Principal component analysis-based statistical well log analysis: A case study from Upper Assam Basin

Siddharth Garia¹, Arnab Kumar Pal², Ravi Karangat³, and Archana M. Nair⁴*

Abstract

Statistical analysis can be used to complement rock physics and well-log-based interpretations. By doing so, a certain amount of uncertainty and bias can be reduced along with the enhancement in rock physics models used for reservoir characterisation. The present study uses a statistical approach to classify the trends between hydrocarbon-bearing and non-hydrocarbon-bearing zones from uncorrected or raw well log data. Such a methodology was adopted to restrict any bias or uncertainty, knowingly or unknowingly, in the datasets obtained from well logs. For the analysis, datasets of three wells from Upper Assam Basin, India were investigated. Cross-plots of different attributes followed by principal component analysis (PCA) were adopted to recognise any conclusive trends. The cross plot between neutron porosity (NPHI) and sonic log (DT) colour-coded with the volume of shale ($V_{\text{shale}}$) values facilitated the identification of lithology for different wells. The nature of trends was investigated and classified based on $V_p$, while PCA indicated that a trend existed for hydrocarbon bearing zone compared with non-hydrocarbon bearing zone datasets. Hence, it can be concluded that PCA is a valuable technique in assessing and deriving meaningful petrophysical interpretations for reservoir characterisation, thereby complementing the results obtained through conventional well log analysis.

Keywords: statistical, well-log, hydrocarbon, principal component analysis, reservoir

Introduction

Petrophysical and geomechanical parameters of a reservoir are generally estimated from seismic, well log, or core data (Mavko et al., 2009). However, uncertainty and error analysis in the quantitative estimation of parameters, especially in seismic and well log data, is generally not well-quantified (Hussain et al., 2017). Several usually ignored uncertainties and unknown biases may affect the rock physics models (Chen and Dickens, 2009). Ignoring such uncertainty can lead to distorted interpretations and may significantly impact the best guess in upstream reservoir engineering (Mukerji and Mavko, 2006). Additionally, understanding the subsurface heterogeneity and the associated uncertainty becomes critical for reliable reservoir characterisation. The heterogeneity, which may be due to several factors, affects the reservoir performance and complicates the understanding of the elastic properties of rocks (Garia et al., 2019). As a result, statistical tools are often employed in rock physics to reduce the associated error and bias in predicting different petrophysical parameters from elastic properties (Garia et al., 2021; 2022). Statistical and physics-based approaches are generally adopted to identify the relationships between different reservoir parameters and well log data (Bhardwaj and Sharma, 2020; Pal et al., 2022; Katre and Nair, 2022).

1 Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India. Email: sidd_41@iitg.ac.in
2 Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India. Email: arnab.pal@iitg.ac.in
3 Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India. Email: ravi.civil@iitg.ac.in
4 Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India. Email: nair.archana@iitg.ac.in
One of the oldest and most well-known techniques for multivariate statistical analysis is principal component analysis (PCA) (Avseth et al., 2010). PCA is a data description technique with the significant advantage that no a priori information is required about the location, lithology, saturation etc. (Prasad et al., 2005). PCA technique is used to reduce the dimensionality of the data without any effect on the existing variation in the dataset (Jolliffe and Cadima, 2016). The data are transformed into a new set of uncorrelated variables, the principal components (PCs), which are linear combinations of the old variables. The principal components (generally first three PC1, PC2 and PC3) retain most of the variation present in the original dataset (around 80% of the data) (Prasad et al., 2005). As a result, instead of analysing the entire dataset, the principal components can be analysed to identify meaningful correlations. Moreover, PCA analysis can be used to interpret possible relationships/dependencies existing among the original dataset (Niculescu and Andrei, 2016). PCA can be successfully employed for a variety of applications, such as in well log analysis for well-to-well correlation by pattern recognition, characterisation of low permeability intervals, aquifer formations, the distinction between hydraulic flow units, etc. (Lim et al., 1998; Morin 2006; Niculescu and Andrei, 2016).

The present study attempts to employ PCA to the well log data of three wells from the Upper Assam Basin, India. Well log data used for the analysis includes caliper, gamma ray, density, and sonic log. We have investigated the potential usefulness of PCA in assessing and deriving meaningful petrophysical interpretations, in addition to the results obtained through conventional well log analysis. The present study observed that the PCA technique can be successfully used to understand different relationships among various parameters such as distinguishing hydrocarbon and non-hydrocarbon bearing zone, thereby aiding in improving rock physics models for reservoir characterisation.

