Porosity inference and classification of siliciclastic rocks from multiple data sets

Luiz G. L. Loures[1] and Fernando S. Moraes[2]

ABSTRACT

We develop a Bayesian formulation for joint inference of porosity and clay volume, incorporating multiple data sets, prior information, and rock physics models. The derivation is carried out considering the full uncertainty involved in calculations from unknown hyperparameters required by either rock physics equations (model coefficients) or statistical models (data variances). Eventually, data variances are marginalized in closed form, and the model coefficients are fixed using a calibration procedure. To avoid working with a high-dimension probability density function in the parameter space, our formulation is derived and implemented using a moving window along the data domain. In this way, we compute a collection of 2D posterior distributions for interval porosity and clay volume, corresponding to each positional along the window's path. We test the methodology on both synthetic and real well logs consisting of gamma-ray, neutron, compressional and shear sonic velocity, and density. Tests demonstrate that integrating the relevant pieces of information about porosity and clay volume reduces the uncertainties associated with the estimates. Error analysis of a synthetic data example shows that neutron and density logs provide more information about porosity, whereas gamma-ray logs and velocities provide more information about clay volume. Additionally, we investigate a change in fluid saturation as a source of systematic error in porosity prediction. A real data example, incorporating porosity measurements on core samples, further demonstrates the consistency of our methodology in reducing the uncertainties associated with our final estimates.

INTRODUCTION

Determination of porosity distribution in a reservoir is usually based on seismic attributes and well information. In that procedure, the interpreter may rely on well-log data and core measurements to determine the relationship between each seismic attribute and porosity, in addition to other reservoir parameters. Most frequently, porosity is derived from selected well-log data, such as neutron, sonic, and density logs or a combination of these, after suitable data corrections. Neutron logs are already presented in terms of porosity, but other types of logs require some kind of transformation to obtain porosity. For example, porosity can be calculated easily from bulk density in clean zones of known matrix and fluid densities. Velocity-to-porosity conversions are made mainly using formulas from rock physics, such as given by Wyllie et al. (1956) [see also Raymer et al.

(1980), Tosaya and Nur (1982), Domenico (1984), Castagna et al. (1985), Han et al. (1986), and Eberhart-Phillips et al. (1989)].

Basically, data correction and transformation form the basis of the most commonly used methods for porosity determination from well logs and sometimes seismic attributes. The main problem with this approach is that it fails to treat the associated uncertainties properly, especially data errors inherent in each type of well log. When a well log is transformed, both signal and noise are converted into porosity. The noise term has contributions from instrument and ambient noise in addition to other formation and well effects.

Data corrections minimize the effects of error propagation in the porosity determination process. For example, in porosity determination from density logs, knowledge of matrix and fluid densities also implies knowing matrix and fluid compositions. But the fluid in the zone of investigation covered by density logging is mud filtrate mixed with residual formation fluid. The presence of residual light

oil and gas in the invaded zone may affect the bulk density strongly. Standard procedures are used to compensate for these effects, usually using information from other available logs, such as resistivity logs. Mud cake is also an issue in determinating porosity from density logs. Calibrations are made to compensate for this effect in a process that becomes less reliable with increasing variations in mud-cake thickness.

In the case of porosity calculation from sonic logs based on Wyllie's equation, the situation can be more complicated because matrix and fluid velocity determination is much less reliable than density. Here again, pore, matrix, and fluid heterogeneities; mud cake; and mud-filtrate composition are sources of uncertainty in porosity determination. When the mud invasion is low in a hydrocarbon-bearing formation, the estimated porosity will be greater than the actual porosity. In an uncompacted, light-hydrocarbon-bearing formation, estimated porosity will remain high even after corrections for lack of compaction are applied.

In this paper, we develop a methodology that addresses the issue of uncertainty in the data and is applicable throughout the reservoir characterization process, beginning with the interpretation of well data. We pursue this idea by addressing porosity determination along the well as an inference problem. The solution is constructed with the focus on integrating multiple data sets and considering selected sources of uncertainty, such as the coupled effect of clay (shale) volume and noise in the data. The formulation follows Bayesian methodology as presented in Jeffreys (1939) and Jaynes (2003) [see also Ulrych et al. (2001) for a tutorial and geophysical applications and Gelman et al. (2004) for a modern treatment].

The Bayesian approach focuses on obtaining a probability density function (PDF) for the parameter under investigation — the posterior distribution — assimilating two kinds of information: from (1) data fitting and (2) prior information on models. Both pieces of information are represented as PDFs, called the likelihood function (or data distribution) and the prior distribution, respectively. Once these distributions are combined to give the posterior distribution, all inferences can be obtained by computing statistics relative to individual parameters (e.g., marginal posterior distributions).

Our strategy is to avoid the complications that arise in multidimensional Bayesian inverse problems [see, e.g., Moraes and Scales (2000)]. The solution is formulated in terms of interval rock properties using a moving window procedure. The final result is a pair of marginal posterior distributions for porosity and clay volume for each depth point. Not only estimates but also uncertainty measures can be obtained straightforwardly from the 1D marginal posteriors. Also, the contribution from each piece of information toward increasing the estimates' reliability can be assessed easily.

Selecting the mathematical models relating well data and reservoir parameters is a key step in constructing the likelihood functions. In fact, existing rock physics models do not depend on porosity and clay volume alone but also depend on other variables such as matrix mineral composition, fluid content, saturation, pressure, and temperature. It is impractical, however, to include all variables as unknowns in the petrophysical inference problem. We choose to restrict our inference to porosity and clay volume as unknown variables, reflecting some of the uncertainty issues discussed above. As a texture indicator, clay volume plays an important role in the classification of siliciclastic reservoirs and in modeling the seismic response (Vernik and Nur 1992; Vernik, 1998). Other variables are either kept out of the selected modeling equations or are treated as fixed values, relying heavily on prior knowledge coming from analysis of core samples and additional well-log information. To emphasize the importance of these additional variables (hyperparameters) in full uncertainty modeling, we derive an expression for a joint posterior distribution for porosity, clay volume, and the set of (unknown) hyperparameters. Then we propose a calibration procedure to fix them as given values.

Tests on synthetic and real data show that the resulting porosity uncertainty is reduced strongly by the integration of independent information contained in multiple well logs. On the other hand, without a more directly related data set such as a gamma-ray log, clay volume is not determined very well. We, however, show that joint inferences of porosity and clay volume can be used to help classify sandstones, following a scheme proposed by Vernik and Nur (1992). In our examples, empirical models express velocity as a function of porosity and clay volume variations, as given by Han et al. (1986). In the real data example, we select a set of well logs from a sandstone channel reservoir of the Blackfoot field, Canada. The rock physics model coefficients are calibrated using porosity measurements from core samples and clay volume from gamma-ray logs as reference values. These values are also used for comparing final estimates of porosity and clay volume.

BAYESIAN FORMULATION

Consider the subsurface represented by a 1D layered model where each layer is homogeneous with constant porosity ϕ and clay volume χ. The problem under investigation is to make inferences about interval porosity and clay volume from a set of well-log data vectors sampled on that interval, from formulas from rock physics studies, and from prior information I. The last is defined as additional information that is independent from the data to be inverted. To derive the solution, we let the set of multiple well-log data be represented by $D = \{\mathbf{d}_1,\ldots,\mathbf{d}_K | \mathbf{d}_i \in \mathcal{R}^{N_i}, i = 1,\ldots,K\}$, where $\mathbf{d}_i$ is a data vector representing one of K types of data (specified in the next section), having N_i as the number of data points in the selected depth interval and where $\mathcal{R}$ is real numbers. To be able to represent the parameter space fully, we also consider the set $M = \{\phi,\chi,H | \phi,\chi \in \mathcal{R}\}$, which includes the subset H of hyperparameters to be defined in the next section. Hyperparameters (or nuisance parameters) are parameters necessary to arrive at a solution but of no direct interest in the inference problem. In our case, hyperparameters arise from specifications of rock physics and statistical models.

In the Bayesian approach, the solution for our main problem is given by the posterior distribution for porosity and clay volume $p(\phi,\chi | D,I)$. But because of the hyperparameters, the full modeling of uncertainties, by application of Bayes' theorem, is given by

$$p(\phi,\chi,H | D,I) = \frac{s(\phi,\chi,H | I)r(D | \phi,\chi,H,I)}{h(D | I)}, \qquad (1)$$

where $s(\phi,\chi,H | I)$ is the joint prior PDF, $r(D | \phi,\chi,H,I)$ is the joint PDF for the data, and $h(D | I)$ is a normalizing PDF, which ensures that the posterior distribution integrates to unity when evaluated at a particular set of observed data values (i.e., $D = D_{\mathrm{obs}}$). The value $r(D | \phi,\chi,H,I)$ is also known as the likelihood function of parameters ϕ, χ, and H to emphasize its inference properties.

Essentially, the joint posterior is the normalized product of the prior distribution and the likelihood function, which can be expressed by

$$p(\phi,\chi,H|D,I) \propto s(\phi,\chi,H|I)r(D|\phi,\chi,H,I). \qquad (2)$$

Expression 2 imposes some requirements that must be met before one can obtain an acceptable posterior PDF as the solution for an inverse problem. First, both the prior and the data distributions must be defined. Then there must be some compatibility between the prior understanding of the model and a data-driven specification of model [i.e., $r(D|\phi,\chi,H,I) > 0$ for some set $\{\phi,\chi,H\}$, where $s(\phi,\chi,H|I) > 0$]. The specification of a likelihood function also implies the existence of a known relationship between D and parameters ϕ and χ, possibly involving additional parameters in H, as discussed in the next section. The definition of the data and prior distributions and the final form of the posterior distribution are discussed after that.

