Identification of stratigraphic formation interfaces using wavelet and Fourier transforms

Shih-Yu Pan\textsuperscript{a}, Bieng-Zih Hsieh\textsuperscript{a}, Ming-Tar Lu\textsuperscript{b}, Zsay-Shing Lin\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}Department of Resources Engineering, National Cheng Kung University, 1 University Road, Tainan 701, Taiwan
\textsuperscript{b}Exploration and Production Business Division, Chinese Petroleum Corporation, No. 140, Jungjeng Road, Miaoli City, Miaoli 360, Taiwan

Received 29 March 2006; received in revised form 12 December 2006; accepted 3 January 2007

Abstract

The purpose of this study was to identify the formation interfaces from geophysical well log data using the wavelet transform, and a combination of the wavelet transform and the Fourier transform methods. In the wavelet transform method, the identification of formation interfaces is based on the wavelet coefficients from the wavelet transform of spontaneous potential (SP) log and gamma ray (GR) log data. In the combination of the wavelet transform and the Fourier transform methods, the wavelet transform, spectrum analysis, and logarithmic transform of well logs were applied to the SP and GR log data successively to obtain clear signals for identifying the stratigraphic formation interface. In this study, a set of ideal log data was first created and analyzed to test the validity of the developed procedures. In analyzing the SP and GR logs from a field, both the wavelet transform method and conventional well log analysis showed similar results. The results from a combination of the wavelet transform and the Fourier transform methods, however, were better than those from the wavelet transform method and the conventional well log analysis.

\textcopyright 2007 Elsevier Ltd. All rights reserved.

Keywords: Formation interface; Stratigraphic interface; Wavelet transform; Fourier transform; Geophysical well log

1. Introduction

Formation interface identification is essential and routine work in interpreting geological or geophysical data in petroleum exploration. The geophysical well log is one of the best sources for obtaining formation properties and identifying interfaces (Crain, 1986; Dewan, 1983). Well log recordings vary as formation lithology or properties change. Generally, the permeable zone logs, such as the spontaneous potential (SP) and gamma ray (GR) logs, are responsive to the lithology of the formation. Well log interpretations include manually or visually discerned formation boundaries to separate adjacent lithologic units. Different interpreters may use subjective criteria for choosing boundaries that may lead to different results (Dewan, 1983; Hsieh et al., 2005). If well log recordings are treated as the signals responding to the specific input energy source from a formation, a signal-process technique, such as signal transforming and filtering, could...
be used to detect the formation interfaces from the well log data. And the results from the signal-processing technique may be somewhat objective.

Lanning and Johnson (1983) were the first to use the Walsh transform, a low-pass filter method, to analyze well log data to detect the rock boundary. Later on, Maiti and Tiwari (2005) applied the Walsh transform to analyze well log data with specific criteria and developed an automated method to detect the lithologic boundaries. The Walsh transform is analogous to the Fourier transform (FT) with a constant window for analyzing stationary signals.

The first modification to the FT came as the short-time Fourier transform (STFT), which segmented the signal by using the windows. The STFT could be used to analyze either high-frequency data components by using narrow windows or low-frequency data components by using wide windows. But both wide and narrow windows cannot exist at the same time. A different window function for different frequency bands to solve this problem was developed—it is called the basis functions “wavelets”, which can contain both time and frequency information simultaneously, and can be used to analyze stationary and non-stationary data (Polikar, 1999).

The wavelet transform (WT) has been applied in many fields, such as mathematics, engineering, medicine, and geology, for many years (Frantziskonis and Denis, 2003; Leung et al., 1998; Polikar, 1999). Panda et al. (1996) demonstrated that several wavelet-decomposition levels of the WT applied to areal permeability data taken from core samples could determine the locations of layer boundaries, faults, and fractures. Alvarez et al. (2003) characterized the lithology of a reservoir using the GR log and seismic traces around the well. After calculating wavelet coefficients and the energy (or power) of the WT, they found that wavelet energy distributions of wells in sandstone were significantly different from those of wells in gravel. Yue et al. (2004) studied the identification of reservoir fluid types by using the wavelet energy from the WT of resistivity logs. They found that the combination of the WT and statistics could be used to determine trends, breakdown points, and discontinuities. Thus, the WT can be combined with other technologies to analyze many kinds of data.
The purpose of this study was to combine the WT and the Fourier technology methods to analyze an SP log and a GR log to obtain low-noise signals in order to identify formation interfaces.

