Research Article

Development and Validation of a Novel Interpretation Algorithm for Enhanced Resolution of Well Logging Signals

Qiong Zhang¹ and Jean-Baptist Peyaud²

¹University of Electronic Science and Technology, China

²Fluid Minerals Interactions, Australia

Correspondence should be addressed to Qiong Zhang; zhanqio@uestc.edu.cn

Received 14 December 2020; Revised 18 December 2020; Accepted 23 December 2020; Published 8 January 2021

Academic Editor: Bin Gao

Copyright © 2021 Qiong Zhang and Jean-Baptist Peyaud. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work presents a novel algorithm that achieves enhanced resolution of well logging signals, e.g., from 1 ft of a pulsed neutron mineralogy tool to 0.04 ft of an imaging tool. The algorithm, denoted as “Digital Core,” combines mineralogical and sedimentological information to generate a high-resolution record of the formation mineralogy which can be consequently applied to thin bedded environments. The keystone to the philosophy of this algorithm is that the spectral information recorded by mineralogy tool is a weighted average of the mineralogy of each lithological component in the analyzed volume. Therefore, by using a high-resolution image log to determine the proportion of each lithological component, their composition can be determined from the mineralogy log data. A field case from a well located in South Australia is presented in this work, and the results validate the feasibility of an integrated core-level petrophysical analysis in a cost-effective and timely manner compared to conventional core measurements.

1. Introduction

Oil well logging has been known for many years and provides an oil and gas well driller with information about the particular earth formation being drilled. Accurate and detailed knowledge of earth formations that may contain reservoirs of the hydrocarbons is required for the exploration and production of hydrocarbons [1–3]. Typically, the recording of rock physical properties (logs) is the only information available in a borehole. These include the resistivity of the formation [4], the acoustic properties [4], the porosity (actually the interaction between neutron and hydrogen in the formation) [4], the electronic density of the formation [4], the natural radioactivity, and the gamma emission induced by neutron stimulation. These data are then interpreted in terms of mineralogy of the formation, matrix density, total and effective porosity, fraction of hydrocarbon in the fluids, etc. For example, it is important to know the lithology of the earth formations as a function of depth, particularly in thinly bedded formations. A lithology characterizing a formation may be determined using one or more of several techniques. A

common technique is to retrieve core samples from an earth formation and perform intensive analysis of the core sample at the surface. Typically, this is conducted off-site at a specialized facility remote from the well site. While core samples can provide the detailed knowledge petroanalysts and geophysicists desire, obtaining the samples from deep within the earth formation and performing the analysis can be quite time intensive [2].

A neutron logging tool may be used to obtain lithology parameters by measuring radiation resulting from neutron irradiation of the earth formation. The measured radiation is indicative of the reaction of the neutrons with constituents of the formation and thus contains information about the earth formation. As one example, interactions between the neutrons and the formation may result in the emission of gamma rays with energy levels characteristic of the materials with which the neutrons have interacted [2]. Measurements are repeated at several borehole depths along the longitudinal axis of the borehole. Each measurement is associated with a borehole depth at which it is taken. Unfortunately, however, neutron logging generally provides a coarse vertical

resolution, i.e., along the borehole axis, between approximately one and two feet. This resolution is insufficient to locate boundaries of thin beds (e.g., beds less than one foot in thickness). Thus, at the current level of technology, a direct measurement is not possible as these objects are commonly far below the vertical resolution of logging tools. For example, a sandstone with 20% clays could correspond to a “dirty” sand with dispersed clay minerals or a “clean” sand with a few argillaceous layers. The properties of the sandstone in terms of flow dynamics, vertical connectivity, porosity, and water content will be very different depending on the actual geological makeup of the formation [3]. This can lead to by-passed pay and/or a misunderstanding of the layer properties. Getting a higher vertical resolution is therefore crucial in complex environments.

