Completed Ensemble Empirical Mode Decomposition: a Robust Signal Processing Tool to Identify Sequence Strata

H Purba, J T Musu, S A Diria, W Permono, O Sadjati, I Sopandi, F Ruzi
1 Research and Development Center for Oil and Gas Technology "Lemigas", Jl. Raya Ciledug, Jakarta Selatan, 12230, Indonesia
2 BOB PT. Bumi Siak Pusako - Pertamina Hulu Energi, Jakarta, Indonesia

Abstract. Well logging data provide many geological information and its trends resemble non-linear or non-stationary signals. As long well log data recorded, there will be external factors can interfere or influence its signal resolution. A sensitive signal analysis is required to improve the accuracy of logging interpretation which it becomes an important thing to determine sequence stratigraphy. Complete Ensemble Empirical Mode Decomposition (CEEMD) is one of non-linear and non-stationary signal analysis method which decomposes complex signal into a series of intrinsic mode function (IMF). Gamma Ray and Spontaneous Potential well log parameters decomposed into IMF-1 up to IMF-10 and each of its combination and correlation makes physical meaning identification. It identifies the stratigraphy and cycle sequence and provides an effective signal treatment method for sequence interface. This method was applied to BRK-30 and BRK-13 well logging data. The result shows that the combination of IMF-5, IMF-6, and IMF-7 pattern represent short-term and middle-term while IMF-9 and IMF-10 represent the long-term sedimentation which describe distal front and delta front facies, and inter-distributary mouth bar facies, respectively. Thus, CEEMD clearly can determine the different sedimentary layer interface and better identification of the cycle of stratigraphic base level.

1. Introduction
The sequence stratigraphy method of stratigraphic interpretation combines the stratal architecture (geometric relationship) of the sedimentary facies within a framework of geometries assumed to be the result of the repeated fill of accommodation by successions of sedimentary strata [1] and the chronological order of their accumulation to enhance the interpretation of depositional setting and prediction of stratatal continuity [2]. Interface identification and sequence partition are the basic of sequence stratigraphy which well log data, core data, outcrop samples and another supporting data (e.g seismic) are used for its object. A problem may arise when information from core data and outcrop are limited and seismic data has poor quality of temporal resolution. The availability of well logging data becomes more important for the partition of sequence stratigraphy [3]. However, during well logging recording, various disruptions can interfere its resolution and quality so that logging data can be non-stationary signal in space.

Complete Ensemble Empirical Mode Decomposition (CEEMD), a robust signal processing approach can handle non-stationary and non-linear process [4, 5, 6]. CEEMD calculates local mean curve which decomposes the complex signal to a series of component function for instantaneous
frequency with apparent physical meaning. In this paper, the CEEMD method is applied for well log data processing and identifying sequence interface in logging.

2. Subsurface Geology
Study area is located on Beruk field, Central Sumatra Basin which consists of 36 wells with the inter-well distance 100 meters up to 2 kilometers (Figure 1).

Correlation model from well log describes eight zones (Zone 1 to 8) of delta front facies model from west to south-east (Figure 2a) and north to south traverse (Figure 2b) which interspersed by distal delta front facies, shale break, and prodelta facies.

**Figure 1.** Study location on Beruk field Central Sumatra Basin (shown by red rectangle) with massive production wells. Inter-well distance varies from 100 meters up to 2 kilometers
Figure 2. Well log correlation identifies 8 zones of delta front facies model and on the top of zone 1 bounded by Telisa Formation from west to south-east (a) and north to south (b). From zone 1 to 6 or 7, delta front facies series interbedded by distal delta front facies and shale break. When getting into zone 7 or 8, there are interval of prodelta facies.

Commonly, the facies comprises mouth bar sandstone, inter-distributary mouth bar sandstone. The thickness of delta front facies varies laterally from north (proximal) to south-east (distal) likewise west.
to east. Zone 1 has good facies distribution from proximal to distal compared to zone 2 and zone 3 (Figure 3a). Zone 2 and 3 become less from proximal to distal and the facies is relative thin or little different in the western and eastern part (Figure 3b and 3c). For zone 4 and 5, delta front facies is thin in the middle but rather thick in proximal and distal (Figure 3d and 3e). Each of thickness distribution of delta front facies of zone 7 and 8 (Figure 3g and 3h) is thin from north to south and west to east but rather thicker in the eastern part of zone 6 (Figure 3f).

