Heterogeneity analysis of geophysical well-log data using Hilbert–Huang transform

Gaurav S. Gairola, E. Chandrasekhar*
Department of Earth Sciences, Indian Institute of Technology Bombay, Powai, Mumbai–400076, India

HIGHLIGHTS

- The data adaptive HHT helps in distinguishing reservoir and non-reservoir zones in wells.
- The HHT together with heterogeneity analysis made with IMFs can quantify heterogeneity in wells.
- Higher shale volume within L-III zone of Well C made it more heterogeneous than Well B.
- HHT and multifractal analysis confirm the role of gas in defining the type of heterogeneity in reservoirs.

ABSTRACT

Geophysical well-log data manifest nonlinear behaviour of their respective physical properties of the subsurface layers as a function of depth. Therefore, nonlinear data analysis techniques must be used to understand and characterize the nature of the subsurface lithologies, vis-à-vis their degree of heterogeneity. One such nonlinear technique is the fully data adaptive Hilbert–Huang Transform (HHT), which constitutes two independent techniques, namely, the empirical mode decomposition technique (EMDT) and the Hilbert spectral analysis (HSA). While EMDT facilitates to decompose the well-log data into oscillatory signals of different wavelengths called intrinsic mode functions (IMFs), which represent different frequency characteristics of the signal and which in turn helps to calculate the degree of heterogeneity in the subsurface, the HSA facilitates to determine the instantaneous amplitudes and frequencies of the IMFs, which can be used to characterize the heterogeneity in the signals. In this study, HHT has been applied to gamma-ray log of the thickest limestone reservoir zones of two different wells: Well B and Well C, located in the western offshore basin of India to determine the respective IMFs. The estimated instantaneous amplitudes and frequencies of the derived IMFs by HSA qualitatively suggested well C to be more heterogeneous than well B. By establishing a relationship between the IMF number \( m \) and its mean wavelength \( l_m \), a heterogeneity index \( \rho \) associated with subsurface layers was determined using \( l_m = k \rho^m \), where 'k' is a constant. \( \rho \) bears an inverse relation with the heterogeneity of the subsurface. The estimated \( \rho \) values confirm our observations from HSA. We attribute the higher degree of heterogeneity in Well C to high average shale volume in the limestone reservoir zone in Well C than in Well B. Interpretation of the results together with those of the heterogeneity analysis of the gas zone in limestone reservoir of Well C is made by comparing with the results of multifractal analysis of the same data carried out earlier.

© 2017 Elsevier B.V. All rights reserved.

* Corresponding author.
E-mail address: esekhar@iitb.ac.in (E. Chandrasekhar).

http://dx.doi.org/10.1016/j.physa.2017.02.029
0378-4371/© 2017 Elsevier B.V. All rights reserved.
1. Introduction