Rock physics models

We follow three standard steps to construct the Bayesian solution: (1) select the data sets, which carry information about ϕ and χ; (2) find mathematical relations between the data and parameters; and (3) define statistical models (PDFs) for prior and likelihood function, based, respectively, on prior parameter information and data uncertainties.

Let $\mathbf{d}_{\phi_N}$, $\mathbf{d}_{V_P}$, $\mathbf{d}_{V_S}$, $\mathbf{d}_\rho$, and $\mathbf{d}_\gamma$ be data vectors representing, respectively, neutron porosity ϕ_N, P-wave velocity V_P, S-wave velocity V_S, density ρ, and gamma-ray γ wireline logs. Because these data vectors are contaminated with noise, they can be expressed as a function of porosity and clay volume plus an error associated with both data acquisition and data modeling (i.e., neglected physical effects). With this in mind and using numbered subscripts for convenience, we can write

$$\mathbf{d}_{\phi_N} \rightarrow \mathbf{d}_1 = \mathbf{f}_1(\phi,\chi) + \mathbf{e}_1, \qquad (3)$$

$$\mathbf{d}_{V_P} \rightarrow \mathbf{d}_2 = \mathbf{f}_2(\phi,\chi) + \mathbf{e}_2, \qquad (4)$$

$$\mathbf{d}_{V_S} \rightarrow \mathbf{d}_3 = \mathbf{f}_3(\phi,\chi) + \mathbf{e}_3, \qquad (5)$$

$$\mathbf{d}_\rho \rightarrow \mathbf{d}_4 = \mathbf{f}_4(\phi,\chi) + \mathbf{e}_4, \qquad (6)$$

and

$$\mathbf{d}_\gamma \rightarrow \mathbf{d}_5 = \mathbf{f}_5(\chi) + \mathbf{e}_5, \qquad (7)$$

where $\mathbf{f}_i$, for $i = 1,\ldots,5$, represent constant-vector valued functions relating the data to ϕ and χ, and where $\mathbf{e}_i$, for $i = 1,\ldots,5$, are the error components in the data. By assuming constant reservoir properties over selected depth intervals and strict dependence of the data on the same properties, one can obtain only constant model responses inside that interval. A variation in observed data is then interpreted as noise.

Next, we specify the functions $\mathbf{f}_i$. According to the above description, these can be represented by $\mathbf{f}_i = f_i\mathbf{u}$, where $f_i = f_i(\phi,\chi)$ is a simple function of porosity and clay volume and $\mathbf{u} = (1,\ldots,1)^T$, with the superscript T standing for a matrix-vector transpose. For $i = 1$, we have the neutron log derived from a nuclear tool that measures the amount of hydrogen in the formation. The amount of hydrogen is not constrained by pore volume alone. The presence of

clay and the type of fluid strongly influence hydrogen content. Additionally, neutron data normally are affected by a calibration effect related to the baseline response of a reference rock, i.e., a matrix effect (see, e.g., Hearst et al., 2000). Considering these effects, a simple form for function f_1 can be written as

$$\phi_N = a_{\phi_N} + \phi + c_{\phi_N}\chi, \qquad (8)$$

where a_{ϕ_N} is the coefficient for a constant calibration effect and c_{ϕ_N} is the coefficient for the shale term, actually corresponding to shale microporosity.

The expressions relating compressional-wave and shear-wave velocities to porosity and clay volume, represented by f_2 and f_3, are both selected from Han et al. (1986). Their paper presents a rock physics study of 75 sandstones samples, with (helium) porosity ranging from 2% to 30% and clay content of 0% to 50% from point counts. Wave-velocity measurements are performed as functions of pressure (up to 50 MPa of confining pressure) and saturation (either dry or water saturated). One of their main results agrees with previous authors (e.g., Castagna et al., 1985) in that velocities can be related linearly to porosity and clay content. This allows us to write, in our notation, the seismic velocities V_P and V_S as given by

$$V_P = a_{V_P} + b_{V_P}\phi + c_{V_P}\chi \qquad (9)$$

and

$$V_S = a_{V_S} + b_{V_S}\phi + c_{V_S}\chi, \qquad (10)$$

where a, b, and c for V_P and V_S subscripts are the empirical model coefficients given in km/s, considering that porosity and clay volume are expressed in terms of volume fractions. For water-saturated samples at 39 MPa differential pressure, Han et al. (1986) obtain 5.59, -6.93, and -2.18 km/s for the coefficients of the compressional velocity model (equation 9) and 3.52, -4.91, and -1.89 km/s for the shear-wave model (equation 10). We discuss the validity of these coefficient values in connection with the real data example.

The f_4 function relating density to porosity and clay volume is described by

$$\rho = \rho_q + (\rho_{fl} - \rho_q)\phi + (\rho_{cl} - \rho_q)\chi, \qquad (11)$$

where ρ_q, ρ_{cl}, and ρ_{fl} are, respectively, the sand grain (mostly quartz), clay (approaching shale), and fluid densities. Finally, the gamma-ray data are described by a function of clay volume alone, given by

$$\gamma = (\gamma_{sh} - \gamma_{sd})\chi + \gamma_{sd}, \qquad (12)$$

where γ_{sh} and γ_{sd} are reference values for the gamma-ray emissions of pure shales and clean sands, respectively.

Notice that none of the model coefficients presented above is known with absolute certainty. This means that a_{ϕ_N}, c_{ϕ_N}, a_{V_P}, b_{V_P}, c_{V_P}, a_{V_S}, b_{V_S}, c_{V_S}, ρ_q, ρ_{cl}, ρ_{fl}, γ_{sh}, and γ_{sd} are unknown quantities, which we include in the set H of hyperparameters.

The likelihood function

Because each well-log data acquisition can be viewed as an independent experiment, one may assume that the noise in the data is also independent. In this case, the joint likelihood function (see equation 1) can be rewritten as the product of five separate likelihoods, depending on its own set of hyperparameters, so that $H = \{H_1,H_2,H_3,H_4,H_5\}$. Then we write

$$r(D|\phi,\chi,H,I) = r_1(\mathbf{d}_1|\phi,H_1,I)r_2(\mathbf{d}_2|\phi,\chi,H_2,I)r_3(\mathbf{d}_3|\phi,\chi,H_3,I)$$

$$\times r_4(\mathbf{d}_4|\phi,\chi,H_4,I)r_5(\mathbf{d}_5|\chi,H_5,I), \tag{13}$$

where the subscripts 1,...,5 are defined as for equations 3–7.

The functional forms of these likelihood functions are specified considering available information about data errors. An available criterion for guiding this process is the principle of maximum entropy, the standard for assigning probabilities in Bayesian inference (see, e.g., Jaynes, 1982). According to this principle, if only the first- and second-order moments are considered for describing errors in the data, then the normal distribution is the appropriate choice for a likelihood function, i.e., it has maximum associated entropy (uncertainty). We accept this result by letting the likelihood functions r_i, $i = 1,...,5$, be given by

$$r_i(\mathbf{d}_i|\phi,\chi,H_i,I) = (2\pi\sigma_i^2)^{-\frac{L_i}{2}}$$

$$\times\exp\left\{-\frac{1}{2\sigma_i^2}[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]^T\right.$$

$$\left.\times[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]\right\}, \tag{14}$$

where σ_i^2 is the error variance corresponding to the ith observed data vector. Realistically, these scaling parameters σ_i^2 of data distributions are also not known a priori. Consequently, they must be treated as unknown parameters to be included in H. This allows for an explicit definition of each independent set of hyperparameters $H_1 = \{\sigma_1,a_{\phi_N},c_{\phi_N}\}$, $H_2 = \{\sigma_2,a_{V_P},b_{V_P},c_{V_P}\}$, $H_3 = \{\sigma_3,a_{V_S},b_{V_S},c_{V_S}\}$, $H_4 = \{\sigma_4,\rho_q,\rho_{cl},\rho_{fl}\}$, and $H_5 = \{\sigma_5,\gamma_{sh},\gamma_{sd}\}$, corresponding to each likelihood model in equation 13 in the form of equation 14.

One can argue that borehole effects and the very subsurface rocks we wish to characterize add correlations to the data, making a full covariance matrix a more adequate description of data errors. In fact, when the true covariance of errors is known, it acts as a filter, eliminating any coherent noise and downweighting data points with higher noise levels. Depending on the assigned covariance structure, the effect can be dramatic. As with the data variance, because covariances are usually not known a priori, we prefer to assign an identity matrix to limit the number of unknown variables. When a sample of very typical noise is available, as described by Gouveia and Scales (1998), it would be appropriate to incorporate it.

The prior distribution

The prior distribution is used to restrict the parameter space, thus reducing the uncertainty involved in calculations. Following Jeffreys (1939), prior distributions may be data based or nondata based. Nondata-based prior distributions are derived from considerations that can come from objective arguments based on theoretical considerations, such as invariance under group transformations and physical bounds on parameters (Jaynes, 1968), or from subjective information, such as the investigator's prior guesses based on background expertise. For a thorough review of issues related to assigning prior probabilities, see Kass and Wasserman (1996).