![Facies maps](image)

Figure 3. Facies map distribution of zone 1, 2, 3, 4, 5, 6, 7, and 8, labeled by (a), (b), (c), (d), (e), (f), (g), and (h), respectively.

3. Theory
CEEMD is an extension method of Ensemble Empirical Mode Decomposition (EEMD) which can separate the high-frequency noise from the raw data by adding the white noise, but the low-frequency noise cannot be reduced. This method is an adaptive signal analysis approach based on the signal
characteristic of local extrema. It decomposes a time signal into series of Intrinsic Mode Function (IMF) components, and each IMF satisfies the following conditions:

- Over the entire time range, the number of zero-crossings must be equal or differ by one at most;
- At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero [7].

CEEMD process should be initiated by Empirical Mode Decomposition (EMD) process which its steps can be shown as follows:

- Find all the maxima and minima of the signal and use cubic splines to match the upper envelope \( U(t) \) and the lower envelope \( L(t) \), respectively. Then, calculate the mean value of the upper and lower envelope \( M(t) \):
  \[
  M(t) = \frac{U(t) + L(t)}{2} \tag{1}
  \]
- Subtract the average envelope \( M(t) \) from the original signal \( S(t) \) to obtain a new series \( N(t) \):
  \[
  N(t) = S(t) - M(t) \tag{2}
  \]
- Use the above IMF decision criteria to determine the sequence \( N(t) \). Repeat 1 and 2 step until sequence \( N_i(t) \) meets the conditions of IMF. \( N_1(t) \) is the first IMF component \( K_1(t) \).
- Subtract \( K_1(t) \) from the signal \( S(t) \) to get a difference signal \( R_1(t) \). The sequence \( R_1(t) \) is treated as a new preprocessing signal to decompose, and repeat step 1, 2, and 3 until got a second-order IMF component. Stop the process if the difference sequence is constant or a monotonic function.

Each IMF correspondingly contains different frequency components of the signal from high to low frequency, representing inherent mode characteristics of the nonlinear signal, and they will vary with the change of the signal. The residual function stands for the average trend of the signal, and the original signal can expressed as a sum of the IMFs and the residual function:

\[
S(t) = \sum_{i=1}^{N} K_i(t) + R_n(t) \tag{3}
\]

The procedure of CEEMD was described as follows [8]:

- Add noise onto raw signal and obtain the first EMD component of the data with noise.
- Repeat the decomposition and add different noise realization
- Compute ensemble average to define it as the IMF-1:
  \[
  K_1(t) = \frac{1}{N} \sum_{i=1}^{N} G_i [S(t) + \alpha T_i] \tag{4}
  \]

Here, \( S(t) \) is the original signal, \( T_i(t) \) is the different white noise, \( \alpha \) is a ratio coefficient, \( G_i [ ] \) denotes generating the \( i \) th IMF component, and \( N \) is the number of realizations. Next, calculate the first different signal \( d_1(t) \):

\[
d_1(t) = S(t) - K_1(t) \tag{5}
\]
Then, set \( d_1(t) + \alpha G_1 [T_i(t)] \) \((i = 1, 2, \ldots, N)\) as the new signal for decomposition. To obtain the first IMF component, the ensemble average needs to be calculated and the result is the second IMF component:

\[
K_2(t) = \frac{1}{N} \sum_{i=1}^{N} G_1 [d_1(t) + \alpha G_1 [T_i(t)]]
\]  

(6)

Finally, repeat the above steps, thus, it can be obtained:

\[
S(t) = \sum_{i=1}^{N} K_i(t) + \text{Res}(t)
\]  

(7)

Where, \( S(t) \) is original data, \( K_i(t) \) is \( i \)th IMF component, and \( \text{Res}(t) \) is the last residual function.

