR measurements of the transverse magnetization relaxation time (T_2) and the diffusion coefficient distributions (D), respectively. In equation (17), the longitudinal magnetization relaxation time (T_1) distribution can be used in place of T_2. T is temperature in degrees Kelvin and $f(gor)$ in equation (17) is an empirically determined function of the gas/oil ratio (Freedman et al., 2001). In these equations the viscosity estimate is inversely proportional to the logarithmic mean of the distribution, i.e., $T_{2,LM}$ in equation (17) and D_{LM} in equation (18). Thus, the correlations do not account for the fact that the shapes of the distributions can affect the viscosity. Furthermore, the empirical constants, a and b, are determined from "best fits" to the assumed regression form for the equations. The empirical constants in these equations are not universally appropriate for all crude oils, and the variance in these constants can cause significant errors in estimated viscosities. Another shortcoming of these correlations is that they do not account for the effects of pressure (Winkler et al., 2004).

The prediction of viscosity from NMR measurements using RBF interpolation can be viewed as constructing the nonlinear mapping from a vector input (e.g., amplitudes in a T_1, T_2, or D distribution) to a scalar output (viscosity), given a suite of input-output examples. The following discussion uses T_2 distributions to illustrate the technique with the understanding that the same methodology works for D and T_1 distributions. Raw spin-echo measurements can also be used since the latter contain the same information as the quantities derived from them. Consider a database of N input-output pairs in which inputs for each live-oil sample consist of T_2-distribution amplitudes ($\vec{A}$), temperatures (T_i), pressures (P_i), gas/oil ratios (gor_i) and corresponding output viscosities (η_i). The viscosity of a crude oil sample not in the database can be predicted using the following RBF mapping function:

$$\eta = \frac{\sum_{j=1}^{N} c_j \exp\left(-\frac{\left\|\vec{A}-\vec{A}_j\right\|^2}{2s_j^2}\right)\cdot\exp\left(-\frac{(T-T_j)^2}{2s_T^2}\right)\cdot\exp\left(-\frac{(P-P_j)^2}{2s_P^2}\right)\cdot\exp\left(-\frac{(gor-gor_j)^2}{2s_g^2}\right)}{\sum_{j=1}^{N} \exp\left(-\frac{\left\|\vec{A}-\vec{A}_j\right\|^2}{2s_j^2}\right)\cdot\exp\left(-\frac{(T-T_j)^2}{2s_T^2}\right)\cdot\exp\left(-\frac{(P-P_j)^2}{2s_P^2}\right)\cdot\exp\left(-\frac{(gor-gor_j)^2}{2s_g^2}\right)} \quad (19)$$

The viscosity of a live crude oil that is not in the database can be predicted using equation (19) and the measured inputs $\vec{A}$, T, P, and *gor*. Note that the measured inputs can be separated in the RBFs because of the factorizability of Gaussian functions. For simplicity, a single width was assigned to the Gaussian functions containing T, P, and *gor*. A borehole fluid-sampling tool equipped with pressure and temperature sensors and an NMR sensor could provide the measurements needed to predict viscosity using equation (19). The gas/oil ratio is attainable if the sampling tool is also equipped with optical sensors for measuring the optical density of the crude oil as a function of wavelength in the NIR region (Fujisawa et al., 2003). The accuracy of the predictions made by equation (19) relies on a database with a large, diverse population of live-oil measurements. Note that only the T_2-distribution amplitudes, not the relaxation times, appear in equation (19). The mapping function does not depend on the relaxation times, provided that all of the amplitudes in the equation span the same range of T_2 values. The maximum amplitude of the T_2 distributions was normalized to unity prior to constructing the mapping function. This step insures that amplitude differences caused by hardware or software calibrations are removed by the normalization. Equation (19) is readily generalized if the database also includes other measurements. For example, if in addition to the T_2-distribution amplitudes the database includes D and T_1 distributions, the amplitude vectors for these measurements would appear as additional Gaussian factors in equation (19).

Viscosity prediction using equation (19) was tested using a small database of T_2 distributions and corresponding measured viscosities acquired on a suite of 16 dead (i.e., *gor* = 0) crude oil samples at a temperature of 30°C at atmospheric pressure. Measured viscosities are shown in Table 1 and measured T_2 distributions are shown in Figure 1. The viscosity was measured for each sample using a laboratory Brookfield Couette-type viscometer with an absolute accuracy of better than five percent.

Observe that although samples 10 and 13 in Figure 1 have virtually identical viscosities their T_2 distributions have quite different log mean values. Note that sample

TABLE 1 Measured Viscosities of 16 Dead Crude Oils

Sample	Measured Viscosity (cp)
1	6.30
2	13.9
3	16.4
4	746.0
5	5.09
6	6.36
7	6.52
8	656.0
9	5.42
10	8.47
11	980.0
12	15.5
13	8.46
14	116.0
15	132.0
16	91.2

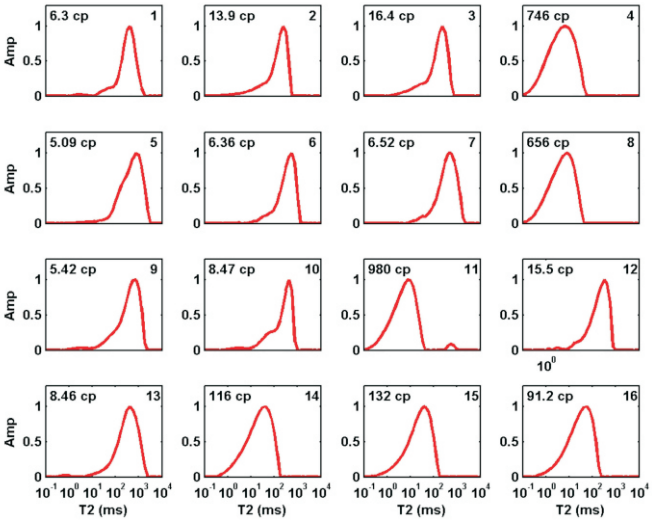

FIG. 1 Database consisting of T_2 distributions and measured viscosities for 16 dead crude oils.

numbers are shown in the upper right corner on each plot. Specifically, sample 10 has a log mean value of 174 ms, whereas sample 13 has a log mean value of 279 ms. This clearly shows the inadequacy of the empirical correlation in equation (17), which would predict viscosities that differ by a factor of 1.6. Figure 2 shows the viscosities predicted from the T_2 distributions using the radial basis mapping function,

$$\eta = \frac{\sum_{j=1}^{15} c_j \exp\left(-\frac{\left\|\vec{A}-\vec{A}_j\right\|^2}{2s_j^2}\right)}{\sum_{j=1}^{15} \exp\left(-\frac{\left\|\vec{A}-\vec{A}_j\right\|^2}{2s_j^2}\right)}, \qquad (20)$$

appropriate for a database of 16 dead oils all measured at the same temperature and pressure. The results in Figure 2 were obtained using the "leave-one-out method" whereby each sample was removed from the database and its viscosity was predicted from the remaining 15 samples in the database. The viscosities predicted from the T_2 distributions using equation (20) are in very good agreement with the viscosities measured in the laboratory. In my view, this result is impressive given the relatively small size of the database. The predicted viscosities shown in Figure 2 were obtained by choosing fixed widths for the Gaussian functions, i.e., s_j = 0.1, 0.5, 1.0, and 5.0. For this database, the predicted viscosities are accurate and show little sensitivity to the width for fixed values selected in the range from about 0.05 to 0.85.

