

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/274271028>

Data Blocking or Zoning: Well-Log-Data Application

Article in Journal of Canadian Petroleum Technology · April 2013

DOI: 10.2118/154017-PA

CITATIONS

0

READS

378

1 author:



Nabil Adani

Kuwait Oil Company

2 PUBLICATIONS 1 CITATION

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Petrophysics Analysis Automation [View project](#)

Data Blocking or Zoning: Well-Log-Data Application

Nabil Al-Adani, Suncor Energy

Summary

Several statistical techniques have been introduced in zoning sequential data such as well-log data or blocking experimental data. In general, these techniques lack the ability to optimize the predicted zones as per the desired number of blocks or zones. In addition, all these techniques' results depend on the amount of data inclusion in the blocking or zonation process. In this paper, a new method has been introduced to address the blocks or zones optimization and data-amount dependency. The new method also helps in establishing a new approach in estimating the data random error, if not known. This paper does not compare all techniques with the proposed approach. One technique has been selected to highlight the advantages of the proposed method.

An example of a Canadian oil-sand well has been used to demonstrate some applications of the new blocking or zoning method.

Introduction

One of the typical approaches in grouping sequential data such as well logs is to assign a window of a certain width (number of sample data) and observe the sudden change in calculated window data averages while sliding the window over the sequential data (Souder and Pickett 1972). The drawback of such a simple method is the lack of control on window width. A larger window will lead to less boundary detection.

Analysis of data variances is a key in automation of pattern recognition. Two methods are compared by applying the different techniques of variance analysis. The objective is to model the random error in the data and zone or block the data on the basis of modeled random error.

This suggests a new method (Method 2) to zone or block data and hints at an approach for estimating random error.

Method 1

Analysis of variance (ANOVA) (Levine et al. 2000) can be used to examine data zonation or blocking. By comparing the variance within the zoned interval with the variance among all zones, the validity of zones can be tested (Lim et al. 1997). In other words, data are zoned initially into two zones. To check the validity of initial data blocking, ANOVA is performed to check the best possible zones break points. Then, each established zone is divided into two zones. This process is repeated until the desired number of zones is achieved or no further zones can be established.

The process of this method can be summarized into the following steps:

- Select a zone break point to divide into two new zones. Each zone should include at least two sample data.
- Calculate the following for the two new zones:
 - Mean variance within zones (MVWZ),

$$\text{MVWZ} = \frac{\sum_i^{n_1} (X_i - \bar{X}_1)^2 + \sum_i^{n_2} (X_i - \bar{X}_2)^2}{n_1 + n_2 - 2}, \quad (1)$$

where n_1 = number of samples in Zone 1, n_2 = number of samples in Zone 2, X = data sample, and \bar{X} = average.

- Mean variance among zones (MVAZ):

$$\text{MVAZ} = n_1 \left(\bar{X}_1 - \bar{\bar{X}} \right)^2 + n_2 \left(\bar{X}_2 - \bar{\bar{X}} \right)^2, \quad (2)$$

where $\bar{\bar{X}}$ = overall average.

- Compute the ratio F_c as follows:

$$F_c = \frac{\text{MVAZ}}{\text{MVWZ}}. \quad (3)$$

(c) Compare the variances using the F -test at a specified significance level $\alpha/2$ ($0 < \alpha < 1.0$). For a selected significance level (the most common value of 0.01) and based on the number of samples in each Zone n_1 and n_2 , the normally distributed F value can be obtained. If F_c is greater than F , then zones are distinct in their variations. If this condition is not met, then change zone division and repeat the above steps. This method does not guarantee that a certain grouping is the optimum zonation of an interval. The F test is valid in this method if the sample data are fully randomized and normally distributed.

The issues associated with this method are the predetermination of zones, unoptimized zones break point, and unknown optimum number of zones. The break point in which data are divided or blocked is an iterative guess until the ANOVA confirms the zonation. There is no criterion by which the optimum break point of a zone can be established. In addition, this method does not specify when to stop zoning the new zones into two further zones.

To address these concerns, the ratio R of variances can be introduced (Gill 1970; Agunwoke et al. 2004), as follows:

$$R = 1 - \frac{\text{MVWZ}}{\text{MVAZ}}. \quad (4)$$

The values greater than zero are considered. R is calculated after zoning a data set. All possible blocking of data is tried, with consideration of the allowable minimum number of samples in each group. In each trial, R is recorded. The blocking with highest positive R value is the criterion to confirm the break point. For every new zonation trial, the R value is investigated independently.

This criterion helps in automatic data blocking, which removes any prezoning requirement. The ratio R also removes the need of F -test analysis.

The process can be summarized as follows (Gill 1970):

$$(a) C_1 = \sum_j^z \sum_i^{n_j} X_{ij}^2, \quad (5)$$

where z is number of zones, n_j is number of data samples in each zone, j is a zone index, and X is the data sample value.

$$(b) C_2 = \left(\sum_j^z \sum_i^{n_j} X_{ij} \right)^2 / \sum_j^z n_j, \quad (6)$$