In this work, a method is proposed to integrate the sedimentological information from a borehole image and the information from a spectroscopic mineralogy log to improve the vertical resolution of the mineralogy log. The image log provides information about layering: depending on the tool used to acquire the data, a layer is marked by a contrast in resistivity or acoustic properties. Layers are commonly associated to sedimentary beds, but they can be occasionally related to the formation of nodules or the development of diagenetic cements [4]. The vertical resolution of the record is of the order of half an inch. The spectroscopic log provides information about the bulk chemistry of the formation from which the mineralogy is derived. Its vertical resolution is of the order of one foot. The first stage involves comparing the image and spectroscopic logs over the same interval to determine the level of complexity of the logged formation: if the formation displays a layering thinner than the resolution of the spectroscopic mineralogy log, the formation is deemed “thinly bedded” [4]. The algorithm proposed in this work applies to these situations. The borehole is split into intervals ranging from one to several feet in length by clustering similar regions based on the log response of both the mineralogy and the borehole image. Within each interval, lithotype facies, identified as layers of similar properties, are determined based on the image log. In the scope of this work, the term “lithotype” refers to a geological unit characterized by a set of parameters, such as, for example, specific lithology, mineralogical composition, porosity, permeability, grain size distribution, sedimentological texture, and sedimentological structures. The term “facies” refers to a specific range of the values of these characteristics that characterize a body of rock and allow discriminating it from its surroundings either by way of measurement, observation, or both. For example, natural gamma radiation logs and neutron-induced radiation logs may provide accurate identification of lithotype facies, but with a coarse vertical resolution of approximately two feet in some tools. A high-resolution borehole image log may provide accurate vertical resolution of resistivity and changes in resistivity as small as a few millimeters but with limited ability to identify lithotype facies or minerals.

By employing a high-resolution image log, the volume of each lithological component in the considered interval can be determined. Several classes of high-resolution image logs are available for this purpose: high-resolution electromagnetic

imager (e.g., resistivity imager), acoustic imager, and optical image log. From the spectroscopic log, we know the overall composition of the interval, but not the detailed composition of each layer. The mineralogical composition associated with each lithotype is determined by solving a constrained optimization problem [3] that maximizes the likelihood of the determined layer composition. For solving the optimization problem, a constrained sampling algorithm specifically generated for this purpose is utilized and presented in this work.

This work presents in detail the developed mathematical algorithm and methods for combining information obtained from a high-resolution log and low-resolution spectroscopic tool. The targeted application of this method is thin-bedded formations. An example is provided from a well in Australia where spectroscopic mineralogy, a resistivity borehole image, and a core were available. The development of this method allows to increase the resolution of the mineralogy log to the level of the image log: what can be recorded on the borehole image can be discriminated in the mineralogy log. This method will enhance our understanding of complex environments, for instance the distribution of organic matter or the distribution of argillaceous beds in reservoir formations. It allows a better understanding of the reservoir properties and their vertical distributions, a more precise location of hydrocarbon-bearing intervals, and can be used to extend core information more accurately over the logged interval. The higher resolution information could also greatly benefit well completion in cases where stimulation is required or in optimizing the completion design. More specific embodiments may also provide mineralogy logs, matrix density logs, and total porosity logs with a high vertical resolution, as well as allowing calculation of net-to-gross and net pay in thin bed formations.

2. Mathematical Theory and Algorithm Implementation

This section provides an overview of the mathematical theory based on which the algorithm is constructed.

A lithology, e.g., sandstone, limestone, or shale, is characterized by a set of measurable parameters such as grain size distribution and composition [4]. The composition can simply be thought of as being characterized by the relative amount of different minerals contained in the lithology. As an example, sandstone contains more quartz than coal but coal contains more organic matter [4]. However, even among lithologies that are categorized into the same group, the composition can vary. Therefore, it is convenient to characterize the content of a certain mineral in a class of lithologies by a probability density functions. These probability density functions characterize the likelihood of finding a certain mineral with a given concentration within the lithology of interest.

Within this work, we are interested in inferring the mineral compositions of geological layers contained in a logging interval from two separate sources of data: (1) from a mineralogical logging tool of coarse resolution; i.e., it gives the average mineralogy over all lithotypes; (2) an image tool that does not provide mineralogical information but has a much finer resolution and can be used to delineate layer

boundaries within the interval. For accomplishing this task, a novel mathematical optimization algorithm is developed to find the most likely combination of lithotypes assigned to the identified layers along with the mineralogical composition of each layer. The solution to the optimization problem must obey several constraints that will be detailed within this section.

The basic idea of the algorithm is to combine information from a facies analysis with the mineralogical information from a mineralogy log derived from spectral data. For each mineral indexed by $m = 1, \dots, M$, the mineralogy log data provides a mineral volume fraction denoted by I_m integrated over the whole borehole interval. The facies analysis provides the depth of the boundaries between different layers measured from the surface that can be used to compute the volume fractions ξ_j of layers $j = 1, \dots, J$. For each possible lithotype $l = 1, \dots, L$, M pdfs denoted by $p_{m,l}(x_{m,l})$ are available, one for each mineral describing the likelihood of finding the mineral volume fraction within this lithology as follows:

$$p_{m,l}(x_{m,l})dx_{m,l} = \text{probability of mineral } m \text{ in lithology } l \quad (1)$$

to have volume fraction $x_{m,l}$.