**Study Area and Applied Methodology**

Upper Assam basin is considered as one of the prolific hydrocarbon-producing regions in India. The present study is based on the evaluation of well log data of three wells from the Upper Assam basin. As per the available well report, well 1 was drilled as a vertical well drilled down to a depth of 2547 m and bottomed out in the granitic basement. Wells 2 and 3 were drilled up to depths of 2614 and 2715 m, respectively. The well penetrated through Post Tipam, Tipam, Bokabil, Barail, Kopili, Sylhet, Tura, Trap, Gondwana and basal sandstone sequences and finally bottomed out in Pre-Cambrian granitic basement. Figure 1 shows the location of the three wells within the study area (Golaghat district, Assam). Table 1 presents the well bore status, measured depth, and the log data used for each well. The responses of all well logs used in the present study for each well is illustrated in Figure 2. The resistivity logs were not available for well 2 and were not considered for the present analysis.

As mentioned earlier, PCA was adopted to identify different patterns based on raw well data without incorporating any unknown bias, uncertainty, and any prior information regarding lithology and saturation. The PCA was performed by constructing a covariance matrix and calculating the eigenvectors. The eigenvector having the largest eigenvalue was used as first PC, as it minimises the distance between the data points and increases the variance between the two parameters (Shen et al., 2019). For a detailed methodology about the
basics involved in PCA as an exploratory tool for data analysis, the readers may refer to a review by Jolliffe and Cadima, (2016). It is to be noted that raw well log data were used for the present analysis, and no corrections were applied. PCA analysis on raw data was carried out to preserve the dataset's true nature and avoid introducing any bias knowingly, or unknowingly.

![Figure 1: Location of the three wells within the study area, the Golaghat district of Assam, a northeastern state in India, as indicated in the inset.](image)

**Table 1: Wells analysed in the present study**

<table>
<thead>
<tr>
<th>Well Name</th>
<th>Well type</th>
<th>Well Bore status</th>
<th>Total MD (m)</th>
<th>Curves used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well 1</td>
<td>Exploratory</td>
<td>Hydrocarbon Producing</td>
<td>2547</td>
<td>Gamma Ray, Resistivity, Sonic, Density, Neutron</td>
</tr>
<tr>
<td>Well 2</td>
<td>Exploratory</td>
<td>Unknown</td>
<td>2614</td>
<td>Gamma Ray, Sonic, Density, Neutron</td>
</tr>
<tr>
<td>Well 3</td>
<td>Exploratory</td>
<td>Unknown</td>
<td>2715</td>
<td>Gamma Ray, Resistivity, Sonic, Density, Neutron</td>
</tr>
</tbody>
</table>
Principal component analysis-based statistical well log analysis

a)

b)
Results and discussion

This study aims to illustrate the application of PCA as an independent technique for pattern recognition and reservoir characterisation using petrophysical data. Crossplots between log datasets using different parameters were used to analyse the applicability of PCA for petrophysical analysis. Crossplots of neutron and sonic porosity for three wells based on shale volume is illustrated in Figure 3.

Figure 3 shows the crossplot between neutron porosity (NPHI) and sonic (DT) curves for wells 1, 2 and 3. The plot is color-coded for different range of $V_{\text{shale}}$, and it shows scattering of data for all the wells (particularly wells 2 and 3). As per the well report, for well 1, the Tura formation consists of sandstone with thin layers of shale. For well 2, the Bokabil formation mainly consists of alternations of sandstone, siltstone, and claystone. The Barail formation consists of alternate bands of shale, sandstone and siltstone with coal and limestone at the bottom. The Kopili formation consists mainly of shale with thin bands of sandstone, siltstone and clay in the upper part and limestone at the bottom. Tura formation consists of sandstone, shale, and thin bands of limestone. For well 3, Bokabil formation consists of dominantly clay/claystone with intercalations of sand and siltstone layers while Barail formation consists of sandstone and siltstone with intercalations of claystone, shale, and coal bands. Thus, from the crossplots, the high $V_{\text{shale}}$ may be attributed to the presence of shale layers. However, a few important observations can be noted: (i) for well 1, the trend seems to be more aligned along the 45-degree line when compared with datasets belonging to wells 2 and 3; (ii) for well 1, the crossplot
indicates sandstone and limestone units with intercalations of shale (due to high value of $V_{\text{shale}}$); and (iii) for wells 2 and 3, the presence of shales can be inferred based on high $V_{\text{shale}}$ value.

Figure 3: Crossplots between neutron porosity (NPHI) and sonic (DT) curves for wells 1, 2 and 3. The different colours of the different cluster points corresponds to the fraction of volume of shale as indicated in the legend.