$$(c) C_3 = \sum_j^z \left(\sum_i^{n_j} X_{ij} \right)^2 / n_j. \quad (7)$$

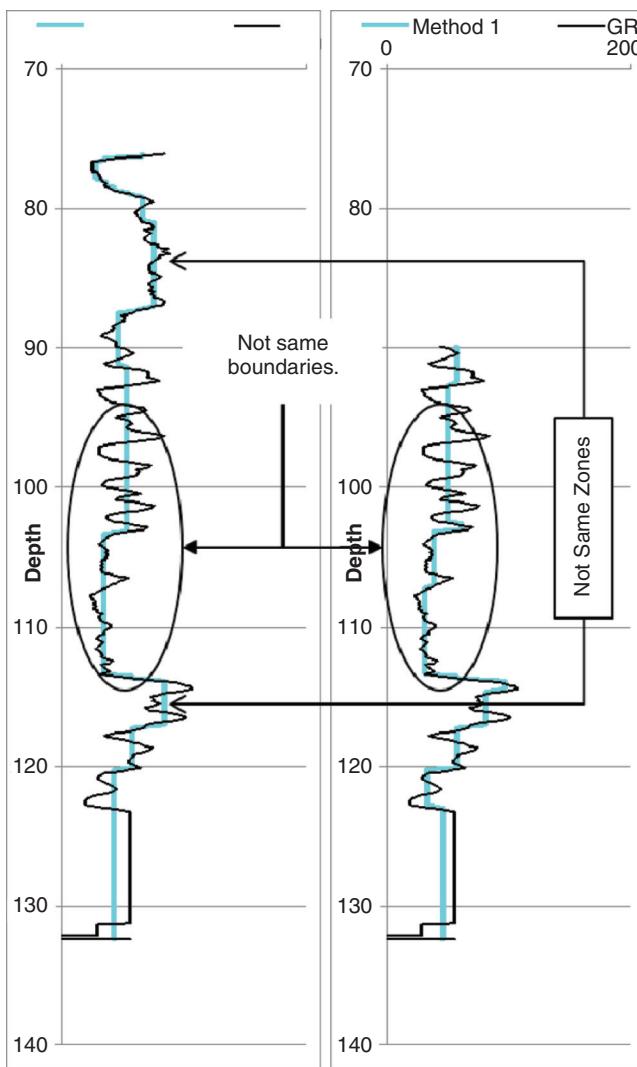


Fig. 1—Method 1: Different results are obtained based on the number of sample data included in the process. In addition, similar zones in sequential data are not identified.

(d) For each old zone j to be divided into two new zones, the following step is performed in each j zone:

- Select a zone break point k , which divides the old zone into two new zones. A minimum of two sample values should be maintained in each new zone.

$$\bullet C_4 = \left(\sum_{i=1}^k X_{ij} \right)^2 / k. \quad \dots \dots \dots \quad (8)$$

$$\bullet C_5 = \left(\sum_{i=k+1}^{n_j} X_{ij} \right)^2 / (n_j - k). \quad \dots \dots \dots \quad (9)$$

$$\bullet C_6 = \left(\sum_{i=1}^{n_j} X_{ij} \right)^2 / (n_j) \quad \dots \dots \dots \quad (10)$$

$$\bullet C_7 = C_3 - C_6 + C_5 + C_4 \quad \dots \dots \dots \quad (11)$$

$$\bullet R = 1 - \left[\frac{(C_1 - C_7)(z-1)}{(C_7 - C_2) \left(\sum_j^z n_j - z \right)} \right], \quad \dots \dots \dots \quad (12)$$

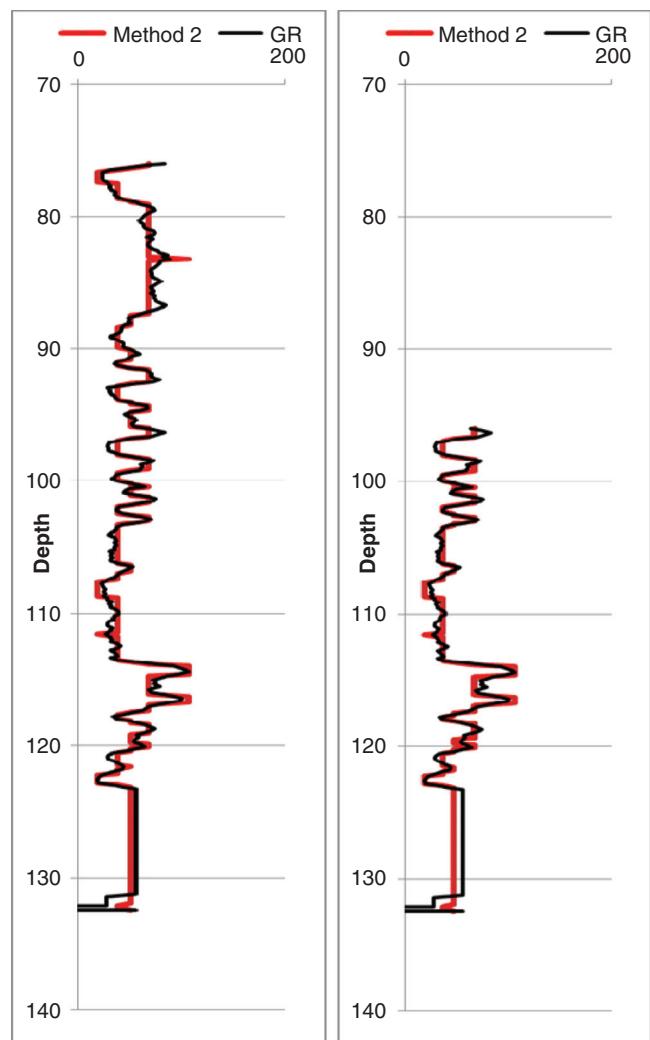


Fig. 2—Method 2: Boundaries detection is independent of amount of data inclusion in the process. In addition, similar zones are identified.

where R is constrained between zero and one and is computed on the basis of C values.

- For every k break point, R is computed. The breakpoint of highest R is recorded as R_x . If the highest R is greater than R_x then the new R_x equals the observed highest R .

(e) Performing Step d in all old zones is considered a single trial. At the beginning of each trial, R_x is assumed to be zero.

(f) Steps d and e are repeated in each trial until no further new break points can be evaluated.

The obvious advantage of this method is full automation of data zonation or blocking based on data variances. This approach is the most common automated technique in petroleum analysis (Lawal and Onyekonwu 2005).

However, in this method the maximum number of zones is controlled by the overall data and their variances. Such dependency does not help in zoning sequential data such as well-log data. Different results are produced depending on the number of samples or interval length. In addition, this method does not provide a means to check the similarities of estimated zones. In Fig. 1, applying the preceding method on the same gamma ray but using different data intervals does not produce unique zones.

Also in this approach, the number of zones or blocks is optimized only if no more break points can be detected. This means that the number of zones cannot be specified earlier. This might be an advantage if the objective is to check the optimum number of zones. One application of the optimum is to determine the random error in a sequential measurement, such as well-log data.

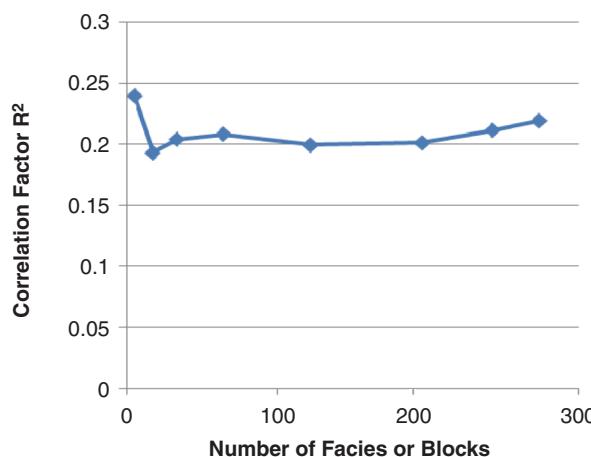


Fig. 3—Optimum number of blocks for gamma ray is 5.

