e-ISSN 2580 - 0752

BULLETIN OF
GEOLOGY
Fakultas Ilmu dan Teknologi Kebumian (FITB)
Institut Teknologi Bandung (ITB)

LINEAR REGRESSION FOR FLOODING SURFACE IDENTIFICATION IN WELL LOG, AND OUTCROP IMAGE

EPO PRESETYA KUSUMAH[1,3], RIDHA SANTIKA RIADI[2], TEGUH SURINO SETIAWAN[2]

[1]Department of Geology, Universitas Pertamina, Email: epo.pk@universitaspertamina.ac.id
[2]Pertamina Hulu Sanga-sanga
[3]SedStrat Analytic

Abstract - Defining parasequences manually would take a huge amount of time. Interpretation subjectivity has also become an issue among stratigrapher when they are dealing with parasequence boundary identification which may resulting in inconsistency of parasequence identification. This paper means to present the use of automation in parasequence boundary identification using simple linear regression method in synthetic data, well log data, as well as outcrop image data.

In stratigraphy, vertical succession of lithology holds a very important meaning. Vertical succession of lithology in paralic setting where deposition occurred in a certain sea level might shows coarsening upward vertical succession. In the event where flooding occurred and sea level abruptly rise, the coarsening vertical succession might be disturbed by sharp change of lithology into finer particle, or simply called vertical discontinuity. Stratigaphers may use vertical discontinuity to identify the presence of flooding surfaces or parasequence boundaries.

Linear regression can be used to identify vertical discontinuity by measuring error occurred due to linear regression prediction. Vertical succession that showing deposition continuity might show small error number in the data where vertical disturbance occurred. The error value might increase significantly. Thus, it would be possible to determine flooding surface using linear regression by applying some threshold. This method has been proven to work using both well log data and outcrop image data which might ease stratigraphy analysis workflow in general.

Key words: Parasequence, automation, quantitative sedimentology, quantitative stratigraphy, sequence stratigraphy, computation.

1. INTRODUCTION

A parasequence set is a set of genetically related parasequences that form a unique stacking pattern bounded by a major marine flooding surface and their contrasting surfaces in most cases (van Wagoner et al., 1988). Defining parasequence boundary is the very first steps that geologist must do in order to conduct sequence stratigraphy correlation. In regressive paralic settings such as deltaic or beach deposit, a parasequence commonly shows shallowing-upward vertical succession and, in many cases showing coarsening-upward succession.

Parasequence boundary may occur when sea-level rise, which may cause the depositional condition to change. As the sea level rise, depositional conditions become relatively deeper thus finer-grained sediment may be deposited in the basin. Parasequence boundary, known as flooding surface (FS), may be identified by abrupt change in vertical succession from shallower deposition into deeper depositional setting. In lithology point of view, flooding surface may be identified by abrupt change lithology from coarse grained sediment into finer grained sediment.

Not all FS are characterized by sharp lithology change from coarse to fine-grained sediment. Deltaic deposits that have strong fluvial influence may show the combination of coarsening-upward and fining-upward succession on the top of parasequence. Coarsening-upward succession signifies the deposition of prodelta – delta front – delta plain facies, while fining upward succession signifies the deposition of fluvial channel facies (Septama et al., 2018). Regardless the condition, this experiment is meant to test the algorithm that being developed in the condition here parasequence where coarsening upward characteristic, and combination between coarsening upward and fining upward succession are presents. Nevertheless, parasequence in carbonate deposition systems are even more difficult to interpret, especially using mere well log data. Almost every depositional element possesses the same mineralogical composition (Wilson, 1975; Wilson & Wilson, 1975) with small contrast in well log respond . Therefore, it is almost impossible to identify flooding surfaces without helps from core or at least image logs.

Geologist in oil and gas company may spend substantial amount of time to interpret flooding surface in geological data, especially in the oil field that possess high-frequency deltaic deposits and thousands of well log data. Inconsistency may become another issue when geologists deal with geological interpretation. An experienced geologist may change their interpretation concept midway when they are doing interpretation which interpretation results at the beginning of the project and at the end of the project may be different. Inconsistency may become much more complex when two or more geologists are working on the same dataset, as it is known that geologists may have different experiences and concepts, resulting in differences in their interpretations.

In order to tackle the interpretation time duration and inconsistency problem, a quantitative solution is required. This paper will present the possibility of using a simple linear regression model to predict flooding surface presence for synthetic data, well-log data, as well as outcrop images. This method may reduce interpretation duration as well as provide more consistent results.

REVIEW ON QUANTITATIVE STUDIES IN STRATIGRAPHIC SURFACES

Quantitative study in stratigraphy has been discussed by many authors. Many of the studies are highlighting the pattern change in stratigraphy that responsible to system tract change, or sequence stratigraphic surface. Unlike parasequence, system tract composed of one or more parasequence that showing sea level change in certain direction which might be sea level drop or sea level rise.

System tract that occurred during sea level rise is called Transgressive System Tract (TST), and system tract that occurred during sea level drop and still sea level is called Regressive System Tract (RST). The boundary between TST and RST is called Maximum Flooding Surface (MFS), and the boundary between RST and TST is called Sequence Boundary (SB) or Maximum Regressive Surface (MRS). Many system tract type that has been popularized by authors (O. Catuneanu et al., 2009; Octavian Catuneanu, 2006, 2017; Miall, 1996), each of the type possess their own characteristic. MFS, SB, and MRS may be identified easily as they exhibits unique lithological features.

The following lists are methods that gives significant impact in quantifying stratigraphic surface in sequence stratigraphy domain:

1. Markov Chain in defining cyclicity in stratigraphic succession. This paper discuss on how markov chain can be used to identify cyclicity in vertical succession. The data that being used in this publication is discrete lithological succession for measuring section in an outctop.

This paper has been referred by many authors that uses similar technique to cyclicity (might be as sequence boundary, or flooding surface), some of them are Mastej, 2002; MIALL, 1973; Schwarzacher, 1969; Sinha S et al., 2015 .

2. Signal processing method in defining sequence stratigraphic boundary. This paper uses signal processing method, like spectral analysis, Continuous Wavelet Transform (CWT) from continuous vertical succession data (i.e. Gamma Ray well log data). The outcome of this methods is mainly a trend log that shows depositional tendency (transgressive and regressive), that can be very useful in defining MFS and MRS (and SB). The derivative of this methods has been discussed in detail in Li & Guo, 2013; Nio et al., 2005; Ye et al., 2017.
3. Machine Learning for lithological prediction and Stratigraphy Analysis. The use of machine learning (and deep learning) require pre-interpreted dataset for the machine to learn and imitate the interpretation process as provided on pre-interpreted dataset. Machine learning in geology are mainly used for lithological prediction using well log data as demonstrated by Hall, 2016; Kusumah et al., 2019; Pratama et al., 2020,. Using more sophisticated machine learning method, some authors shows the possibility on using machine learning approach for Stratigraphic surface identification Gerald, 2020; Gosses, 2020.

