Predicting variability in well performance using the concept of shale capacity: An application of machine learning techniques

Ritesh Kumar Sharma[1] and Satinder Chopra[1*]

Abstract

The production from a shale well depends on how accurately that well has been placed through zones with superior reservoir quality and completion quality. To be able to locate such pockets, an integration of different types of reservoir properties such as organic richness, frackability, fracture density and porosity is essential. One way of achieving this is by using cut off values for the different reservoir properties and generating a *shale capacity volume*. However, in this regard, a couple of things need to be borne in mind. Not only seismic data provide indirect measurements of the individual reservoir properties, but different seismically derived attributes may contribute to the computation of the individual components of the shale capacity volume. Therefore, the definition of the different reservoir property cut off values is subjective and difficult to finalize. What may be desirable is that the different seismically derived attributes are interpreted simultaneously for predicting the performance of a well, which may not be a straightforward task. The application of machine learning techniques seems like an attractive alternative here, which we explore in this article. For doing so, data examples from the Delaware Basin as well as the Western Canadian Sedimentary Basin (WCSB) have been considered. The dataset from the Delaware Basin is used to gain confidence in applying unsupervised machine learning techniques as one of them is used thereafter on the dataset from WCSB, where production data associated with different wells were available.

Introduction

The ultimate goal of reservoir characterization work carried out for a shale play is to enhance hydrocarbon production by identifying the favourable drilling targets. The drilling operators have a perception that in organic-rich shale formations horizontal wells can be drilled anywhere in any direction, and hydraulic fracturing at regular intervals along the length of the laterals can then lead to better production. Given that this understanding holds true, all fracturing stages are expected to contribute impartially to the production. However, studies have shown that only 50% of the fracturing stages contribute to overall production (Miller et al., 2011). This suggests that repetitive drilling of wells and their completions without attention to their placement must be avoided and smart drilling needs to be followed by operators. Smart drilling consists of optimally placing a horizontal well and thereafter stimulating it in such a way that more uniform production across fracturing stages occurs, which leads to a better overall production.

Miller et al., (2010, 2011, 2013) showed that reservoir quality (RQ) and completion quality (CQ) must be considered in planning of strategic hydraulic fracturing stimulation for increasing the total production of a horizontal well. While organic richness,

effective porosity, fluid saturation, pore-pressure and gas-in-place define RQ, the CQ is represented by in-situ stress field, mineralogy (clay content and type), and the presence of natural fracture along with their orientation. Therefore, an integration of these properties should be considered for identifying the sweet spots. Studies have compared production cost between factory drilling and smart drilling and shown significant advantages of the latter (Schuenemeyer and Gautier, 2014). Smart drilling may also lead to prolific to mediocre producers, rather than poor producers. The above discussion puts emphasis on the integration of different reservoir properties for predicting the potential of a shale play, which is usually not followed by geoscientists as they are still engaged in correlating individual seismically-derived attributes with the production data, which must be avoided. Different reservoir properties can be integrated by introducing the concept of *shale capacity* which is defined as the ability of a shale play to produce hydrocarbons once it gets fractured (Ouenes, 2012). Mathematically, *shale capacity* (SC) is defined as a function of *TOC* (total organic content), natural fracture density (*FD*), brittleness (*BRT*), and porosity (∅) as follows

$$SC = TOC_{net} \ FD_{net} BRT_{net} \varnothing_{net} \qquad (1)$$

where $TOC_{net} = 0$ when $TOC < TOC_{cut-off}$, $FD_{net} = 0$ when $FD < FD_{cut-off}$, $BRT_{net} = 0$ when $BRT < BRT_{cut-off}$ and $\varnothing_{net} = 0$ when $\varnothing < \varnothing_{cut-off}$.

The importance of the concept of shale capacity in predicting a well's performance for different shale plays has been shown by Ouenes (2014) and Newgord et al. (2015). From the equation of shale capacity given above, it is obvious that an optimal combination of all four parameters could lead to a higher shale capacity, i.e. the shale capacity exists only in case all four parameters are above their cut-off values. In other words, an ideal shale well must be drilled in a high TOC zone, which is brittle enough to be fractured, and a natural fracture system must be intercepted by the induced hydraulic fractures to develop a high porosity system. Therefore, due attention should be devoted to all these parameters for determining the potential of a shale play.