According to equation 1, the prior distribution s is defined over the whole domain of porosity, clay volume, and set H. To represent the prior distributions, we need a different partition of H than that presented in the previous section, which is based on data types in the likelihood functions (i.e., sets of data variance and model coeffi-

cients). This time we need a partition based on parameter types, separating one set for statistical distribution hyperparameters and another set for model coefficients. Thus, let $H = \{H_c,H_\sigma\}$, where H_c and H_σ respectively represent the sets of all rock physics model coefficients and error variances, i.e.,

$$H_c = \{a_{\phi_N},c_{\phi_N},a_{V_P},b_{V_P},c_{V_P},a_{V_S},b_{V_S},c_{V_S},\rho_q,\rho_{cl},\rho_{fl},\gamma_{sh},\gamma_{sd}\},$$

$$H_\sigma = \{\sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_5\}.$$

It is reasonable to assume that the data variances are independent from other parameters (i.e., porosity, clay volume, and model coefficients). In this way, we can write the prior distribution as

$$s(\phi,\chi,H|I) = s_0(\phi,\chi,H_c|I)\prod_{i=1}^{5} s_i(\sigma_i|I). \tag{15}$$

We follow one of the most conservative Bayesian practices in assigning prior probabilities, which is to avoid accounting for any uncertain information in the prior distribution. This corresponds to Jeffreys' strategy for assigning uninformative prior distributions. Data variances are nonnegative parameters, varying between 0 and ∞, so the use of the improper (not normalizable) distribution $1/\sigma_i$ is consistent with expressing complete unawareness of such parameters (see Ulrych et al., 2001). Next, we need to specify $s_0(\phi,\chi,H_c|I) = w(H_c|\phi,\chi)q(\phi,\chi)$. An uninformative prior distribution for porosity and clay volume can be prescribed by letting q be constant over the region where the likelihood shows any appreciable value. Following these considerations, the prior distribution is expressed as

$$s(\phi,\chi,H_c|I) = \frac{w(H_c|\phi,\chi)}{\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5}, \quad \text{for } 0 < \sigma_{i=1,...,5} < \infty. \tag{16}$$

This approach for specifying q, suggested by Box and Tiao (1973, their section 1.5), corresponds to extending porosity and clay volume to an unbounded domain by allowing them to vary on $[-\infty, +\infty]$. This makes it possible to derive an analytical expression for the posterior distribution (next section). After that, the posterior distribution is renormalized over a truncated domain, corresponding to known limits for porosity and clay volume. Porosity is set to vary on $[0,\phi_{\text{crit}}]$, where ϕ_{crit} is the critical porosity (usually 0.4 for sandstones), and clay volume is set to vary on $[0, 1]$.

The posterior distribution

The posterior distribution can be found by application of Bayes' theorem (equation 2), using equations 13, 14, and 16 to obtain

$$p(\phi,\chi,H|D,I) \propto w(H_c|\phi,\chi)\prod_{i=1}^{5} \sigma_i^{-(L_i+1)}$$

$$\times\exp\left\{-\frac{1}{2\pi\sigma_i^2}[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]^T\right.$$

$$\left.\times[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]\right\}, \tag{17}$$

where $L_i, i = 1,...,5$, are the sizes of each data vector $\mathbf{d}_i$.

To obtain the final form of the posterior distribution, we return to the discussion of how to treat the hyperparameter set. A standard

treatment for hyperparameters is to eliminate them by integration, a procedure known in statistics as marginalization of the joint distribution. This most often cannot be done analytically, but the standard deviation in normal distributions is an exception (see Box and Tiao, 1973). Based on this idea, it is possible to represent the marginal posterior PDF for porosity, clay volume, and the set of rock physics model coefficients (H_c) by writing

$$p(\phi,\chi,H_c|D,I)$$
$$\propto \int_{\sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_5} p(\phi,\chi,H|D,I)d\sigma_1 d\sigma_2 d\sigma_3 d\sigma_4 d\sigma_5.$$

(18)

Upon substituting equation 17 for the joint posterior into equation 18, it becomes clear that integration can be performed on separate pairs likelihood prior, containing a single variance variable σ_i as defined by the product term on the right-hand side of equation 17. The marginalization of variance of a normal distribution with Jeffreys' noninformative prior distribution is a well-known result in statistics, yielding a Student's t-distribution when the forward model is linear and parameters vary on $[-\infty, +\infty]$. This is not our case because we have the hyperparameters and a bounded domain for porosity and clay volume. After integration, however, it is possible to write the posterior distribution in the form

$$p(\phi,\chi,H_c|D,I) \propto w(H_c|\phi,\chi)$$
$$\times \prod_{i=1}^{5} \{[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]^T[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]\}^{-L_i/2}.$$

(19)

Detailed analytical derivations for equation 19 can be found in Box and Tiao (1973, sections 1.3.6 and 2.7.2) and Bretthorst (1990, p. 16).

The final goal is to find $p(\phi,\chi|D,I)$. This can be done in connection with modern Bayesian analysis using Markov-chain Monte Carlo (MCMC) methods, which provide numerical computations of the marginal posterior distribution over a large number of hyperparameters. This represents a full treatment of associated uncertainties. An alternative approach is to perform a calibration procedure, also relying on prior information and additional data sets (e.g., core measurements), and then to fix hyperparameters to their calibrated values during inversion. Whereas the MCMC approach introduces additional computational complexity, the calibration approach may generate artificially lower uncertainties or biased results. On the other hand, model calibration analysis is an essential stage of reservoir characterization, generally yielding useful interpretations about the reservoir properties and adequate estimates of hyperparameters.

To fix hyperparameters prior to inversion is equivalent to assuming that the prior distribution for hyperparameters $w(H_c|\phi,\chi)$ is concentrated strongly around the point of the calibrated set of coefficients $H_c = H_{\text{cal}}$. Consequently, we can write

$$p(\phi,\chi|D,I) = \int \delta(H_c - H_{\text{cal}})p(D|\phi,\chi,H_c,I)dH_c$$
$$= p(\phi,\chi|D,H_c = H_{\text{cal}},I), \quad (20)$$

where δ is the Dirac delta distribution. In this case, the marginal pos-

terior distribution $p(\phi,\chi|D,I)$ is identical to the conditional distribution $p(\phi,\chi|D,H_c = H_{\text{cal}},I)$, given by

$$p(\phi,\chi|D,I) \propto \prod_{i=1}^{5} \{[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]^T[\mathbf{d}_i - \mathbf{f}_i(\phi,\chi)]\}^{-L_i/2}.$$

(21)

Following the strategy outlined in the previous section, equation 21 is normalized over a bounded domain defined by $0 < \phi < \phi_{\text{crit}}$ and $0 < \chi < 1$ to yield the final joint posterior distribution for porosity and clay volume.

PRACTICAL IMPLEMENTATION

Equation 21 is a 2D PDF for interval porosity and clay volume, given a set of well-log measurements and prior information, corresponding to a given model discretization unit. Such formulation is suitable for implementation using a moving window — in this case, running along a 1D space as defined by the well trajectory. The goal is to obtain a collection of posterior PDFs corresponding to each depth interval defined by the window position. In this way, it becomes straightforward to make any desired inference about the parameters (porosity and clay volume) from the 2D posterior PDFs. In particular, we compute the marginal distributions to obtain either the mean, median, or mode as estimates and the lower and upper limits for the 0.95 probability interval as a measure of the associated uncertainty.

The whole procedure is summarized in Figure 1. First, we select the well-log data and the size of the moving window (left). Then, for each depth corresponding to the central position of the window, we evaluate the joint posterior PDF (equation 21) and compute the corresponding pair of marginal distributions for porosity and clay volume (center). These marginal PDFs are gathered by running the window over the desired depth range, as in the two plots on the right side of Figure 1. The grayscale represents probability density (PD) for porosity or clay volume, as indicated on each plot. Three solid lines overlay the gray area in both porosity and clay volume plots. The curve in the center represents the posterior mode, and the other two lines represent the limits of the 0.95 probability intervals, consisting of a lower limit (left gray line) and upper limit (right gray line).

Another inference application implemented in this work is a lithological classification based on the scheme proposed by Vernik and Nur (1992). For that, we choose four classes and ranges of parameters: (1) clean arenite — $(0.22 \leq \phi \leq 0.35)$ and $(\chi \leq 0.04)$; (2) arenite — $(0.14 \leq \phi \leq 0.22)$ and $(0.04 \leq \chi \leq 0.15)$; (3) wacke — $(0.06 \leq \phi \leq 0.14)$ and $(0.15 \leq \chi \leq 0.35)$; and (4) shale — $(0.0 \leq \phi \leq 0.07)$ and $(\chi \geq 0.35)$. Class limits are not rigorous and should be adjusted to match target lithology at known locations. We use a naive decision criterion to assign a class for a given depth location to compute the total probability, over class ranges, and to select the class corresponding to the maximum probability. More elaborate criteria are available in the literature of Bayesian classification and risk analysis (see, e.g., MacKay, 2003).