From the mineralogy and image log data, it is unknown which lithotypes are present in an interval. A proposition of lithotypes is defined as an assignment of a lithotype to each layer in the interval, i.e., the map $j \leftarrow l$ for $j = 1, \dots, J$. There are a total of $L!/(L-J)!$ combinations all of which have to be analyzed. It will later be demonstrated that the vast majority of these combinations is infeasible because they cannot satisfy the constraints. Nevertheless, a smart preselection of relevant lithotypes by experienced users might be necessary if L is large.

In Section 2.1, the probability density functions $p_{m,l}(x_{m,l})$ are discussed with particular focus on where the pertinent data can be obtained from. For using a MCMC method [5] to find the most likely $x_{m,j}$ for all m and j , we need to be able to create uniformly distributed samples over the space of all permissible $x_{m,j}$. To this end, constraints and an efficient sampling algorithm are described in Sections 2.2 and 2.3, respectively.

2.1. Probability Density Functions. The goal of the presented algorithm is to find the most likely combination of mineral volume fractions $x_{m,l}$ that satisfies all constraints detailed in Section 2.2. The likelihood of a combination is given by the joint probability density function $p(\vec{x})$, where for convenience of notation, we collect the $x_{m,l}$ in the vector $\vec{x}$. Within this work, the joint probability function is simply the product of the single mineral probability density functions:

$$p(\vec{x}) = \prod_{m,l=1}^{M,L} p_{m,l}(x_{m,l}). \quad (2)$$

The single probability density functions $p_{m,l}(x_{m,l})$ are input to the presented algorithm and must be specified by

the user. This appears to be an equally or maybe more difficult task than to infer the $x_{m,l}$ manually using the geophysicist's experience and knowledge. For making the developed method practical, it is necessary to create a database of $p_{m,l}(x_{m,l})$ that can be reused for subsequent analysis. We decided to implement a bootstrapped-learning algorithm. At first, the geoscientist analyzing the dataset creates probability density functions following geological insight and ensures that the outcome matches the expectations. Once a sufficiently reliable database of probability density functions has been obtained, the algorithm described in Section 2 can be used. As more log data is analyzed, the process is a two-way street: existing probability density functions can be used to analyze results, and actual results can be used to enhance the probability density functions. Missing probability density functions are successively added.

For creating a useful library of probability density functions, we group sets of probability density functions by similarity with respect to general geological indicators, e.g., lithotypes. These geological indicators are known before logging commences at a particular location. A suitable set of probability density functions is retrieved from a database using the set of indicators for the logging location. Note that depending on the general geological features of a location, some minerals or lithologies may be present in one data set but absent in another. It is important to point out that the set of pdfs associated with the particular set of indicators is extended and possibly enhanced as analysis progresses by adding missing probability density functions and fixing inconsistencies of analysis results with observed mineral compositions.

2.2. The Setup of Constraints. Both equality and inequality constraints apply to the solution of the optimization problem. The obvious inequality constraints simply state that the volume fractions need to be between zero and one:

$$0 < x_{m,j} < 1. \quad (3)$$

However, most pdfs are zero almost everywhere, and therefore, it makes sense to set somewhat tighter bounds on the volume fraction $x_{m,j}$:

$$\alpha_{m,j} < x_{m,j} < \beta_{m,j}. \quad (4)$$

There are two types of equality constraints present in the optimization problem. The first type of constraint incorporates the knowledge of the interval integrated mineralogy into the optimization problem. It states that the sum of mineral contents over all layers weighted by the volume fraction of each layer must equal the mineral content determined by the mineralogy tool:

$$\sum_{j=1}^J \xi_j x_{m,j} = I_m. \quad (5)$$

FIGURE 1

Secondly, the volume fractions of all minerals within each layer must sum to unity:

$$\sum_{m=1}^M x_{m,j} = 1. \quad (6)$$

In addition, a lithology l assigned to layer j might have a zero probability to have any traces of mineral m in it and in such cases makes sense to strictly enforce

$$x_{m,j} = 0, \quad (7)$$

to avoid numerical issues in the solution process. There are a total of $M + J$ equality constraints of types given by Equations (5) and (6) and P equality constraints of type Equation (7). The total number of equality constraints is $K = P + M + J$.