Linear regression to identify MFS or MRS in a lithological vertical succession has been briefly discussed by Nio et al., 2005. The method however lack of description on FS identification prior to MFS and MRS identification. In contrast with Nio et al., 2005, this paper meant to highlight the possibility of using linear regression for FS identification rather than MFS and MRS in lithological vertical succession.

2. DATA AND METHODOLOGY

PRINCIPLE OF FLOODING SURFACE IDENTIFICATION WITH LINEAR REGRESSION, AND ITS APPLICATION TO SYNTHETIC DATASET

This research utilizes vertical succession (or sediment profile) that provided form synthetic data, well log, and outcrop image. Those data are converted into stratigraphic vertical succession data. Vertical succession holds two important information, they are vertical index (V) and grain size index (Gs) (**Figure 1**). Vertical index is value that represent vertical location, the value should be in integer or decimal data format, and Grain size is number that represent the size of grain with the value that ranging from 0 to 1 in decimal format, where 0 represent coarse grained sediment, and 1 represent fine grained sediment. Linear regression model and prediction conducted by using V and Gs in a windowed section (F**igure 1**).

Asuming that Gs = m*Vi + C, thus to get m and c value, linear regression modelling is required:

$$f(V_{(i-i+n)}, Gs_{(i-i+n)}) = (m, C)$$

where:

$f(V_{(i\text{-}i+n)}, Gs_{(i\text{-}i+n)})$	= linear regression model
$V_{(i\text{-}i+n)}$	= Vertical index in windowed section
$Gs_{(i\text{-}i+n)}$	= Grain size index in windowed section
i	= windows number index
n	= sliding window size
m	= linear regression coefficient
C	= linear regression intercept

Linear regression model are used to predict Gsx, which are grain size value in V(i+n+1), and err (error) are calculated by calculating the differences between predicted value (Gsx) with real value in the same vertical location index (Gs(i+n+1))

$$Gsx = m * V(i+n+1) + C$$
$$err = Gsx – Gs(i+n+1)$$

where:

Gsx	= Predicted Gs value using linear regression in V(i+n+1)
err	= Difference between predicted Gs and GSx

The data that show good linearity may have err value close to 0, any err value that deviates far from 0 may indicate nonlinear V and Gs relationship (Figure 1). Positive err value indicate abrupt Gs change from coarser sediment to finer sediment, while negative err value indicate abrupt Gs change from finer sediment to coarser sediment. Flooding surface (FS) signified by presence of abrupt change in lithology (Grain size (Gs)) that signified by large negative value of err.

In order to extract FS from err value, a cutoff must be applied. Apart from linearity of windowed data, err value will vary with other factors, such as window size and inherited noises that present on the data. To better understand on how window size affect, err value, an experiment has been conducted by calculating err on many window size and assess how much err value changes **Figure 2**.

Figure 2 show synthetic data set that consist of 700-point data that consist of Gs and V. On second column showing err value with window size of 65, and third column is err value map on different window size over the data from window size 5 to 95. The color in third column represent err magnitude, where yellow represent high positive value and dark blue represent low negative value. As can be seen it the picture, larger window set yield wider err contrast. The fourth column is err value map with -0.15 cutoff.

According to this experiment, it can be concluded that small window size only gives small err number and did not capture possible FS. Larger window on the other hand, may give contrast in err value which can be used to determine FS on vertical succession data with -0.15 cutoff. Caution must be risen on determining cutoff. Cutoff value may be different from one dataset to another dataset due to difference in inherited noise, average parasequence thickness and data range. One solution to this issue is by conducting trial and error experiment on defining cutoff using err map.

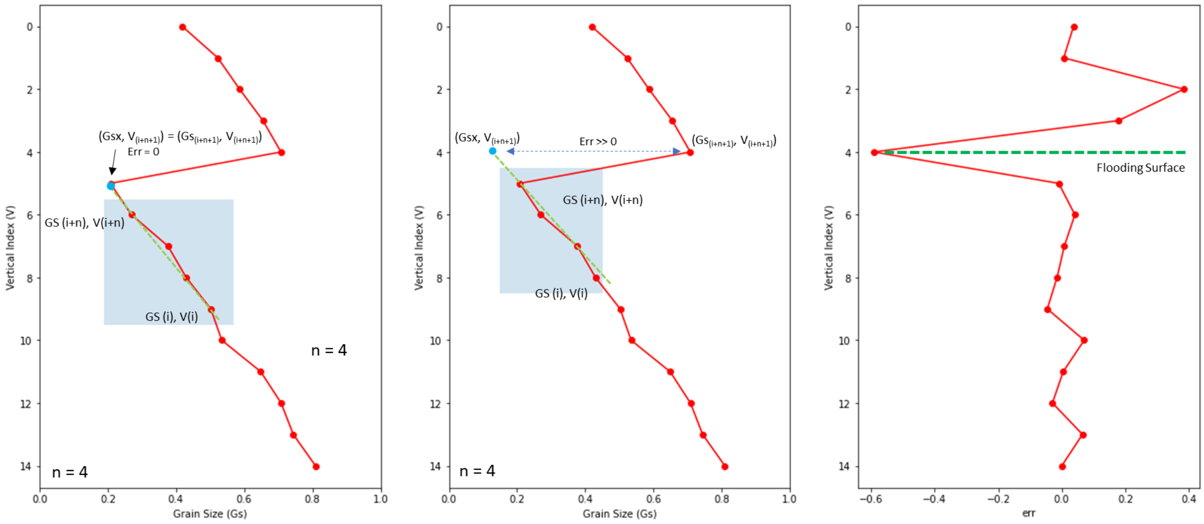

Figure 1 (left, middle) Linear regression modelling and prediction on different window location. The window on the left picture shows small prediction error due to good data linearity, while picture in the middle shows huge error that occurred due to bad data linearity. (Right) Every single error (err) calculated in different window will be recorded.