SYNTHETIC DATA EXAMPLES

This section presents the analysis of an application to synthetic well-log data, consisting of neutron porosity, sonic velocity (compressional and shear), density, and gamma-ray logs. We are particularly interested in evaluating our methodology with respect to the
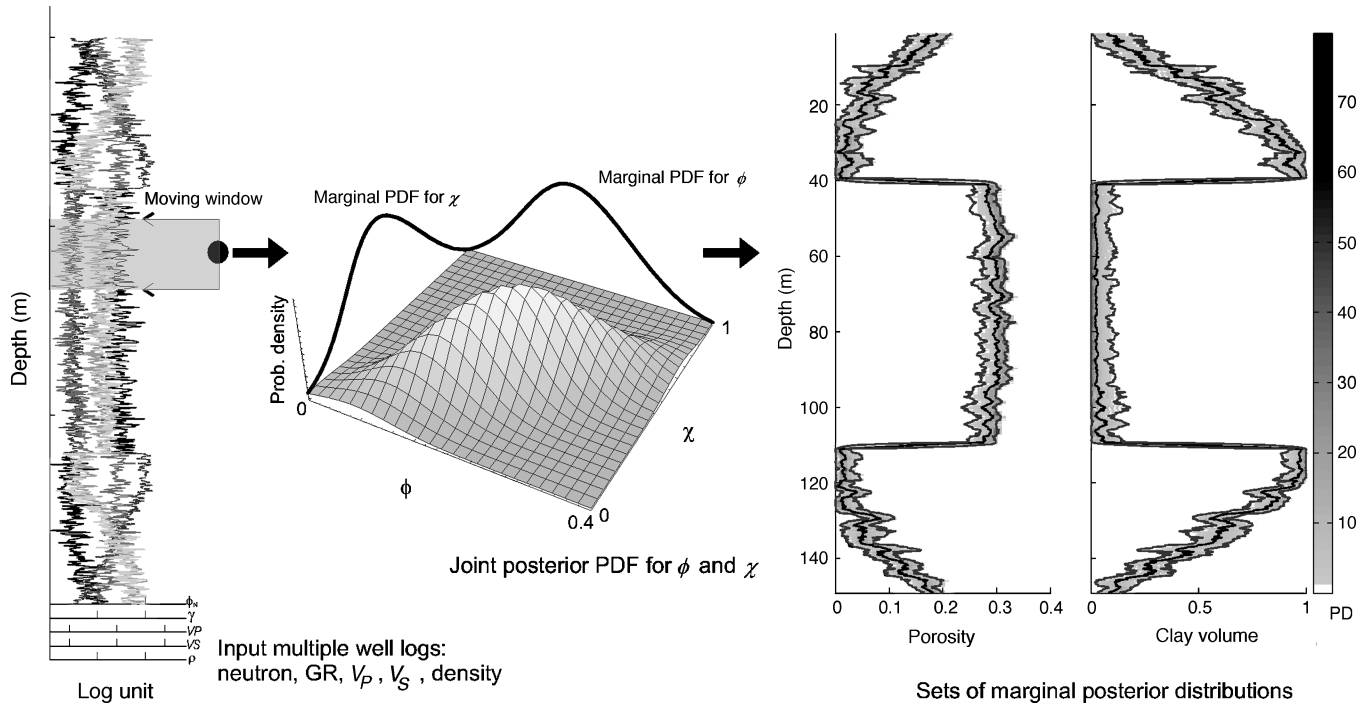
Figure 1. Schematic (not to scale) illustrating the inversion using a moving window (left). The shaded area indicates the data interval spanned by the window, which is used to evaluate equation 21 for the joint posterior distribution for porosity and clay volume (center). Marginalization of the joint posterior PDF for all depth locations yields two sets of marginal posterior PDFs for porosity and clay volume (right). The grayscale represents probability density (PD), and the lines over the images represent the mode (center black), lower (left gray), and upper (right gray) limits for a centered 0.95 probability interval of the marginal posterior distributions.
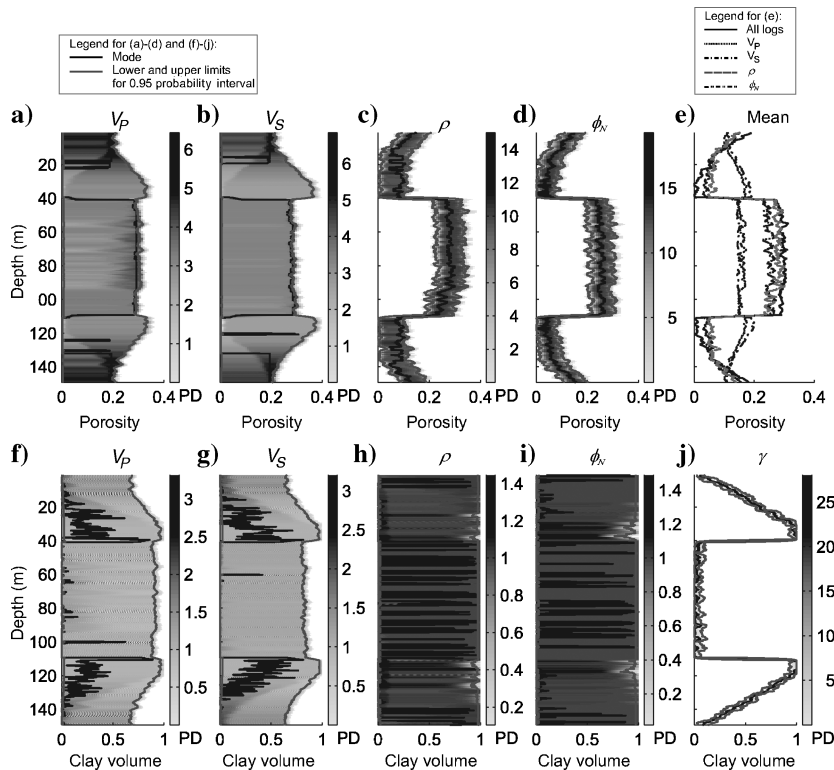
Figure 2. Synthetic data inversion results using single logs. (a)–(d) Collections of marginal posterior PDFs for porosity ϕ, except for (e) showing the mean of each set of porosity PDFs. (f)–(j) Collections of marginal PDFs for clay volume χ. Titles in each PDF plot indicate the input well-log data for inversion. (See main text and Figure 1 for a detailed description of image plots.)

use of empirical velocity equations 9 and 10, uncertainty and resolution analysis, and random and systematic noise sensitivity and its practical applicability in classifying sandstones. The results are illustrated in Figures 2–4.

The data are generated based on a three-layer synthetic model (Figure 4), including a central reservoir layer consisting of a 70-m-thick sandstone layer ($L2$). The sequence is constructed beginning with a sandstone model of 0.15 porosity and 0.05 clay volume. The original sandstone (facies A) is gradually replaced by a combination of shale and dispersed clays until it becomes pure shale (facies B). At the top of the reservoir (facies C), the clay volume is 0.03, decreasing to 0.015 below (facies D). The oil-water contact is positioned at a 110-m depth. The effective porosity, resulting from the sand-shale combination, is computed as described by Hearst et al. (2000, their p. 367).

The synthetic neutron porosity, density and gamma-ray data are generated using equations 8, 11, and 12 respectively. To compute velocities, we apply Gassmann-type petroelastic modeling at constant pressure for the entire model. Effective elastic moduli are computed from a hypothetical sand-shale matrix composition and Voigt-Reuss-Hill averaging. The idea is to contrast Han's models with Gassmann's and to analyze the sensitivity with respect to changes in satura-

tion as a source of systematic error. To test for random noise sensitivity, we add zero mean pseudo-random Gaussian noise to all logs. The standard deviation is set as either 5%, 10%, or 15% of the corresponding mean data value. The data spacing is 30 cm.

Several tests are performed to evaluate the importance of each type of well-log data in the presence of different levels of random noise and a variation in fluid saturation in the reservoir layer. First, we consider each kind of log separately and then test the inversion using different combinations of logs. In Figure 2, we present the results for single log inversions using a seven-point (2.1 -m) window, which is appropriate because of the smooth nature of the model. Figure 2a–d shows resulting marginal posterior PDFs for porosity, corresponding to each data set parameterized with respect to porosity (V_P, V_S, density, and neutron porosity ϕ_N). Figure 2e shows porosity posterior mean curves computed from all PDFs in each set, in addition to final porosity estimates (solid black curve) from combined data inversion (see Figure 3). The rms error for porosity estimates from single log inversions are, respectively, 0.1203 (V_P), 0.1312 (V_S), 0.0333 (density), and 0.0337 (ϕ_N). At the reservoir layer, V_P and density also show noticeable effect from a change in fluid saturation, degrading porosity rms error to a maximum of 0.1393 and 0.0443, respectively.

The bottom row of Figure 2 presents collections of marginal posterior distributions for clay volume, corresponding to inversions of V_P, V_S, density, ϕ_N, and gamma-ray data. Note the extremely large uncertainty associated with inversions for porosity or clay volume, using velocities or using a single log with all inversions for clay volume, except in the obvious case of gamma-ray inversion. Estimates of clay volume from individual log inversions (not shown) produce rms errors of 0.3836 (V_P), 0.3697 (V_S), 0.4147 (density), 0.4181 (ϕ_N), and 0.0497 (gamma ray). The uncertainty observed in the porosity marginal posterior densities is largely the result of coupling with clay volume through the linear models. As a result of this effect, one can only determine an upper boundary for porosity from velocities (i.e., lower boundary is obviously $\phi = 0$). Figure 2j reflects high uncertainty levels by showing completely different porosity estimates, depending on the type of data used as input. Best porosity estimates (lower rms errors) are given by the density and neutron-log inversions. For clay volume, lower rms errors are produced from gamma-ray inversion, followed by V_S and V_P inversions.