Over the past several decades, traditional signal processing methods relied heavily on Fourier transform (FT) for extraction of hidden information from various signals, particularly, the frequency. However, although the FT has time and again been proved to be the best technique for analysing linear and stationary signals alone, they fail to provide information on the time and space variation of the identified frequencies in the signal (see [1–3]). In case of geophysical well-log data, particularly, such information on space–frequency localization is essential and important for an improved understanding of the data (see [1] and references therein). Since most geophysical data are aperiodic, nonstationary and nonlinear in nature, use of FT is therefore definitely not the right choice to unravel the spatio-temporal characteristics of the signal. The wavelet transformation (WT) has been proved to be one of the efficient techniques for identifying the spatio-temporal characteristics of a variety of signals ([2, 4–6], to cite a few). In case of well-log data, WT has found its wide applications in effectively describing the inter-well relationship [7], determining the sedimentary cycles [8], reservoir characterization [9, 10] and for determining the space-localization and identifying the depths to the tops of reservoir zones ([1] and references therein). Despite having its application on broad ranges of a variety of signals, the perennial problem with wavelet analysis is the identification of a suitable mother wavelet that best suits to analyse the signals. This leaves one to search for an alternative recourse, which can provide required information from the signal under investigation. Recently, Subhakar and Chandrasekhar [11, 12] studied fractal and multifractal behaviour of geophysical well-log data used in the present study to understand the inherent presence of long-term correlations in the data, vis-à-vis their degree of multifractality of heterogeneous subsurface of different wells. Another efficient and fully data adaptive signal analysis technique is the Hilbert–Huang Transform (HHT), which constitutes two independent techniques, namely, the empirical mode decomposition technique (EMDT) and the Hilbert spectral analysis (HSA), due to Huang et al. and Huang and Wu [13, 14]. EMDT decomposes the data into oscillatory signals of different wavelengths, known as intrinsic mode functions (IMFs). EMDT has found several applications in a variety of specializations in geophysics, such as magnetotellurics ([15–17], seismics [18, 19], fault detection in mechanical devices [20], atmospheric sciences [21], sequence stratigraphy [22], besides financial data [23] and medical science [24]. Datig and Schlurmann [25] discussed the limitations of HHT, while applying it to study irregular water waves. On the other hand, HSA facilitates to determine the instantaneous amplitudes and frequencies of the derived IMFs, which can be used to characterize the heterogeneity in the data. The IMFs provide space–frequency resolution of the signal, just the similar way the discrete wavelet transform (DWT) decomposes the signal into approximate and detailed wavelet coefficients in its transformation (see [2, 5]). However, the major advantage of EMDT over the DWT is that in case of the former, the decomposition of the signal into IMFs can easily be done through simple sifting operations, which is totally data adaptive, unlike in the latter, where a similar operation needs an arduous task of choosing suitable wavelet and scaling functions. As a result, EMDT is the best suited technique to analyse nonlinear and nonstationary signals and also acts as an efficient filter bank [26]. Gaci and Zaourar [27] provided a method to carryout heterogeneity analysis of the data using the generated IMFs.

In the present study, we applied HHT to gamma-ray log and neutron porosity log corresponding to a 60 m thick reservoir zone within the thickest L-III limestone pay zones (see Fig. 2 of [1]) of two different wells: Well B and Well C, of Bombay offshore basin, off the west coast of India. The entire L-III reservoir consists of six limestone layers, separated by dense impermeable shales and tight limestones acting as potential barriers, preventing the flow of hydrocarbons [28]. Based on this observation, Subhakar and Chandrasekhar [12] suggested that a zone-wise analysis of well-log data would provide clues on the degree of heterogeneity present within this major hydrocarbon reservoir in Bombay offshore basin. The main objective of the study is to provide a quantitative study of the degree of heterogeneity present within the L-III reservoir zone and perform heterogeneity analysis of data of only L-III reservoir zone using HHT and compare the results with those obtained by multifractal studies of the same data sets [12]. The organization of the paper is as follows. Section 2 briefly describes the geology of the study area and the data used in the present study. Section 3 discusses the basic theory and mathematical details of EMDT and HSA techniques and the heterogeneity analysis. Section 4 describes the application of EMDT, HSA and heterogeneity analysis to the well-log data sets considered for the present study. Section 5 provides the results and discussion and Section 6 provides the conclusions of the present study.

2. Geology of study area and dataset

The Bombay offshore basin is the most prolific basin of India and is geographically located in the western continental shelf of India (Fig. 1). The subsurface consists of multiple reservoir zones dominated by carbonate reservoirs. The major production from the basin comes from the L-III reservoir, which contains more than 80% of the total reserves in the entire Mumbai Offshore basin [28]. Cyclic sedimentation of shale and limestone is identified in L-III, which is composed mainly of wackestone, mudstone and minor packstone. The pores are mainly intergranular, intra-granular, moldic, vuggy and micro fissures. The excellent permeability of the reservoir is due to interconnection of solution cavities by micro fissures [28]. Further details about the geology and lithostratigraphy of the study area can be found in [1]. The deposition of shale bands in the limestone reservoir zone acts as a vertical permeability barrier. Due to the cyclic sedimentation, the shale beds present in the limestone reservoir zone degrades the reservoir quality by reducing the porosity and permeability. Such a geological and stratigraphic set up forms a natural laboratory, which aptly suits to carry out heterogeneity analysis of the high-yield oil and gas wells in the study region.
3. Methodology

3.1. Empirical Mode Decomposition Technique (EMDT)

EMDT is the most suitable data adaptive technique to analyse nonlinear and non-stationary data [14]. The end result of EMDT is decomposition of different frequency components of the signal (from highest to lowest) till the signal becomes monotonic, which is called residue and represents the overall trend of the signal.