The availability of core data, well log curves such as dipole sonic with azimuthal measurements, and image logs could potentially arm the reservoir engineers or petrophysicists with direct measurements of different reservoir properties for estimating shale capacity. However, direct measurements of such properties are not possible with surface seismic data, and so, indirect methods are used. Even when the seismically-derived properties do become available, it is not advisable to use equation 1 directly for shale capacity determination. This is because the definition of cutoffs of different desired reservoir properties is subjective and difficult to finalize. In addition, there is no universally accepted definition of brittleness (Sharma et al., 2019). Furthermore, the indirect measurements determine a physical relationship between a seismically derivable attribute and reservoir property of interest, e.g. impedance-porosity, density-*TOC* and Poisson's ratio-quartz content, etc. However, the same relationship may not be applicable for all shale plays as it depends on mineralogy, porosity, pore shape and saturation fluid (Qian, 2013) which are usually found to vary from basin to basin. Accordingly, different kinds of attributes must be considered in the process of defining shale capacity, which is not an easy task to tackle manually. However, machine learning techniques could be quite useful for the purpose and are worth exploring. As many machine learning techniques are available, they are discussed next.

Application of machine learning techniques

There are three broad families of machine learning algorithms namely, visualization and dimensionality reduction, clustering and supervised learning. More details about these individual classes and their application on well-log data as well as seismic data have been demonstrated by the authors (Chopra et al., 2019). While supervised learning techniques are preferred over unsupervised techniques, it is always challenging to collect adequate data for the former and an abundance of data are required for training and validation purposes. On the other hand, unsupervised machine learning techniques do not carry such requirements and hence have immense potential for being used in the geophysical domain, where the lack of appropriate data is quite common. For example, if we wish to predict shale capacity using one of the supervised machine learning techniques say based on neural networks, it is required to have enough training data points for *SC* which can be derived as per equation 1. As discussed earlier, it is an arduous task to compute *SC* in the geophysical domain which thus restricts the use of supervised machine learning techniques. Therefore, an attempt has been made here to extract *SC* information with the help of unsupervised machine learning techniques. In order to increase our confidence in the application of unsupervised techniques, it is essential to calibrate it with the application of supervised or semi-supervised techniques. To do so, a comparison of *k*-means clustering and Bayesian classification for defining different facies was followed on a dataset from the Delaware Basin.

Comparison of facies determination using unsupervised and supervised machine learning techniques

In order to demonstrate the application of machine learning techniques, well log data for five wells from the Delaware Basin have been considered. To start with, different well-log curves such as sonic, gamma ray, density, neutron-porosity and resistivity were used to define different facies using *k*-means clustering. As we need to specify the number of data clusters for executing the analysis, eight clusters were used in the process considering our broad zone of interest (Bell Canyon to Devonian Carbonate) in the Delaware Basin. Figure 1 shows the result of *k*-means clustering carried out on the five different wells. Different facies are noticed for different intervals at individual wells. Also, spatial variation in the clustered facies can be noticed within Bone-Spring formation, Wolfcamp formation and Barnett formation when going from one well to another. It is encouraging that *k*-means clustering
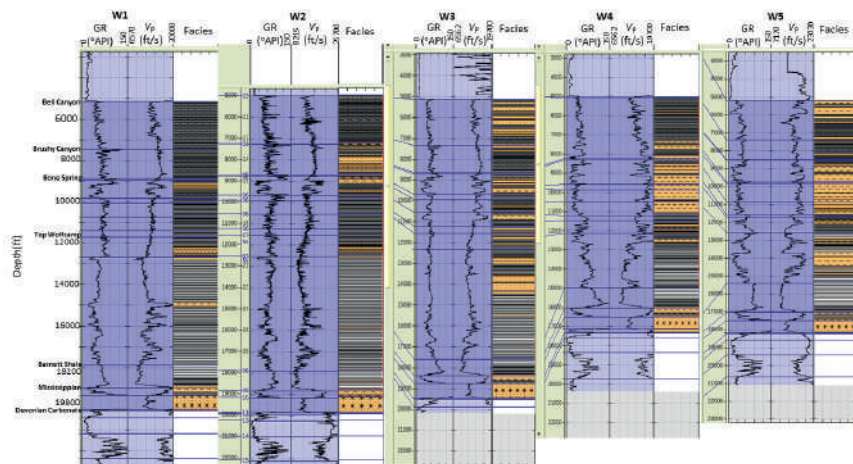