4. CEEMD-Well Log Curve Type

The result of well log data decomposition in IMF series reflects different sedimentary cycle grades. If the inter-correlation of sequence interfaces can be identified, it can be considered as the standard to split the sequence of well log data. Base level occurs when the river reaches a balance under the dynamics (potential energy surface). Sedimentary boundary will extend towards the base level through sedimentary or erosion action to get a new balance which is called sequence formation cycle [9, 10]. Various cycle of sequence stratum can be seen in well log data as regular or irregular pattern of mineral grains granularity and lithology. This pattern can be considered as different amplitude and frequency characteristic in well log curve likewise signal pattern. Then, the break points of curve are fixed as a sequence interfaces [11]. Besides, it is useful to make the mirror image on response characteristic of CEEMD curve of well log parameters reflecting all kinds of sedimentary environment which is summarized in five basic types (See on Table 1).

| Curve type               | Indicating type | Analysis of curve type                                                                 |
|--------------------------|-----------------|---------------------------------------------------------------------------------------|
| Gradually varied type    | Top             | Poorly sorted. To the top, amplitude becomes less gradually and the curve turns graded bed sequence |
|                          | Bottom          | Poorly sorted. To the bottom, amplitude becomes less gradually and the curve turns reverse grain size order |
| Abrupt change type       | Top             | Well sorted. An abrupt change at the top. Stacked upward-finishing cycle.               |
|                          | Bottom          | Well-sorted. An abrupt change at the bottom. Stacked downward-finishing cycle            |
| Fluctuating type         | Top             | Granularity turns from coarser below to finer and the amplitude becomes less drastically. It belongs to typical transgression and shows positive grain size order |
|                          | Bottom          | Granularity turns from finer below to coarser and the amplitude becomes bigger drastically. It belongs to typical regression and shows reverse grain size order |
| Thin interbedding type   | Inter-bedding   | Multi-lithology alternately appears and changes quickly. It often can be seen in thin inter-bedding |
| Block combination type   | Simplex lithology | The lithology is simplex, so the grain size of sediment changes are small and stable |

Table 1. Basic type of sedimentary environment corresponding to CEEMD curve of well log
5. Methodology
CEEMD decomposes the original data into IMF with different frequency and residual. Each IMF describes characteristic dimension and information. First and second IMF tend to high frequency reflecting the short-term cycle characteristic while the next IMFs (fourth and so on) have middle- low frequency reflecting the middle and long term stable cycle. Residual refers to the data trend or zero drift of instrument [12]. Based on this rule, correlation of IMF obtained after CEEMD can be determined as follows:

- Calculate the correlation coefficient for every IMF(i) and its original signal. Mostly, those IMFs which have a good correlation with the original signal are the best signal components reflecting stable sedimentary environment.
- Analyze the correlation between different IMFs and combine those IMF(i) with good correlation. This indicates that good correlation reveals will be a consistent in variation style. Combining them can reduce interference factors and strengthen signals that reflect geological characteristic.

Gamma ray (GR) log is a common logging method to measure intensity of natural gamma ray in rock along the well bore. GR curve reflects the granularity, sorting and clay minerals content, which can be used to judge lithology, compare stratum, and estimate the shale content. Spontaneous potential (SP) log is an effective logging method for analyzing geological profile in borehole, which can be used to measure shaft spontaneous potential changes in an open hole. SP curve is often used to divide the sandy-shaly profile into sandstone and shale lithology, compare the stratum, define the interface position of filtering layer, calculate the content of shale of the stratum, and spread the sedimentary face to study, etc. The presented analysis indicates that these two well logging data (GR and SP) are the sensitive parameters that are suitable for sequence division of the study area. At the same time, its curve indication has consistent rule: at the profile of sandstone and shale section well, the value of these two curves is low at sandstone while the value is high at shale. So, these parameters can be combined. Normalization treatment on natural gamma ray and spontaneous potentials curves was done to eliminate the influence of units and select linear normalization formula [13]:

\[
\text{GR}_{\text{norm}} = \frac{\text{GR} - \text{GR}_{\text{min}}}{\text{GR}_{\text{max}} - \text{GR}_{\text{min}}} \tag{8}
\]

\[
\text{SP}_{\text{norm}} = \frac{\text{SP} - \text{SP}_{\text{min}}}{\text{SP}_{\text{max}} - \text{SP}_{\text{min}}} \tag{9}
\]

Average values of the normalized parameter (GR and SP) were combined to gain a new parameter marked GR – SP:

\[
\text{GR} - \text{SP} = \frac{\text{GR}_{\text{norm}} + \text{SP}_{\text{norm}}}{2} \tag{10}
\]