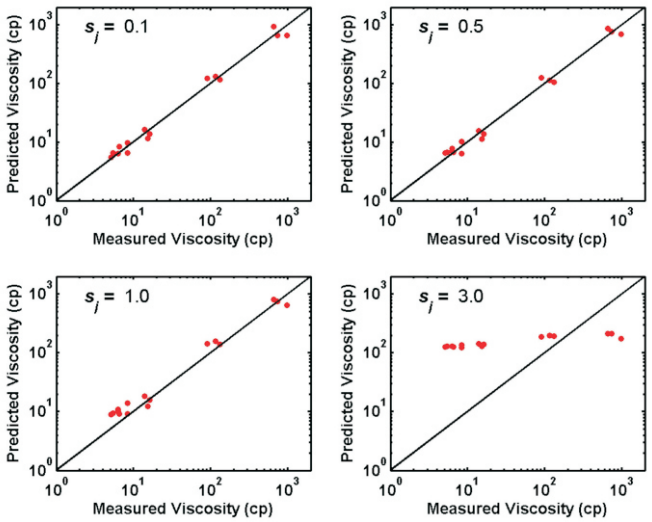

FIG. 2 Viscosities predicted from NMR T_2 distributions compared to laboratory-measured viscosities. The predicted viscosities in the four subplots above are computed using the values shown on the graphs for the widths of the Gaussian functions. The predicted viscosities are accurate and insensitive to the width of the Gaussian functions for a range of values from about 0.05 to 0.85.

TABLE 2 Nearest Neighbor (NN) Distances

Sample No.	NN Distances
1	0.39
2	0.24
3	0.24
4	0.35
5	0.58
6	0.39
7	0.31
8	0.31
9	0.40
10	0.65
11	0.31
12	0.66
13	0.31
14	0.26
15	0.26
16	0.46

As seen in Figure 2, if the widths are increased beyond 0.85 then the errors in the predicted viscosities start to increase gradually and eventually, as expected, all of the predicted viscosities approach the mean of the database output values. These observations are consistent with the intuitive notion that the optimal widths should be of the order of the NN distances in the input space. Table 2 shows the Euclidean NN distances for the T_2 distributions in the database. The NN distances provide a good starting point for choosing the widths. In this example, choosing the Gaussian widths to be of the order of the NN distance provides accurate predictions. The coefficients in equation (20) were computed using two methods: (1) the NWRE approximation was used to replace the coefficients with the database outputs (i.e., the laboratory-measured viscosities) and (2) the coefficients were computed from the interpolation equations. The NWRE approximation provided good results; in this case, very little if any improvement was achieved by computing the coefficients.

The viscosities of the samples in Table 1 were also estimated from the log mean values of the T_2 distributions shown in Figure 1. A value of a = 9.558 was used for the correlation parameter in equation (17). In the literature, a range of a values (from 4.0 to 9.558) has been cited for crude oils. This large spread in correlation parameters leads

to uncertainties with a factor of two for viscosities predicted from the log mean values of crude-oil T_2 distributions. Figure 3 shows viscosity values predicted from the log mean values compared to the measured viscosity values. One can see from Figures 2 and 3 that the RBF predicted viscosities are in much better agreement with the measured viscosities than are those predicted from the log mean values. The agreement between predicted and measured viscosities for the lower viscosity range in Figure 3 can be improved by reducing the value of the parameter, a, used in equation (17). However, this causes poorer agreements at the higher-viscosity range. The real advantage of the RBF method is that there are no model-dependent parameters and that the method lends itself to measurements made on live oils at elevated temperatures and pressures.

Predicting the molecular composition of crude oils from NMR measurements

It is well established that NMR measurements of relaxation time and diffusion coefficient distributions provide information about the molecular composition of crude oils (Freedman et al., 2001). Previous work on predicting composition from NMR measurements attempted to derive physics-based models that relate composition to relaxation time and molecular diffusion coefficient distributions (Freed, 2004; Heaton and Freedman, 2005). The derivation of accurate physics-based models is difficult because crude oils are complex mixtures containing hydrocarbon molecules with a wide range of shapes, weights, and sizes. This problem is exactly the kind of problem for which RBF interpolation offers an elegant model-independent method of solution. The prediction of composition was tested on the same database of 16 dead crude oils used above for the prediction of viscosity. For the prediction of composition, the database inputs are the T_2 distributions shown in Figure 1, and the outputs are the corresponding compositions, which were measured in the laboratory using gas chromatography.

The radial basis mapping function for predicting crude-oil molecular composition is similar to the one used for predicting viscosity except that the mapping function and the coefficients are both m-dimensional vectors, where m is the number of constituents in the compositions measured by the gas chromatography. The compositions measured for the database oil samples consisted of the molar fractions for carbon numbers C-1 through C-29 and C-30+. The compositions predicted from the T_2 distributions using RBFs are shown in Figure 4. The NWRE approximation was used for the coefficient vectors. Predictions were essentially insensitive to the widths of the RBFs over a range of width values from approximately 0.1 to 1.0. Note the good agreement between the predicted compositions and those measured by gas chromatography. This agreement is especially impressive because of the relatively small size of the database. Composition was predicted for each of the 16 samples by sequentially removing that sample from the database and then using the remaining 15 samples to construct the mapping function. The poorer predictions for the three most viscous samples in the database are

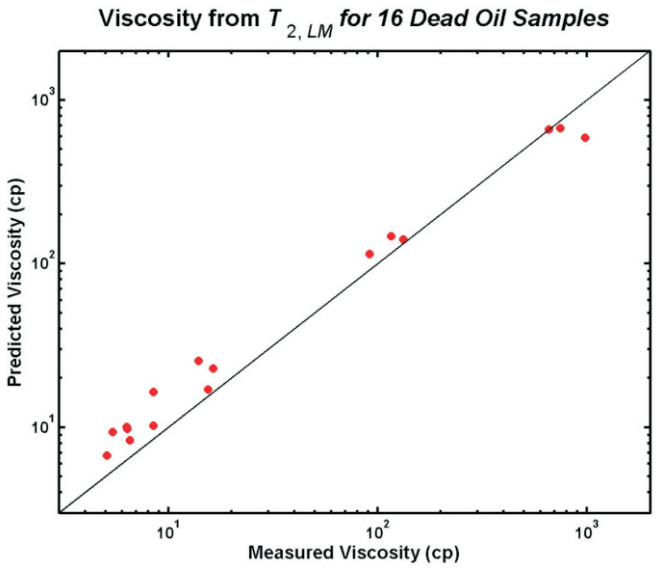

FIG. 3 Viscosities predicted from the log mean value of the NMR T_2 distributions using equation (17) compared to laboratory-measured viscosities.

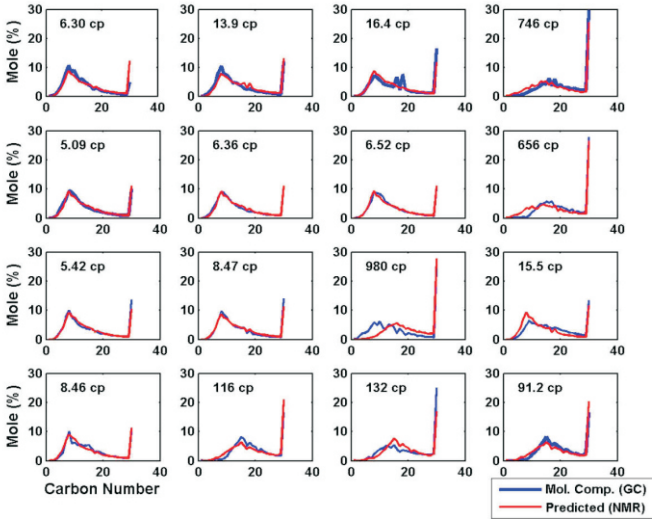

FIG. 4 Comparison of compositions predicted from T_2 distributions using RBFs with those measured in the laboratory using gas chromatography (GC).

explained by the fact that there are only three viscous oils in the database.

Predicting borehole corrections for a 3D induction tool

The previous sections showed how RBF interpolation could be applied to predict reservoir fluid properties from NMR measurements. The following sections show how RBF interpolation could be applied for predicting borehole corrections for measurements acquired with a 3D induction tool.