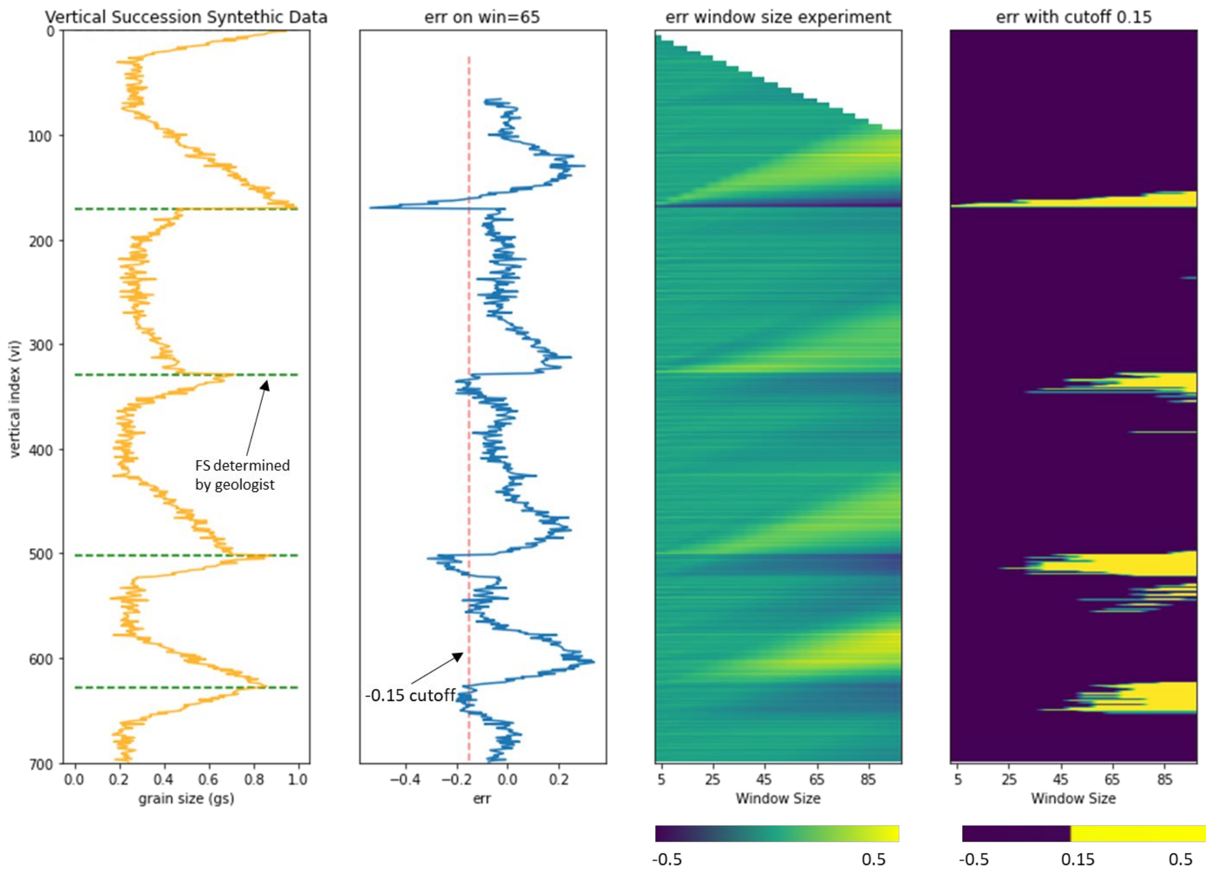

Figure 2 (1st col) Synthetic lithological data, where 0 represent sandy lithology and 1 represent shally lithology. (2nd col) Err calculation on synthetic dataset with -0.15 cutoff. (3rd col) The experiment result on applying different window size on the same dataset, it can be seen that magnitude of error and width of error changes over the windows size change. (4th column) Same as figure in 3rd col with -0.15 cutoff applied.

After cutoff has been applied, some interval data that identified as possible FS has been provided. In order to convert interval into actual Vertical Index, 1D data clustering and Min bagging method is required. 1D data clustering is a method to segregate serial data that only have one dimension into different groups. Data that possess relatively similar value will be grouped into a same group and there are minimal requirements number on how much data required to create a group. Should the data be below minimum required number, then possible FS will not be considered in any group and will be erased. Min bagging in method to find smallest vertical index in a same FS group. **Figure 3** illustrate how 1D data clustering work, and Min bagging for every single cluster.

Figure 4 shows FS prediction with linear regression result from synthetic dataset that has been clustered and Min bagging method has been applied on the data, as can be seen in the result that presented on fourth columns fits really wells with FS predicted by geologist.

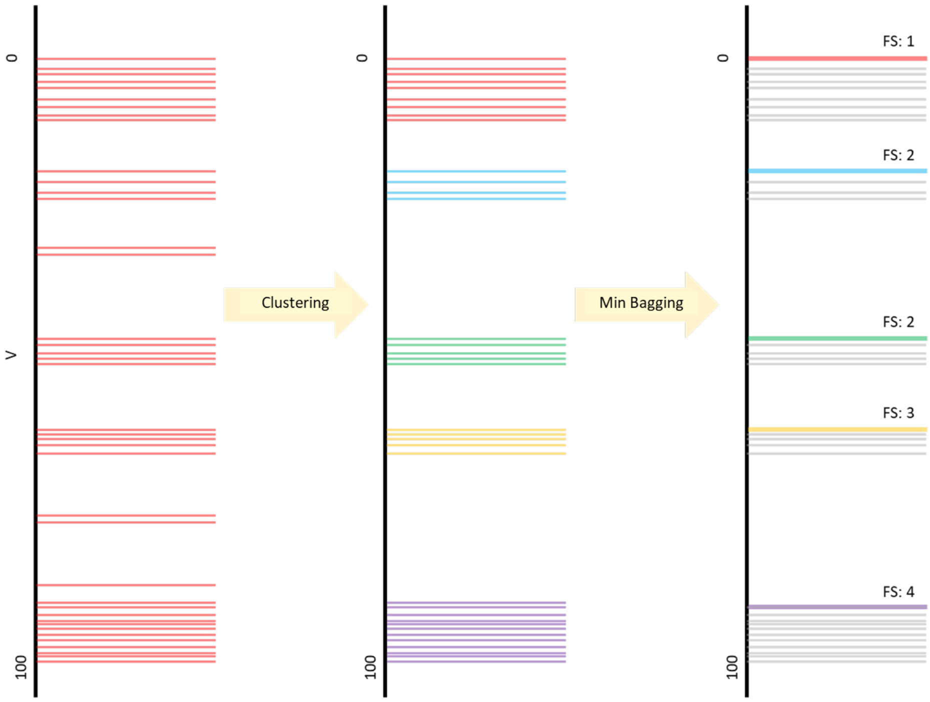

Figure 3 An illustration on how clustering and min bagging work.