2. Basic theory

2.1. Fourier transform

The data in the SP and GR well logs are functions of the depth. The depth of the well logs can be treated as “time” in processing the SP or GR log as a signal. These data can be processed as discrete signals and transformed into another domain, depending upon the kernels used (Soliman et al., 2003). The transformed results may reveal special frequencies of the data and can be used to filter the data. The most popular transformation is the FT. If the discrete signal is the periodic function and the sampling frequency is greater than double the band-limited frequency, the discrete Fourier transform (DFT) can be used. Then the signal in the time domain \( x(t) \) is transformed into the frequency domain \( X(f) \) in the DFT, which can be expressed as (Oppenheim et al., 1999; Wickerhauser, 1994):

\[
X[k] = \sum_{n=0}^{N-1} x[n] e^{-i2\pi nk/N},
\]

\[
x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i2\pi nk/N}.
\]

To efficiently calculate Eq. (1) or (2), the fast Fourier transform (FFT) is used (Oppenheim et al., 1999). The FFT dramatically reduces the size of the multiplier and can be used even when \( N \) is not a power of two but can be expressed as the product of two or several integers (Oppenheim et al., 1999). In the present study, we used the FFT to calculate the spectrum of raw data (SP and GR logs) during analyzing data in the combination of WT and FT methods.

2.2. Wavelet transform

The WT provides varying time and frequency resolutions by using windows of different lengths (Polikar, 1999). The kernel of the WT includes two variables, phase (or location) and scale, instead of only one, as in the FT (Boggess and Narcowich, 2001; Misiti et al., 2000). The kernel of the WT is called the wavelet function \( \psi \). The result derived from the WT, \( W_\psi \), is called the wavelet coefficient and can be expressed as follows (Boggess and Narcowich, 2001; Misiti et al., 2000; Wickerhauser, 1994):

\[
W_\psi(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \overline{\psi} \left( \frac{t - b}{a} \right) dt,
\]

\[
x(t) = \frac{1}{K_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} \psi \left( \frac{t - b}{a} \right) W_\psi(a, b) \frac{db \, da}{a^2},
\]

where \( \psi \) is a wavelet function, \( \overline{\psi} \) is a conjugate complex of \( \psi \), and \( a \) and \( b \) are scale and phase (or location) variables.

The type of the WT depends on the wavelet functions used (Misiti et al., 2000). The Haar function (Fig. 1) is the first, simplest, discontinuous and resembles a step function (Misiti et al., 2000). Other kernels commonly used in the WT are the Coiflet, the Daubechies, and the Morlet (Fig. 1). In this study, we used both Daubechies (db4 used in the MATLAB software) and Haar (db1 used in the MATLAB software) wavelet functions to analyze SP and GR log data, respectively.

The wavelet function \( \psi \) in Eq. (3) is a twin-scale relation, including scale \( a \) and phase \( b \). Thus, the

![Fig. 1. Four wavelet functions used in the wavelet transform (Misiti et al., 2000).](image-url)
wavelet coefficients \((W_c)\) are also twin-scale functions and can be represented in a three-dimensional graph. If wavelet coefficients are obtained by continuously shifting the wavelet function from the variable phase \((b)\) and stretching the shape of the wavelet function from the scale \((a, \text{ where } a = 1, 2, 3, 4, \ldots)\) the process is called the continuous wavelet transform (CWT) (Misiti et al., 2000). A signal can be recalculated from the wavelet coefficients \((W_c)\) using the inverse WT (Eq. (4)) (Boggess and Narcowich, 2001; Misiti et al., 2000).

In the present study, we also used the discrete wavelet transform (DWT), whose scale \((a)\) is in the form \(2^n\), and which is applied in filtering (Leung et al., 1998; Lu and Li, 1999; Mallat, 1989). In the DWT, a signal is decomposed into several lower-resolution components constructed using the approximation (cA) and detail wavelet coefficients (cD), and can be expressed as follows (Misiti et al., 2000; Yue et al., 2004):

\[
cA_k^0 = \sum_n x(n) \cdot \phi(n - k),
\]

\[
cD_k^j = \sum_n h_{1(n-2k)} cA_{n}^{j-1},
\]

\[
cA_k^j = \sum_n h_{0(n-2k)} cA_{n}^{j-1},
\]

where

\[
h_{0k} = \frac{1}{\sqrt{2}} \int \phi\left(\frac{t}{2}\right) \overline{\phi}(t - k) \, dt,
\]

\[
h_{1k} = \frac{1}{\sqrt{2}} \int \psi\left(\frac{t}{2}\right) \overline{\phi}(t - k) \, dt.
\]