Description of the 3D induction tool borehole-correction problem

This application of RBFs shows how they can be used to correct the 3D tri-axial induction tool measurements for borehole effects. The same approach can be used to correct other types of open- and cased hole logging tool measurements (e.g., resistivity, sonic, nuclear) for borehole effects.

The 3D induction tool provides an especially good example of the power of RBF interpolation as a tool for solving difficult inverse problems because of the complexity and magnitude of the borehole effects on the tool response. The 3D induction tool acquires 234 signals at each sampling interval in the borehole; this large number of signals increases the degree of difficulty for borehole-correction problems.

Transverse magnetic dipole (TMD) transmitter coils induce axial borehole currents that produce receiver signals with very large borehole effects. This is why correcting the 3D induction tool signals for borehole effects is challenging. TMD transmitter coils can excite long-range longitudinal (i.e., axial) currents in the borehole and these can inductively couple strongly with the receiver coils. By contrast, conventional induction tools have longitudinally-oriented transmitter and receiver coils that produce borehole and formation currents that flow in planes transverse to the axial or borehole direction. Therefore, the receiver signals they excite have smaller borehole signals by comparison.

The 3D induction tool is designed to exhibit azimuthal, radial, and axial sensitivity so that the measured signals are sensitive to the conductivity anisotropy and the radial and axial conductivity distributions of the earth formations penetrated by the borehole. An experimental version of the 3D induction tool was discussed at a recent SPE meeting (Barber et al., 2004). In a Cartesian coordinate system fixed in the logging sonde, the axial direction is parallel to the direction of the sonde axis (i.e., parallel to the borehole). Multiple depths of investigation are achieved by employing nine receivers spatially separated in the axial direction from the transmitter. Each antenna coil has an associated magnetic-dipole-moment vector whose magnitude is proportional to the product of the cross-sectional area of the coil and the amplitude of the electrical current in the coil. The direction of the magnetic-moment vector is normal to the plane of the coil. Directional sensitivity to the formation conductivity distribution is achieved with antennas whose magnetic dipole moments are oriented in both the longitudinal (i.e., axial direction) and transverse directions.

The nine transmitter-receiver spacings included in the 3D induction tool are 6, 9, 12, 15, 21, 27, 39, 54, and 72 inches. The transmitter consists of a tri-axial antenna (e.g., solenoidal coils with dipole moments in the longitudinal and two orthogonal transverse directions). The short-spacing receivers, located at 6, 9, and 12 inches from the transmitter, each have a single longitudinally oriented antenna, whereas the six remaining receivers are tri-axial. The short-spacing receivers acquire signals at a single frequency (26 kHz), whereas the six other receivers acquire signals at two frequencies (13 and 26 kHz). Each of the receivers in the 3D induction tool is a mutually balanced pair. In induction logging tools, the transmitter is energized by an alternating current that induces alternating currents in the conductive formation and the borehole surrounding the logging sonde. The currents induce voltages in the receiver coils that are in-phase (i.e., resistive) and 90° out-of-phase (i.e., reactive) with respect to the transmitter current. The in-phase component is called the R-signal and the out-of-phase component is called the X-signal. A phase-sensitive detector is used to measure both components. A complex or phasor voltage can represent the R- and X-channel signals. However, this paper considers only real signals. Therefore, the number of distinct measured signals is doubled compared to the number of complex signals. The set of measured receiver voltages are sensitive to formation conductivity anisotropy, to the radial and axial distributions of formation conductivity, and to the borehole signal.

The borehole signal for each data channel depends in a nonlinear and complex fashion on numerous quantities including: borehole radius, mud conductivity, near-wellbore formation conductivity, formation conductivity anisotropy factor, and tool position or standoff for an eccentered tool. Borehole effects may also depend on the direction of the anisotropy in a dipping formation or deviated wellbore.

If all of the aforementioned parameters upon which the borehole signal depends were known during logging operations, a forward model consisting of a formation penetrated by a borehole could be used to invert the 3D induction tool raw measurements and determine the formation electrical properties. This approach is not practical because some of the parameters upon which the borehole effect depends are typically either not known (e.g., conductivity anisotropy, standoff) or only known approximately. Alternatively, an

inversion can be used to determine both borehole and formation properties. The latter approach is not practical because the computations are too elaborate to be performed in real time during logging operations.

It will be shown that the RBF interpolation can be used to correct the 3D induction tool raw tool measurements for borehole effects without having knowledge of the borehole parameters. In the process, it will be shown that a very difficult and possibly unstable inverse problem can be reduced to a relatively simple estimation problem.

Construction of the database and RBF mapping function

The 3D induction tool measures 234 raw voltages that are induced in the receiver coils. After applying downhole electronic calibrations and gain corrections, the measured voltages are converted to fully calibrated raw apparent conductivities. It is the measured, raw, apparent conductivities that must be borehole corrected before further data processing is performed to estimate formation conductivity distributions.

A database of raw (i.e., not corrected for borehole effects) apparent conductivities, $\vec{\sigma}_{a,i}^{(raw)}$ for $i = 1,2,\ldots,N$, can be computed by solving Maxwell's equations for the 3D induction tool in boreholes with radii (r_i), standoffs ($\vec{d}_i$), and mud conductivities ($\sigma_{m,i}$) which penetrate formations with vertical ($\sigma_{v,i}$) and horizontal ($\sigma_{h,i}$) conductivities. Note that the database includes tool eccentering (standoff) in different directions. The modeled database cases are for infinitely homogeneous and "transversely anisotropic" media for which the horizontal conductivities in the planes perpendicular to the borehole are different from the vertical conductivities in the planes parallel to the borehole. The anisotropy parameter for a transversely anisotropic formation is defined by

$$\lambda_a = \sqrt{\frac{\sigma_h}{\sigma_v}}\,. \tag{21}$$

The components of the raw apparent conductivity vectors, $\vec{\sigma}_{a,i}^{(raw)}$, are the *R*- and *X*-channel apparent conductivities determined either from all or from a subset of the 234 transmitter-receiver couplings measured by the tool. The *N* database cases correspond to choosing different values for the three borehole parameters and two formation parameters in the model. These parameters should each be selected to span a set of physically feasible values.

Borehole-corrected apparent conductivities, $\vec{\sigma}_{a,i}^{(c)}$, for $i = 1,2,\ldots,N$, can be computed by solving Maxwell's equations for the 3D induction tool in an infinite transversely anisotropic homogeneous medium (e.g., without a borehole). From the raw and borehole-corrected apparent conductivities one can compute a borehole correction, $\Delta\vec{\sigma}_{a,i}$, which by definition is given by

$$\Delta\vec{\sigma}_{a,i} \equiv \vec{\sigma}_{a,i}^{(raw)} - \vec{\sigma}_{a,i}^{(c)}\,. \tag{22}$$

The RBF mapping function for predicting the 234 borehole-corrected 3D induction signals can be written in the form,

$$\vec{\sigma}_{a,i}^{(c)} = \frac{\displaystyle\sum_{j=1}^{N} \vec{c}_i \exp\left(-\frac{\left\|\vec{\sigma}_a^{(raw)} - \vec{\sigma}_{a,i}^{(raw)}\right\|^2}{2s_i^2}\right)}{\displaystyle\sum_{i=1}^{N} \exp\left(-\frac{\left\|\vec{\sigma}_a^{(raw)} - \vec{\sigma}_{a,i}^{(raw)}\right\|^2}{2s_i^2}\right)}\,. \tag{23}$$

The coefficient vectors, $\vec{c}_i$, can be determined from equations similar to equations (12) through (16). Equation (23) can be used to predict borehole-corrected data directly from the raw 3D induction signals without having to know the six parameters on which the borehole effect depends (i.e., borehole radius, conductivity anisotropy, standoff, direction of standoff, borehole fluid conductivity, and near-wellbore formation conductivity). If desired, one can also use the database to construct an RBF mapping function to predict the aforementioned parameters from the raw tool measurements (or likewise, construct a mapping function to predict the borehole correction defined in equation 22). However, the beauty of the RBF approach for solving the borehole correction problem is that knowledge of these parameters is not required in the estimation.