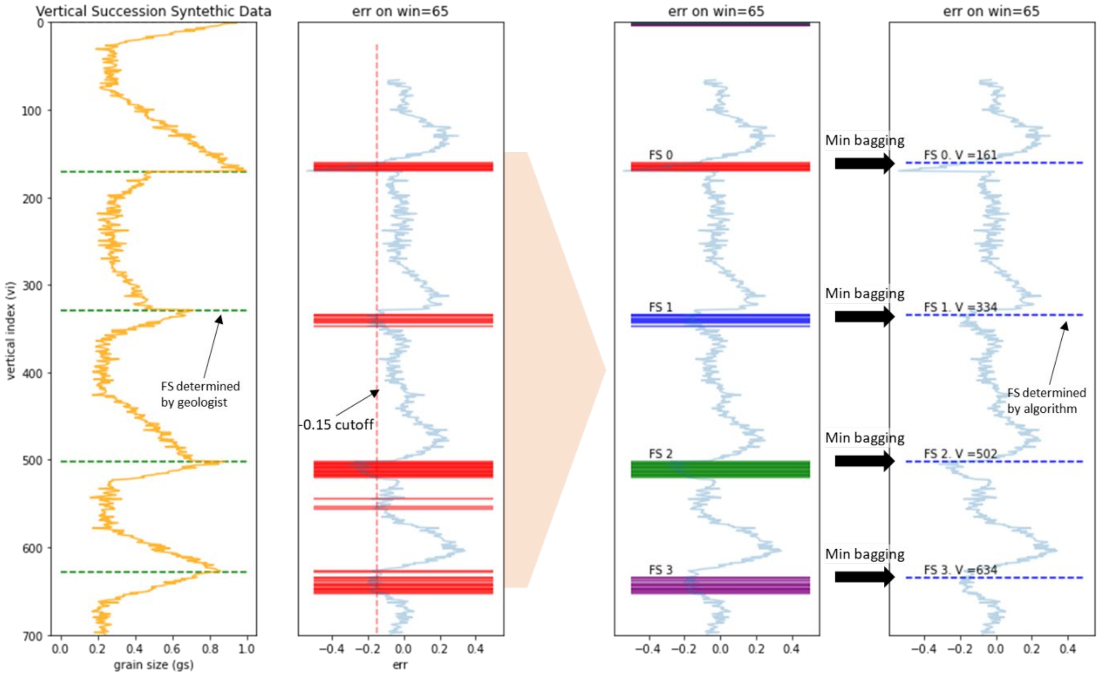

Figure 4 FS prediction result on synthetic dataset. This figure also illustrate how possible FS interval being clustered and after min bagging as been applied.

Figure 5 is summary of FS identification workflow using linear regression model for different type of data. The key different between these different data are the origin. Image data can also be treated as 1D data by extracting pixel value vertically, then continue the workflow as if it is a 1D data.

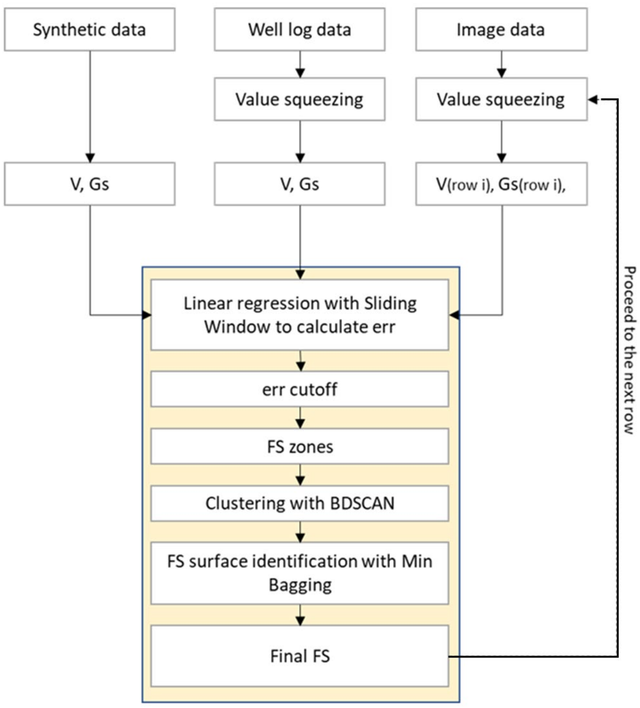

Figure 5 General workflow on how FS identification work.

3. RESULTS

IMPLEMENTATION ON WELL LOG DATA SET: B-01 KUTAI BASIN

Badak field is a giant gas producing field in Indonesia that located in Kutai basin, East Kalimantan. This field discovered in 1972 and still producing gas until this paper written. The structure of Badak field characterized by four-way dip closure that span 75km along the anticline line that probably formed in Pliocene time (Huffington, 1990).

Stratigraphically, Badak Field consist of prograding deltaic succession that dominated with fluvio-deltaic depositional environment. The deposition starts in Middle Miocene with prodelta setting with gradual change into more proximal facies, with some fluvial incision on the upper part of stratigraphy unit. Sand/shale ratio changes following the change of depositional setting, in general the sand/shale ratio showing thickening upward succession.

Figure 7 show typical well Gamma Ray log respond that present in B-01 well with its associated facies. This figure shows variability of parasequence type presents in B-01 well, there are some parasequence that shows coarsening upward succession at the bottom that followed by coarsening upward succession on top of it. This type of parasequence signify proximal deltaic deposition, allowing deposition of channel (fluvial or distributary channel).

Majority part of the well shows distinctive flooding surfaces that can be identified by abrupt lithology change from coarse lithology into finer lithology. This well present very good example of parasequence definition on well log data. **Figure 8** and **Figure 9** shows the highlight of B-1 well Gamma ray log respond, it can be observed that the vertical succession possess many FS. Presently, hundreds of well has been drilled in this Badak Field, which obviously needs significant time in order to identify FS with manual interpretation.

Gamma ray log in B-01 consist of 21201 points data reading with data spacing of 0.5 ft, **Figure 6** show statistical property of Gamma ray log data. Since the data is stretching more than 1, gamma ray log value normalization is required in order to squeeze the data range to fit 0 to 1 value to fit Gs criterion. Normalization applying these following equation:

$$Gs = (GR - P5) / (P95 - P5)$$

Once normalization has been implemented, FS can be identified by following workflow as has been described above.

Figure 8 and **Figure 9** shows the result of FS identification with using developed algorithm at depth of 2000 – 2500 ft and 6000-6500 ft respectively. It can be seen there are some inaccuracy on FS prediction that occurred due to value contrast over a small distance. Other than that, majority of FS prediction are quite good.