Note that \(cA_k^0\) in Eq. (5) is calculated from \(x(n)\), raw data from either the SP or GR log in this study. \(\phi\) is a scale function, which is orthogonal to the wavelet function \((\psi)\) used in the WT. \(k\) in Eqs. (5)–(9) is equivalent to the phase \((b)\) in Eqs. (3) and (4). The detail wavelet coefficient (cD) was iteratively decomposed from the approximation wavelet coefficient (cA) (Eqs. (6) and (7)). By using the inverse WT of the approximation and detail coefficients, the signal, \(x\), was reconstructed as follows (Misiti et al., 2000):

\[
x(t) = cA_f + \sum_{j \leq f} cD_j.
\]

The original signal could be obtained from the sum of its approximation (cA) and its detail (cD). In this study, the DWT of the well log data was used to decompose the detail wavelet coefficient (cD). The inverse WT of the detail wavelet coefficient was then used to reconstruct the high-frequency signal.
Thus, raw data can be filtered for further analysis. The advantage of the WT is that a variable window (short or long time-interval) can be used to observe the characteristics of the data. For a signal with rapid variation, a small window can be applied. For a signal with slow variation, a larger window can be used. Both the approximation response (equivalent to the low-frequency response using low-pass filtering) and the detail response (equivalent to the high-frequency response using high-pass filtering) can be calculated at the same time (Mallat, 1989; Misiti et al., 2000). In the present study, the detail response ($R_{cD}$) was analyzed and interpreted.

### 3. Procedures for analyzing the well log

The WT method and the combined WT and FT method were developed to analyze the SP or GR log in order to extract a specific signal to identify stratigraphic interfaces. In the present study, we used the Signal Processing Toolbox and Wavelet Toolbox in MATLAB software version 6.1 to perform the FT and WT calculation.
Fig. 4. (a) Raw ideal data, (b) continuous scales distribution, and (c) discontinuity discrimination of an ideal signal.

Fig. 5. (a) Raw SP log data and (b) raw GR log data.
3.1. Wavelet transform method

Both the CWT and DWT were used to identify the formation interfaces that lay between different formation characteristics. In the CWT, the wavelet coefficients \( W_c \) were calculated from the SP or GR log as functions of scale \( a \) and phase \( b \). We used the Haar function of all scales to form a three-dimensional plot with different wavelet coefficients, which was then used to identify formation interfaces (Fig. 2).

In the DWT, well log data, such as the SP or GR log data, were decomposed to approximation coefficients \( c_A \) from the low-pass filter, and detail coefficients \( c_D \) from the high-pass filter. In this step, the detail wavelet coefficients \( c_D \), representing the high-frequency signals and including probably the apparent variations of interfaces, were used for defining the stratigraphic interfaces. We also used the Haar function with scales in the form of \( 2^n \) in this step (Fig. 2).

3.2. Combination of wavelet transform and Fourier transform methods

In the combination of the WT and FT methods, the raw data of the SP or GR log were first decomposed into a detail wavelet coefficient \( c_D \) and an approximation wavelet coefficient \( c_A \). For the data with less noise, such as the SP log, the detail wavelet coefficient from a single-level decomposition of the DWT was used for further spectrum analysis (Fig. 3). For the noisy data, such as GR log data, the detail wavelet coefficient from three-level decomposition of the DWT was applied before doing spectrum analysis (Fig. 3). Then the inverse WT was applied to the detail wavelet coefficient to obtain reconstructed data \( R_{cD} \). Later, the \( R_{cD} \) was transformed to obtain the spectrum of reconstructed data \( R_{cDF} \) using the fast Fourier transform. By choosing a proper frequency band from the spectrum of the \( R_{cDF} \), modified reconstructed data \( MR_{cDF} \) were obtained. We then used the

Fig. 6. (a) Continuous scales distribution and (b) raw SP log data.
inverse FFT to transform the MRcDF into the time domain to obtain the modified reconstructed data in time domain (MRcD). After the raw data were filtered twice using the combined WT and FT method, the logarithmic distribution of filtered data was calculated using Eq. (12) (Jang and Jang, 2003):

\[ \text{LRcD} = \log_2 \left( \frac{\text{MRcD}}{440} \right) + 69. \]  

4. Results

4.1. Analysis of ideal well log data

To use the theory and procedures described above, we first analyzed a set of ideal well log data.
(either SP or GR log data) with a plateau waveform (Fig. 4(a)), representing a clean sand formation in the interval between the depths of 7420 and 7440 ft, and shale formations above 7420 ft and below 7440 ft. By applying the CWT to the raw ideal data (Fig. 4(a)), wavelet coefficients showing two depths at 7420 and 7440 ft with sharp gradients (Fig. 4(b)) were obtained. Note that the boundaries of the sand formations are consistent with the places of sharp gradients.