Testing the borehole corrections using a database

To demonstrate use of RBF interpolation to borehole correct raw 3D induction tool measurements, I constructed a sparsely populated database of 3D raw induction tool responses for the 234 data channels. The database was constructed using five values of σ_m, six values of σ_h, three values of anisotropy parameter (λ_a), four borehole radii, and three values of tool standoff. The standoff direction for the database is parallel to the *x*-direction. Raw tool responses represent the database inputs and the anisotropy parameter is defined in equation (21). Induction tool responses for the 234 data channels were also computed for the same formation parameters without the presence of a borehole. These represent the homogeneous medium responses and are the database outputs. After removing duplicate cases and cases that were not considered to be within the physical limits of the tool or formation parameters, I tested the reduced database, which consisted of 916 cases. Each test involved removing a sample (i.e., a case) from the database and then

using equation (23) to predict the borehole-corrected responses for the 234 data channels of the 3D induction tool. The coefficients used in equation (23) were computed from the database interpolation equations. Widths (s_j) used in equation (23) were computed from the nearest neighbor distances (NN_j) of the input measurements. For the results shown here, widths were heuristically computed using the equation

$$s_j = 0.2 * NN_j \ . \tag{24}$$

Borehole-corrected data predicted from equation (23) were compared to the homogeneous media (i.e., the database formations without a borehole) tool responses for all 234 data channels. In the vast majority of cases, the agreement between the predicted borehole corrections and the target (i.e., the homogeneous medium responses) responses was within four or more decimal places. The nearest neighbor distances for the 916 raw input measurements varied widely from a minimum distance of 0.32 mmhos to a maximum distance of 1241 mmhos. This is clearly a case where using a single width for the RBF widths would be incorrect. For the raw-measurement input vectors in equation (23), only a subset consisting of 177 of the 234 data channels was used. The 117 *R*-channel signals were all used, but only 60 *X*-channel signals were used in the estimation. This is a nice feature of the method, because it allows one to exclude data that might be corrupted by noise or other factors. In the examples shown here, input raw measurements are assumed accurate. However, some of the data were excluded to emphasize the point that all of the 3D induction tool raw measurement channels are not needed to perform the borehole corrections.

Some typical results are shown in Figures 5 through 11 for a few cases picked at random from the database. The raw and borehole-corrected conductivities in the plots are in units of mmhos. Note that the 117 *R*-channel results are shown in the plots except for Figure 8 in which *X*-channel results are shown. The *X*-channel results are not displayed to save space because wherever the *R*-channel borehole corrections agree with their target responses, the *X*-channel responses agree equally well with their target responses (e.g., Figures 7 and 8). Note that in Figures 5-7 and 9-11, the nonzero values of the 117 borehole-corrected apparent conductivities are the 39 diagonal (e.g., σ_{xx}, σ_{yy}, σ_{zz}) *R*-channel couplings. In Figure 8, the nonzero borehole-corrected apparent conductivities are the 39 diagonal *X*-channel tool couplings. The off-diagonal couplings are all correctly predicted to be zero, as expected for the tool response in a homogeneous medium. In general, off-diagonal couplings are nonzero only if azimuthal symmetry is broken, for example, by vertical fractures, dipping beds, or tool standoff.

One of the examples, shown in Figure 10, provides insight into how RBF interpolation would fail if the underlying database were not adequately populated. Database sample no. 25 represents one of the extreme cases in the

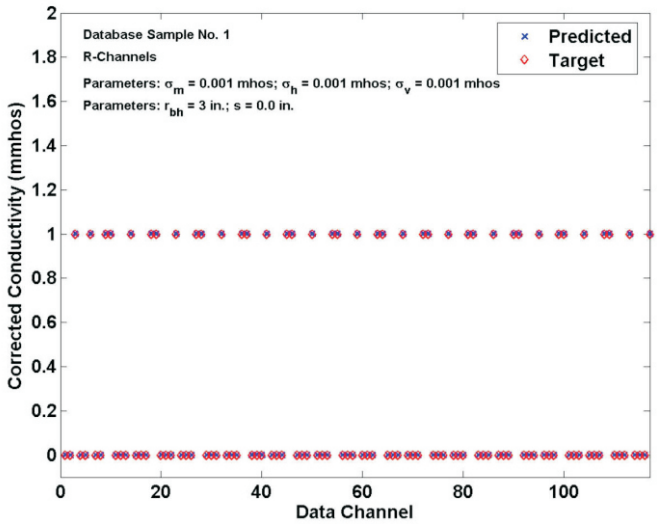

FIG. 5 Comparison of the 117 predicted borehole-corrected *R*-channels to the target (no borehole) tool responses for database sample no. 1. The predicted (x-marks) and target (diamonds) responses agree to four or more decimal places and cannot be distinguished from one another in the plot. Only the 39 diagonal *R*-channel couplings are nonzero as expected.

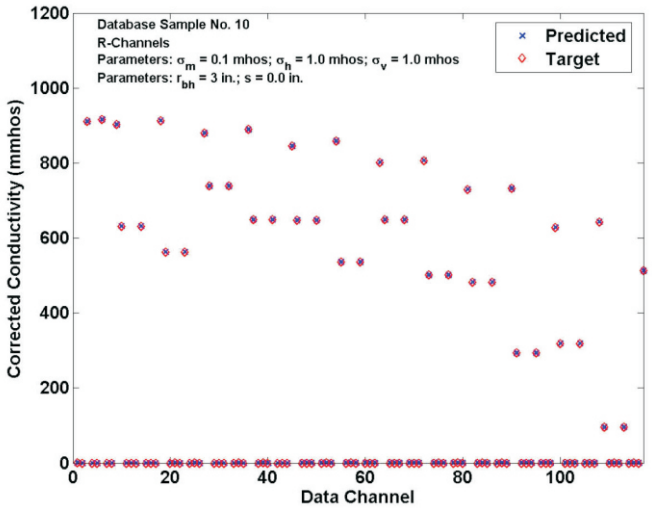

FIG. 6 Comparison of the 117 predicted borehole-corrected *R*-channels with the target (no borehole) tool responses for database sample no. 10. The predicted (x-marks) and target (diamonds) responses agree to four or more decimal places and cannot be distinguished from one another in the plot. Only the 39 diagonal *R*-channel couplings are nonzero as expected.

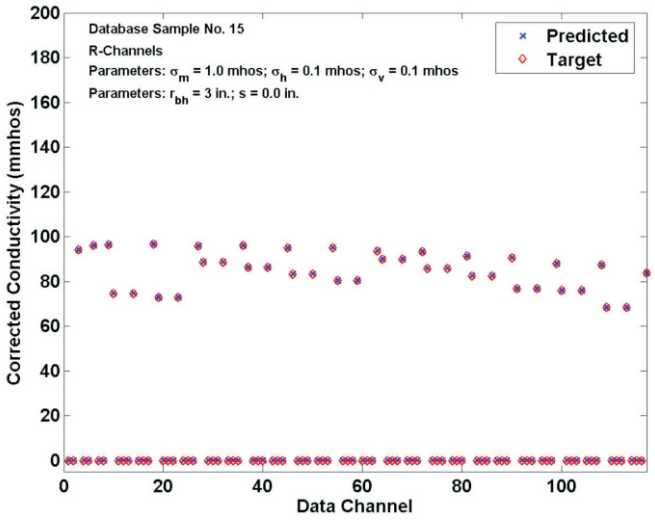

FIG. 7 Comparison of the 117 predicted borehole-corrected *R*-channels with the target (no borehole) tool responses for database sample no. 15. The predicted (x-marks) and target (diamonds) responses agree to four or more decimal places and cannot be distinguished from one another in the plot. Only the 39 diagonal *R*-channel couplings are nonzero as expected.