Figure 1 Illustrating the different facies at individual wells for different intervals as a result of k-mean clustering carried out on sonic, gamma-ray, resistivity, density and neutron porosity well-log curves.
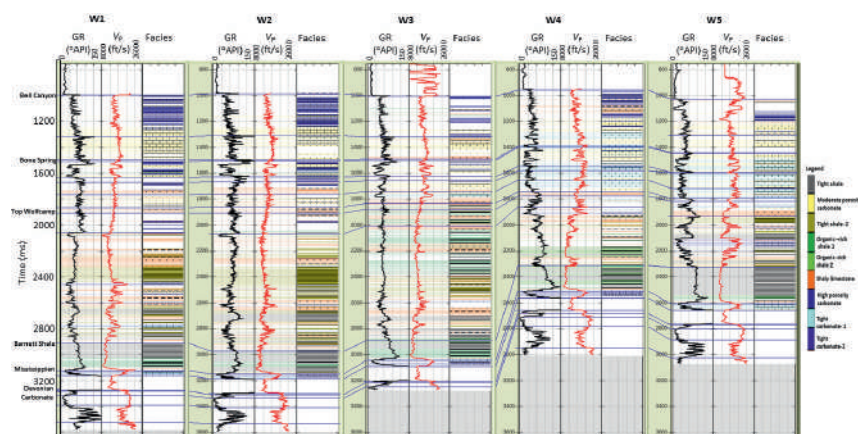
Figure 2 Interpretation of litho-classification based on well log neutron and density porosities by restricting their values.
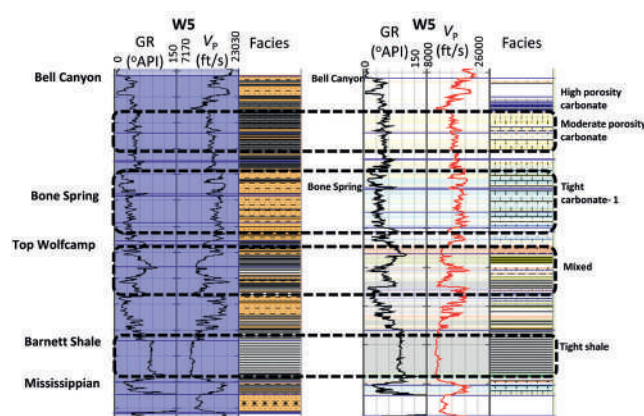

Figure 3 The mapping of unsupervised facies (left) with supervised facies (right) for well W5.

emphasizes the vertical as well as lateral variation in terms of different facies, which is obvious if different well-log curves are analysed together. However, the problem with such clustering is that it does not say what those facies are.

To increase confidence in such clustering, Bayesian classification carried out on the five wells used in the clustering analysis has been considered. In order to define the different facies a statistical approach was followed, entailing a graphical cross-plot method for determination of the volume of shale and effective porosity in a formation. The approach starts with cross-plotting of neutron-porosity ($\boldsymbol{\emptyset}_N$) and density-porosity ($\boldsymbol{\emptyset}_D$) curves covering a broad zone of interest, where different facies can be defined based on the volume of shale and effective porosity information. The details of this approach can be found in Sharma et al., 2019. The validity of interpreted facies was ascertained using processed mud-log data as well as other petrophysical analysis. The interpretation of Bayesian classification for different litho-units equivalent to the one shown in Figure 1 has been illustrated in Figure 2. On examining this figure, it is obvious that the carbonate content within the Bone-Spring formation increases on going from well W1 to W5 which is expected as wells W4 and W5 are close to the Central Basin Platform. The k-mean clustering also exhibits different facies at individual wells within the Bone-Spring formation. For the Mississippian interval, while Bayesian classification shows the presence of limestone, k-mean clustering yields consistent facies over this interval. Clay-rich shale (tight shale) and organic-rich shale interpreted within the

Barnett formation show the correspondence with the exhibited facies when k-mean clustering is followed. Additionally, there is reasonable correlation between Bayesian classification and k-mean clustering outcomes over Wolfcamp interval. Consequently, different facies generated using k-mean clustering could be mapped on Bayesian classification. For example, such a mapping at well W5 is shown in Figure 3. Notice, different cluster facies shown on the left side can be correlated reasonably well with Bayesian classification displayed on the right side. Such a resemblance noticed among different facies defined from supervised and unsupervised techniqueslends confidence in the latter approach.