Using the DWT, the detail wavelet coefficients were obtained from ideal well log data. A negative wavelet coefficient at 7420 ft and a positive wavelet coefficient at 7440 ft show the upper and lower boundaries of the formation (Fig. 4(c)). When the value of the ideal well log signal changed from 0 to 100 (at 7420 ft), the wavelet coefficient was negative. When the value changed from 100 to 0 (at 7440 ft), the wavelet coefficient was positive. Thus, the zone between the two corresponding but opposite (positive vs. negative) wavelet coefficients was shown to have a specific formation lithology, i.e., a sand formation.

For ideal well data with no noise, it was not necessary to filter the data with the combination of the WT and FT methods. When the well log data has noise, however, this method is useful for filtering noisy signals.

4.2. Analysis of field data

4.2.1. Descriptions of field data

The SP and GR logs used in this study were from an oil well (well X-1) in which the main oil-bearing
zone is in the M-formation between 7604 and 7805 ft (Fig. 5). From the core analysis and geological interpretations, there were four stratigraphic units in the M-formation, such as A-sand, B-sand, C-sand, and D-sand (Fig. 5), representing different deposit environments and characteristics.

From the SP log (Fig. 5(a)) of well X/C01, it can be observed that there is a pattern of three curved responses representing three zones or formations (Zone 1, Zone 2, and Zone 3) (Fig. 5(a)). Two shale breaks exist near 7655 ft and near 7735 ft. By comparing this with the core data, the first permeable bed (Zone 1) was called A-sand, which is in the interval from 7604 to 7664 ft. The second thick permeable bed (Zone 2 from 7664 to 7720 ft) indicated by the SP log corresponds to the B- and C-sand identified from the core data. There was no apparent shale break in the SP log at the interface between the B- and C-sand. The last permeable bed (Zone 3), from 7730 to 7805 ft, is indicated by the SP log as D-sand. The responses of the GR log in well X/C01 were similar to those of the SP log. However, well log responses in the GR log present a pattern of more than four curves in the depth between 7604 and 7805 ft (Fig. 5(b)).

4.2.2. Analysis of the SP log data

Three zones (Zone 1, Zone 2, and Zone 3 in Fig. 5) can be identified from conventional SP log interpretation. The scales distribution (Fig. 6(a)) can be obtained by applying the CWT to the raw SP log data (Fig. 6(b)). From the result of the scales distribution (Fig. 6(a)), three major zones in the M-formation can not be clearly observed.

Based on wavelet coefficients of the SP log response (Fig. 7(a)) from the DWT of the one-level decomposition using the Haar function, three oblique lines are drawn from three positive–negative pairs of wavelet coefficients, showing three formation stratigraphies (Fig. 7(b)). Three zones or units can be identified from wavelet coefficients, and the results are the same as those from the conventional SP log analysis. It is difficult to recognize the interface between B- and C-sand in

Fig. 11. (a) Spectrum analysis of reconstructed SP log signal from detail wavelet coefficients. (b) The frequency band chosen from (a).
Zone 2 when analyzing the SP log response (Fig. 7) with a conventional well log interpretation and the WT.

In the combined WT and FT method, the spectrum of the SP log data was calculated and showed that there was almost no apparent frequency (Fig. 8). Because SP log data were less noisy (Fig. 3), detail coefficients (cD) (Fig. 9) from single-level decomposition of the DWT were used. The inverse WT of detail wavelet coefficients was then applied to reconstruct the \( R_{cD} \), i.e., the high-frequency signal of the raw SP log (Fig. 10).

After the spectrum of the \( R_{cD} \) was analyzed using the fast Fourier transform, there was only one distinct frequency band in the reconstructed signal, 20–40 nHz (Fig. 11(a)). The frequency band in the center of the spectrum, 30–31 nHz, was chosen (Fig. 11(b)). Note that time in the FT or inverse FT was replaced by depth in this study. Then the inverse FFT was used to transform the chosen frequency band, 30–31 nHz, to obtain modified reconstructed data, which is similar to the wave of a human voice (Fig. 12) showing four zones equivalent to the four sands in the M-formation. The logarithmic distribution data (Fig. 13) from the logarithm transform showed two bigger and two smaller waveforms between 7600 and 7800 ft in the M-formation; these four waveforms are the actual four formation stratigraphic zones.