The algorithm of EMDT is as follows:

**Step 1:** Identify the local maxima and minima of the signal (say, $x(s)$) under investigation.
Fig. 3. (a) Representation of the test signal comprising of a mixture of three sinusoidal waves of 50 Hz, 100 Hz and 150 Hz frequencies and its decomposed intrinsic mode functions (IMFs) together with the residue. (b) The reconstructed signal and the error between the original and reconstructed signals.

Step 2: Interpolate (preferably cubic spline (see [13])) the local maxima and minima to form upper envelope and lower envelope and calculate its mean, $m_1$.

Step 3: Calculate the first proto-mode, $x_1$ given by $x_1 = x(s) - m_1$.

Step 4: Since $x_1$ may consist of new extrema, it should be considered as the signal and steps 1–3 above should be repeated to get $x_{11} = x_1 - m_{11}$. This procedure is repeated, say, up to ‘k’ iterations, after which, we get $x_{1k} = x_{1(k-1)} - m_{1k}$. The stopping criterion for these iterations, defining the normalized squared difference between two successive sifting operations is given by $SD_k = \frac{\sum_{i=0}^{N} |x_{1k-1}(i) - x_{1}(i)|^2}{\sum_{i=0}^{N} |x_{1}(i)|^2}$. $N$ denotes the number of data points. Theoretically, the stopping criterion, $SD_k$ explains that the sifting process stops only if (a) the number of extrema and the number of zero
crossings are equal or at most differ by one in the entire signal and (b) \( m_1 \) is zero for the entire signal (see [13] for further details). While the condition (a) ensures that there should be no local fluctuations within the signal, the condition (b) ensures estimation of well-behaved Hilbert transform for extraction of meaningful instantaneous frequencies in the data. In the present study, we have considered \( SD_k \leq 0.1 \) as the stopping criterion to determine the first IMF. This implies that if the stopping criterion is satisfied after ‘k’ iterations, then \( x_{1k} \) becomes the first IMF (I_1), which signifies the highest frequency (shortest wavelength) oscillations. It is important to note here that the condition, \( SD_k \leq 0.1 \) is rather arbitrary and largely depends on data quality, in the sense that higher \( SD_k \) values are generally set for poor-quality data and vice-versa for better estimates of IMFs.

**Step 5:** To determine the second IMF \( (I_2) \), calculate the residue, \( r_1 \) by subtracting \( I_1 \) from the original signal \( x(s) \) and treat \( r_1 \) as a new signal and repeat steps 1–4. Frequency of \( I_2 \) will be lower than that of \( I_1 \).
Fig. 6. Reconstructed gamma-ray log from the IMFs for hydrocarbon zone in L-III limestone of (a) Well B and (b) Well C. Note the extremely small error between the original and reconstructed data, signifying the chosen stopping criterion of $SD_k \leq 0.1$ to be aptly justified (see Section 3).

**Step 6:** The entire decomposition process described above (steps 1–5) must be continued till all IMFs are determined and till the residue $r_n$ becomes a monotonic function, where no further IMFs can be extracted from the data. All the estimated IMFs can be synthesized to get back the original signal as $x(s) = \sum_{i=1}^{n} I_i + r_n$.

Fig. 2 summarizes the complete EMD algorithm depicted in the form of a flow chart. As can be understood from the above algorithm, the selection of stopping criterion plays a major role in the computation of each IMF. In the next section, we show how the chosen standard deviation-based stopping criterion is aptly justified in deriving IMFs, when the above algorithm is tested on synthetic data.

3.1.1. Application of EMD analysis to a test signal

Before applying the EMD technique to the original well-log data, we have tested the algorithm on a test signal and verified the correct decomposition of the constituent frequencies present in the test signal. Fig. 3(a) (top panel) shows the test signal, $f(x)$, representing a mixture of three different frequencies, given by $f(x) = \sin(50x) + \sin(100x) + \sin(150x)$. Fig. 3(a) also represents the decomposed signal in the form of IMFs. The least IMF number signifies the highest frequency component present in the signal and vice-versa. The residue signifies the inherent long-period trend present in the signal and will always be monotonic. The reconstructed signal together with the error (estimated to be of the order of $10^{-15}$) is shown in Fig. 3(b).