For well 1, the outliers belong to Tura formation (2061 – 2068 m) with lithology -sandstone along with thin layers of shale. For well 2, the outliers in the cross plots may be attributed to the sandstone, shale, limestone of Bokabil, Tura and Gondwana formations. For well 3, the outliers in the crossplots may be attributed to clay, siltstone, and claystone of Tipam, Bokabil, Barail formations. The crossplot between NPHI and DT in Figure 3 shows a trend for both wells 2 and 3. The range of $V_{\text{shale}}$ values showing the trend is characteristic of shale. The well report also suggests the presence of shale in those wells. Subsequently, PCA was performed to derive more patterns and infer correlations.
The correlation matrix generated from PCA as in Table 2, indicates that for well 1, DT correlates to GR and NPHI. Similarly, for well 2, DT shows a greater correlation to NPHI and RHOB, while for well 3, RHOB is correlated to GR and NPHI. In PCA, mainly the first three principal components are used to determine the correlation in the data. Therefore, in this study, the first three PC values were determined for all the wells (Table 3). From Table 3, it can be observed that depth, NPHI and RHOB influence the value of PC1, while depth influences PC2 and DT influences PC3.
Table 3: PC1, PC2 and PC3 values for wells 1, 2 and 3

<table>
<thead>
<tr>
<th>Variables</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>0.755</td>
<td>0.823</td>
<td>-0.277</td>
</tr>
<tr>
<td>DT</td>
<td>-0.142</td>
<td>0.148</td>
<td>0.908</td>
</tr>
<tr>
<td>GR</td>
<td>0.507</td>
<td>0.435</td>
<td>0.186</td>
</tr>
<tr>
<td>NPHI</td>
<td>0.575</td>
<td>-0.306</td>
<td>-0.136</td>
</tr>
<tr>
<td>RHOB</td>
<td>0.621</td>
<td>-0.138</td>
<td>0.216</td>
</tr>
</tbody>
</table>
Figure 4: Crossplot between PC1 and PC2 for (a) well 1, (b) well 2, and (c) well 3, and between PC1 and PC3 between (d) well 1, (e) well 2, and (f) well 3. All these plots show a conclusive trend for shale.

Figure 4 presents PCA of each well separately and Figure 5 represents all wells together colour-coded with $V_p$ values derived from the DT logs) to understand the scatter and derive any meaningful correlations. From Figure 5, it was evident that for well 1, a correlation existed (points aligned along a 45-degree line) in both single well data plots (PC2 versus PC1 and PC3 versus PC1) and combined well data plots, while for well 2 and 3 no trend emerged. The plot between PC3 vs PC1 for well 2 indicates a trend which can be matched with that shown in the Figure 5 (yellow colour indicating $V_p$ between 2.5 and 3 km/s). This range of $V_p$ values may be characteristics of shale present at depth 1900-2585 m. Moreover, the plot for PC2 and PC1 for well 3 (green colour indicating $V_p$ ranging from 3.5 to 4.5 km/sec) may be indicative of shale present at depth 2450-2690 m.
Principal component analysis-based statistical well log analysis

Figure 5: Crossplots between (a) PC1 and PC2, and (b) PC1 and PC3 plotted for wells 1, 2 and 3 colour-coded with $V_p$ values. The $V_p$ values were used for pattern recognition with a dominant trend for shale in the cross-plot.

Figure 6: Crossplot of PC2 and PC3 for wells 1, 2 and 3 plotted together, with cluster points in red representing well 1 data, in cyan from well 2 and in green from well 3. Cluster points from the hydrocarbon zone in well 1 exhibits a distinct trend not exhibited by the water bearing zones from wells 2 and 3. As per the well report, well 1 passes through Bokabil (1557-1694 m), Barail (1694-1737 m), Kopili (1737-2001 m) formation. Well 2 passes through Bokabil (1501-1766 m), Barail (1766-1820 m), Kopili (1820-2126 m), Sylhet (2126-2191 m), Mikir Trap (2191-2273 m) and Gondwana (2273-2569 m) formation. Well 3 passes through Bokabil (1575-1946 m), Barail (1946-2059.5 m), Kopili (2059.5-2318 m), Sylhet (2318-2400 m), Trap (2400-2441.5 m), and Gondwana (2441.5-2674 m) formation.

GEOHORIZONS, December 2022
© SPG India. All rights reserved.
The colour-coded segregated trends based on PC2 and PC3 for all three wells are shown in Figure 6 (PC2 versus PC3 for all three wells). Colour coding complements the investigation and derivation patterns from the scattered dataset. Well 1 data points are aligned along a 45-degree line, while well 2 and 3 data points are aligned almost perpendicular to well 1 for PC2 versus PC3. From further investigation, it was clear that the well 1 pattern represents the hydrocarbon-bearing zone while well 2 and well 3 pattern indicates water-bearing zones.