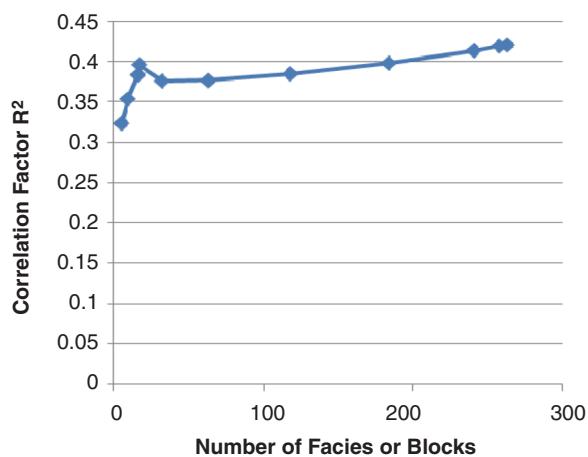


Fig. 4—Optimum number of blocks for bulk density is 17. Number of blocks should be as few as possible.

Method 2

An alternative method is a data-binning approach. In the first step, data are binned into two bins defined by the minimum (\bar{X}_1), arithmetic average (\bar{X}_2), and maximum (\bar{X}_3). The maximum and minimum are considered average values of zones, which consist of one sample in each zone. This provides pseudoupper and -lower zones for the process. Each bin or block is divided further into two blocks by evaluating an arithmetic average of each block. To ensure the validity of such a break point, the average value is compared with the upper and lower block average values to check population (block average) similarities. If the new blocks' averages are not similar, then division is valid. This process is repeated for each of the newly established blocks until no more division can be validated.

To check the validity of a new division, the following student *t*-test is recommended at a selected significance level:

(a) Divide a block into two new blocks.

(b) Estimate the variance of one of the new blocks S_1^2 and an upper or lower block variance S_2^2 , as follows:

$$S_1^2 = \frac{1}{n_1-1} \left[\sum_{i=1}^{n_1} X_i^2 - \frac{\left(\sum_{i=1}^{n_1} X_i \right)^2}{n_1} \right] n_1 > 1, \quad (13)$$

where S_1^2 = variance of the new block average and n_1 = number of data used in the new average estimation (\bar{X}_1).

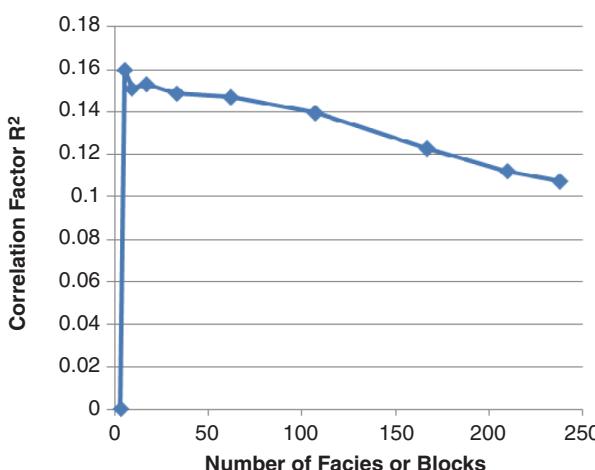


Fig. 5—Optimum number of blocks for neutron porosity is 5.

$$S_2^2 = \frac{1}{n_2-1} \left[\sum_{i=1}^{n_2} X_i^2 - \frac{\left(\sum_{i=1}^{n_2} X_i \right)^2}{n_2} \right] n_2 > 1, \quad (14)$$

where S_2^2 = variance of the selected upper or lower boundary and n_2 = number of data used in the upper or lower average estimation (\bar{X}_2).

(c) Evaluate t_c value for two independent samples *t*-test, as follows:

$$t_c = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (15)$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}. \quad (16)$$

(d) Check the t_c value using the student *t*-test to ensure the validity of the division.

The main advantage of this method over the previous one is the consistency of determined boundaries regardless of selected intervals or amount of sample data included in the process. In addition, because of the nature of this method, the similar zones in sequential data such as well-log data are identical and unique. Fig. 2 shows gamma ray zonation or blocking. The same block or zone boundaries are obtained independently of the samples number used in the process.

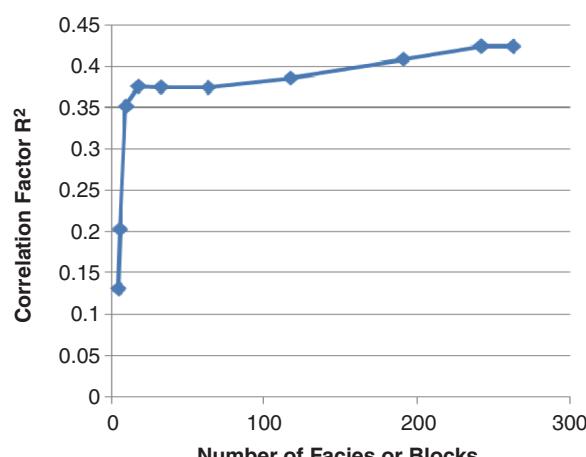


Fig. 6—Optimum number of blocks for photoelectric is 17. Number of blocks should be as few as possible.

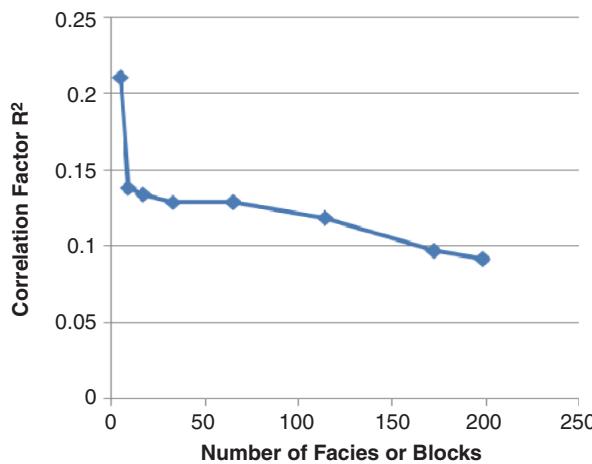


Fig. 7—Optimum number of blocks for compressional slowness is 5.