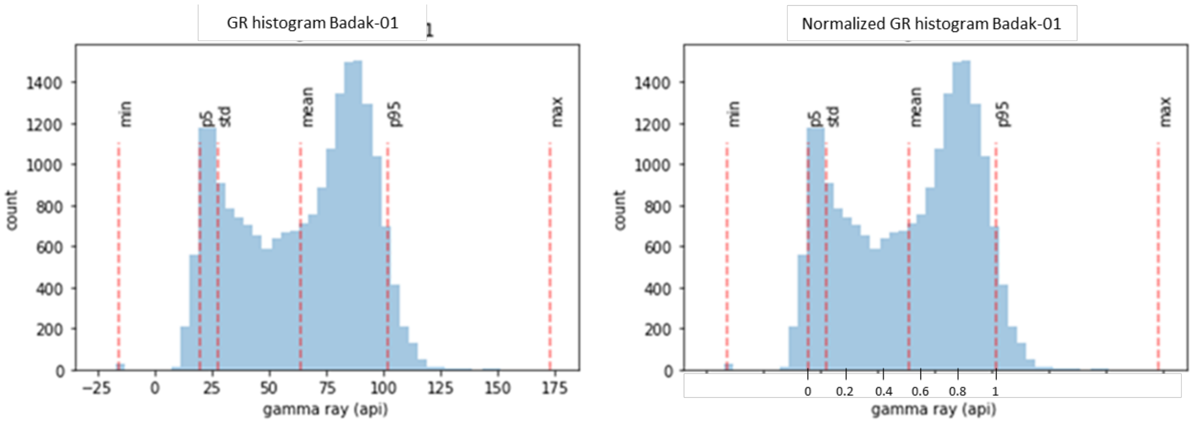

Figure 6 Histogram of B-01 Gamma Ray (left) before normalization. (right) after normalization.

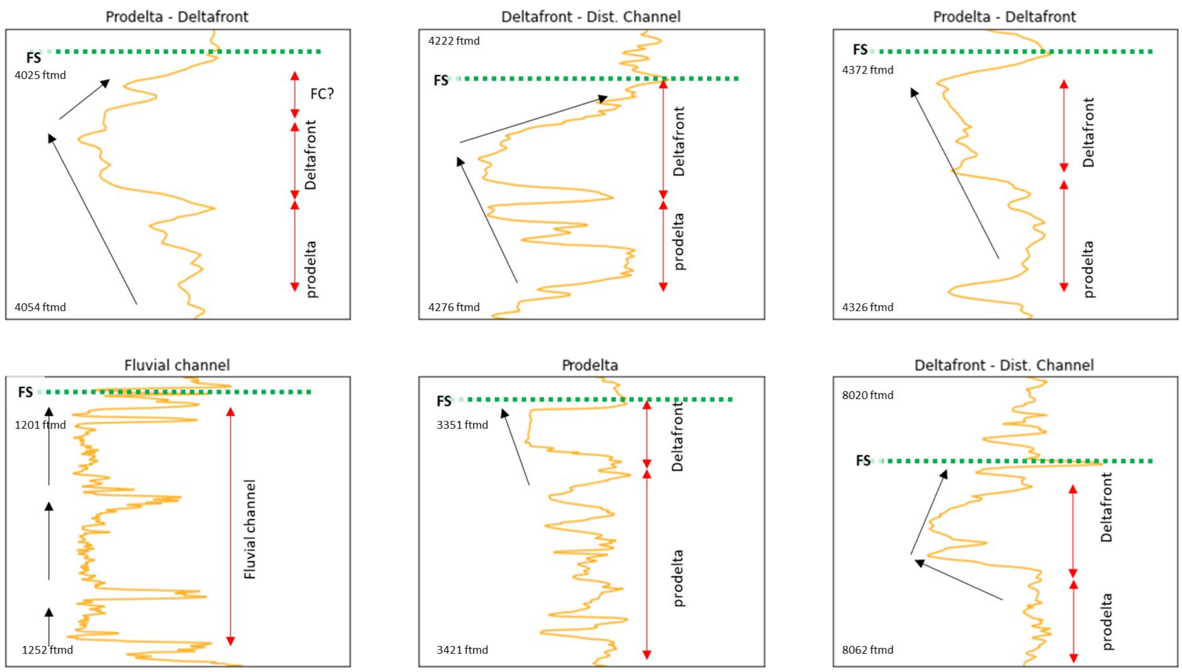

Figure 7 Characteristic as well as simplified depositional facies labeling using Gamma Ray log data in B-01 well.