The fast Fourier transform was performed on the test signal and its each IMF to identify the frequencies of the decomposed IMFs. Fig. 4 shows the frequency spectra of the test signal (Fig. 4(a)) and those of IMFs (Fig. 4(b)–(d)), clearly identifying the three frequencies considered in the test signal. This strongly confirms that the chosen stopping criterion, $SD_k \leq 0.1$ is aptly justified to compute the IMFs.

3.2. Hilbert Spectral Analysis (HSA)

The HSA was included as a part of HHT to extract the instantaneous frequency, instantaneous amplitude and instantaneous phase from each IMF. In case of well-log data, these instantaneous attributes provide information about the frequencies present in the data, their respective amplitudes and phases with respect to depth. The IMFs extracted by EMDT have a well-defined Hilbert transform and thus HSA of IMFs provides a more meaningful space–frequency–amplitude description of a signal. The analytic function obtained from the Hilbert transform is given by $A(s) = x(s) + i\hat{x}(s)$, where
Fig. 7. Hilbert spectra of IMFs of gamma-ray log data of (a) Well B and (b) Well C (see Fig. 5). The colour scale represents the normalized instantaneous amplitude. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

\[ \hat{x}(s) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(u)}{s-u} du, \]

defines the Hilbert transform of each IMF. \( P \) denotes the Cauchy principal value of singular integral.

The instantaneous amplitude is given by \( a(s) = (\hat{x}(s)^2 + \hat{x}(s)^2)^{1/2} \) and the instantaneous phase by \( \theta(s) = \tan^{-1} \left( \frac{\hat{x}(s)}{\hat{x}(s)} \right) \). The instantaneous frequency is \( \omega(s) = \frac{d\theta(s)}{ds} \) [13, 14].

3.3. Heterogeneity analysis

Heterogeneity analysis of well-log data quantifies the qualitative observations made in the HSA exercise and provides a quantitative estimate of the degree of heterogeneity in the subsurface. A careful observation of the derived IMFs through EMDT explains that EMDT acts as a filter bank [26], much the similar way the discrete wavelet transform does [2, 4]. Since the EMDT functions as a filter bank, there exists a non-linear relationship between the IMF number \( m \) and its respective mean wavelength \( (l_m) \) of each IMF, given by \( l_m = k\rho^m \), where \( \rho \) designates the heterogeneity index and \( k \) is the constant [27]. As the above relation between \( l_m \) and \( m \) explains, the estimated \( \rho \) values bear an inverse relation with the heterogeneity of the subsurface. Smaller \( \rho \) values designate higher heterogeneity of the subsurface and vice-versa. The two-step procedure to determine \( \rho \) is as follows.

**Step 1:** For each IMF, calculate the mean wavelength \( (l_m) \) defined as the ratio of total number of points to the total number of peaks [29].

**Step 2:** Draw a plot between the IMF number, \( m \) and logarithm of \( l_m \). The antilogarithm of the slope of this line defines the heterogeneity index, \( \rho \).
Fig. 8. Linear regression between the logarithm of mean wavelength of each IMF and the respective IMF number for gamma-ray log of hydrocarbon zone in L-III limestone of (a) Well B and (b) Well C. See Section 3.3 for more details.

Fig. 9. Histogram representation of the average shale volume (in percentage) corresponding to the hydrocarbon reservoir zone in L-III limestone of (a) Well B and (b) Well C, indicating higher shale volume in Well C (34%) than in Well B (28%).

Fig. 10. Cross-over between the neutron porosity log (solid line) and density log (dashed line) seen at the top of the L-III zone in Well C. The highlighted portion of the data (shown in dashed box) signifies a characteristic of the presence of a continuous thick 20 m gas zone (after [12]) within the L-III zone. The dotted line within the dashed box represents the gas-corrected neutron porosity.

4. Application of EMDT and HSA to well-log data

4.1. EMD of gamma-ray log of L-III zones of Well B and Well C

The major reservoir in the study area is limestone (identified as L-III in the lithostratigraphy) having a thickness of about 470 m and located at an average depth of about 1.3 km below the sea floor (see [1]). The hydrocarbon zone within L-III in both the wells is approximately 60 m thick, whose data are sampled at a rate of 0.1524 m. The EMDT was applied to this 60 m portion of gamma-ray logs of Well B and Well C for reservoir characterization. The EMDT produced 8 IMFs and a residue
for gamma-ray logs of both the wells (Fig. 5). The reconstructed gamma-ray logs, obtained by synthesizing all the IMFs and the residue of respective wells produced the original signals with least error of the order of about $10^{-13}$ between the original and synthesized logs (Fig. 6).