We want to create uniformly distributed samples of $x_{m,j}$ that satisfy Equations (5)–(7). To this end, we collect the equality constraints into the matrix E :

$$E \vec{x} = \vec{r}, \quad (8)$$

where $\vec{x}$ collects the $x_{m,j}$, E collects the coefficients of Equations (5)–(7), and $\vec{r}$ collects the right hand sides of Equations (5)–(7).

The matrix E is rank deficient nominally containing more columns than rows (short and wide) and has a nullspace of rank $N = M \cdot J - K$. Usually, the matrix is made square by adding N rows consisting of all zeros. As there always exists a vector $\vec{x}$ that satisfies Equation (8), it follows that there

are infinitely many solutions to Equation (8) as opposed to none if $\vec{r}$ was not in the range of E .

Basis vectors spanning the null space of E can be determined by using SVD [6]. These orthonormal basis vectors are denoted by $\vec{z}_n, n = 1, \dots, N$, and are collected in the matrix Z of dimensions $M \cdot J \times N$ (it is tall and skinny). The complete set of solutions of Equation (8) can then be written as follows:

$$\vec{x} = \vec{x}_0 + Z\vec{q}, \quad (9)$$

where $\vec{x}_0$ is a particular solution that can, e.g., be computed by using the Moore-Penrose inverse [7]:

$$\vec{x}_0 = E^{-1} \vec{r}. \quad (10)$$

The vector $\vec{q}$ has fewer entries than the vector $\vec{x}$ and denotes a space of reduced dimensionality in which the solution lives. In Figure 1, a three-dimensional space, in which there are three $x_{m,j}$, is depicted that has a single constraint. In this case, the null space is of dimension two permitting samples to be located on a plane. In general settings, the space of permissible samples is a N -dimensional hyperplane in a $M \cdot J$ dimensional space.

Using Equation (9), we can use uniform samples of $\vec{q}$ for creating uniform samples of $\vec{x}$ [8]. However, these samples do not necessarily satisfy the inequality constraints, because it is unclear within which limits we need to sample q_n to limit the $x_{m,j}$ to within $\alpha_{m,j}$ and $\beta_{m,j}$ [3].

The inequality constraints Equation (4) define a $M \cdot J$ -dimensional hyperbox that the valid samples living on the

N-dimensional hyper-plane given by Equation (9) are constrained to. Figure 1 depicts the hyperbox and the hyperplane intersecting it. Valid samples are located on the fraction of the hyperplane that intersects the box and are colored in red.

2.3. Constrained Sampling Algorithm Development. The simplest possible algorithm would be to sample $\vec{q}$ in a large enough space enclosing the hyperbox, e.g., the smallest ellipsoid that fully includes the box, and then use rejection sampling. However, as the dimensionality of the space is much larger than three (usually about 15–20), rejection sampling will be extremely inefficient. This is fairly typical for high-dimensional spaces and is usually referred to as the curse of dimensionality [3].

The algorithm described here is a novel and highly efficient algorithm that does not require rejection and therefore has great capability in addressing high dimension problems as are typical in the described mineralogy analysis.

First, let us assume that we have a seed vector $\vec{x}$ satisfying Equations (4) and (9). A bootstrapping algorithm for obtaining this seed vector is detailed in Section 2.4. Then, we can sample another random vector $\vec{y}$ satisfying the equality constraints Equation (5) through (7) but not the inequality constraints Equation (4). This can be accomplished by first randomly sampling $\vec{q}$ followed by using Equation (9). The difference vector $\vec{d} = \vec{x} - \vec{y}$ lies within the hyperplane defined by the inequality constraints, and all points located on the line drawn through the end points of $\vec{x}$ and $\vec{y}$ are given by

$$\vec{p} = \vec{x} + s \vec{d}, \quad (11)$$

where s is the step size that needs to be determined. The limits on the step size are computed by first computing $s_{m,j,\min}$ and $s_{m,j,\max}$:

$$\begin{aligned} s_{m,j,\min} &= \min \left(\frac{\alpha_{m,j} - x_{m,j}}{d_{m,j}}, \frac{\beta_{m,j} - x_{m,j}}{d_{m,j}} \right), \\ s_{m,j,\max} &= \max \left(\frac{\alpha_{m,j} - x_{m,j}}{d_{m,j}}, \frac{\beta_{m,j} - x_{m,j}}{d_{m,j}} \right). \end{aligned} \quad (12)$$

The constraint for s are given by

$$\min(s_{m,j,\min}) < s < \max(s_{m,j,\max}). \quad (13)$$

The simplest Algorithm 1 that yields uniformly distributed samples is to perform a random walk as follows:

However, in practice, we found that the following approach performed better given the same execution time:

2.4. Obtaining the Seed Vector. Within this section, an algorithm for obtaining the first seed vector denoted by $\vec{x}$ in Algorithm 2 is introduced. However, a seed vector satisfying inequality and equality constraints does not always exist. It only exists if the hyperplane defined by the equality constraints intersects the box defined by the inequality con-

1. Uniformly sample $\vec{q}$.
2. Compute $\vec{y} = \vec{x} + Z\vec{q}$.
3. Compute $\vec{d} = \vec{x} - \vec{y}$.
4. Uniformly sample s in between $s_{m,j,\min}$ and $s_{m,j,\max}$.
5. Compute $\vec{x}' = \vec{x} + s\vec{d}$.
6. Set $\vec{x} = \vec{x}'$ and go to 1.

ALGORITHM 1: Random walk algorithm.

straints. Therefore, we first check if a solution exists which is detailed in Section 2.4.1. If a solution exists, obtaining the seed vector is trivial from the previously obtained information. Nevertheless, for the sake of completeness, we describe this procedure in Section 2.4.2.

2.4.1. Check Solution Existence. To facilitate the existence check, we first introduce a coordinate transformation such that limits $\alpha_{m,j}$ and $\beta_{m,j}$ are mapped to zero and one, respectively. Quantities expressed in these new coordinates are denoted by a tilde to distinguish them from the original coordinate system. The coordinate transformation is given by

$$\tilde{x}_{m,j} = \frac{1}{\beta_{m,j} - \alpha_{m,j}} x_{m,j} - \frac{\alpha_{m,j}}{\beta_{m,j} - \alpha_{m,j}}. \quad (14)$$

For compactness of notation, we collect all the coefficients of the first term into the matrix T and the second term of the different into the vector $\vec{b}$ such that we can write

$$\vec{\tilde{x}} = T\vec{x} + \vec{b}. \quad (15)$$

All permissible solution, i.e., all solutions on the hyperplane, can now be expressed in this new coordinate system:

$$\vec{\tilde{x}} = TZ\vec{q} + T\vec{x}_0 + \vec{b}. \quad (16)$$

The transformation transforms the hyperbox to be the unit cube around the midpoint $\vec{l} = [1/2, \dots, 1/2]$. The unit cube corresponds to the locus of points whose distance from $\vec{l}$ is less than or equal to 1/2 in the maximum norm:

$$\text{Interior of the unit cube : } \left\| \vec{l} - \vec{\tilde{x}} \right\|_{\infty} \leq \frac{1}{2}. \quad (17)$$

The vector $\vec{\tilde{x}}$ is the locus of all points on the hyperplane as given by Equation (9). Hence, by solving the linear optimization problem,

$$g = \min_{\vec{q}} \left\| \vec{y} - P\vec{q} \right\|_{\infty}, \quad (18)$$

$$\vec{y} = \vec{l} - T\vec{x}_0 - \vec{b}, \quad (19)$$

$$P = TZ, \quad (20)$$

1. Uniformly sample $\vec{q}$.
2. Compute $\vec{y} = \vec{x} + Z\vec{q}$.
3. Compute $\vec{d} = \vec{x} - \vec{y}$.
4. Perform a line search between $s_{m,j,\min}$ and $s_{m,j,\max}$ to find the maximum $\vec{x}^*$ of the joint probability density $p(\vec{x})$.
5. Set $\vec{x} = \vec{x}^*$ and go to 1.

ALGORITHM 2: Line search algorithm.

FIGURE 2

the condition for the hyperplane to intersect the hyperbox simply is follows:

$$g \leq \frac{1}{2}. \quad (21)$$

The solution to the minimization problem Equation (20) is obtained by transforming it into the corresponding linear programming problem [9, 10]:

Minimize t subject to

$$\begin{aligned} \vec{y} - P\vec{q} &\leq t\vec{1}, \\ \vec{y} - P\vec{q} &\geq -t\vec{1}, \end{aligned} \quad (22)$$

where $\vec{1}$ is a vector of the correct length of all ones.