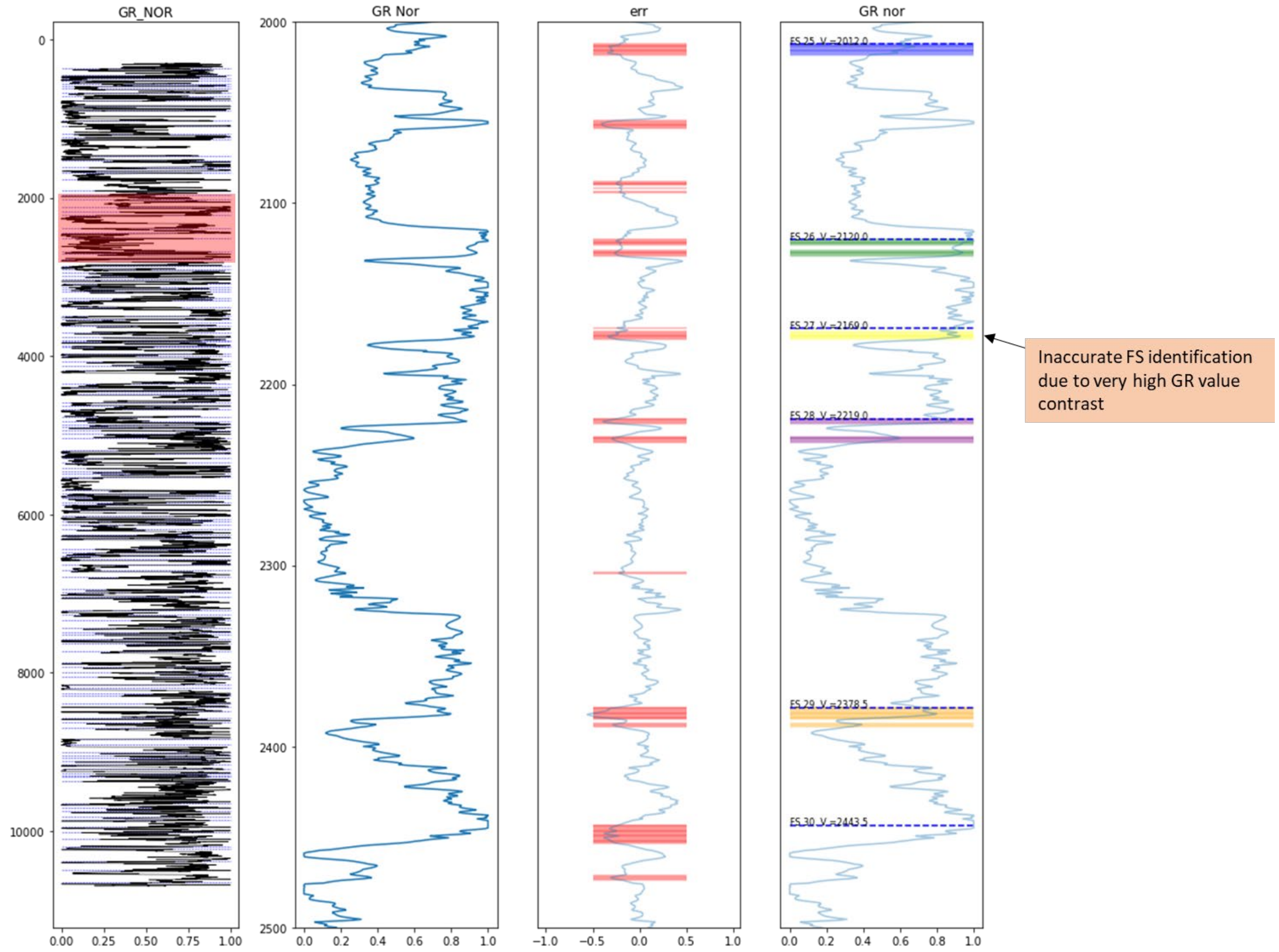

Figure 8 FS identification result on B-01 well interval 2000 – 2500 mMd

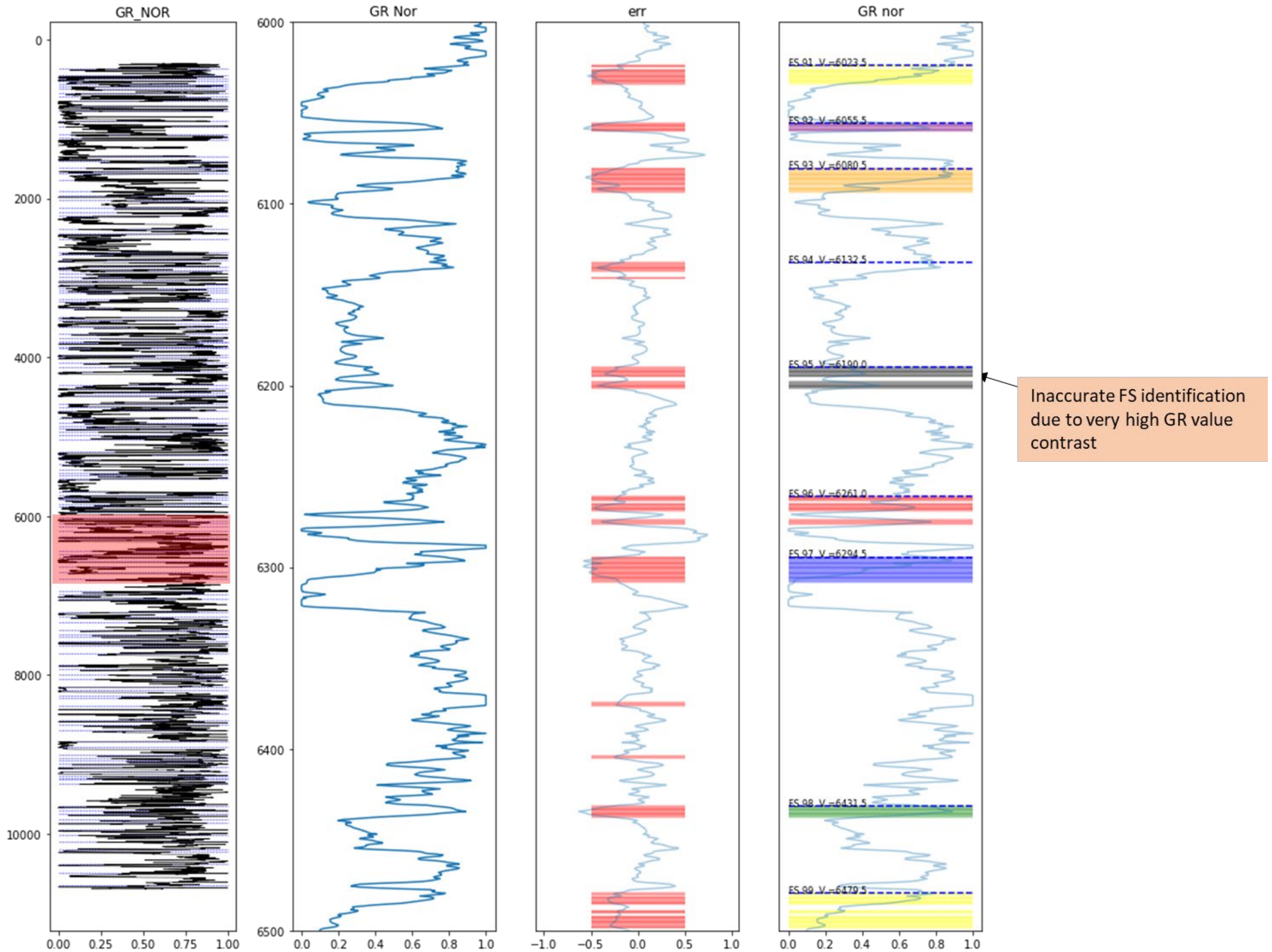

Figure 9 FS identification result on B-01 well interval 6000 – 6500 mMd.

DOI: 10.5614/bull.geol.2022.6.2.6

4. DISCUSSION

IMPLEMENTATION ON OUTCROP IMAGE

The workflow developed for FS identification in 1D data (synthetic and well log) may also be implemented in Outcrop Image. The key difference between 1D data compared to image data is that image data is composed of series of 1D data.

The image that being used in this experiment shows variation of lithology. Lithology can be easily identified through visual observation in the picture, the color in the picture usually represent lithology. In general understanding, brighter color may represent coarser grained sediment, and darker color may represent finer grained sediment. In data perspective, color will be represented by number in digital format, and thus it will be possible to extract the Vertical index (V) and Grain Size (Gs) from an image.

Digital image format store colors information in 16bit data format with minimum value of 0 and maximum value of 255 in four different color channel (RGBA). Although 4 channels are available, only three channels that will be used for this experiment (Red channel, Green channel, and Blue channel). Data squeezing is required for the algorithm to work.

Figure 10 statistical property of image that being used in this research after data squeezing.

Once data squeezing has been applied, FS identification workflow will be conducted using column by column basis (one column equals to one vertical succession extracted from the image). The amount of time may be different from one to another image, bigger image require a longer time to do the workflow compare to smaller one.

Figure 11 shows the result of FS identification using the proposed workflow with cutoff of -0.15 and linear regression window size of 40. **Figure 12** shows FS identification on single extracted Gs from image column number 200. Using one vertical succession, some obvious FS has been identified with good confident. **Figure 12** shows FS identification using all columns extracted from the image that marked by cyan colored cross markers. This experiment resulting in some markers that shows lateral continuity that shows presence of FS, and some scattered marker that occurred due to false FS prediction. Key factors that presumably affect the result are presence of shadow on the image and presence of object that are not geological object on the image. This experiment shows more manual works need to be done in order to refine FS identification on image basis. Work on balancing the shadow and masking non geological object are subjects for further research.