6. Result and Discussion
The decomposition of GR – SP data, for instance well BRK-30, are manifested to each of IMF presented in Figure 4. IMF-1 to IMF-5 reflects strong vibration and high frequency while IMF-6 up to residual (R) signal more low frequency. A different IMF component can be obtained through the decomposition of CEEMD. Superposing of decomposed components, the original signal may be reproduced. Using such characteristic, the universe accumulation of the IMF-1 up to IMF-10 and trend item R of each stages of GR – SP is repeated, namely, the combination from the component with low frequency to high frequency (see on Figure 5).
Figure 4. The composition of GR – SP (well BRK-30) into IMFs 1-10 which is the new combined parameter between GR and SP. RES panel represents the final residual which showed the data trend.

Figure 5. Reconstruction of GR – SP (well BRK-30) components. Blue solid line is GR – SP and red broken line is correlation coefficient of reproduced sequence with GR – SP. (a) Residual (red broken line with $R^2 = 0.755$) (b) Residual + IMF10 (red broken line with $R^2 = 0.819$). (c) Residual + IMF-10 + IMF-9 (red broken line with $R^2 = 0.837$). (d) Residual + IMF-10 + IMF-9 + IMF-8 (red broken line with $R^2 = 0.849$). (e) Residual + IMF-10 + IMF-9 + IMF-8 + IMF-7 (red broken line with $R^2 = 0.871$). (f) Residual + IMF-10 + IMF-9 + IMF-8 + IMF-7 + IMF-6 (red broken line with $R^2 = 0.926$). (g) Residual + IMF-10 + IMF-9 + IMF-8 + IMF-7 + IMF-6 + IMF-5 (red broken line with $R^2 = 0.955$). (h) Residual + IMF-10 + IMF-9 + IMF-8 + IMF-7 + IMF-6 + IMF-5 + IMF-4 (red broken line with $R^2 = 0.969$). (i) Residual + IMF-10 + IMF-9 + IMF-8 + IMF-7 + IMF-6 + IMF-5 + IMF-4 + IMF-3 (red broken line with $R^2 = 0.984$). (j) Residual + IMF-10 + IMF-9 + IMF-8 + IMF-7 + IMF-6 + IMF-5 + IMF-4 + IMF-3 + IMF-2 (red broken line with $R^2 = 0.996$). (k) Residual + IMF-10 + IMF-9 + IMF-8 + IMF-7 + IMF-6 + IMF-5 + IMF-4 + IMF-3 + IMF-2 + IMF-1 (red broken line with $R^2 = 1$).
IMF-5, IMF-6, IMF-7 and IMF-10 acted importantly in original sequence (Figure 5). The changes of original sequence were mainly caused by the oscillation of these four functions. They all have consistent change pattern. This pattern reflects the characteristics of stable sedimentary environment of the stratum. Specifically, the wave scope of IMF-5, IMF-6, and IMF-7 covered several meters to tens of meters and the wave width of IMF-10 is over one hundred meters. The cycles with different base levels commonly are divided according to time distribution. In practical layers, the thickness of short-term cycle is of several meters to dozens of meters. The middle term cycle is dozens of meters to nearly one hundred meters and the long term cycles are nearly one hundred meters to several hundred meters. Therefore, IMF-5, IMF-6, and IMF-7 reflect the short-term circle characteristics, while IMF-10 the middle and long term characteristics.

According to sedimentology, the stratum in different periods may have different sedimentation rules while the stratum in the same period may have similar characteristic. At well BRK-30, combination of IMF-5 plus IMF-6, IMF-7, IMF-10 describes a short-term sedimentation rules of the stratum (Figure 6a to 6c) while IMF-6 plus IMF-7, IMF-10 or IMF-7 plus IMF-10 most likely middle-term sedimentation (Figure 6d to 6f). For long-term sedimentation is illustrated on Figure 6g which IMF-10 describes the stratum boundary of inter-distributary mouth bar. At well BRK-13, short-term sedimentation and middle-term can be depicted on Figure 6h and Figure 6i or 6j, respectively. The long-term or period lies on IMF-10 (Figure 6k) and consistently showing the boundary of inter-distributary mouth bar.