4.2.3. Analysis of GR log data

The scales distribution (Fig. 14) from the CWT showed that there was too much noise. No specific interface can be identified from this analysis, probably because the GR log data contained too much noise.
In the DWT, the results of the wavelet coefficients also showed too much noise (Fig. 15), indicating more than four permeable layers in the M-formation. When the stratigraphic oblique lines with some noise are remarked, it is confusing and difficult to analyze these lines. And if there is no core data, it may be difficult to interpret the formation interface using the raw GR log with the WT method.

In the combined WT and FT method, the spectrum of the GR log data (Fig. 16) was calculated and showed that the signal had a small amount of high-frequency noise and no apparent frequency. Then the three-level decomposition of the DWT was used to obtain detail wavelet coefficients (cD) (Fig. 17). The inverse WT of detail wavelet coefficients was applied to obtain the reconstructed signal of the $R_{cD}$ (Fig. 18), which is the high-frequency signal of the raw GR log data. There are two distinct frequency bands, 0–20 and 30–45 nHz, in the spectrum of the $R_{cD}$ (Fig. 19(a)), and the strongest (or the most apparent) frequency band is 8.5–10 nHz (Fig. 19(b)). Then the inverse FT was applied to transform the chosen frequency band (8.5–10 nHz) signals from the frequency domain to the time domain (in the depth domain). The results were similar to the wave of a human voice (Fig. 20) showing four zones equivalent to the four sands in the formation. Next, the logarithmic distribution of this formation was calculated (Eq. (12)). The final results (Fig. 21) also revealed four distinct waveforms in the M-formation, in the range of 7600–7800 ft. These four clean waveforms corresponded to the four formation stratigraphic zones shown in the raw GR log data (Fig. 5(b)).
Fig. 16. Spectrum analysis of raw GR log of well X–1.

Fig. 17. Detail wavelet coefficients of raw GR log.

Fig. 18. Reconstructed signal from detail wavelet coefficients of raw GR log.
5. Discussion

Using the Haar function (db1) in the WT method in the SP log response, only three layers in the M-formation can be identified (Figs. 6 and 7). The M-formation, however, actually contains four layers. The interface between B-sand and C-sand (Zone 2 in the M-formation) is difficult to recognize. This difficulty is also encountered in a conventional well log analysis. Using the WT of the GR log response, more than four layers can be obtained (Figs. 14 and 15). But the responses of the GR log, both in a conventional well log and in the wavelet coefficients, contain noise. The noise in the wavelet coefficients of the GR log response is dominant and increases the difficulty of analyzing the formation stratigraphies.

When analyzing the field data using the combined WT and FT method, single-level decomposition of the WT was used to analyze the SP log data because the SP log data relatively contained some noise. According to the characteristics of the various kernels (Misiti et al., 2000), the discontinuous Haar function (db1) was used to analyze the GR log data because of the noise in the GR log, and the Daubechies function (db4), being a more smooth wavelet function, was used to analyze the SP log data because there was less noise in the SP log data. In this study, when filtering out noise to obtain a modified spectrum of reconstructed data ($R_{OD}$) in the combined WT and FT method, the choice of the suitable frequency (band) was also important. If the frequency band of 26–28 nHz in the spectrum had been chosen from the spectrum of the reconstructed SP log signal, the result of the logarithm distribution data would have been different from those obtained in the 30–31 nHz band. The frequency in the center portion of the spectrum (Fig. 11(b)), i.e., the frequency band of 30–31 nHz, should be chosen so that the corresponding result (Fig. 13) is correct.

Fig. 19. (a) Spectrum analysis of reconstructed GR log signal from detail wavelet coefficients. (b) The frequency band chosen from (a).
When another band, such as 11–13 nHz, was chosen instead of 8.5–10 nHz (Fig. 19(b)) to filter the GR log data, the results differed completely from those obtained in the 8.5–10 nHz band. Thus, the strongest or most distinct frequency in the spectrum for the GR log data should be chosen to obtain the best results.

6. Conclusions

The following conclusions are made for this study.

We developed procedures for the WT method and for the combined WT and FT method to analyze SP and GR log data.

The results of identifying formation interfaces from conventional well log interpretation and from the WT method are similar, and unable to permit objective identification of all the formation interfaces for the case study. When using the combined WT and FT method, all the formation interfaces can be identified.

When analyzing SP and GR log data, the frequency (band) in the center and strongest portion of the spectrum of reconstructed data should be chosen to obtain valid results.

References