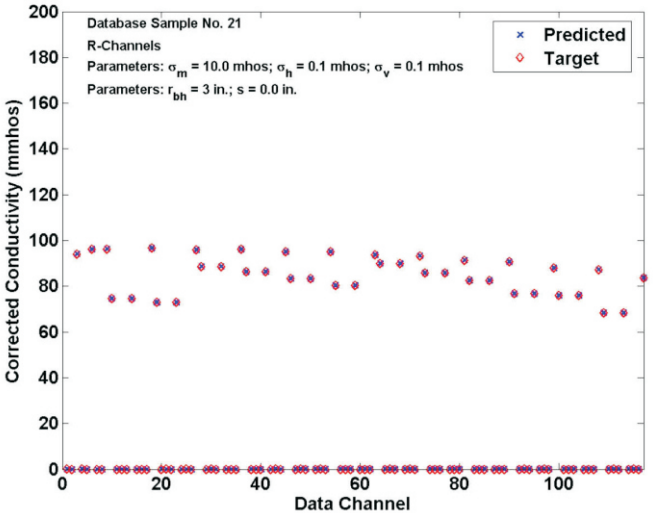

FIG. 9 Comparison of the 117 predicted borehole-corrected *X*-channels with the target (no borehole) tool responses for database sample no. 21. The predicted (x-marks) and target (diamonds) responses agree to 4 or more decimal places and cannot be distinguished from one another in the plot. Only the 39 diagonal *R*-channel couplings are nonzero as expected.

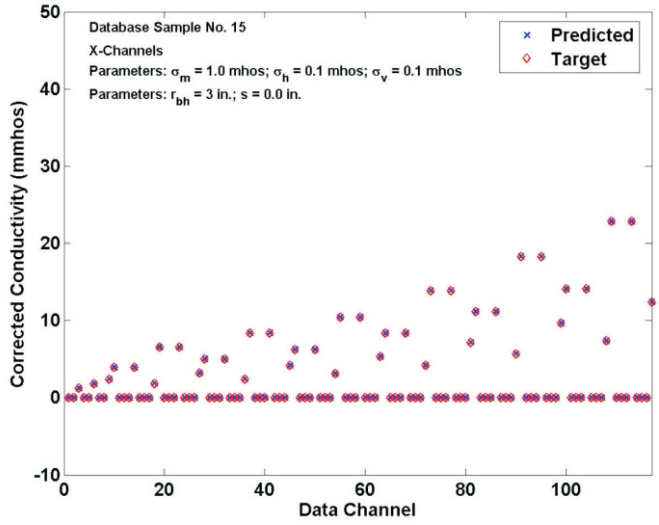

FIG. 8 Comparison of the 117 predicted borehole-corrected *X*-channels with the target (no borehole) tool responses for database sample no. 15. The predicted (x-marks) and target (diamonds) responses agree to four or more decimal places and cannot be distinguished from one another in the plot. Only the 39 diagonal *X*-channel couplings are nonzero as expected.

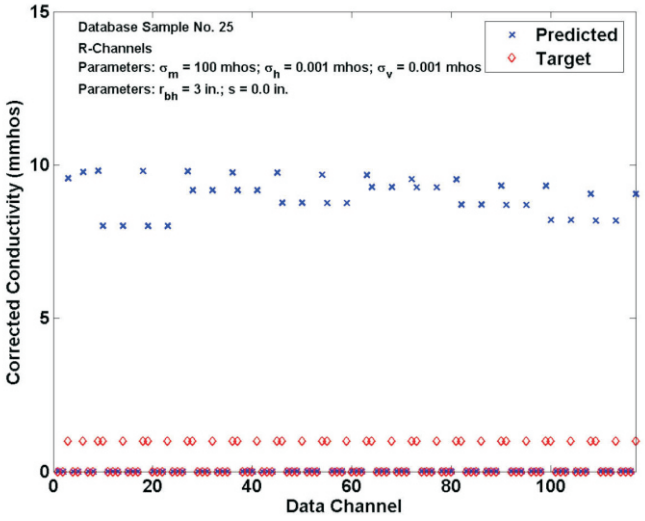

FIG. 10 Comparison of the 117 predicted borehole-corrected *R*-channels with the target (no borehole) tool responses for database sample no. 25. The predicted diagonal *R*-channel couplings do not agree with the target responses because the database is too sparsely populated as discussed in the text. This can be corrected by adding a few samples to the database that are similar to sample 25. For example, these would have borehole parameters identical to sample no. 25 and formation conductivity values close to and both above and below 0.001 mhos.

database. For this case, the borehole and formation parameters are: σ_m = 100 mhos, σ_h = 0.001 mhos, σ_v = 0.001 mhos, and the borehole radius is 3 inches. The tool was centered in the borehole (i.e., $s = 0$). To compute the borehole-corrected tool responses, case no. 25 was removed from the database and the remaining 915 samples in the database were used in equation (23) to predict the borehole corrections. The "nearest" sample in the database is sample no. 26 which has identical borehole parameters to sample no. 25 but has different formation properties; e.g., σ_h = 0.01 mhos and σ_v = 0.01 mhos. As expected, the borehole-corrected predictions for sample no. 25 are close to 0.01 mhos because sample no. 26 makes the dominant contribution to the summation in equation 23. To correct this problem, one must increase the number of samples in the database. The additional samples would have the same borehole parameters as sample no. 25 but would include samples with lower formation conductivities; e.g., samples with conductivity values close to and above and below 0.001 mhos.

Figure 11 shows estimation results for a database case with a large value for tool standoff parallel to the x-direction. When the standoff is in the x-direction, raw tool measurements have nonzero values for the off-diagonal couplings, σ_{xz} and σ_{zx} (the other four off-diagonal couplings are zero). For the example shown in Figure 11, only the 39 diagonal R-channel couplings are nonzero, all the off-diagonal couplings are zero for the borehole-corrected data, as they should be for an anisotropic, homogeneous medium without a borehole. This last example is particularly impressive because of the accurate removal of the large borehole effects on the off-diagonal raw measurements.

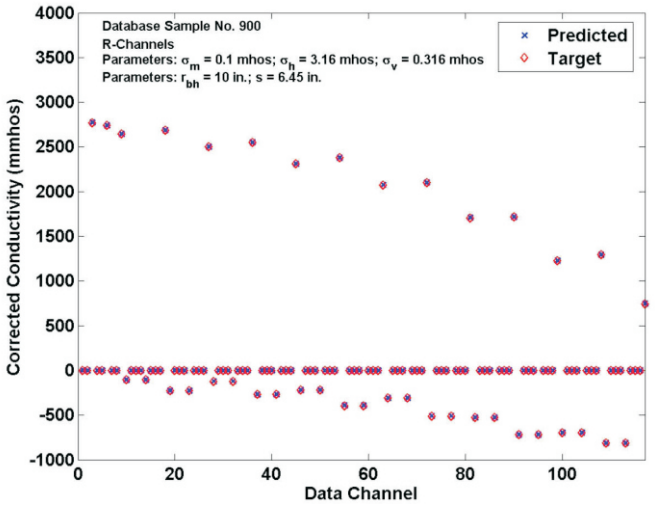

FIG. 11 Comparison of the 117 predicted borehole-corrected R-channels with the target (no borehole) tool responses for database sample no. 900. The excellent agreement with sample no. 900 for which there is a large standoff is impressive because of the large borehole effects on the raw measurements. The large standoff, which is in the x-direction for the database cases, produces large values of raw off-diagonal couplings, i.e., σ_{xz} and σ_{zx}. The nonzero borehole-corrected responses are the 39 R-channel diagonal couplings as expected. The large off-diagonal couplings on the raw measurements have been removed from the borehole-corrected signals as required.