Next, we run the inversion using two different combinations of logs. The first combination only considers V_P, V_S, and density logs. The other combination includes all logs, which are related
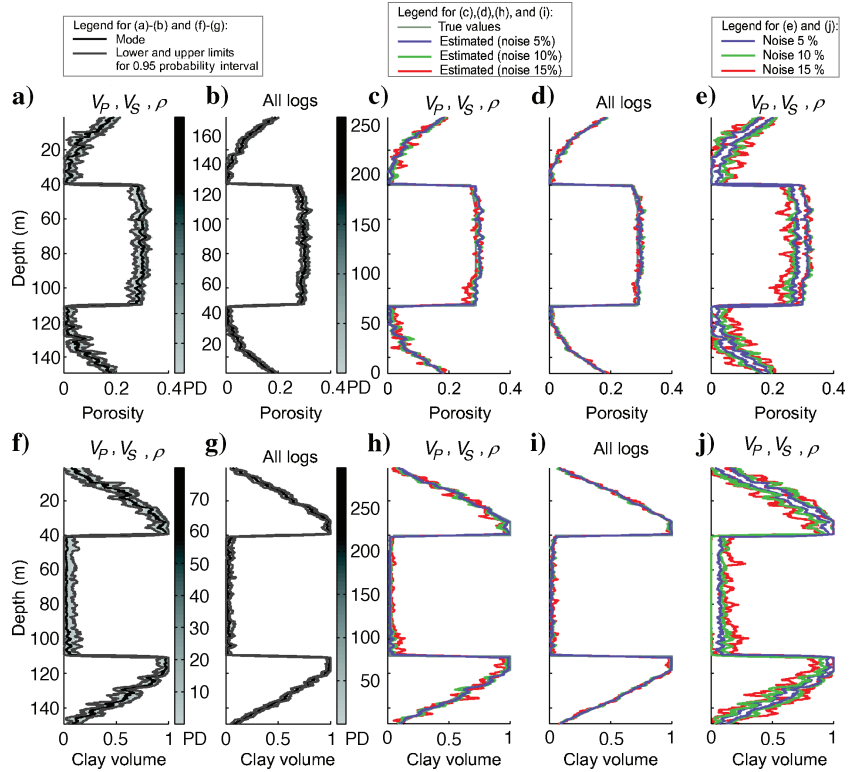

Figure 3. Inference results for (top) porosity and (bottom) clay volume from the synthetic data example. (a), (b), (f), (g) Marginal posterior PDFs for two combinations of input data, as indicated by the titles of each plot. (c), (d), (h), (i) Comparison of the true porosity or clay volume with estimates obtained from the posterior distributions for progressively higher noise content (5%, 10%, and 15%). (e), (j) Lower and upper limits for 0.95 probability interval, using the same noise variation.
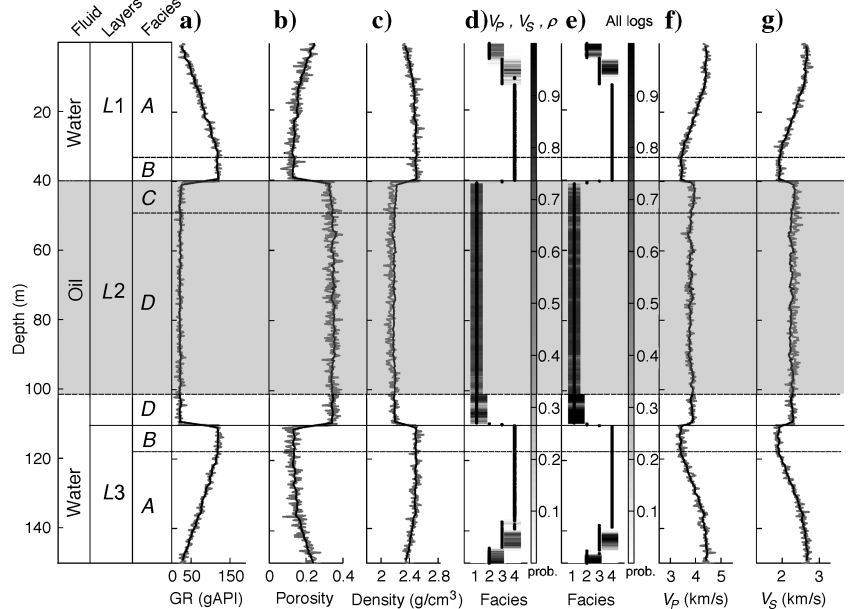

Figure 4. Data fit and lithological classification resulting from the inversion of synthetic data. (a), (b), (c), (f), (g) Synthetic noisy (gray) and computed data (black), respectively, corresponding to gamma-ray (GR), neutron porosity, density, V_P, and V_S data. (d), (e) Classifications with respect to four facies (see text for class description). Grayscale represents class probabilities, and points indicate the class corresponding to maximum probability. Horizontal lines and annotations on the left side of the figure show the details of the three-layer ($L1$, $L2$, and $L3$) synthetic model, with varying clay volume (A, B, C, and D) and fluid saturation (water and oil).

to porosity and clay volume, as presented in the previous section and in Figure 2. The motivation behind using a partial combination of V_P, V_S, and density logs is to draw some correspondence with seismic attributes, which may be available throughout the reservoir.

The results for the combined data inversion are shown in Figure 3, with corresponding input data combinations indicated by the plot titles. Following the same layout as Figure 2, the top row presents the results for porosity, and the bottom row presents the results for clay volume. The first two columns of plots show the marginal posterior distributions for each data combination, with Figure 3a and 3f, corresponding to velocities and density with Figure 3b and 3g corresponding to a combination of all possible logs. The solid lines overlying these four images are the same, as described in Figures 1 and 2. We run multiple tests to evaluate the noise sensitivity, using data contaminated with progressively higher-noise content (5%, 10%, and 15%). We only show the posterior PDFs for the 10% noise level (Figure 3a, 3b, 3f, and 3g), but we also compare resulting estimates (Figure 3c, 3d, 3h, and 3i), and limits for the 0.95 probability interval (Figure 3e and 3j) for all three noise levels.

Note that estimates are stable with respect to noise in both data combinations. Porosity is well determined (low associated uncertainty), whereas the error associated with clay volume may be large. The change in saturation within the central layer produces a still noticeable effect on porosity estimates, but the effect is greatly reduced when all logs are combined. The limits for the 0.95 probability interval show progressively wider spreads as noise is increased but respect well-resolved quantities such as an upper bound for porosity. This is noted in the asymmetrical nature of the limits at the reservoir layer in Figure 3e.

The rms errors for the first data combination (V_P, V_S, and density) and noise levels of 5%, 10%, and 15% are, respectively, 0.0163, 0.0181, and 0.0219 for porosity and 0.0493, 0.0564, and 0.0699 for clay volume. The data combination involving all logs together pro-

duces rms errors of 0.0136 (5% noise), 0.0147 (10%), and 0.0151 (15%) for porosity estimates and 0.0435 (5%), 0.0452 (10%), and 0.0477 (15%) for clay volume.

Figure 4 presents the resulting data fit and lithological classifications, in addition to details of the three-layer ($L1$, $L2$, and $L3$) synthetic model. These details include variations in lithology (porosity and clay volume) and water saturation S_w, as represented, respectively, by facies A, B, C, and D and by annotations of the fluid content (water or oil). Figure 4a–c, 4f, and 4g represents the final fit to the synthetic data (gray curve) by modeled data (black curve) using the posterior means. Notice in Figure 4f and 4g that a single set of coefficients for Han's equation produces a response very close to the one by Gassmann modeling (synthetic data). In the central part of Figure 4, one can find two classifications (Figure 4d and 4e) corresponding to each input data combination. The grayscale represents class probabilities and points indicate the class corresponding to maximum probabilities. Despite the larger uncertainties associated with estimates from the first data combination (V_P, V_S, and density) compared with the second combination (all logs), both combinations produce nearly the same classification result.

The synthetic tests presented here show that integrating information from independent data sets in posterior PDFs can reduce the effect of noise, as verified by comparing Figures 2 and 3. Analysis of rms errors indicates porosity errors decrease by 85%, when going from a single log inversion of V_P to a log combination of V_P, V_S, and density. For the same tests, clay volume also presents an rms error decrease of 85%. If we consider tests combining all logs (V_P, V_S, density, neutron porosity, and gamma ray), the rms error decreases by 88% for porosity and by 96% for clay volume. Alternatively, when considering porosity estimation from the density log (best case for a single log inversion), we observe an error reduction of 46% for the V_P, V_S, and density combination and of 56% for all logs combined. Also notice that whenever saturation is changing (i.e., in Figure 4, layer $L2$), it affects porosity and clay volume estimates, but the effect is relatively small after combining of all sources of information.