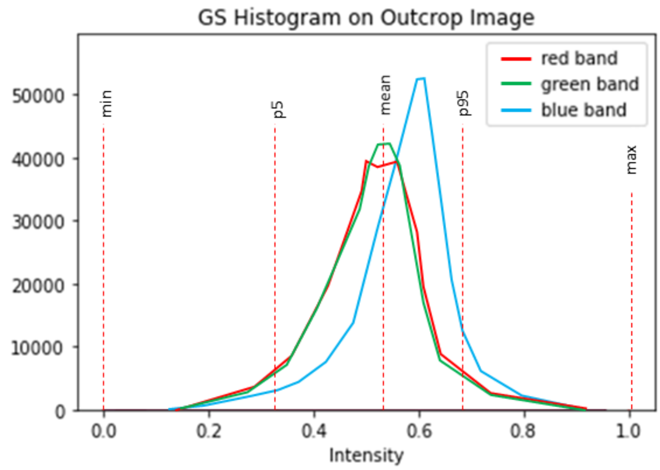

Figure 10 Statistical property of Image data after data squeezing

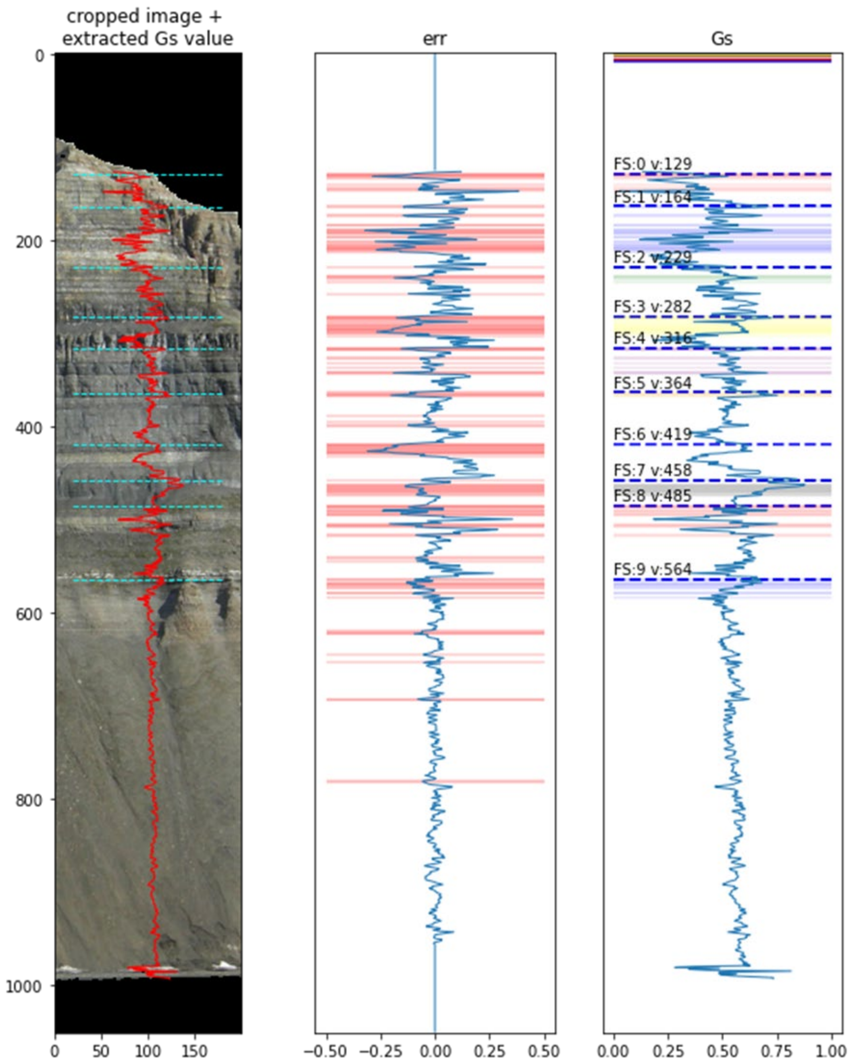

Figure 11 FS identification on single extracted Gs from image column number 200.