Predicting density and composition from NIR absorption measurements

This section considers mass density and molecular compositional information predicted from NIR absorption measurements on live crude oils. The database consisted of 14 live crude oils on which NIR absorption versus wavelength, gas/oil ratios, gas chromatography composition from C-1 to C-7+, and laboratory density measurements were available. Unfortunately, the temperatures and pressures at which the NIR measurements were performed were not available. Although this lack of data somewhat compromises the analysis, results are nevertheless encouraging. One of the motivations for applying RBF interpolation to the inversion of NIR data is that current inversion methods are based on Beer's law, which is not strictly applicable to fluid mixtures. The hope is that it might be possible to extract more and better fluid property information from NIR data using RBF interpolation than is possible with conventional methods.

Standard chemometric models and solutions (e.g., principal component analysis) based on Beer's law are used to predict compositions and gas/oil ratios of reservoir fluids from NIR absorption measurements made by formation fluid sampling tools. This application has been discussed in numerous papers. A recently published Society of Petroleum Engineers (SPE) paper on this subject (Fujisawa et al., 2003) discussed the application of principal component analysis to invert models based on Beer's law and shows that weight percentages of molecular species with carbon numbers C-1 (methane), C-2 to C-5 (ethane through pentane), C-6+ (hexane plus heavier components) and CO_2 (carbon dioxide) could be predicted from downhole NIR absorption measurements.

It is instructive to discuss Beer's law and its limitations. Let wide band NIR radiation be incident upon a fluid sample. The absorbance (also known as the optical density) at wavelength, λ, of the sample is given by

$$A(\lambda) \equiv -\log_{10}\left(\frac{I(\lambda)}{I_o(\lambda)}\right) = \alpha \cdot c \cdot l, \qquad (25)$$

where I_o and I are the incident and transmitted power of a beam of radiation that has traversed l centimeters of a homo-

geneous absorbing medium that contains c moles per liter of an absorbing substance with molar absorptivity equal to α. The equality on the right-hand side of equation (25) is known as Beer's law.

Note that the optical density is always ≥ 0. An optical density equal to zero indicates that the sample is totally nonabsorbing. An optical density equal to 1.0 indicates that 90% of the incident energy is absorbed, 2.0 means that 99% is absorbed, etc. NIR spectrometers usually measure the absorbance as a function of wavelength in the range from approximately 700 to 2,500 nm. Some spectrometers measure the reflected radiation.

The chemistry textbook by Skoog and West (1976) presents a clear derivation and discussion of Beer's law and its limitations. Chemometric models for predicting composition of mixtures that are based on Beer's law assume that the absorption of radiation by a complex mixture is equal to the sum of the absorptions that would be measured on the individual constituents of the mixture (Heise and Winzen, 2002). This assumption ignores the fact that the absorption measured in a mixture can differ from the sum of the individual component absorptions because molecular interactions between different constituents in the mixture are non-linear.

NIR spectra of hydrocarbons

NIR radiation occupies that part of the electromagnetic (EM) spectrum in the wavelength range from approximately 800 nanometers (nm) to 2,500 nm. The adjacent region of the spectrum at longer wavelengths from approximately 2,500 nm to 25,000 nm is called the mid-infrared (MIR) range. The energy of the radiation in the MIR part of the EM spectrum corresponds to fundamental mode excitations of molecular vibrations. Fundamental mode excitations are those from the ground state to the first excited level. For a polyatomic molecule with N atoms there are 3N-6 (3N-5 for a linear molecule) normal modes or fundamental vibration frequencies corresponding to various stretching, bending, and rocking modes.

The absorption peaks observed in the NIR part of the spectrum for reservoir fluids correspond to excitations from the ground state to the second or third excited states. These absorptions are called first and second overtones. The second overtone mode has lower probability of excitation and, therefore, is less intense than the first overtone mode. Other absorption peaks known as combination modes are also observed in the NIR region. Combinations are usually caused by the excitation of a C-H stretching mode plus one or more bendings or rocking modes. Many different combination modes can be excited which contribute to the richness and complexity of the NIR region.

In the NIR region, the absorption spectra of hydrocarbons are broader and have lower optical densities (i.e., are more transparent) than in the MIR region. As a result, the association of peaks and structure in an NIR-absorption spectra with specific functional groups or molecules is less straightforward than for MIR spectra.

A paper by L. G. Weyer (1985) discusses, among other things, the NIR spectra of aliphatic and aromatic hydrocarbons. It is useful to summarize briefly some of the information in this paper. The first overtones of the C-H stretch mode for aliphatic hydrocarbons are found in the wavelength range from approximately 1,600 to 1,800 nm, and the second overtones are in the range from approximately 1,100 to 1,250 nm. The first combination bands for aliphatic hydrocarbons are observed in the wavelength range between 2,000 and 2,400 nm, and a second weaker band is observed from approximately 1,300 to 1,450 nm. For aromatic hydrocarbons the first overtone of the C-H stretch mode occurs at approximately 1,685 nm, and the second overtone occurs at approximately 1,143 nm. Combination bands are found at approximately 2150 and 2460 nm with much weaker bands at 1,420 to 1,450 nm. Water has an O-H stretch first overtone at approximately 1,440 nm, a second overtone at approximately 960 nm, and an intense combination band from approximately 1,960 to 2,100 nm.

Prediction of density and composition

The use of RBF interpolation to predict reservoir fluid properties from NIR spectroscopy involves a straightforward application of the material presented in the foregoing sections. The first step is the construction of a database of NIR absorbance and fluid properties pressure-volume-temperature (PVT) laboratory measurements (including composition, density, gas/oil ratio) made on a representative suite of oils measured at different temperatures and pressures. The database is then used to construct a mapping function to predict fluid properties from NIR absorbance measurements on oils that are not in the database.

The database used for this paper consisted of 14 live crude oils on which molecular compositions, gas/oil ratios, and densities were measured along with optical densities (i.e., absorbances). The viscosities of oils in the database were not known. However, the oils are all believed to exhibit low values of viscosity. Figure 12 shows the optical density data, acquired in the NIR wavelength range from 1,000 to 2,100 nm. Note that optical densities are negative for some of the samples; this was caused by a measurement offset. The molecular composition data for the 14 oils in the database were measured using gas chromatography (GC). Composition data consisted of mole percentages of molecules with carbon numbers C-1 (methane), C-2 (ethane), C-3 (propane), i-C-4 (iso-butane), n-C-4 (normal butane), i-C-5 (iso-pentane), n-C-5 (normal pentane), C-6 (hexane),

and C-7+ (heptane plus molecules with higher carbon numbers). This database was not ideal because temperature and pressure data were not available. Also, the accuracy of the RBF predictions of the heavier components could have been tested if the GC measurements had included higher carbon numbers.

The mapping function used to predict composition from NIR measurements is obtained from equation 19 by replacing the scalar viscosity output by a composition output vector. A vector containing the measured optical density at each wavelength replaces the amplitudes in the T_2 distribution. The scalar coefficient in equation (19) is replaced by a coefficient vector with dimension equal to the number of carbon numbers in the GC compositions. Figure 13 shows the compositions predicted from the NIR absorption measurements along with those measured by GC. Carbon numbers C-1 through C-6 and C7+ are on the x-axis in Figure 13. These results were obtained using the NWRE approximation for the coefficient vectors. NN distances were used to define the RBF widths. These results are very encouraging and show that RBF interpolation is a practical method for predicting composition from NIR absorption measurements. Figure 14 shows the oil densities predicted from NIR measurements. Note that they agree with the measured values to within 0.02 g/cc. It should be noted that there are no well-established correlations that relate NIR absorption spectra to mass densities of crude oils. Further work is needed to explore the range of crude oil viscosities for which NIR absorption spectra can be used to predict accurate mass densities.