With such confidence gained, the next question addressed is if it is possible to predict the variability of wells performance in an area of study using unsupervised machine learning techniques. The prerequisite for following any of the techniques is that the production data associated with different wells must be available. Since well production data are rarely available in the public domain for any dataset from USA, another dataset from the WCSB is taken into consideration to answer the question of implementing a machine learning approach for estimating shale capacity volume. As has been mentioned earlier, an ideal shale well would be drilled in a formation that has the following properties:

• It is organically rich and porous and thus can produce enough quantity of hydrocarbons.
• It is brittle/frackable enough to get fractured easily.
• The induced fractures resulting from hydraulic fracturing intercept the existing natural fracture system in the formation.
• The induced fractures are compatible with type of fracture network pattern desired for the development.

Knowing the limitations of seismic data in measuring these properties directly, their proxies only are derived from seismic data. For example, porosity variations may give rise to changes in impedance, density, velocity ratio etc. Similarly, high quartz or carbonate content may have impact on Young's modulus and Poisson's ratio in addition to other rock properties. Consequently, it should be possible to detect changes in different properties required for shale capacity volume from the surface seismic response. However, coupling of reservoir properties with seismically derived attributes are complex and not easy to understand. Therefore, different seismic attributes should be

analysed simultaneously to derive the individual components of shale capacity volume. For example, organic richness and porosity have prominent impact on P-impedance, density, V_p/V_S, and Lambda-rho and so these attributes can be treated as their proxies. Furthermore, fracture toughness, strain energy density (Sharma et al., 2019), and fracture intensity computed using velocity variation with azimuth (VVAz) and the proposed fracture toughness (FT) approach can be considered as a proxy of frackability and fracture/stress induced anisotropy in addition to curvature attributes. Next, all these attributes should be integrated using machine learning techniques to predict the shale capacity volume as illustrated in the cartoon shown in Figure 4.

This workflow has been implemented on a dataset from the WCSB where the Montney and Duvernay formations represent the zones of interest and are shown in Figure 5. In order to
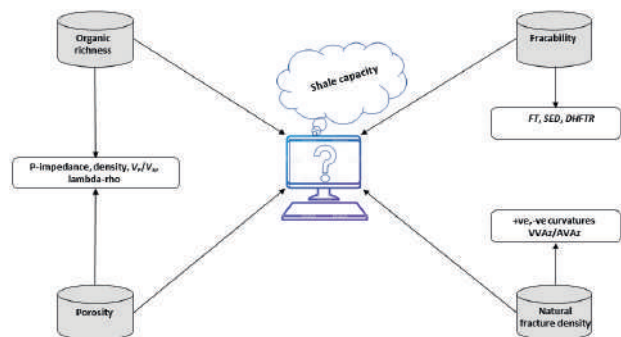

Figure 4 A cartoon illustrating the machine learning approach of predicting shale capacity volume by integrating different required elements.