The advantage of this method is that in each level of division of old blocks into new blocks, the zones' boundaries are optimized. This means that a certain number of zones or divisions can be attempted but the process optimizes the number of zones or blocks through the student *t*-test at a selected significance level (usually

0.01). In other words, with increasing attempts to divide further, an optimum unique number of zones is achieved.

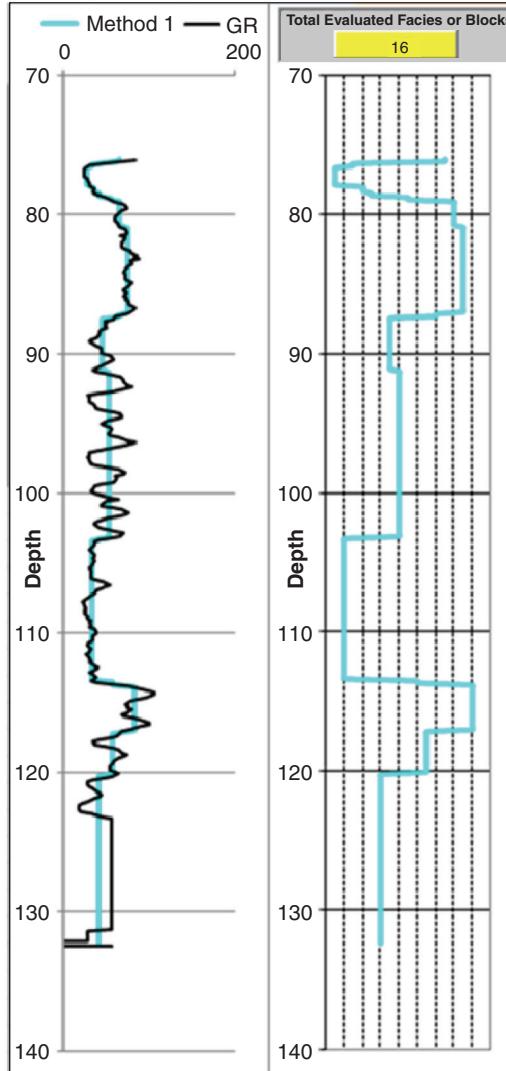
The maximum possible number of blocks that can be achieved in iteration is 2^n , where n is an iteration index. The final estimated number of blocks is always equal to or less than the maximum number of iterations or attempts.

Method 2—Blocks Optimization and Random-Error Estimation. In Method 2, the number of blocks is initially assumed. This number is further minimized through variance-analysis optimization at a selected significance level in the student *t*-test. One of the applications of the measured data samples zones or blocks estimation is to evaluate existing zones that have been measured. In sequential data (such as well data), the log-data estimated zones may be related to the actual geological facies.

Based on the purpose of data zonation, more criteria can be added to optimize the number of zones or blocks. The determination coefficient r^2 can be used to compare actual geological facies and estimated zones or blocks from log data. If the estimated number of zones or blocks is higher than the desired blocks, Method 2 is attempted with a lower initial guess of zones.

$$r^2 = \frac{b_0 \sum Y_i + b_1 \sum X_{ii} Y_i + \dots - \frac{(\sum Y_i)^2}{N}}{\sum Y_i^2 - \frac{(\sum Y_i)^2}{N}}, \quad \dots \dots \dots \quad (17)$$

Method 1



Method 2

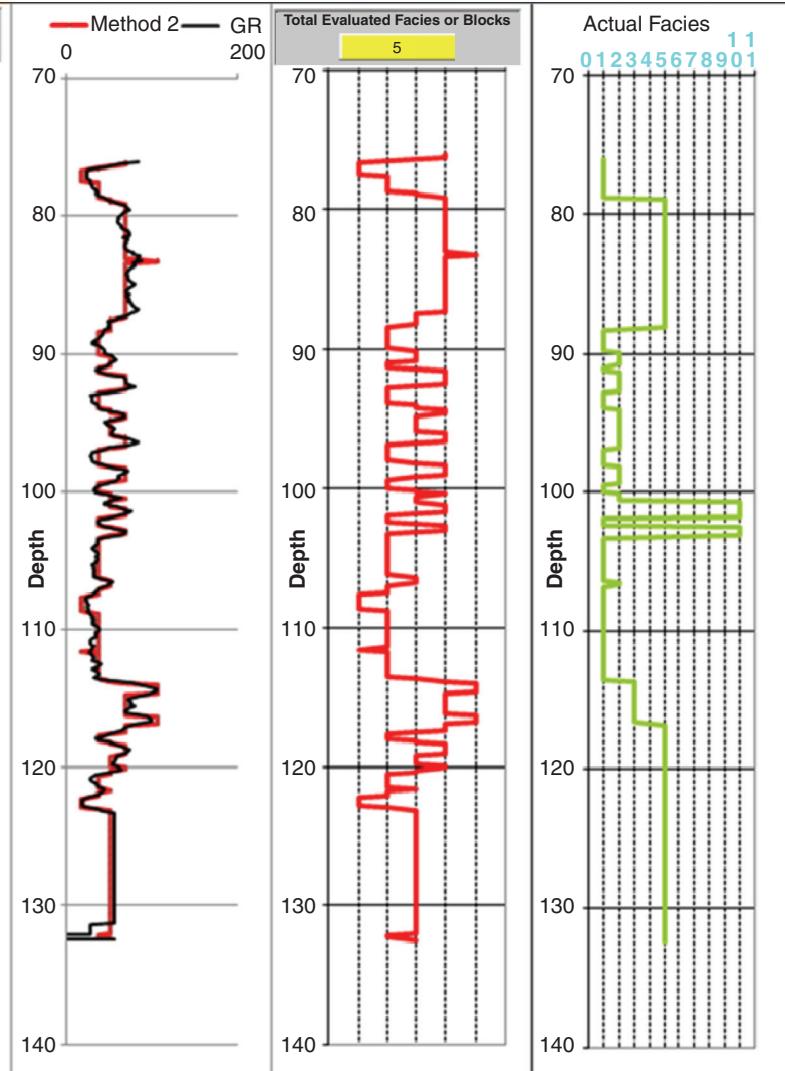


Fig. 8—Autoblocking of gamma ray using Methods 1 and 2. An optimum number of facies (16) was achieved using Method 1, with determination coefficient r^2 of 18.52% and random error of 30.93%, and 5 facies were achieved using Method 2, with determination coefficient r^2 of 23.97% and random error of 11.02%.