**Figure 6.** Illustration and identification of short, middle, and long-term cycle sequence stratum of well BRK-30 correlated to well BRK-13 using CEEMD method (a). Well BRK-30 : IMF-5 + IMF-6 (b). IMF-5 + IMF-7 (c). IMF-5 + IMF-10 (d). IMF-6 + IMF-7 (e) IMF-6 + IMF-10 (f) IMF-7 + IMF-10 (g) IMF-10 (h) Well BRK-13: IMF-5 + IMF-6 (i) IMF-7 + IMF-10 (j) IMF-8 + IMF=9 (k) IMF-10

7. Conclusion

CEEMD is robust signal analysis method to decompose the inherent characteristics in non-linear and non-stationary signal which this trend also occurs in well log recording. This method is helpful for identifying sequence strata with different frequency intrinsic (IMF). Combination of these IMFs can represent short, middle, and long-term sedimentation through characteristic curve (Table 1). In this study, short-term and middle-term describes the facies of distal front and delta front. These cycles were identified at the combination of IMF-5, IMF-6, and IMF-7 and calibrated to facies analysis at well. Commonly, the low frequency (IMF-9 and IMF-10) trend represents the long-term sedimentation
and the cycle was detected at the boundary of inter-distributary mouth bar facies. Thus, the combination of IMFs determines the different sedimentary layer interface and better identification of the cycle of stratigraphic base level. Furthermore, downscaling sequence characteristic can be applied to the seismic data for lateral distribution.

Acknowledgments
The authors thank to BOB PT. Bumi Siak Pusako - Pertamina Hulu Energi for the permission to publish this study results

References
[1] Neal J and Abreu V 2009 Sequence stratigraphy hierarchy and the accommodation succession method Geology 37 p 779-782
[2] Catuneanu O, Abreu V, Bhattacharyya J P, Blum M D, Dalrymple R W, Eriksson P G, Fielding C R, Fisher W L, Galloway W E, Gibling M R, Giles K A, Holbrook J M, Jordan R, Kendall C G, Macurda B, Martens J O, Miall A D, Neal J E, Nummedal D, Pomer L, Posamentier H W, Pratt B R, Sarg J F, Shanley K W, Steel R J, Strasser A, Tucker M E and Winker C 2009 Towards the standardization of sequence stratigraphy Earth-Science Reviews vol. 92 p 1-33.
[3] Van Wagoner J C, Mitchum R M, Campion K M and Rahaman V D 1990 Siliciclastic Sequence Stratigraphy in Well Logs, Cores, and Outcrops: Concepts for High-Resolution Correlation of Time and Facies AAPG Methods in Exploration Series, No. 7 (The American Association of Petroleum Geologists, Tulsa, USA)
[4] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C and Liu H H 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. A 454 p 903-995.
[5] Huang N E, Shen Z, and Long S R 1999 A new view of nonlinear water waves: the Hilbert spectrum Ann. Rev. Fluid Mech. 31 p 417-457
[6] Huang N E and Wu Z 2004 A review on Hilbert–Huang transform: Method and its applications to geophysical studies Rev. Geophys. 46 2 RG2006
[7] Li Y J, Hu X Y and He Z X 2010 Application situation and prospects of Hilbert-Huang Transform in geophysics, Chin.J.Eng.Geophys. 7 p 537-543
[8] Torres M E, Colominas M A, Schlotthauer G, and Flandrin P 2011 A complete ensemble empirical mode decomposition with adaptive noise: ICASSP IEEE p 4144–4147.
[9] Zheng R C, Yin S M and Peng J 2000 Sedimentary dynamic analysis of sequence structure and stacking pattern of base-level cycle, Acta Sediment. Sin. 18 3 p 369-375 (in Chinese).
[10] Zheng R C, J. Peng and Wu C R 2001 Grade division of base-level cycles of terrigenous basin and its implications, Acta Sediment. Sin. 19 2 p 249-255 (in Chinese).
[11] Serra O and Abbott H T 1982 The contribution of logging data to sedimentology and stratigraphy, Soc. Petrol.Eng.J. 22 1 p 117-131
[12] Zhang T and Nie C Y 2011 Information extraction of acoustic logging for oil layers based on EMD time-frequency analysis, J. Changchun Univ. 21 8 p 15-18 (in Chinese).
[13] Zhao N and Li R 2015 EMD Method Applied to Identification of Logging Sequence Strata. Acta Geophysica Vol.63 5 pp 1256-75