determine the input attributes required for shale capacity as per the cartoon shown in Figure 4, simultaneous impedance inversion was performed on 5D PSTM pre-stack data by following proper data conditioning, robust low-frequency models and accurate inversion parameters. Such inversion yields P- and S-impedance volumes. Thereafter, other attributes such as Lambda-rho, Mu-rho, E-rho, Poisson's ratio, fracture toughness (FT), strain energy density (SED) were computed using impedance volumes. Next, simultaneous inversion was performed on the individual azimuth-sectored gathers to determine FT volumes for different azimuths. Having computed these volumes, the magnitude of seismic anisotropy was estimated from FT azimuthal variation via differential horizontal fracture toughness ratio (DHFTR). The analysis of variation of velocity with azimuth (VVAz) and the variation of amplitude with azimuth (AVAz) were also performed on the dataset used here. For executing these analysis, common-offset and common-azimuth (COCA) gathers were first generated for adding stability and bringing in more foldage to the analysis. Thereafter, residual azimuthal travel-time shifts required to align all azimuths on the individual reflection events, were determined using dynamic trim statics after following some data conditioning. Having gained the confidence in the data conditioning, the determined time shifts are then inverted for RMS anisotropic attributes ($V_{NMO\text{-}fast}$, $V_{NMO\text{-}slow}$, and $\varnothing_{fast}$) for each CMP. Next, intermediate RMS attributes are converted to $V_{INT\text{-}fast}$, $V_{INT\text{-}slow}$ and fracture orientation using the Generalized Dix equation (Grechka et al., 1999). Subsequently, flattened gathers are used in the curve fitting analysis based on the Rüger equation
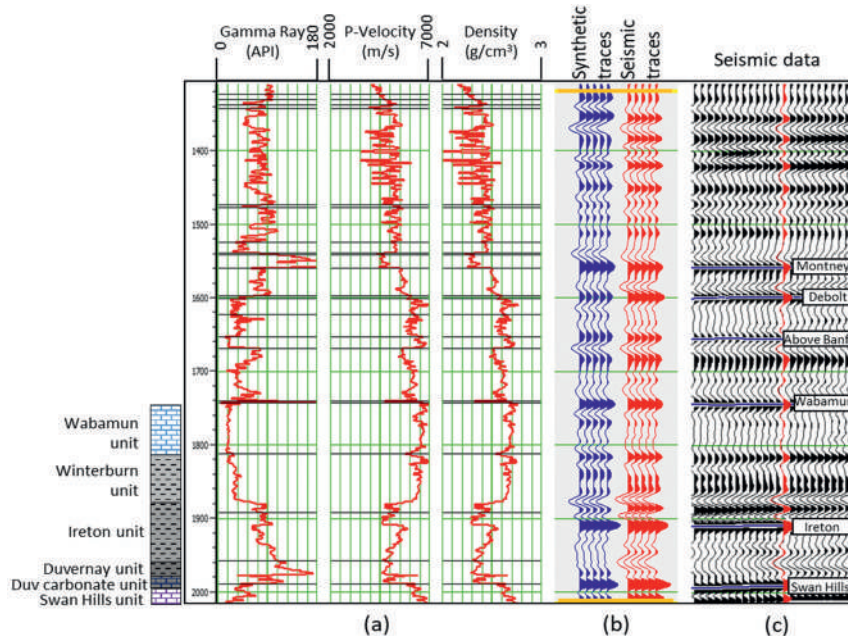

Figure 5 Correlation of well log curves with seismic data. Blue traces represent the synthetics (generated with the wavelet shown above) while the red traces represent the seismic data. The overall correlation between the synthetic and red traces is good. (Data courtesy: TGS, Canada).
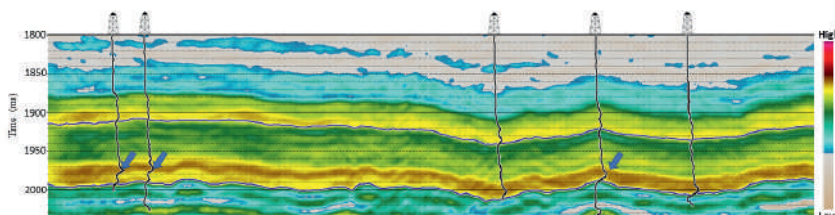