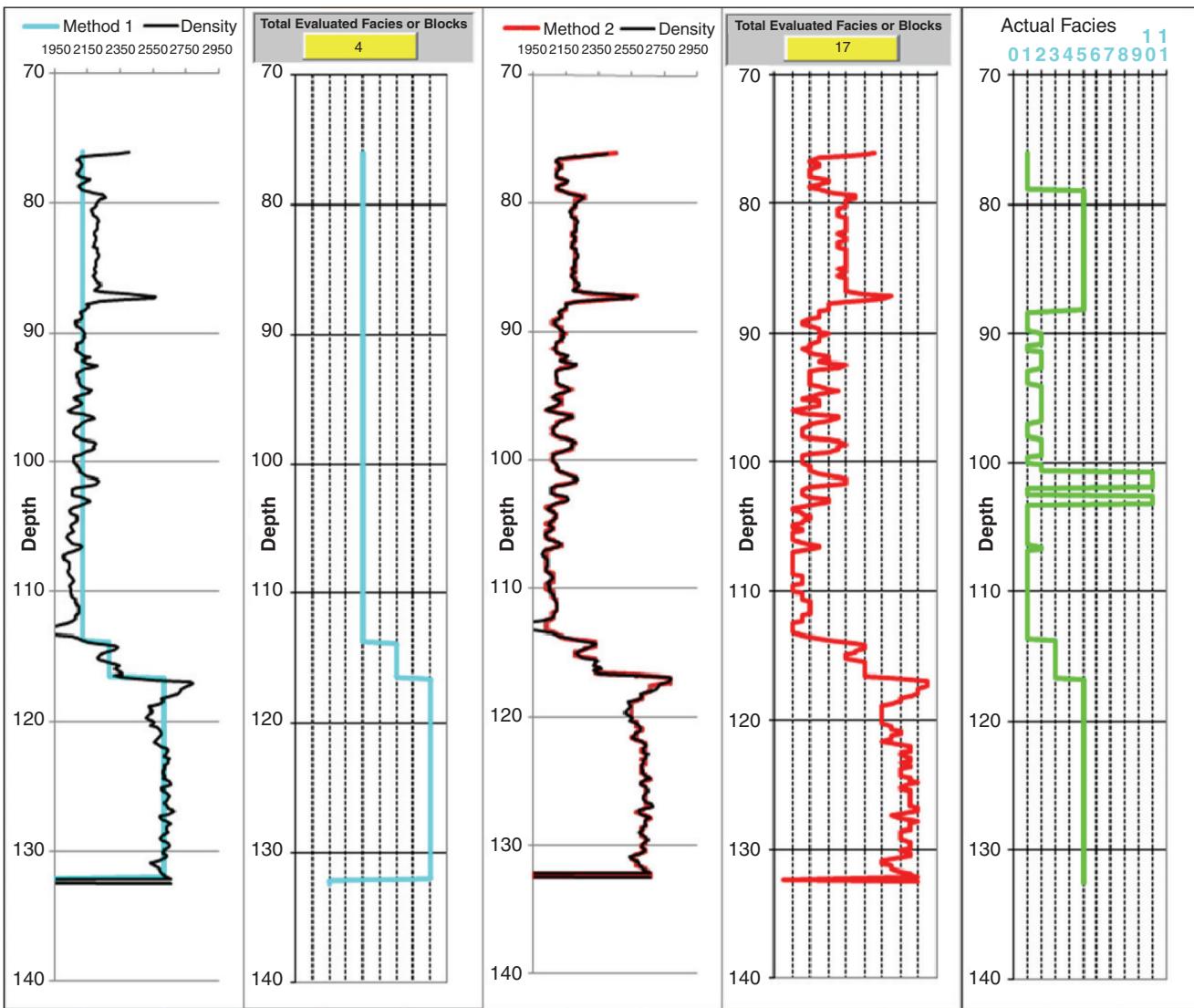


Fig. 9—Autoblocking of bulk density using Methods 1 and 2. An optimum number of facies (4) was achieved using Method 1, with determination coefficient r^2 of 19.43% and random error of 17.9%, and 17 facies were achieved using Method 2, with determination coefficient r^2 of 39.59 % and random error of 1.05%.

where $b_0, b_1 \dots b_n$ are the linear regression coefficients, X_i is the estimated blocks, N is the total number of samples and Y is the geological-facies values.

Figs. 3 through 7 show the block-optimization process based on the estimated determination coefficient (r^2) applied on depth sequential data (gamma ray, neutron porosity, bulk density, photoelectric factor and compressional slowness). Bulk-density and photoelectric-log data are more in line with geological facies. Increasing the number of blocks improves the number of log samples explained by actual facies until the optimum number of blocks is achieved. It is essential to minimize the number of estimated blocks to match the number of facies determined by geologists.

To estimate the random error in any measurement, actual values in comparison with some standard reference values should be performed. However, in most cases, reference values are not available and source of measurements could also be from several sources. This complicates the issue and makes it difficult to establish any reliable understanding of the random error in measured data. Usually repeat measurements are performed through different sources of data to infer random error, assuming that the random error can be established by means of test repeatability.

An alternative approach is to investigate what is the reference value. Because this is the value under investigation through a measurement, then the actual value can be assumed as the average of the measured values in the estimated zone. The concept is that

by estimating the optimum zones or blocks seen by a measurement, the average of the measured value in a zone may represent the reference value. By comparing the measured values with the estimated zones or blocks averages, the standard error is estimated. The standard-error value for each zone may represent the random error of the measured values within zones or blocks.

$$\text{Random Error} = \text{Standard Error} = \sqrt{\frac{\sum(Y_i - aX_i - b)^2}{n-2}}, \quad (18)$$

where n = number of data samples, $a = \frac{n\sum Y_i X_i - \sum Y_i \sum X_i}{n\sum X_i^2 - \sum X_i \sum X_i}$, $b = \frac{\sum Y_i - a\sum X_i}{n}$, X_i = log data, and Y_i = blocks or zones averages.