2.4.2. Algorithm for Obtaining a Seed Vector. Once existence of a solution is established, one can simply use the point on the hyperplane with the smallest distance in the infinity norm to the center of the box and transform it back to the original basis. To this end, let us denote the solution as the minimization problem as $\vec{q}^*$. Then, we substitute $\vec{q}^*$ into Equation (16) and apply the inverse transformation:

$$\vec{x} = T^{-1} \left(TZ\vec{q}^* + T\vec{x}_0 + \vec{b} - \vec{b} \right) = \vec{x}_0 + Z\vec{q}^*. \quad (23)$$

While $\vec{x}$ is a valid seed vector for the random walk algorithm, it was found that it could be far away from the maximum. A better initial guess can be obtained by moving it closer to where we expect the maximum. For each one-

dimensional pdf used for forming the joint pdf in Section 2.1, we can easily infer where it assumes its maximum value. Let us denote this point as $x_{m,j}^{\max}$ and collect all of these abscissas in the vector $\vec{x}^{\max}$. Then, we estimate the maximum of the joint pdf to be close to $\vec{x}^{\max}$, but we note that it does not satisfy either the equality or inequality constraints. Therefore, we project the vector onto the hyperplane described by Equation (9) and denote the projection as $\vec{x}_{\perp}^{\max}$. While $\vec{x}_{\perp}^{\max}$ is guaranteed to satisfy the equality constraints, it could possibly be in violation of the inequality constraints. For approximating the best possible estimate close to $\vec{x}_{\perp}^{\max}$, we solve the following one-dimensional maximization problem for obtaining the best step size s_{step} :

$$s_{\text{step}} = \max_s p \left(\vec{x} + s \left(\vec{x}_{\perp}^{\max} - \vec{x} \right) \right) \quad (24)$$

and then update the initial seed for the random walk process as follows:

$$\vec{x} \leftarrow \vec{x} + s_{\text{step}} \left(\vec{x} + s \left(\vec{x}_{\perp}^{\max} - \vec{x} \right) \right). \quad (25)$$

Figure 2 shows an overview of the mathematical solver flowchart.

2.5. Digital Core Interpretation Model. This algorithm, presented above, has been consequently implemented into an interpretation model denoted as Digital Core, because this algorithm has the potential to replace the conventional,

expensive coring process in well logging and thus provides a comprehensive characterization of formation core.

This interpretation model executes the algorithm based on a multistep workflow as follows:

- (1) Determine zones of consistent mineralogy and resistivity response
- (2) Determine the volume of each lithology from image logs using facies
- (3) Execute the algorithm as described in Section 2.2
- (4) Obtain the lithology and the mineralogical compositions to their corresponding facies
- (5) Conduct quality control: a moving average at the resolution of the pulsed neutron log. The average should fit with the pulsed neutron measurement
- (6) Produce a set of deliverables that typically include:
 - (i) Lithology corresponding to each volumetric fraction in the analyzed section
 - (ii) High-resolution mineralogy corresponding to each lithology layer

3. Manufactured Test Case

In this section, the Digital Core model is applied to a manufactured test case and is consequently validated as discussed below.

3.1. Manufacturing the Reference Solution and Preparing Input Data. We selected a three-layer test case and used three lithologies, sandstone, shale, and coal (note: usually, the number of layers and the total number of tested lithologies are not identical). The pdfs for these lithologies are known. The volume fractions for the three layers are selected and listed in Table 1. Along with these volume fractions, we selected that layer 1 was sandstone, layer 2 was shale, and layer 3 was coal.

From the pdfs, the most likely composition for each lithology is selected. Then, we use the volume fraction to compute the averaged mineralogy that typical pulsed neutron geochemical tool measures. This data is detailed in Table 2.

3.2. Numerical Results. To simulate noise, we perturb the mineralogy data, the layer volume fractions (i.e., the image log data), and both the pdfs' abscissa and ordinate values using Gaussian noise with noise levels chosen as the Gaussian's standard deviation being $x\%$ of the mean value of the perturbed quantity. We selected x to be 0, 1, 2.5, 5, and 10.

Using the 3 lithologies and 3 layers, there are a total of 6 ways of assigning the lithologies to the different layers. The algorithm always found the correct permutation, i.e., 1: >sandstone, 2: >shale, and 3: >coal, for $x < 10\%$. Moreover, all other permutations were rejected. For $x = 10\%$, out of 5 trials, the VC returned:

TABLE 1: Volume fraction and lithology in each layer.

Layer	Volume fraction	Lithology
1	0.55	Sandstone
2	0.3	Shale
3	0.15	Coal

TABLE 2: Mineralogy of each layer and the computed average mineralogy.