Figure 6 An arbitrary line section passing through different wells from PCA-1. Overlaid GR curves show the location of the Duvernay formation as it is associated with a high GR response highlighted by blue arrows. (Data courtesy: TGS, Canada).
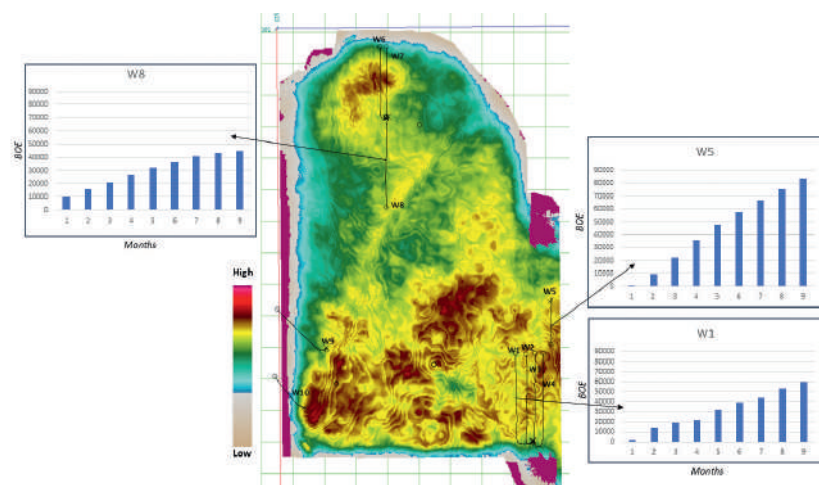
Figure 7 Horizon slice extracted from the PCA-1 volume averaged over a 30ms window to map the variability within the Duvernay formation. (Data courtesy: TGS, Canada).
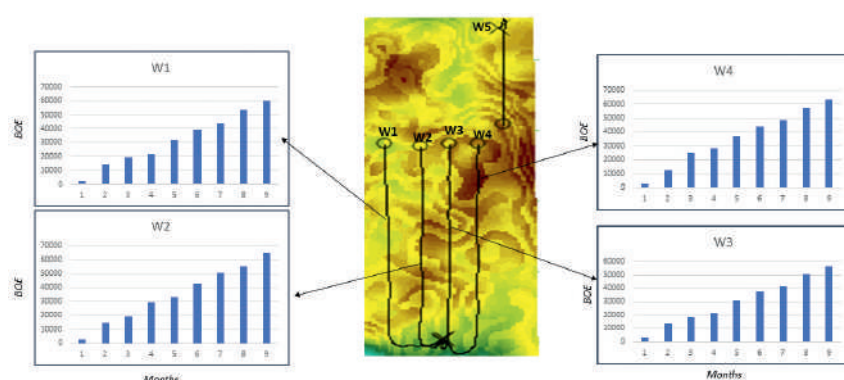

Figure 8 Zoom of south eastern part of horizon slice shown in Figure 7. (Data courtesy: TGS, Canada).

(Ruger and Tsvankin, 1997) which provides the estimation of fracture intensity and orientation.

Volumetric curvature attributes are valuable in mapping subtle flexures and folds associated with fractures in deformed strata (Chopra and Marfurt, 2007). In addition to faults and fractures, stratigraphic features such as levees and bars and diagenetic features such as karst collapse and hydrothermally altered dolomites also appear to be well-defined on curvature displays. By first estimating the volumetric reflector dip and azimuth that represents the best single dip for each sample in the volume, followed by computation of curvature from adjacent measures of dip and azimuth, a full 3D volume of curvature values is produced. There are many curvature measures that can be computed, but the principal most-positive and most-negative curvature measures are the most useful in that they tend to be most easily related to geologic structures. These attributes were generated for the dataset under investigation.

These seismic attributes were put through the unsupervised machine learning technique to predict the shale capacity volume. For the sake of simplicity, we begin with principal component analysis (PCA), so as to figure out the patterns and relationships in them. Usually the first three principal components carry almost all of the information contained in the input attributes, with PCA-1 containing a large part of that. Consequently, PCA-1 can be treated as a proxy for the shale capacity volume. Figure 6 shows an arbitrary line passing through different wells from the PCA-1 volume from a 3D seismic survey in central Alberta. The display exhibits both the lateral and temporal variations. To

capture the lateral variations in the data, Figure 7 shows a horizon slice averaged over a 30ms window covering the Duvernay Formation. The hot colours on the display represent higher values than the greenish/bluish colours, but as PCA is an unsupervised machine learning technique, it is difficult to conclude as to which colour conveys what information. To gain some insight into this dilemma, the nine-month cumulative barrels of oil equivalent (BOE) production data available for those wells are brought in and found to be associated with different colours. Notice the productivity of a well increases in going from the greenish colour to hot colours. It may therefore be concluded that hotter colours are preferable for the delineation of sweet spots.

Consequently, we look for correlation of the variation in production for the different horizontal wells drilled from the same pad with the colours on the display. Figure 8 exhibits a zoom of the four wells seen to the southeast part of the horizon slice in Figure 7, along with their nine-month cumulative BOE production data. The general trend of increased productivity with the intensity of hot colours holds true if wells W1 and W4 are compared. However, it falls short of explaining the higher production associated with well W2.