REAL DATA EXAMPLE

Our methodology is applied to a well-log data set from a lower Cretaceous sandstone of the Glauconitic Formation in the Blackfoot field, located 15 km southeast of Strathmore, Alberta, Canada. The target rocks are incised valley-fill sediments within the Glauconitic Formation. In the Blackfoot area, the Glauconitic sand thickness varies from 0 to more than 35 m. Three discontinuous phases of valley incision have been identified. The lower and upper members are made of quartz sandstone with 0.18 average porosity, while the middle member is characterized by low-porosity, tight lithic sandstone, and high clay volume. The lithic sandstone sequence is found from 1681–1689 m, indicated by the two horizontal lines in Figure 5a. Figure 5 also shows (from left to right) gamma-ray, neutron and core (gray points) porosity, density, V_P and V_S data from well 08-08, used for this example. The well logs are available with sample spacing of 0.125 m. Note in the lithic interval that the high
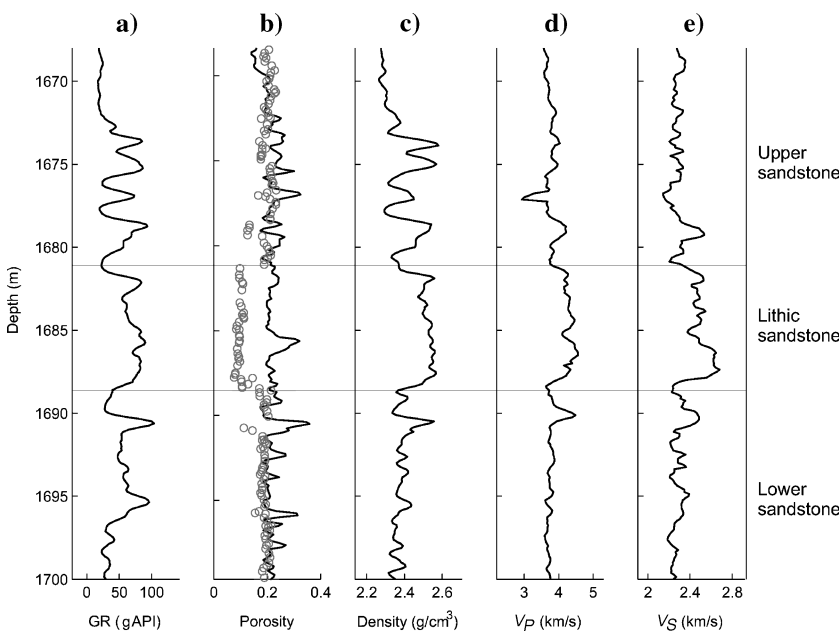

Figure 5. Well logs used in the real data example, consisting of (a) gamma ray (GR), (b) neutron porosity, (c) density, (d) V_P, and (e) V_S logs. The gray circles on the porosity plot represent the porosity derived from core analysis. Horizontal lines separate the three main lithological sequences.

clay volume leads to a higher porosity from the neutron log than the porosity derived from core samples. This shift clearly represents the effect of hydrogen in the clays on the neutron-log response.

One key step of this methodology is to define the rock physics models and the associated hyperparameters, which are required to construct the likelihood function. We use the same equations as in the synthetic example, but the problem now is how to determine all coefficients. In particular, the coefficients for velocity equations 9 and 10, as determined by Han et al. (1986), apply only to the set of rocks studied in their work. Although these authors show that the results may be extended to other consolidated sandstones, it is always important to validate the model by calibration, considering the specific set of rocks and data under investigation. To calibrate the model for this example, we use data from core sample analysis, which is presented in the next section.

Hyperparameters: Calibration of model coefficients

In a previous section, we gathered 13 unknown rock physics model coefficients corresponding to P- and S-wave velocities, and to neutron porosity, density, and gamma-ray logs. Here we estimate these hyperparameters using a model calibration procedure. The calibration is conducted on water-saturated intervals, where we assume that fluid composition and properties are known. In addition to well-log data, independent measurements of porosity from core analysis are used as reference values for the calibration.

The calibration is conducted in two stages. First, we analyze the relationship between clay volume, porosity, and density by standard cross-plotting (Figure 6) and interactive modeling. From the crossplot we take grain density ρ_q and the porosity for pure sand and then interactively adjust γ_{sh} and γ_{sd} for best estimates of a matrix effect a_{ϕ_N}, shale microporosity c_{ϕ_N}, clay density ρ_{cl}, and fluid density ρ_{fl} by regression analysis. The known fluid density is also used for quality control. Next, we use resulting clay volume from optimal gamma-ray transformation and core porosities, as reference data, to estimate the velocity-model coefficients of equations 9 and 10, i.e., a_{V_P}, b_{V_P}, c_{V_P}, a_{V_S}, b_{V_S}, and c_{V_S}. In this calibration procedure, we employ standard Bayesian regression analysis (Zellner, 1971), using normally distributed errors with unknown variance.

Following the same steps described for the Bayesian methodology, we obtain the posterior distributions for the V_P and V_S coefficients when porosity and clay volume are given. We then compute estimates for the coefficients to update Han's equations to comply with the lithologies under investigation. Figure 7 shows resulting marginal posterior distributions for (pairs and single) coefficients, corresponding to the V_P model (equation 9). Figure 7a–c shows the 2D joint marginal posterior PDFs $p_{ab}(a_{V_P}, b_{V_P}|\mathbf{d}_{V_P}, \phi, \chi)$, $p_{ac}(a_{V_P}, c_{V_P}|\mathbf{d}_{V_P}, \phi, \chi)$, and $p_{bc}(b_{V_P}, c_{V_P}|\mathbf{d}_{V_P}, \phi, \chi)$. The
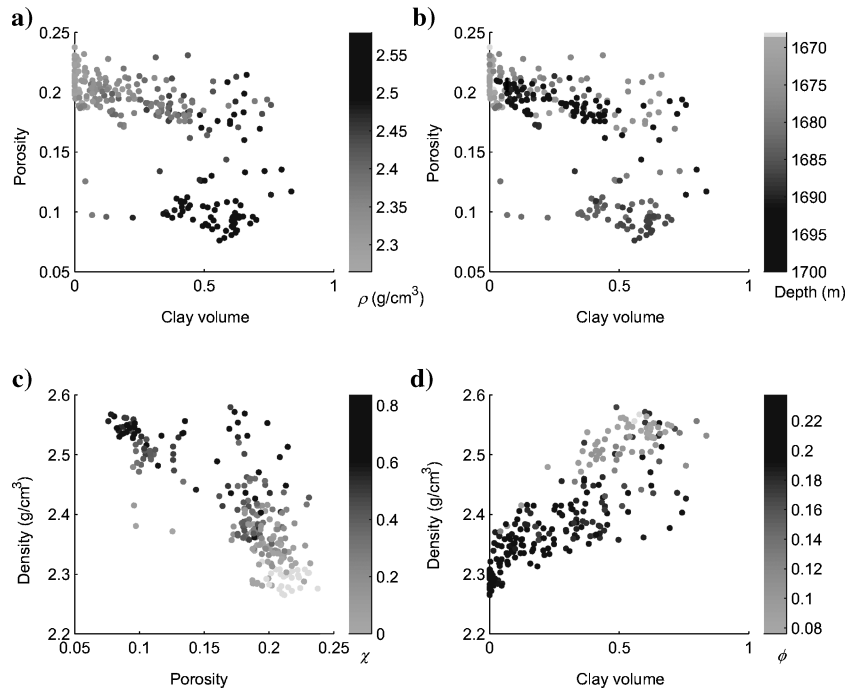

Figure 6. The first step in the calibration procedure is the analysis of crossplots depicting the relationships between clay volume χ, porosity ϕ, and density ρ.
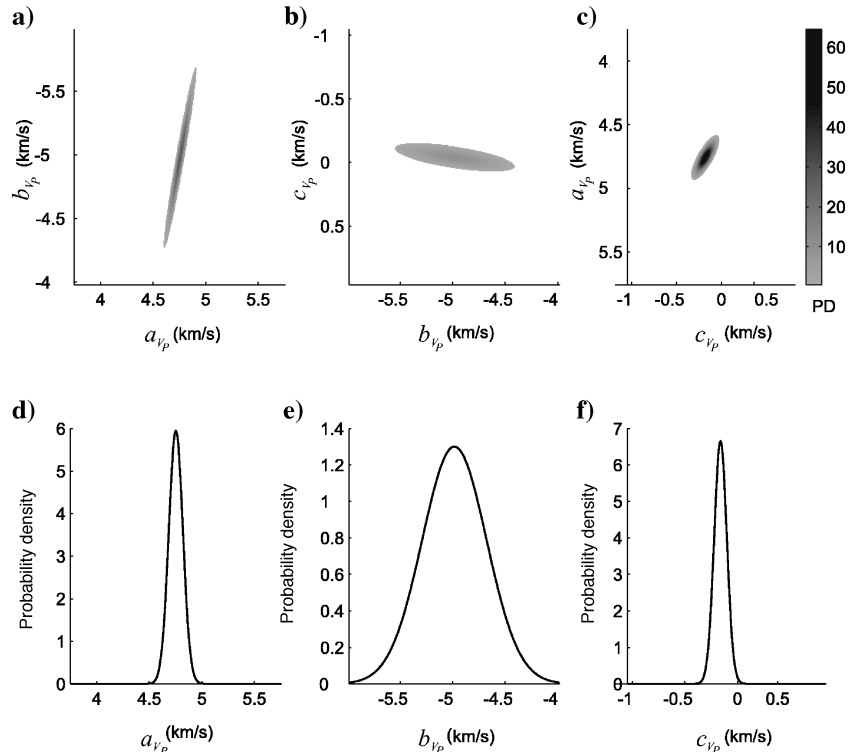