	Layer 1	Layer 2	Layer 3	Measured
Organic	0.00000	0.00937	0.88791	0.13600
Kaolinite	0.03107	0.10146	0.02823	0.05176
Quartz	0.90045	0.35264	0.02823	0.60527
Siderite	0.01850	0.03883	0.05046	0.02939
Illite	0.05125	0.49746	0.00677	0.17844

- (i) It rejected all permutation once
- (ii) It found permutations (1: >sandstone, 2: >shale, and 3: > coal) and (1: >sandstone, 2: >coal, and 3: > shale) to be viable but found that the correct permutation was 50% more likely
- (iii) It accepted the correct permutation and rejected all others 2 times
- (iv) It accepted the incorrect (1: >sandstone, 2: >coal, and 3: > shale) and rejected all other combinations

Note that 10% standard deviation about the mean could make the volume fractions of layer 2 and layer 3 "switch places" in which case the algorithm cannot distinguish (1: >sandstone, 2: >coal, and 3: > shale) would be the correct answer!

For $x = 0, 1, 2.5$, and 5, selected predicted and reference mineralogies are depicted in Figure 3.

4. Field Data Case Study

A representative field data example is presented in this section and compared to core data to demonstrate the feasibility of the Digital Core interpretation model. This case study proves that the introduced method is readily capable of generating a mineralogical record with high vertical resolution for a formation logged with a pulsed neutron geochemistry tool and a high-resolution borehole imager.

4.1. Field Data Collection. A well from Australia where all log information and a core section of 10 ft was collected. This section was used to test the method with the result presented in Figure 4. The mineralogy was recorded with a pulsed neutron tool and was calibrated to XRD measurements. The borehole image was recorded with a resistivity imaging tool, processed, and flattened, and a high-resolution resistivity curve was generated. At its conventional resolution, the spectroscopic mineralogy shows relatively constant weight fractions for illite (37.5%), kaolinite (10.5%), siderite (about

FIGURE 3

FIGURE 4

7%), quartz (44%), and organic carbon (about 1%), which are consistent with values generally encountered in this formation for this part of the basin.

The resistivity image log is used to determine facies over the cored interval. It shows a series of thin beds with varying resistivities that were discriminated in three facies: less than 140 ohm-m, assumed to be more argillaceous silts identified as “shale,” 140 to 180 ohm-m, expected to contain more quartz and labelled as “silt,” and higher than 180 ohm-m, expected to be siderite-cemented silts identified as “cemented shale.” These labels indicate which type of formation would have a similar mineralogical composition: for instance, “shale” means that the formation has a mineralogical composition similar to that of a shale. The core shows that the formation consists of a siltstone with variations of mineralogical composition.

4.2. Field Data Processing and Interpretation Analysis. Based on XRD data available for this well, PDFs are generated for

TABLE 3: Comparison between the weighted average mineralogical composition (WAC) and the average composition determined from the spectroscopic mineralogy log for the 2 solutions determined by the algorithm. 70%: solution with 70% probability; 30%: solution with 30% probability. Mineral compositions expressed as % of the formation weight; EF: electrofacies, in ohm-m.

(a)						
70%	Illite	Kaolinite	Organic	Siderite	Quartz	EF
Shale 1	50.0	10.0	2.3	4.0	33.7	<140
Silt	26.6	10.5	0.4	2.4	60.2	<140<180
Shale 2	35.0	10.0	1.5	24.3	29.2	180<
WAC	37.1	10.2	1.3	6.8	44.6	
Log	37.3	10.3	1.2	7.0	44.2	

(b)						
30%	Illite	Kaolinite	Organic	Siderite	Quartz	EF
Silt	25.0	10.5	0.1	2.0	62.3	<140
Shale 1	48.7	10.0	2.4	4.0	34.8	<140<180
Shale 2	35.0	10.0	1.2	24.7	29.1	180<
WAC	37.1	10.2	1.3	6.8	44.6	
Log	37.3	10.3	1.2	7.0	44.2	

each of these lithological components and are consequently used in the algorithm. The algorithm determined two possible solutions, each characterized by the probability (the most likelihood) that this solution could occur: only the solution with higher probability is presented in this communication.

FIGURE 5

Table 3 shows that the average mineralogical compositions for each solution are consistent with the average mineralogical composition determined from the spectroscopy log over the interval, indicating that both solutions are mathematically acceptable. The solution with the highest probability is presented in Figure 1 along with the core. Aside from depth matching issues, the highly resistive layers observed on the image and associated with siderite-cemented intervals correspond to brownish streaks marking the occurrence of siderite cements on the core. The distribution of the “shale” facies (more argillaceous sediments) also generally corresponds to the darker intervals in the core. The high-resolution mineralogy, after applying a 1 ft mobile average filter, shows a good agreement with the spectroscopic mineralogy recorded by the logging tool (Figure 5).