Applications

Method 2 is applicable in automatically blocking data on the basis of their variances. Method 2 can be applied in generating histograms, estimating random errors in measured data, and studying the expected experimental factors through blocking results. In this

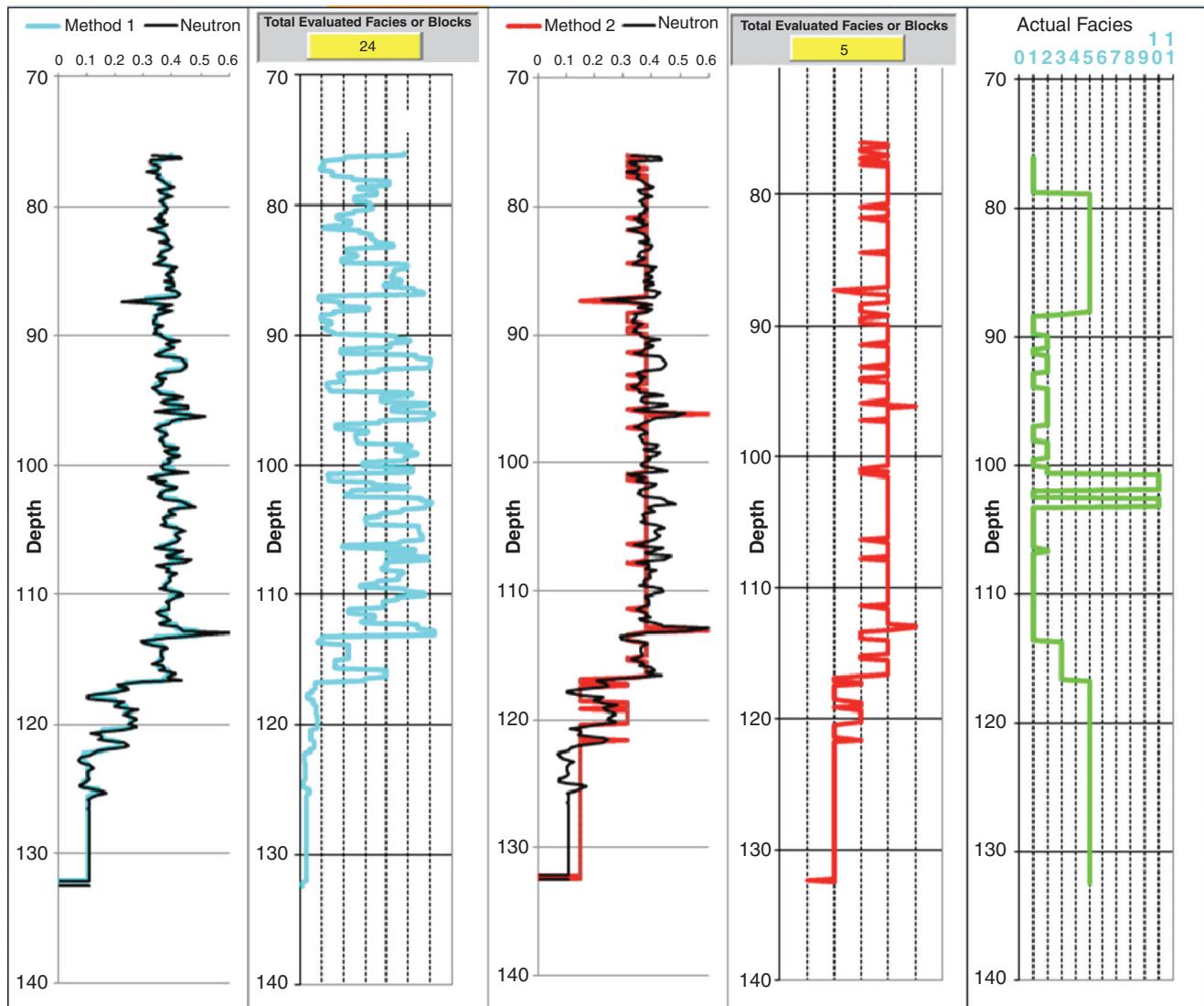


Fig. 10—Autoblocking of neutron porosity using both Methods 1 and 2. An optimum number of 124 facies was achieved using Method 1 with determination coefficient r^2 of 11.98% and random error of 33.48%, and 5 facies were achieved using Method 2, with determination coefficient r^2 of 18.32% and random error of 8.4%.

paper, application on sequential data in the depth domain is studied. Specifically, Method 2 is applied on well-log data.

Case Study. In this case, the number of required log measurements to predict geologically described facies was studied. An example of a Canadian oil-sand well (1AA080409312W400) in Mackay River field (MacMurray formation) is selected to study the minimum number of logs required to reproduce geologist-described facies. Common techniques used in industry are cluster analysis, neural network, probability plots, multivariate regression, fuzzy logic, multilinear graphical clustering, and nonlinear optimization. This paper does not discuss these techniques (Lee and Datta-Gupta 1999; D'Windt 2007).

The geological facies in the example well are driven on the basis of observed rhythmic bedding, composed of inter laminated to thinly interbedded silty and sandy mud. This represents mud-dominated inclined heterolithic stratification. In other words, the described geological facies are represented to some degree by clay content and grain size. Because it is a visual description of core, the facies are biased by the core texture and color appearance. In the subject well, geologists described five facies, and facies number increases as the visual clay content increases. Facies 10 is used to identify breccia, which does not follow the same logic as the rest of the facies numbers.

On the basis of this information, gamma ray, bulk-density, neutron-porosity, compressional-slowness, and photoelectric log

measurements were selected in the study because these logs are more dependent on rock mineral composition in the case of Canadian Fort Hills oil sands.

Methods 1 and 2 of blocking were applied on all log data. Figs. 8 through 12 shows the results of blocking using both methods with determination coefficient, random error, and number of facies. Method 2 clearly results in the minimum number of facies or blocks, with better determination coefficient and minimum random error.

By comparing the blocking on all logs, bulk density and photoelectric have produced the best correlation with geological facies as individual data. However, an integrated approach should be considered to address the minimum number of logs needed for the best representation of the geological facies.

To test the probability of combination, a linear combination of the form was assumed:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots , \quad \dots \dots \dots \quad (19)$$

where $X_1, X_2, X_3 \dots$ are the estimated facies or blocks from selected log data and Y is the actual geological facies.

On the basis of the preceding fit-determination coefficient, the best combination is determined. In this case, a combination of gamma ray, bulk density, and photoelectric has resulted in a determination coefficient of 49.29%.