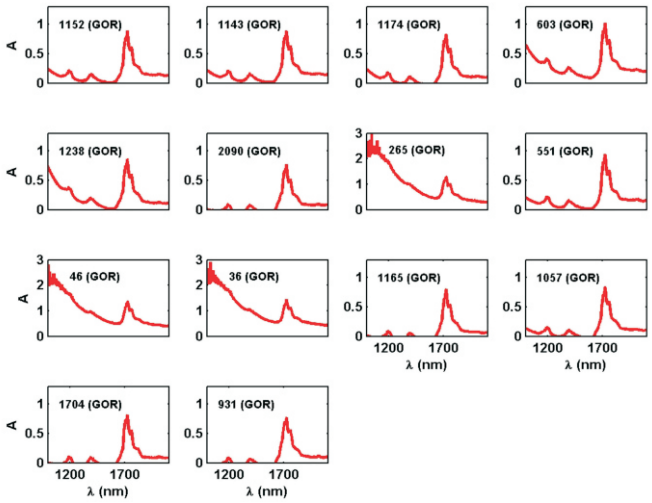

FIG. 12 Measured NIR absorbances (optical densities) versus wavelength for a database of 14 live oils.

SUMMARY AND CONCLUSIONS

This paper has proposed a new method for solving difficult inverse problems for which it is possible to construct a large calibration database. The proposed method does not require iterative training and is simple to implement once the database is established from available measurements or computations. Three example problems of current interest in well logging application have been used to demonstrate

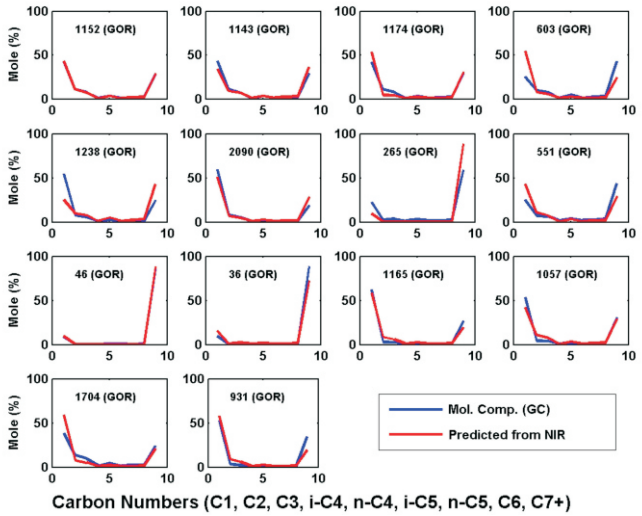

FIG. 13 Comparison of compositions predicted from measured NIR absorbances with those measured in the laboratory using GC for a database of 14 live crude oils.

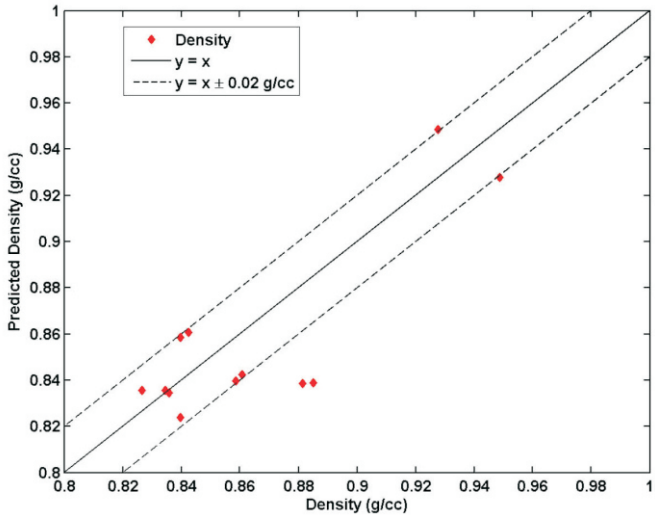

FIG. 14 Comparison of mass densities predicted from NIR absorbance measurements with those measured in the laboratory for a database of 14 live crude oils.

the flexibility, accuracy, and reliability of the method. Good benchmarking results have been obtained for all three example problems except for those data that are at the boundaries of the databases used in this paper. To remedy this problem the databases can be expanded so that their boundaries encompass the range of data encountered in practice.

Gaussian RBFs were used as basis functions in this paper, because they are the only multivariate RBFs that can be factored into products of univariate functions. This is an attractive feature for problems where input data from multiple sensors are required (e.g., see equation (19)).

The inversion method developed in this paper can be used to approach complex inverse problems for which accurate forward models are unknown and can also be used to obtain fast solutions of inverse problems associated with computationally intensive forward models. Moreover, the methodology proposed in this paper can be used to solve a broad range of estimation problems encountered in well-logging and geophysical interpretation.

The same method can also be used as a fast forward model, i.e., the RBF mapping function becomes a forward model if one exchanges database inputs and outputs. There are many interesting questions that remain for future work. For instance, is there an optimal technique for selecting the widths of the RBFs? If the database inputs and outputs are noisy, what is the optimal way to choose the coefficients of the RBF basis functions? How does the size and distribution of samples in the database affect the accuracy of the predictions?

NOMENCLATURE

		[Units]*
$\vec{A}$	amplitudes of a T_1, T_2, or D distribution of crude oil in equation (19)	[arbitrary]
$\vec{A}_j$	amplitudes of a T_1, T_2, or D for j-th database case in equation (19)	[arbitrary]
$A(\lambda)$	NIR optical density at wavelength λ	[dimensionless]
a	parameter in T_2-viscosity empirical correlation in equation (17)	[s cp $\cdot$ K^{-1}]
b	parameter in D-viscosity empirical correlation in equation (18)	[cm$^2 \cdot$ s$^{-1} \cdot$ cp $\cdot$ K^{-1}]
c	concentration of an absorbing medium in equation (25)	[moles/liter]
$\vec{c}_i$	coefficient vector multiplying RBF function (e.g., see equation (2))	[units of $\vec{F}(\vec{x})$]
C	matrix containing the database coefficient vectors in equation (12)	[units of $\vec{F}(\vec{x})$]
$\vec{d}_i$	standoff for i-th database case for 3D induction tool	[in.]
D	molecular diffusion coefficient	[cm$^2 \cdot$ s^{-1}]
D_{LM}	logarithmic mean value of D distribution	[cm$^2 \cdot$ s^{-1}]
$\vec{f}(\vec{x})$	multivariate function to be approximated from its sample values	
$\vec{F}(\vec{x})$	RBF approximation to $\vec{f}(\vec{x})$ (see equation (2))	
$f(gor)$	empirically determined function of gas/oil ratio in equation (17)	[dimensionless]
gor	gas/oil ratio in equation (17)	[m^3/m^3]
$I(\lambda)$	transmitted power of NIR radiation at wavelength λ	[watts]
$I_o(\lambda)$	incident power of NIR radiation at wavelength λ	[watts]
l	optical path length in equation (25)	[cm]
N	number of cases in database	
NN_j	nearest neighbor distance for database input $\vec{x}_j$, i.e., $\min\|\vec{x}_j - \vec{x}_i\|$ for all $i \neq j$	[units of $\vec{x}_j$]
P	pressure of live oil in equation (19)	[psi]
P_j	pressures of database measurements in equation (19)	[psi]
r_i	radius of borehole for i-th database case for 3D induction tool	[in.]
s_j	width of Gaussian RBF centered at database input	[units of $\vec{x}_j$]
s_g	width of Gaussian RBFs for GOR measurements in equation (19)	[m^3/m^3]
s_P	width of Gaussian RBFs for pressure measurements in equation (19)	[psi]
s_T	width of Gaussian RBFs for temperature measurements in equation (19)	[°F]
T	temperature of live oil in equation (19)	[°F]
T_j	temperatures of database measurements in equation (19)	[°F]
T_1	longitudinal relaxation time	[ms]
T_2	transverse relaxation time	[ms]
$T_{2,LM}$	logarithmic mean value of T_2 distribution	[ms]
$\vec{x}$	n-dimensional input vector	
$\vec{x}_i$	n-dimensional input vectors in database for $i = 1,2,\ldots,N$	
$x_{i,m}$	m-th component of $\vec{x}_i$	
x_m	m-th component of $\vec{x}$	
$\vec{y}_i$	m-dimensional output vector in database corresponding to input $\vec{x}_i$	
Y	matrix containing the database output vectors in equation (12)	