Figure 7. Marginal posterior distributions for the V_P model coefficients. (a)–(c) For pairs of coefficients $[(a_{V_P}, b_{V_P}), (a_{V_P}, c_{V_P}),$ and (b_{V_P}, c_{V_P}), respectively]. (d)–(f) For a single coefficient $[a_{V_P}, b_{V_P},$ and c_{V_P}, respectively]. Analogous results are obtained for V_S.
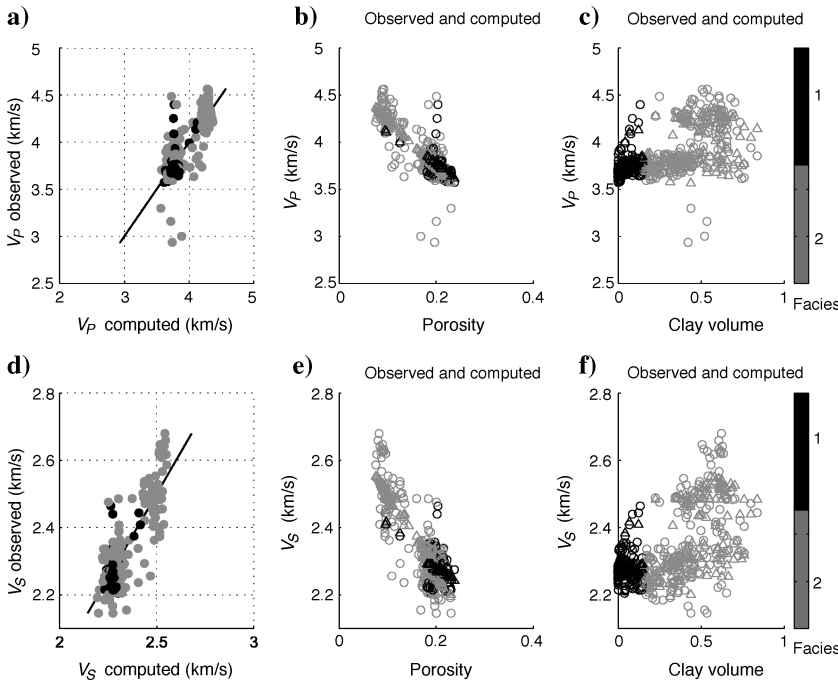
Figure 8. Calibrated models for (top) V_P and (bottom) V_S, using data from well 08-08. (a), (d) Computed wave velocities (equations 22 and 23) against the observed (sonic) velocities. (b), (e) Porosity against wave velocities (observed — circle; computed — triangle). (c), (f) Clay volume against wave velocities (observed — circle; computed — triangle). The grayscale separates samples above and below 0.20 of clay volume.
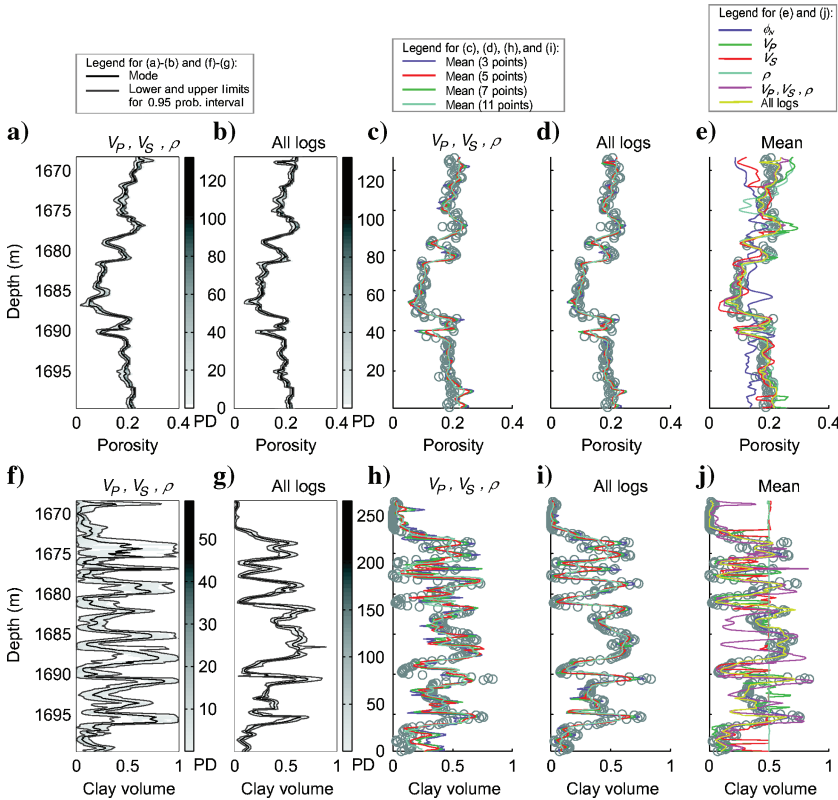

Figure 9. (Top) Porosity and (bottom) clay volume inversion results for the two input data combinations of logs, as indicated by plot titles (a), (b), (f), (g) marginal posterior PDFs. (c), (d), (h), (i) The effect of different moving window sizes (3, 5, 7, and 11 points, respectively). Gray circles represent the reference porosity (from core data) or clay volume (from gamma-ray log) values. (e), (j) Comparison of estimates using individual and combined input data sets.

1D marginal posteriors $p_a(a_{V_P}|\mathbf{d}_{V_P}, \phi, \chi)$, $p_b(b_{V_P}|\mathbf{d}_{V_P}, \phi, \chi)$, and $p_c(c_{V_P}|\mathbf{d}_{V_P}, \phi, \chi)$ are shown in Figure 7d–f. Notice the elliptical form in the joint PDF images, which indicates the existence of some degree of correlation resulting from the linear relationship among the velocity coefficients. The uncertainty associated with each coefficient is best seen on the 1D marginals for the individual coefficients. Very similar results are obtained for S-wave velocity. The final velocity models for the sands are given by

$$V_P = 4.40 - 3.32\phi + 0.33\chi \quad (22)$$

and

$$V_S = 2.56 - 1.32\phi - 0.17\chi. \quad (23)$$

As before, porosity and clay volume are expressed in volume fractions, and model coefficients are given in kilometers per second.

We evaluate equations 22 and 23 for the data from well 08-08 using the calibration data, which are porosity from core measurements and clay volume from gamma-ray data. To verify the calibration, we plot observed and computed velocities (Figure 8a and 8d) and velocities against either porosity (Figure 8b and 8e) or clay volume (Figure 8c and 8f). On the last four plots, observed values are represented by circles and computed values are noted by triangles. The calibration procedure separates data with clay volume above a given threshold value, yielding a separate set of model coefficients for sand and shales. We use a threshold value of 0.20, for obtaining the shales $(4.78, -5.21, -0.03)$ km/s for V_P and $(2.68, -2.21, 0.08)$ km/s for V_S.

Next, we perform the inversion using the calibrated rock physics models in the likelihood functions for the joint inference of porosity and clay volume.

Porosity and clay volume inferences

In the real data inversion, we follow the same procedures as in the synthetic example, running several inversions using different combinations of (or individual) data sets and moving window sizes. In Figure 9, the first two plots in each row show resulting posterior densities obtained from the inversion of two data combination, which are velocities and density [Figure 9a is $p(\phi|\mathbf{d}_{V_P}, \mathbf{d}_{V_S}, \mathbf{d}_\rho)$ and Figure 9f is $p(\chi|\mathbf{d}_{V_P}, \mathbf{d}_{V_S}, \mathbf{d}_\rho)$] and all data [Figure 9b is $p(\phi|\mathbf{d}_{\phi_N}, \mathbf{d}_{V_P}, \mathbf{d}_{V_S}, \mathbf{d}_\rho)$ and Figure 9g is $p(\chi|\mathbf{d}_{\phi_N}, \mathbf{d}_{V_P}, \mathbf{d}_{V_S}, \mathbf{d}_\rho, \mathbf{d}_\gamma)$]. Additional plots show the estimates in comparison with the corresponding reference values (gray circles) used in the calibration procedure, i.e., core porosity and clay volume from gamma-ray data. The first set of additional plots (Figure 9c, 9d, 9h, and 9i) show the effect of the window size. Inversions are per-

formed using three (0.375 m), five (0.625 m), seven (0.875 m), and eleven (1.375 m) data points per window. Notice the estimated curves become smoother as the window size increases. We select the seven-point window for computing the posterior distributions. At the 11-point window (1.375 m), resolution starts to decrease considerably. Figure 9e and 9j show estimates using the mean of the marginal posterior distributions, considering a variety of input data including individual and multiple logs. These plots show, for example, how porosity estimates from different data types are affected by their respective sources of uncertainty. Final estimates are given by the posterior mean (yellow curve) using all logs for the inversion.

Figure 10 shows the resulting fit to the well-log data (gray) generated by computed data (black) using the posterior mean. We also run the lithological classification (Figure 10d and 10e), as presented in the synthetic data example, using the class limits similar to those suggested by Vernik and Nur (1992).