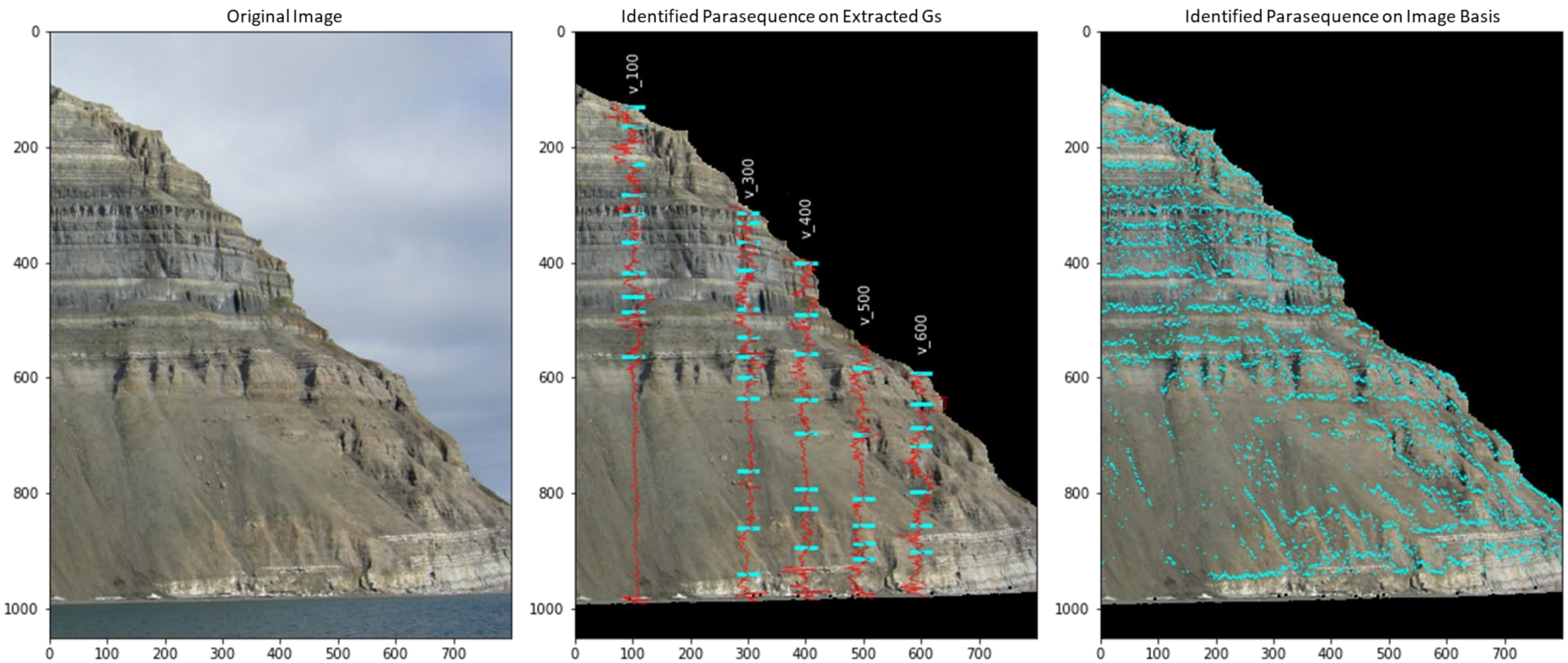

Figure 12 (Left) The original image. (Middle) FS identification using 4 extracted Gs on image data, where cyan represent FS identified with the workflow. (Right) Identified FS using all extracted Gs on image data.

5. CONCLUSION

According to the example provided in the discussion above, this method may provide geologist a quantitative solution on FS identification using well log and image data. This quantitative approach may reduce ambiguous interpretation due to geologist subjectivity as well as reduce interpretation time in identifying FS.

As shown on the experiment, the cutoff might be different from one data to another data. In order to standardize the cutoff, implementation using multi well basis works on window size and cutoff sensitivity is required. And on top of that, good database system to store well log data is also required in order to reduce performance lag occurred due to data query execution time.

FS prediction using image data on the other hand, may providing some good FS identification that marked by laterally continuous marker. This experiment also resulting in ambiguous result FS marker that occurred due to presence of shadow and non-geological object. Manual work especially in shadow balancing, and non-geological masking is required in order to improve this method.

ACKNOWLEDGEMENT

Authors would like to express gratitude to Pertamina Hulu Sanga-sanga (ex. VICO Indonesia) for providing well log data from B-01 well. Without the data, this experiment will not be able to be done and won't be published. Author would like to acknowledge Universitas Pertamina for giving hardware and financial support for this research.

REFERENCES

Catuneanu, O., Abreu, V., Bhattacharya, 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. S. C., Macurda, B., Martinsen, O. J., Miall, A. D., Neal, J. E., Nummedal, D., Winker, C. (2009). *Towards the standardization of sequence stratigraphy*. In Earth-Science Reviews (Vol. 92, Issues 1–2, pp.1–33).

Catuneanu, Octavian. (2006). *Principles of sequence stratigraphy*. Elsevier.

Catuneanu, Octavian. (2017). *Sequence Stratigraphy: Guidelines for a Standard Methodology*.

Gerald, B. F. (2020). Automated Well Log Correlation.

Gosses, J. (2020). Justin GOSSES / predictatops: v0.0.5 smallSfixed in directory name in configurationplugfiles.py.

Hall, B. (2016). *Facies classification using machine learning*. The Leading Edge, 35(10), 906–909.

Kusumah, E. P., Lestari, W., & Susanty, M. (2019). Wireline Log Data Conditioning to Improve Machine Learning Based Facies Identification Across the Well. *AAPG Asia Pacific Technical Symposium, The Art of Hydrocarbon Prediction: Managing Uncertainties.*

Li, X., & Guo, Y. (2013). *Application of Activity and Wavelet Analysis on Well Logging Sedimentary Cycle Division*. Open Journal of Geology, 03(02), 62–65.

Mastej, W. (2002). *An Application Of Markov Chain Analysis To Studies On Lithofacies Sequences In The Alluvial Fans From The "Bełchatów" Lignite Deposit (Poland)*. In Annales Societatis Geologorum Poloniae (Vol. 72, Issue 3).

Miall, A. D. (1973). *Markov chain analysis applied to an ancient alluvial plain succession.* Sedimentology, 20(3), 347–364.

Miall, A. D. (1996). *Principles of sequence stratigraphy. In Principles of sequence stratigraphy*.

Nio, S. D., Brouwer, J. H., Smith, D., de Jong, M., & Böhm, A. R. (2005).

Spectral trend attribute analysis: applications in the stratigraphic analysis of wireline logs. first break, 23(4).

Pratama, H., Syahputra, L. A., Fauzan Albany, M., Abdullah, A., Kusuma, E., Astuti, W., Mujihardi, B., & Suhardja, S. K. (2020). *Early Results of Comparison between K-Nearest Neighbor and Artificial Neural Network Method for Facies Estimation Hasil Awal Perbandingan antara K-Nearest Neighbor dan Artificial Neural Network Method untuk Facies Estimation.* In Jurnal Geofisika (Vol. 18, Issue 01).

Huffington, R. M., & Helmig, H. M. (1990). Badak Field--Indonesia--Kutai Basin/East Kalimantan (Borneo).

Schwarzacher, W. (1969). *The use of Markov chains in the study of sedimentary cycles.* Journal of the International Association for Mathematical Geology, 1(1), 17–39. https://doi.org/10.1007/BF02047069

Septama, E., Putra, C. M. E., Pambudi Suseno, A. M., Hasani, N., Putra, G. N., Iswanto, F., & Novantyono, S. (2019). *How did Channel Systems Survive the Transgressive Period? New Sedimentological Insights from Outcrop, Modern Analog and Subsurface Data in the Lower Kutai Basin*. Berita Sedimentologi.

Sinha, S., Das, S., & Sahoo, S. R. (2015). *Application of Markov chain and entropy function for cyclicity analysis of a lithostratigraphic sequence–a case history from the Kolhan Basin, Jharkhand, Eastern India.* Jharkhand, Eastern India Geology & Geophysics (Electronic Materials 4 2-7.

van Wagoner, J. C., Posamentier, H. W., Mitchum, R. M., Vail, P. R., Sarg, J. F., Loutit, T. S., & Hardenbol, J. (1988). An Overview Of The Fundamentals Of Sequence Stratigraphy And Key Definitions. In Sea-Level Changes (pp. 39–45). SEPM (Society for Sedimentary Geology).

Wilson, J. L. (1975). *Carbonate Facies in Geologic History*. In Carbonate Facies in Geologic History. Springer New York.

Wilson, J. L., & Wilson, J. L. (1975). P*rinciples of Carbonate Sedimentation*. In Carbonate Facies in Geologic History (pp. 1–19). Springer New York.

Ye, S. J., Wellner, R. W., & Dunn exxon mobil, P. A. (2017, March 6). Rapid and consistent identification of stratigraphic boundaries and stacking patterns in well-logs - An automated process utilizing wavelet transforms and beta distributions. *Proceedings - SPE Annual Technical Conference and Exhibition*.