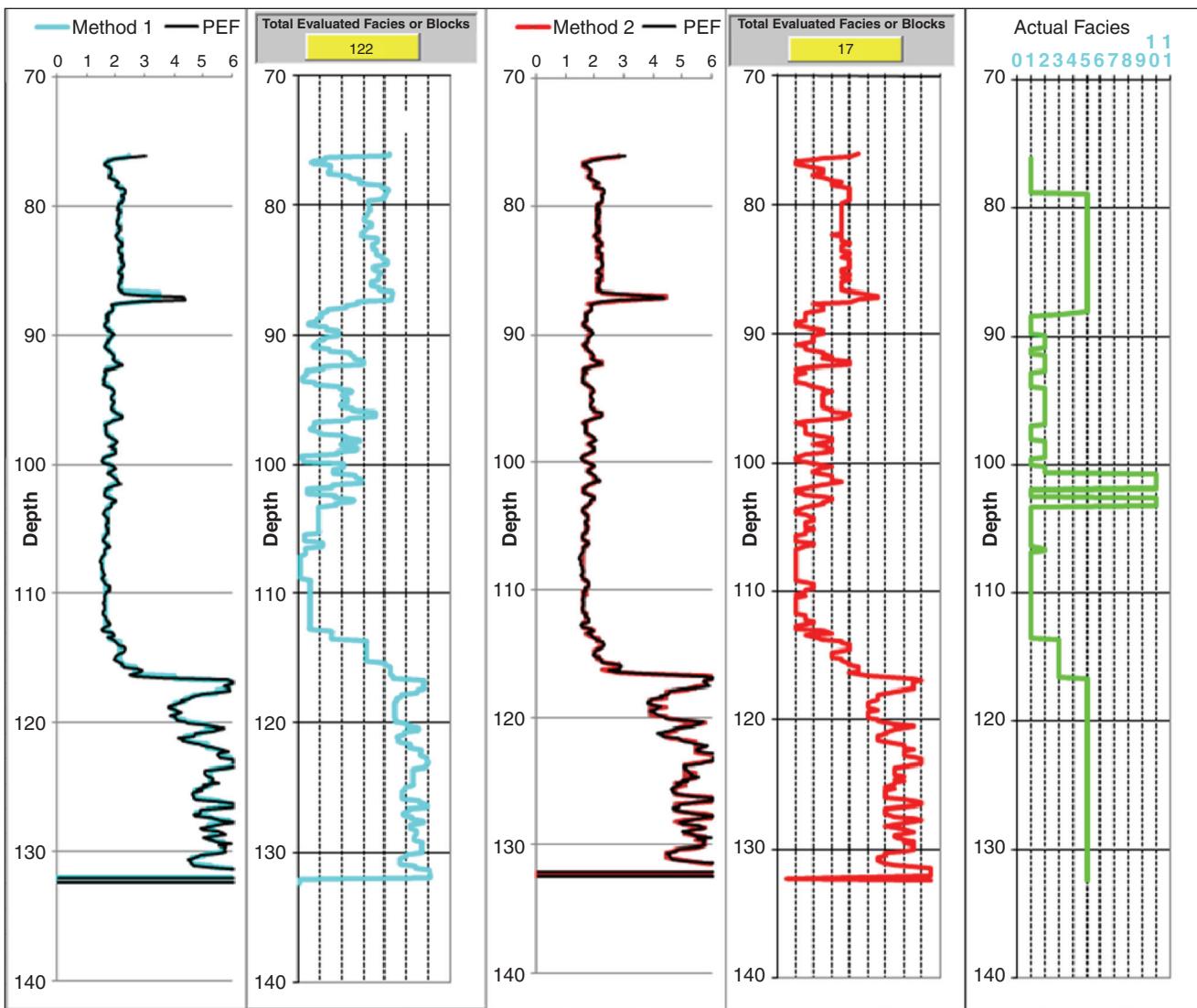


Fig. 11—Autoblocking of photoelectric factor (PEF) using Methods 1 and 2. An optimum number of 122 facies was achieved using Method 1, with determination coefficient r^2 of 42.04% and random error of 31.84%, and 17 facies were achieved using Method 2, with determination coefficient r^2 of 38.53% and random error of 0.2%. This shows that PEF alone cannot be correlated to the geologists criteria of the facies.

From this test, it is concluded that the best representation of geologically interpreted facies by log is achieved with gamma ray, bulk density, and photoelectric. In addition, these logs do not predict all geological facies. The log prediction fails for Facies 10 (**Figs. 13 and 14**). This is in line with what gamma ray, bulk-density, and photoelectric measurements are for. These measurements do correlate with clay content. However, breccia does not have a unique clay signature.

Conclusion

We suggest that Method 2 is better than Method 1 in controlling the number of desired blocks or zones, and it also performs the block or zone similarity test. These two features of Method 2 help in sequential-data application, such as well-log data. In addition, Method 2 can be applied further to investigate the random error, if it is not known.

References

- Agunwoke, G.O., Egbele, E., and Onyekonwu, M. 2004. A Statistical Approach to Reservoir Zonation. Paper SPE 88962 presented at the Nigeria Annual International Conference and Exhibition, Abuja, Nigeria, 2–4 August. <http://dx.doi.org/10.2118/88962-MS>.
- D'Windt, A. 2007. Reservoir Zonation And Permeability Estimation: A Bayesian Approach. Paper 2007_UUU presented at the SPWLA Annual Logging Symposium, Austin, Texas, USA, 3–6 June.
- Gill, D. 1970. Application of a Statistical Zonation Method to Reservoir Evaluation and Digitized Log Analysis. *AAPG Bull.* **54** (5): 719–729. <http://dx.doi.org/10.1306/5D25CA35-16C1-11D7-8645000102C1865D>.
- Lawal, K.A. and Onyekonwu, M.O. 2005. A Robust Approach to Flow Unit Zonation. Paper SPE 98830 presented at the Nigeria Annual International Conference and Exhibition, Abuja, Nigeria, 1–3 August. <http://dx.doi.org/10.2118/98830-MS>.
- Lee, S.H. and Datta-Gupta, A. 1999. Electrofacies Characterization and Permeability Predictions in Carbonate Reservoirs: Role of Multivariate Analysis and Nonparametric Regression. Paper SPE 56658 presented at the SPE Annual Technical Conference and Exhibition, Houston, 3–6 October. <http://dx.doi.org/10.2118/56658-MS>.
- Levine, D.M., Ramsey, P.P., and Smidt, R.K. 2000. *Applied Statistics for Engineers and Scientists: Using Microsoft Excel & Minitab*. Upper Saddle River, New Jersey, USA: Prentice Hall.
- Lim, J.-S., Kang, J.M., and Kim, J. 1997. Multivariate Statistical Analysis for Automatic Electrofacies Determination from Well Log Measurements. Paper SPE 38028 presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Kuala Lumpur, 14–16 April. <http://dx.doi.org/10.2118/38028-MS>.
- Souder, W.W. and Pickett, G.R. 1972. A Computerized Method for the Zonation of Digitized Well Logs. Paper SPE 4019 presented at the Fall Meeting of the Society of Petroleum Engineers of AIME, San Antonio, Texas, USA, 8–11 October. <http://dx.doi.org/10.2118/4019-MS>.