We focus our analysis of the results on the interval of interest, ranging from 1675 m to 1680 m (reservoir sands), and the zone of high clay volume from 1683 m to 1692 m. On the latter, the neutron log shows the effect of high clay content, which leads to overestimation of porosity with respect to core porosity. This effect, which is clearly noticeable by the blue curve in Figure 9e, disappears after combining the various data sets in the inversion. Additionally, the overall porosity estimates are also good outside this interval for both data combinations. In contrast, clay volume inversion displays large associated uncertainty in Figure 9f and 9g, especially when using only velocities and density. The improvements on inference of clay volume, after incorporating the gamma-ray log, is somewhat artificial because the clay volume from gamma-ray data has been used as a reference in the calibration stage, given the lack of core information about clay volume. Despite these negative aspects observed in the inversion of clay volume, the resulting classification, performed using the entire collection of marginal posterior distributions for porosity and clay volume, closely agrees for both data combinations. This can be verified by comparing Figure 10d, which shows classification from velocities and density, and Figure 10e, which is all logs combined.

Figure 10. Comparison of observed (gray) and final computed (black) well-log curves for (a) gamma ray, (b) neutron porosity, (c) density, (f) V_P, and (g) V_S. Plots (d) and (e) show classifications performed using marginal distribution resulting from inversion of different well-log combinations.

DISCUSSION

The cornerstone for the development of our methodology is selecting the modeling relations between the data, porosity, and clay volume. The integration of mathematical models derived from rock physics studies, in an inversion context, constitutes an alternative to the traditional practice of using those models as a simple formula for transforming variables. It also offers advantages over conventional regression methods, given extra inference questions that can be addressed to the marginal posterior distributions. The examples presented include the uncertainty analysis and lithological classification. However, the choice of the most appropriate modeling equations from rock physics is still an issue to be investigated. Empirical
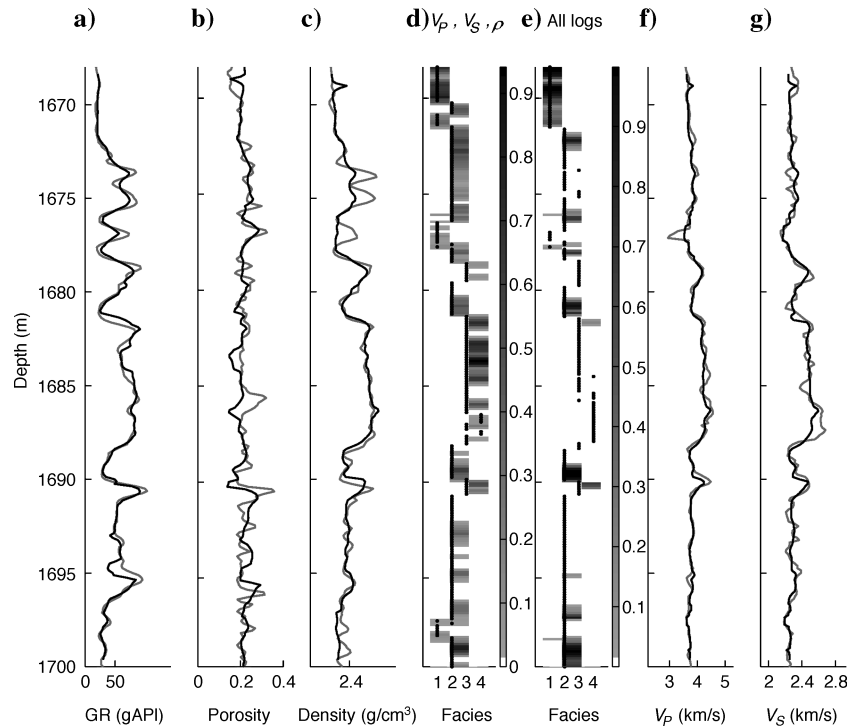
models require a data-based calibration procedure for adjusting the regression coefficients and more complex computational algorithms for full modeling uncertainties. Our calibration approach can be considered an entry-level solution to the full problem. The next level would be to find approximate solutions, such as given by the empirical Bayesian approach of Carlin and Louis (2000) [see also Malinverno and Briggs (2004) and Mitsuhata (2004) for geophysical applications].

There are formation effects that have been neglected in this study because of modeling limitations. Some of these effects are the result of pressure, fluid saturation, cementation, and borehole effects. Currently, it is impractical to include in the inference problem a large number of physical parameters that contribute to well-log response. Limitations come either from the modeling equations, which may not be developed enough to include these additional effects, or from the strong ambiguity involved in some multiparameter inverse problem. Our methodology, however, exploits the complex combination of errors from different sources by integrating multiple data sets. This leads to a great number of cancellations and the enhancement of common information, minimizing the impact of unmodeled effects.

CONCLUSION

Our analysis of overall rms error associated with the estimates in the synthetic data example substantially improves the quality of estimates because we incorporate more data sets into the inversion. Sources of systematic errors are present in both of our examples, represented by changes in saturation and lithology. Their corresponding effects are reduced greatly by combining information from different data sets. Such an ability to reduce the effects of random and system-

atic errors makes our methodology useful for real-time applications of well-log data interpretation. In addition, examples involving multiple data sets demonstrate that the methodology offers resources for handling a considerable amount of information. In fact, our formulation can be implemented readily in two or three dimensions with relative efficiency. Consequently, the methodology can also be used in a reservoir characterization workflow incorporating core, wireline, and seismic (attribute) data.

The results presented in the form of marginal posterior PDFs provide adequate elements for quantitatively and qualitatively evaluating the information content (or uncertainty) in each well-log data set and the improvements resulting from combining different logs. Our analyses show that neutron porosity and density data are the most informative about porosity, whereas gamma-ray and velocity data are the most informative about clay volume. The marginal PDFs for clay volume show that the uncertainty associated with clay volume may be too high. However, resulting information about clay volume can be used reliably, in combination with porosity information, to classify sandstones.

The size of the moving window is an important parameter; it must be selected to maintain the balance between noise reduction and vertical resolution. Best moving-window sizes must be selected on a case-by-case approach using a similar analysis of different window-size results.

ACKNOWLEDGMENTS

REFERENCES

Box, G. E. P., and G. C. Tiao, 1973, Bayesian inference in statistical analysis: Addison-Wesley Publishing Company.

Bretthorst, L., 1990, An introduction to parameter estimation using Bayesian probability theory: Proceedings of the Maximum Entropy and Bayesian Method Conference, 53–79.

Carlin, B. P., and T. A. Louis, 2000, Bayes and empirical Bayes methods for data analysis: Chapman and Hall/CRC Press.

Castagna, J. P., M. L. Batzle, and R. L. Eastwood, 1985, Relationships between compressional-wave and shear-wave velocities in elastic silicate rocks: Geophysics, **50**, 571–581.

Domenico, S. N., 1984, Rock lithology and porosity determination from shear and compressional wave velocity: Geophysics, **49**, 1188–1195.

Eberhart-Phillips, D., D. Han, and M. D. D. Zoback, 1989, Empirical relationships among seismic velocity, effective pressure, porosity, and clay content in sandstone: Geophysics, **54**, 82–89.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004, Bayesian data analysis: Chapman and Hall/CRC Press.

Gouveia, W., and J. A. Scales, 1998, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis: Journal of Geophysical Research, **103**, 2759–2779.

Han, D., A. Nur, and D. Morgan, 1986, Effects of porosity and clay content on wave velocities in sandstones: Geophysics, **51**, 2093–2107.

Hearst, J. R., P. H. Nelson, and F. L. Paillett, 2000, Well logging for physical properties: John Wiley & Sons, Inc.

Jaynes, E. T., 1968, Prior probabilities: Institute of Electrical and Electronics Engineers Transactions on Systems and Cybernetics, **SSC-4**, 227–241.

——, 1982, On the rationale of maximum-entropy methods: Proceedings of the Institute of Electrical and Electronics Engineers, **70**, 939–952.

——, 2003, Probability theory — The logic of science: Cambridge University Press.

Jeffreys, H., 1939, Theory of probability: Oxford University Press.

Kass, R. E., and L. Wasserman, 1996, The selection of prior distributions by formal rules: Journal of the American Statistical Association, **91**, 1343–1370.

MacKay, D. J. C., 2003, Information theory, inference, and learning algorithms: Cambridge University Press.

Malinverno, A., and V. A. Briggs, 2004, Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes: Geophysics, **69**, 1005–1016.

Mitsuhata, Y., 2004, Adjustment of regularization in ill-posed linear inverse problems by the empirical Bayes approach: Geophysical Prospecting, **52**, 213–239.

Moraes, F. S., and J. A. Scales, 2000, Local Bayesian inversion: Theoretical developments: Geophysical Journal International, **141**, 713–723.

Raymer, D. S., E. R. Hunt, and J. S. Gardner, 1980, An improved sonic transit time-to-porosity transform: 21st Annual Meeting, Society of Professional Well Log Analysts, paper P.

Tosaya, C., and A. Nur, 1982, Effects of diagenesis and clays on compressional velocities in rocks: Geophysical Research Letters, **9**, 5–8.

Ulrych, T. J., M. D. Sacchi, and A. Woodbury, 2001, A Bayes tour of inversion: A tutorial: Geophysics, **66**, 55–69.

Vernik, L., 1998, Acoustic velocity and porosity systematics in siliciclastics: The Log Analyst, **39**, 27–35.

Vernik, L., and A. Nur, 1992, Petrophysical classification of siliciclastics for lithology and porosity prediction from seismic velocities: AAPG Bulletin, **76**, 1295–1309.

Wyllie, M. R. J., A. R. Gregory, and L. W. Gardner, 1956, Elastic wave velocities in heterogeneous and porous media: Geophysics, **21**, 41–70.

Zellner, A., 1971, An introduction to Bayesian inference in econometrics: John Wiley & Sons, Inc.