4.2. Hilbert spectral analysis

The IMFs obtained from the empirical mode decomposition of the gamma-ray log of L-III reservoir zone in both the wells were subjected to Hilbert spectral analysis. The instantaneous amplitude and the instantaneous frequency estimated as a function of depth for each IMF are shown in Fig. 7 for Well B (Fig. 7(a)) and Well C (Fig. 7(b)).
Fig. 12. Linear regression between the logarithm of mean wavelength of each IMF and the respective IMF number corresponding to (a) original neutron log (solid line in Fig. 10) and (b) its gas corrected equivalent (dotted line in Fig. 10), clearly explaining the lower degree of heterogeneity (high $\rho$ value) in the original neutron log compared to the higher degree of heterogeneity (low $\rho$ value) in the gas-corrected neutron porosity log. The multifractal singularity spectra (c) of the original neutron log of L-III pay zone of Well C (open circles) and its gas-corrected equivalent (stars) confirm the heterogeneity analysis results (see [12] for more details).

4.3. Heterogeneity indices for L-III reservoir zone using IMFs

As explained earlier, while the instantaneous amplitude and instantaneous frequency determined for each log serve as qualitative indicators of heterogeneity in the subsurface, the heterogeneity analysis provides a quantitative estimate of the degree of heterogeneity within the subsurface. Fig. 8 depicts the graph drawn between the logarithm of mean wavelength of each IMF and the IMF number, plotted for gamma-ray log of L-III reservoir zone of Well B (Fig. 8(a)) and Well C (Fig. 8(b)). The antilogarithm of the slope of the best fit line represents the heterogeneity index, $\rho$.

5. Results and discussions

Knowledge of porosity, net thickness and quantity of shale present in the reservoir zone is essential for effective characterization of reservoirs [30]. An intercomparison of such information between different wells facilitates to correctly assess their heterogeneous behaviour. The EMDT (Fig. 5(a), (b)) and the HSA (Fig. 7(a), (b)) performed on gamma-ray logs of Wells B and C show different frequency characteristics (IMFs), suggesting the presence of a certain degree of heterogeneity in both the wells. To assess the degree of heterogeneity in a well, the contribution of each IMF can be separately studied using its normalized instantaneous amplitude and instantaneous frequency obtained from HSA. A general understanding is, if a signal is generated from a larger heterogeneous system, it will definitely produce more IMFs than a signal generated from a smaller heterogeneous system. While this may be true, sometimes, it will be difficult to ascertain the heterogeneity in different systems, when signals generated from them produce equal number of IMFs, as in the present case (Figs. 5 and 7). In such situations, it becomes difficult to establish, which system is more heterogeneous than the other?

Among the Hilbert spectra of several IMFs shown in Fig. 7, the spectra of IMF 2 (representing the second highest frequency component present in the gamma-ray log) of Well B (Fig. 7(a)) and Well C (Fig. 7(b)), suggest that the normalized amplitude of the representative instantaneous frequency is smaller in Well B than in Well C. Further, corresponding to the spectra of IMF 2, while the amplitudes of the instantaneous frequency are small (blue colour) throughout the data length in Well B, they are relatively higher (red colour) in Well C. It is important to understand here that the existence of many thin beds in the reservoir zone will reflect in the form of high frequency components, which can be easily seen in the spectra of initial IMFs representing high frequencies in any log. So, within the 60 m thick carbonate reservoir zone of both the wells, such large amplitude high frequency components are more in Well C than in Well B, which in turn signifies higher degree of
heterogeneity in Well C than in Well B. Similarly in the successive IMFs also, the amplitudes of the respective instantaneous frequencies are relatively higher in Well C than in Well B. This provides a first-hand qualitative assessment suggesting Well C to be more heterogeneous than Well B. This is further corroborated by the heterogeneity index, \( \rho \) (see Fig. 8), which clearly suggests that Well C (\( \rho = 1.68 \)) is more heterogeneous than Well B (\( \rho = 1.96 \)) (cf. Section 3.3). The cause for such a high heterogeneity in Well C is attributed to the higher shale content in Well C (34%) than in Well B (28%), as determined from the histogram analysis of the shale volume (see Fig. 9) estimated using gamma-ray logs of both the wells. Thus it became evident that among the hydrocarbon reservoirs within the L-III zones of both wells, Well C is more heterogeneous than Well B. The HSA has been very effective in presenting the distinct differences in the heterogeneity in the reservoirs within the L-III zone (Fig. 7).