Figure 6 shows that the trend of the hydrocarbon bearing zone (well 1) is different from that of the water bearing zone (wells 2 and 3). In other words, with the help of PCA, it becomes easier to distinguish between the trends. These results were validated from the well reports that suggested well 1 to be hydrocarbon bearing, wells 2 and 3 were found to be water bearing. As per the well report, well 1 indicated hydrocarbon shows while passing through Bokabil (1557-1694 m), Barail (1694-1737 m) and Kopili (1737-2001 m) formations while well 2 and well 3 turned out to be water bearing.

Conclusions

The following important conclusions can be drawn from the present study:

- From the cross plot between NPHI and DT for well 1, sandstone and limestone units with shale intercalations may be inferred, while for wells 2 and 3, the presence of shales can be inferred based on high GR value.
- From well 1 PCA, it was observed that NPHI and GR have a positive correlation with DT. Also, depth, NPHI, RHOB influences the value of PC1.
- From well 2 PCA, it was observed that: DT has a negative correlation with RHOB, while NPHI has a positive correlation with DT. Also, depth influences the value of PC2.
- From well 3 PCA, it was observed that: RHOB has a positive correlation with GR, while a negative correlation was observed between RHOB and NPHI. Also, DT influences the value of PC3.
- The nature of trends investigated and classified based on $V_p$ indicates that for well 2, $V_p$ value ranging between 2.5 to 3 km/sec is characteristic of shale present at depth 1900-2585 m. Consequently, for well 3, $V_p$ values ranging from 3.5 to 4.5 km/sec indicate shale present at depth 2450-2690 m.
- On investigating colour code segregated PC values, it was observed that a trend existed for the hydrocarbon bearing zone in well 1 while PC datasets were scattered for non-hydrocarbon bearing zone of well 2 and 3. These inferences were also validated from the well reports that suggested significant hydrocarbon shows in well 1. Hence, PCA can distinguish hydrocarbon-bearing zones from non-hydrocarbon-bearing zone.
- Thus, PCA analysis can be recommended as a quick look evaluation tool to analyse the well log response without performing rigorous conventional analysis of the data. This type of study will help carry out rock physics modelling.
The significance of using PCA is highlighted in the present study. Future studies may focus on utilising other statistical techniques, such as Monte Carlo simulation and neural network/machine learning algorithms, to complement the well-log analysis.

Acknowledgements

The authors are thankful to NDR (National Data Repository), DGH (Director General of Hydrocarbon), India and IIT Guwahati for providing technical support for research.

References


Jolliffe, I. T., and J. Cadima, 2016, Principal component analysis: a review and recent Developments, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.


Lim, J. S., J. M. Kang, and J. Kim, 1998, Artificial intelligence approach for well-to-well log correlation, in SPE India Oil and Gas Conference and Exhibition. OnePetro.


Pal, A. K., S. Garia, S., K. Ravi, and A. M. Nair, 2022, Pore scale image analysis for petrophysical Modelling, Micron, 154, 103195.


Shen, C., S. Asante-Okyere, Y. Yevenyo Ziggah, L. Wang, and X. Zhu, 2019, Group method of data handling (GMDH) lithology identification based on wavelet analysis and dimensionality reduction as well log data pre-processing techniques. Energies, 12(8), 1509.

**Siddharth Garia** received his Ph.D. degree from Indian Institute of Technology, Guwahati, Assam, India, in 2022.

He is currently working as a guest faculty in Dr. BR Ambedkar National Institute of Technology, Jalandhar. His research interest includes rock physics, exploration geophysics, seismic inversion, soil structure interaction and geotechnical engineering.

---

**Arnab Kumar Pal** received his M. Tech degree from Indian Institute of Technology, Guwahati, Assam, India, in 2018.

He is working towards his Ph.D. with Indian Institute of Technology, Guwahati. His research area includes rock physics, near surface geophysics, digital image analysis and digital rock physics.
**Ravi Karangat** received his Ph.D. degree from the Indian Institute of Science, Bangalore, India in 2013.

He is currently an Associate Professor in the Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India. His research interests include geo environmental engineering, geo-energy systems, engineering behaviour of unsaturated soils, research on hazardous waste management and experimental soil mechanics.

**Archana M. Nair** received her Ph.D. degree from Indian Institute of Technology, Bombay, India, in 2008.

She is currently an Associate Professor in the Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India. Her research interests include remote sensing in planetary exploration, rock physics modelling, groundwater hydrology, isotope hydrology, critical zone, near-surface geophysics, and quantitative geomorphology.