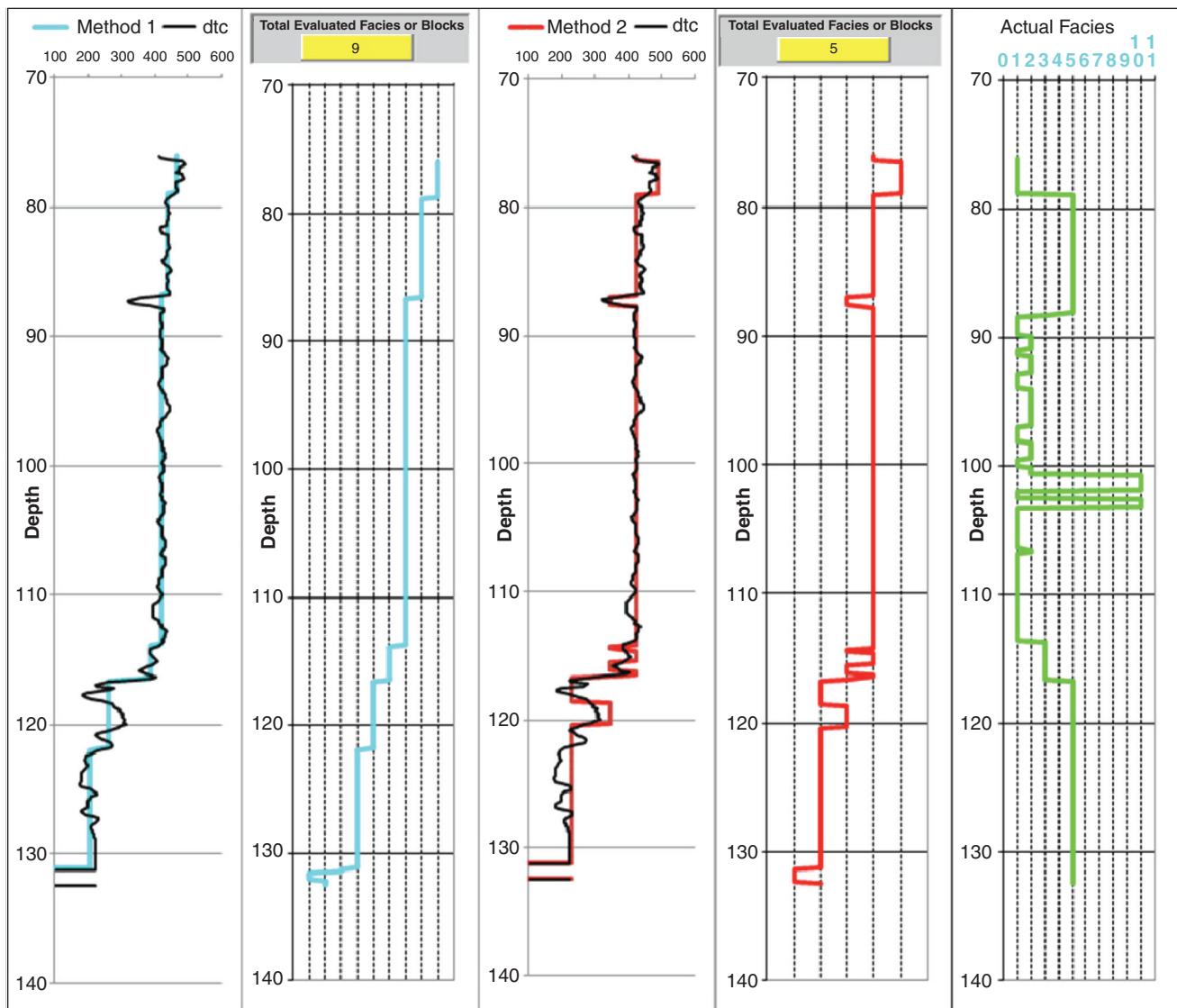


Fig. 12—Autoblocking of compressional slowness using Methods 1 and 2. An optimum number of 9 facies was achieved using Method 1, with determination coefficient r^2 of 14.3% and random error of 11.79%, and 5 facies were achieved using Method 2, with determination coefficient r^2 of 23.64% and random error of 4.32%.

Nabil Al-Adani holds a BSc degree in petroleum engineering from King Fahd University of Petroleum & Mining (1995) and a masters degree in engineering from the University of Calgary (2010). For the past 15 years, he has worked in the Middle East and North America as a senior petrophysicist, with a focus on porosity, permeability, and saturation analyses. Al-Adani

invented a Stoneley permeability evaluation method for fractured reservoirs and developed several interpretation techniques. He currently works for Suncor Energy in Canada where he helps in developing porous-media engineering solutions for bitumen recovery in sandstone and carbonate.

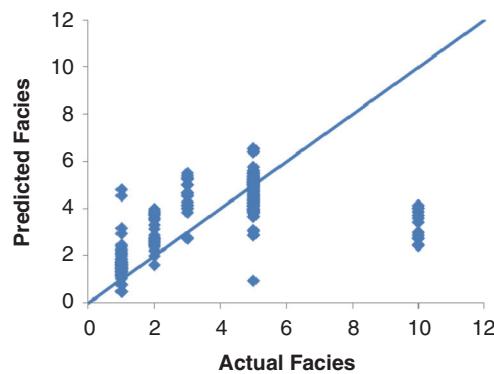


Fig. 13—Combined estimated and actual facies or blocks.

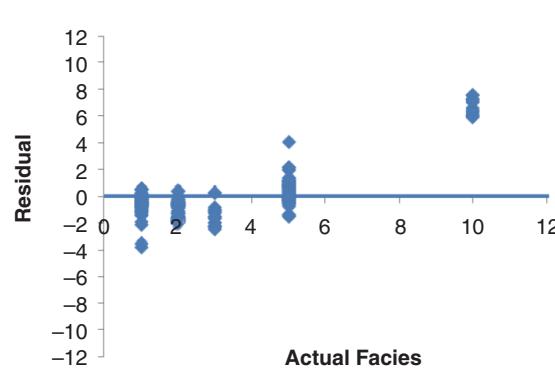


Fig. 14—Residual plot of combined estimation.