Greek Symbols

α	molar absorbtivity in Beer's law (equation (25))	[liters/(mole $\cdot$ cm)]

γ	regularization parameter in equation (4b)	[dimensionless]
η	viscosity	[cp]
η_i	viscosity of i-th oil in database	[cp]
λ_a	formation resistivity anisotropy factor	[dimensionless]
λ	wavelength of NIR radiation	[nm]
$\vec{\sigma}_a^{(raw)}$	apparent raw conductivities measured by 3D induction tool	[mhos]
$\vec{\sigma}_a^{(c)}$	apparent borehole corrected conductivities for 3D induction tool	[mhos]
$\vec{\sigma}_{a,i}^{(raw)}$	apparent raw conductivities for i-th case in database	[mhos]
$\sigma_{i,j}$	coupling between transmitter dipole in direction i with receiver dipole in direction j	[mhos]
σ_h	horizontal conductivity of infinite homogeneous formation	
$\sigma_{h,j}$	horizontal conductivity of formation for i-th case in database	[mhos]
σ_m	conductivity of borehole fluid	[mhos]
$\sigma_{m,j}$	conductivity of borehole fluid for i-th case in database	[mhos]
σ_v	vertical conductivity of infinite homogeneous formation	[mhos]
$\sigma_{v,i}$	vertical conductivity of formation for i-th case in database	[mhos]
$\varphi(\lvert\vec{x}-\vec{x}_i\rvert)$	RBF centered at $\vec{x}_i$	[dimensionless]
Φ	interpolation matrix whose elements are RBFs evaluated at the database inputs	

*Note that units for some quantities are not given because the units are problem dependent. This includes units for the database input and output vectors (e.g., $\vec{x}$ and $\vec{y}$ and the matrices constructed from these vectors.

ACKNOWLEDGMENTS

I would like to thank Steve Bonner, Chris Hopkins, and Jan Smits for supporting this work, Tracy Broussard for computing the 3D tri-axial induction database, and Nikhil Joshi for providing the oil samples, GC compositions, and NIR absorption measurements. Arjun Kurup of Rice University provided the laboratory measurements of viscosity and NMR relaxation times. Finally I would like to acknowledge Prof. R. L. Parker, Dr. Richard Rosthal, Dr. Mike Sheppard, Dr. Reza Taherian, and Prof. C. Torres-Verdin for helpful comments on the manuscript.

REFERENCES

Barber, T., Anderson, B., Abubakar, A., Broussard, T., Chen, K. C., Davydycheva, S., Druskin, V., Habashy, T., Homan, D., Minerbo, G., Rosthal, R., Schlein, R., and Wang, H., 2004, Determining formation resistivity anisotropy in the presence of invasion, SPE paper 90526: Society of Petroleum Engineers, presented at the SPE Annual Technical Conference and Exhibition.

Buhmann, M. D., 2003, *Radial Basis Functions: Theory and Implementation*: Cambridge University Press, Cambridge, UK.

Freed, D., 2004, Molecular composition from diffusion or relaxation measurements, U.S. Patent Application Publication, US 2004/0253743 A1, Published Dec. 16.

Freedman, R., Cao Minh, Chanh, Gubelin, G., Freeman, J. J., McGinness, Thal, Terry, Bob, and Rawlence, David, 1998, Combining NMR and density logs for petrophysical analysis in gas-bearing formations, paper II *in* 39th SPWLA Annual Logging Symposium Transactions: Society of Professional Well Log Analysts (Appendix A).

Freedman, R., Lo, S., Flaum, M., Hirasaki, G. J., Matteson, A., and Sezginer, A., 2001, A new NMR method of fluid characterization in reservoir rocks: Experimental confirmation and simulation results: *SPE Journal*, vol. 6, no. 4, p. 452–464.

Fujisawa, G., Mullins, O. C., Dong, C., Carnegie, A., Betancourt, S. S., Terabayashi, T., Yoshida, S., Jaramillo, A. R., and Haggag, A., 2003, Analyzing reservoir fluid composition in-situ in real time: Case study in a carbonate reservoir, SPE paper 84092: Society of Petroleum Engineers, presented at the SPE Annual Technical Conference and Exhibition.

Haykin, S., 1999, *Neural Networks: A Comprehensive Foundation*: Prentice Hall, Hamilton, Ontario, Canada (Chapter 5).

Heaton, N. and Freedman, R., 2005, Method for determining molecular properties of hydrocarbon mixtures from NMR data, U.S. Patent 6,859,032.

Heise, H. M., and Winzen, R., 2002, Fundamental chemometric methods, *in* H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise, eds., *Near-Infrared Spectroscopy: Principles, Instruments, and Applications*, Wiley-Vch., Chapter 7, Weinheim, Germany.

Micchelli, C. A., 1986, Interpolation of scattered data: distance matrices and conditionally positive definite functions: *Constructive Approximation*, vol. 2, p. 11–22.

Powell, M. J. D., 2001, Radial basis function methods for interpolation to functions of many variables, presented at the Fifth Hellenic-European Conference on Computer Mathematics and its Applications, p. 1–23.

Skoog, D. A., and West, D. M., 1976, *Fundamentals of Analytic Chemistry*: Holt, Rinehart, and Winston, p. 505–510.

Tarantola, A., 2005, *Inverse problem theory: methods for data fitting and methods for model parameter estimation*: SIAM Books, Philadelphia, PA.

Weyer, L. G., 1985, Near-infrared spectroscopy of organic substances: *Applied Spectroscopy Reviews*, vol. 21, no. 1 and 2, p. 1–43.

Winkler, M., Freeman, J. J., and Appel, M., 2004, The limits of fluid property correlations used in NMR well logging: An experimental study of reservoir fluids at reservoir conditions, paper DD *in* 45th SPWLA Annual Logging Symposium Transactions: Society of Professional Well Log Analysts.

ABOUT THE AUTHOR

Robert (Bob) Freedman graduated summa cum laude in physics from the University of Houston and holds a PhD in condensed matter physics from the University of California at San Diego. Bob joined Shell Development Company's Bellaire Research Center in Houston as a senior petrophysicist in 1976 and transferred to Shell Oil Company's operations as a senior petrophysical engineer in 1980. In 1981, he founded Petroleum Physics Corp., a consulting company specializing in formation evaluation. Bob joined Schlumberger in 1986 and is presently a Scientific Advisor and Manager in the Formation Evaluation Department at the Schlumberger Sugar Land (Texas) Product Center. During a 30-year career in the petroleum industry he has published papers on physics, mathematical modeling, data processing, and interpretation of NMR, induction, and dielectric tools; computerized log interpretation; modeling of deep-reading resistivity tools for relief well drilling; and mathematical inverse methods. He has been an active member of the SPWLA for 25 years and won a Best Paper Award for papers co-authored in 1994, 1997, and 1998. He has served on the SPWLA Technology Committee and has been an SPWLA Distinguished Lecturer. Bob has also been an active member of SPE for 28 years and has served as Associate Editor of the *SPE Formation Evaluation Journal*, SPE Distinguished Lecturer and Author, and is a member of the SPE Petroleum Professional Certification Subcommittee. He is a recipient of the SPE 2004 Cedric K. Ferguson Award. Bob has 44 industry publications and has been awarded 21 patents on well-logging technology.