A careful observation of the neutron porosity log and the bulk density log (see Fig. 10) shows a cross-over of about 20 m thick from the top of L-III limestone in Well C. In other words, the density and porosity logs show clear phase differences within the top 20 m in the L-III limestone in Well C. This cross-over occurs as a result of low density and low hydrogen index due to presence of gas in the reservoir. To understand the role of gas on the heterogeneity in this 20 m thick zone (see the highlighted portion of the data in the 1360 m–1380 m range in Fig. 10), the neutron porosity log has been corrected for gas (represented with dotted line in Fig. 10), using the formula, 
\[
\Phi_C = \frac{\Phi_N^2 + \Phi_D^2}{2},
\]
where \( \Phi_C \) represents the gas-corrected neutron porosity, \( \Phi_N \) denotes the original neutron porosity and \( \Phi_D \) is the density derived porosity, given by 
\[
\Phi_D = \frac{\rho_{matrix} - \rho_{fluid}}{\rho_{matrix} - \rho_{bulk}}
\]
[30], using the limestone matrix density \( (\rho_{matrix}) \) value as 2.71 g/cc, fluid density \( (\rho_{fluid}) \) as 1.0 g/cc and \( \rho_{bulk} \) designates the density log value at every sampled depth. Interestingly, when EMDT was performed on \( \Phi_N \) and \( \Phi_D \) data, the number of IMFs obtained for \( \Phi_N \) is five (Fig. 11(a)), and for \( \Phi_D \) is six (Fig. 11(b)). This gave first-hand information suggesting the latter to be more heterogeneous than the former. The \( \rho \) value for \( \Phi_N \) (1.86) is more than \( \Phi_D \) (1.76), confirming that \( \Phi_D \) is more heterogeneous than \( \Phi_N \) (Fig. 12(a),(b)). Through multifractal studies of the same 20 m thick zone of Well C, Subhakar and Chandrasekhar [12] also showed that \( \Phi_D \) is more heterogeneous than \( \Phi_N \) (Fig. 12(c)) suggesting that the presence (absence) of gas in the subsurface reservoir zone makes the reservoir homogeneous (heterogeneous).

Application of different signal analysis techniques, such as wavelet analysis [1], multifractal analysis [12] and the present study, to the same well-log data sets has improved the overall understanding of the characterization of the reservoir under study. While wavelet analysis [1] helped to identify the depths to the tops of pay zones, the multifractal analysis [12] and the present study helped to clearly distinguish the heterogeneity in the wells and the gaseous and non-gaseous zones within the reservoir. Therefore, application of multi data analysis techniques to well-log data will be quite worthwhile to improve the interpretation and reservoir characterization.

6. Conclusions

1. The fully data adaptive and nonlinear HHT proves its worthiness over the other conventional signal analysis techniques in unravelling the heterogeneous spatial distribution of the reservoir and non-reservoir zones within the main L-III reservoir zone in both the wells.
2. The combined EMD and HSA techniques, together with the heterogeneity analysis have been important and essential to quantify the degree of heterogeneity within the L-III reservoir zones of both the wells of the Bombay High region.
3. L-III zone of Well C is more heterogeneous than that of Well B due to higher shale volume in the former.
4. Heterogeneity analysis results of neutron log agree with those of multifractal analysis and confirm that the presence (absence) of gas in the subsurface reservoir zone makes the reservoir homogeneous (heterogeneous).

Acknowledgements

The authors express their sincere thanks to Oil and Natural Gas Corporation (ONGC) Ltd., India, for providing the necessary data and their permission to publish the work. They also express their sincere thanks to the handling editor, Prof. Eugene Stanley and two anonymous referees for their kind words of encouragement and critical comments, which have helped to improve the paper.

References