In our attempts to find an answer here, PCA-2 and PCA-3 volumes are also examined, in the hope of extracting additional information. It was found that PCA-3 is dominated by the curvature attribute, which is as a discontinuity attribute different from the other input attributes. Although only ranking third in PCA it is a key component in defining shale capacity volume (Ouenes et al., 2015).
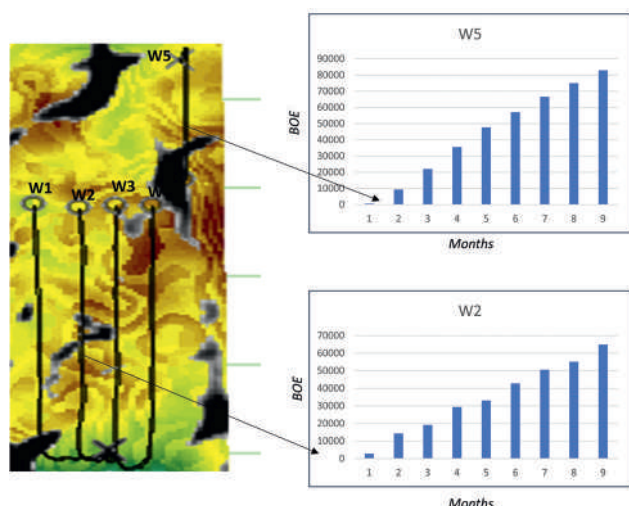
Figure 9 Equivalent display to the one shown in Figure 8 after overlaying the lineaments interpreted on the curvature attribute using transparency. (Data courtesy: TGS, Canada).
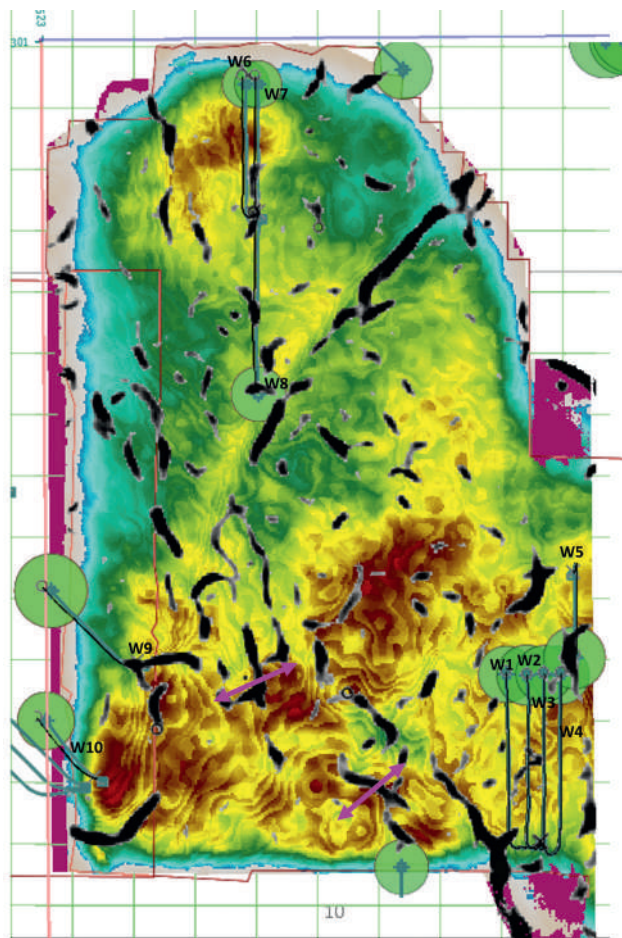

Figure 10 Composite horizon slice display from the PCA-1 volume averaged over a 30ms window covering the Duvernay interval, overlaid with curvature lineaments using transparency and cumulative nine-month BOE production. (Data courtesy: TGS, Canada).

Therefore, the different lineaments in the zone of interest are identified on the curvature attribute and then overlaid on the horizon slice from PCA-1 attribute using the transparency as shown in Figure 9. The well W2 seems to have been drilled into a naturally fractured zone as it is passing through a lineament that could be

the possible reason of exhibiting higher production. Similarly, the enhanced production from well W5 is seen to be associated with a zone exhibiting hot colours and crossing a large lineament that probably exhibits higher permeability. In a similar fashion, the variation of different wells from the northern side (Figure 7) can be explained in terms of amplitude on PCA-1 being associated with lineaments interpreted on the curvature attribute. Well W7 seen on the display in Figure 7 is associated with higher production than well W6, though they are both drilled from the same pad and exhibit similar amplitudes in terms of their colours. There are, however, curvature lineaments that well W7 traverses, but not well W6. In a similar vein, well W8 drilled from the same pad but in the southerly direction yields higher production than well W6, even though it is not associated with hot colours, but traverses more lineaments.