Some discrepancies are observed in the distribution of “silt” facies in the upper part of the log where an interval with argillaceous composition corresponds to light-colored levels in the core (86–87 ft) (Figure 6). The occurrence of the “shale” can be explained by the dark streak on the image, probably related to data acquisition that affected the determination of the facies. At 87 ft, a siderite-cemented interval could not be located with confidence on the core. This might be caused by the lighter color of the silt, next to that of the siderite-cemented shale in the picture. However, the complex variation of mineralogy between 87.5 and 88.5 ft, as well as below 90 ft, displays a good correspondence to the color variations observed on the core.

In complex environments as the one in the example, where the grain size distribution does not change significantly and the laminations mostly correspond to variations in mineralogy, this result demonstrates that the method is globally working: the high-resolution mineralogy displays a

FIGURE 6

fairly representative account of the variability observed in the core. The level at 86–87 ft shows that the method is sensitive to thresholding while defining the facies. A good knowledge of the mineralogy for the logged formation and the geology of the area would naturally be important as lateral variation can occur: this ensures the PDF can take into account the regional variations of facies. The process itself, including the workflow and the algorithm, appears to be both robust and versatile to account for complex formations such as siltstone, thinly bedded sandstones, and most probably thinly bedded limestone or partially dolomitized limestone.

Further data from other wells have also been explored, and similar performance is achieved from the algorithm. Due to sensitivity issue, this work only presents one field data case as discussed above in this section.

5. Conclusion

The presented method enables a direct quantification of mineralogy at the scale of the lamination in thinly bedded formations and thus allows for determining organic content, matrix density, and total porosity at high resolution, resulting in a more accurate calculation of net pay and net-to-gross ratio. The method provides great advantage over conventional mineralogy log interpretation by revealing the full vertical variability of a formation that would otherwise appear very homogeneous, revealing by-passed potentially targets. The interpretation methodology, once fully implemented and validated for the mineralogy log, could also be applied to other nuclear logs such as density and porosity logs, as well as NMR and acoustic logs. Furthermore, a future work will combine machine learning into various aspects of the Digital Core, e.g., facies recognition and log prediction.

Data Availability

Underlying data is available upon request.

Disclosure

The initial work has been presented in 2016 SEG Conference with an abstract submitted to the 2016 SEG Technical Program.

Conflicts of Interest

The authors declare that there are no potential conflicts of interest, such as financial interests, affiliations, or personal interests or beliefs, that could be perceived to affect the objectivity or neutrality of the manuscript.

Acknowledgments

The authors thank Alberto Mezzatesta and Lena Thrane for their support and input towards this work. This work is funded by research funds from the University of Electronic Science and Technology of China.

References

- [1] Q. Zhang, F. Mendez, J. Longo, S. Gade, and J. Peyaud, "Development of reduced order modelling techniques for downhole logging sensor design," in *SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition*, Bali, Indonesia, Oct. 2015.
- [2] Q. Zhang, "Development of a surrogate model for elemental analysis using a natural gamma ray spectroscopy tool," *Applied Radiation and Isotopes*, vol. 104, pp. 5–14, 2015.
- [3] D. P. de Farias and B. van Roy, "On constraint sampling in the linear programming approach to approximate dynamic programming," *Mathematics of Operations Research*, vol. 29, no. 3, pp. 462–478, 2004.
- [4] J. B. Peyaud, M. Pagel, J. Cabrera, and H. Pitsch, "Mineralogical, chemical and isotopic perturbations induced in shale by fluid circulation in a fault at the Tournemire experimental site (Aveyron, France)," *Journal of Geochemical Exploration*, vol. 90, no. 1–2, pp. 9–23, 2006.
- [5] P. Diaconis, "The Markov Chain Monte Carlo Revolution," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 179–205, 2009.
- [6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins, 3rd ed edition, 1996.
- [7] R. Penrose, "A generalized inverse for matrices," *Proceedings of the Cambridge Philosophical Society*, vol. 51, no. 3, pp. 406–413, 1955.
- [8] K. Van den Meersche, K. Soetaert, and D. V. Oevelen, "xsample(): AnRFunction for sampling linear inverse problems," *Journal of Statistical Software*, vol. 30, no. Code Snippet 1, 2009.
- [9] E. Stiefel, "Note on Jordan elimination, linear programming and Tchebycheff approximation," *Numerische Mathematik*, vol. 2, no. 1, pp. 1–17, 1960.
- [10] G. B. Dantzig and M. N. Thapa, *Linear Programming 1: Introduction*, Springer Verlag, 1997.