Finally, in Figure 10, besides overlaying the curvature lineaments using transparency, the production data has been posted on the horizon slice display shown in Figure 7, with the higher producing wells having bubbles of a bigger radius. Based on the discussion mentioned above, the following are the take-away points from this analysis:

- Production of a well is seen to increase with the intensity of colours.
- The presence of intersecting lineaments in the hot coloured zones leads to the higher production as noticed for the wells W5 and W9.
- As shown in Figure 10, the location of double-sided magenta arrows could be considered as hot spots for future drilling.

Conclusions

Considering the importance of shale capacity in defining the potential of a shale play, we have attempted to extract this information from seismic data using machine learning techniques. Different unsupervised machine learning techniques and their implementation in identifying the different facies were discussed. The confidence in using unsupervised machine learning techniques was gained by calibrating their outcomes with the supervised learning technique. Next, principal component analysis was carried out for integrating different components of the shale capacity such as organic richness, porosity, natural fracture density, and frackability. The variability in well performance was addressed by using the PCA-1 along with the curvature attribute (PCA-3). Such an analysis suggests that the combination can be used for identifying future drilling locations.

Acknowledgements

References

Chopra, S. and K.J. Marfurt. [2007]. *Seismic attributes for prospect identification and reservoir characterization*. Tulsa, Oklahoma, USA.

Chopra, S., R.K. Sharma, and K.J. Marfurt. [2019]. Unsupervised machine learning facies classification in the Delaware Basin and its comparison with supervised Bayesian facies classification. *89th Annual International Meeting, SEG,* Expanded Abstracts, 2619-2623.

Grechka, V., I. Tsvankin and J.K. Cohen. [1999]. Generalized Dix equation and analytic treatment of normal-moveout velocity for anisotropic media. *Geophysical Prospecting*, **47**, 117-148.

Miller, C., E. Rylander, and J.L. Calvej. [2010]. Detailed rock evaluation and strategic reservoir stimulation planning for optimal production in horizontal gas shale wells. *AAPG International conference and exhibition,* Abstracts.

Miller, C., G. Water, and E. Rylander. [2011] Evaluation of production log data from horizontal wells drilled in organic shales. *SPE Annual Technical Conference and Exhibition*, SPE 144326.

Miller, C., D. Hamilton, S. Sturm., G. Waters, T. Taylor, J.L., Calvez, and M. Singh. [2013]. Evaluating the impact of mineralogy, natural fractures and in situ stresses on hydraulically induced fracture system geometry in horizontal shale wells. *SPE Annual Technical Conference and Exhibition,* SPE 163878.

Newgord, C., M. Mediani, A. Ouenes, and P. O'Conor. [2015]. Bakken well performance predicted from shale capacity. *Unconventional Resources Technology Conference (URTeC),* 2166588.

Ouenes, A. [2012]. Seismically driven characterization of unconventional shale plays. *CSEG Recorder,* **2**, 23-28.

Ouenes, A. [2014]. Distribution of well performances in shale reservoirs and their predictions using the concept of shale capacity. *SPE Annual Technical Conference and Exhibition*, SPE 167779.

Ouenes, A.N. Umholtz, and Y. Aimene. [2015]. Using geomechanical modelling to quantify the impact of natural fractures on well performance and microseismicity: Application to the Wolfcamp, Permian Basin. *Unconventional Resources Technology Conference (URTeC),* 2173459.

Qian, Z. [2013]. *Geophysical responses of organic-rich shale and the effect of mineralogy*. Ph.D. thesis, University of Houston.

Rüger, A., and I. Tsvankin. [1997]. Using AVO for fracture detection: Analytic basis and practical solutions. *The Leading Edge,* **10**, 1429–1434.

Schuenemeyer, J.H., and D. Gautier. [2014]. Probabilistic resource costs of continuous oil resources in the Bakken and Three Forks Formations, North Dakota and Montana. *Unconventional Resources Technology Conference (URTeC),* 1929983.

Sharma, R.K., S. Chopra and L. Lines. [2019]. Replacing conventional brittleness indices determination with new attributes employing true hydrofracturing mechanism. *89th Annual International Meeting, SEG,* Expanded Abstracts, 3235- 3239.