Multivariate statistical applications for quantifying contribution of seismic data to production prediction in the Wolfcamp, Midland Basin

William M. Bashore[1] and Christian W. Grant[1]

Abstract

A workflow with the objective of quantifying the added value of seismic-based information to improved prediction of oil production in the Wolfcamp section of the Midland Basin is described. A variety of multivariate statistical algorithms are employed in the workflow: unsupervised and supervised classification, categorical-based nonlinear regression, and spatial regression for well log estimation. Porosity is estimated from elastic inversion volumes within the context of lithotype classification originating with well logs. Porosity feet becomes the primary driver in production prediction. Two parallel analyses are performed. In the first, only well logs are used to predict porosity along all vertical wellbores. As a proxy for the impact of well spacing, different minimum search distances are investigated for porosity estimation error and associated error in prediction of cumulative oil production at six months. The second analysis is similar except, along each wellbore, porosity is extracted from the seismic porosity volume and then used as soft data. These two analysis flows are then applied to horizontal wellbores, which have no original porosity logs, to compare how well each predicts known cumulative production. The incorporation of seismic-based information does show significant error reduction.

Introduction

Questions were raised as to whether seismic-based information could be used to improve prediction of production in horizontal wellbores and, if so, could this improvement be quantified in barrels. This study addresses those questions and provides a graphical workflow that includes a variety of multivariate statistical algorithms fit for different purposes. The workflow's fundamental premise is that improved prediction of porosity by integrating seismic data ultimately leads to an improved prediction in production performance.

The seismic control consists of a series of elastic inversion volumes from a 250 mi^2 survey northwest of Midland. Production within this area of interest is primarily from the Spraberry and Wolfcamp stratigraphic formations. The Wolfcamp sequence was selected for this paper, but similar results were obtained for the Spraberry. The final well control within the active seismic traces nominally includes 740 vertical wells and 67 horizontal wells targeting the Wolfcamp.

Figure 1 shows the study area location, and the areal grid of the seismic volumes used in the quantitative analysis is outlined by the map of the top of the Wolfcamp section. The background paleogeographic map of the Early Permian highlights the vicinity of the study area as being east of the Central Basin Platform. The complete set of wells, both vertical and horizontal, is also shown.

The following workflow uses a number of multivariate statistical approaches, each to accomplish a different objective. The approaches include unsupervised classification using *k*-means (MacQueen, 1967), semisupervised classification using a branching or hierarchical approach based on a dynamically growing self-organizing tree (DGSOT) (Luo et al., 2004), supervised classification using quadratic discriminant analysis (QDA) (Hastie et al., 2008), nonlinear regression using alternating conditional expectation (ACE) (Breiman and Friedman, 1985), and linear regression using ordinary kriging (Isaaks and Srivastava, 1989) and collocated cokriging (Xu et al., 1992).

Well log classification and porosity distributions

In the study area, the Wolfcamp is a complex of mainly shale with interbedded argillaceous carbonates and calcareous sandstones with mixed lateral continuity and extent. These facies are interpreted as hemipelagic deposits and sediment density-flow deposits reworked, locally, by bottom currents (Baumgardner et al., 2014). Because of these characteristics, two complications are observed. First, as will be seen, porosity distributions are highly dependent on lithotype; second, seismic interpretation of laterally discontinuous events is problematic. To perform statistical analyses on porosity and to facilitate the seismic interpretation, it is necessary to formulate a lithotype classification within the wells based on a diverse suite of logs. Ideally, it would have been useful to have such a classification scheme from core descriptions for control and calibration, but because one wasn't available, a statistical approach was employed to model one.

All vertical wells selected for statistical analysis have a well log suite of at least a gamma ray, a density, and a neutron porosity log. Most of these also have a deep resistivity log (typically deep induction log), and a small number have a P-sonic log. Tops picked and used for stratigraphic zonation include the Dean Lime, Upper Wolfcamp, Lower Wolfcamp, and the Strawn. Using wells with the full suite of these five logs and limiting the analysis from the Base San Andres to the Strawn, a classification model for this sequence was generated using a semisupervised multivariate approach.

Figure 2 contains an illustration of results from the application of DGSOT. This method was selected over other unsupervised classification approaches, such as *k*-means, because it provides user guidance and genetic relationships of class to class. There are two implementations of the algorithm: fining-downward and coarsening-upward. These are akin to the way core descriptions are usually handled; that is, when in doubt, create a new class as it can be possibly lumped subsequently with one or more other classes.

DGSOT includes a "smart" engine for how to grow the tree both laterally and vertically. Typical machine learning classification methods, such as a neural network and *k*-means, are used for this engine. Figure 2a shows the results of growing the tree to completion allowing two to four child nodes at each vertical split, where

[1]Drillinginfo.

https://doi.org/10.1190/tle37030190.1.

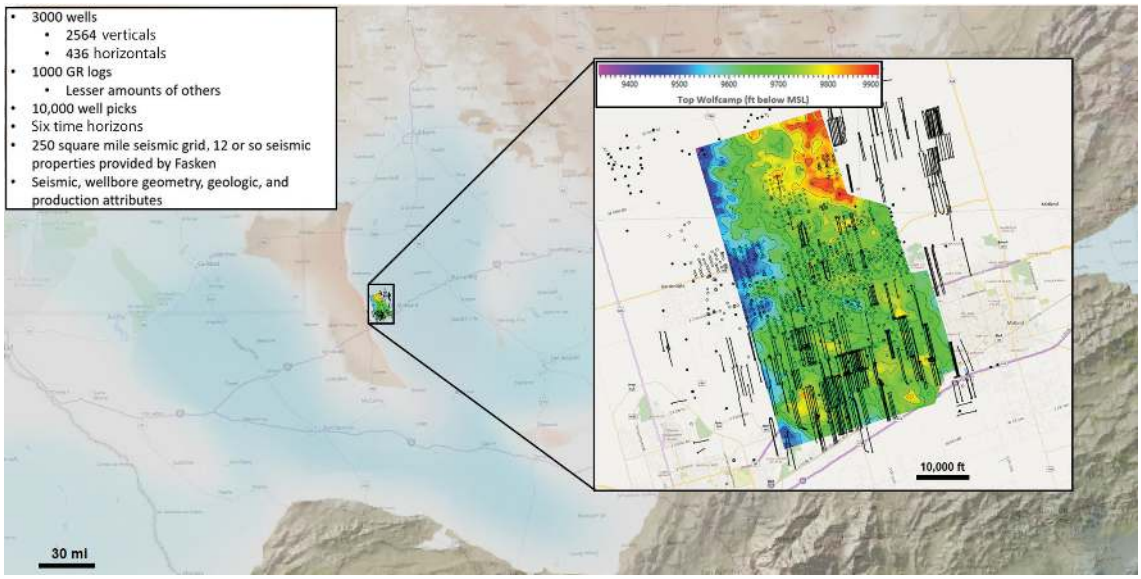

Figure 1. Study area in the Midland Basin with Wolfcamp paleogeography and a map of seismic data outline and vertical and horizontal well control.

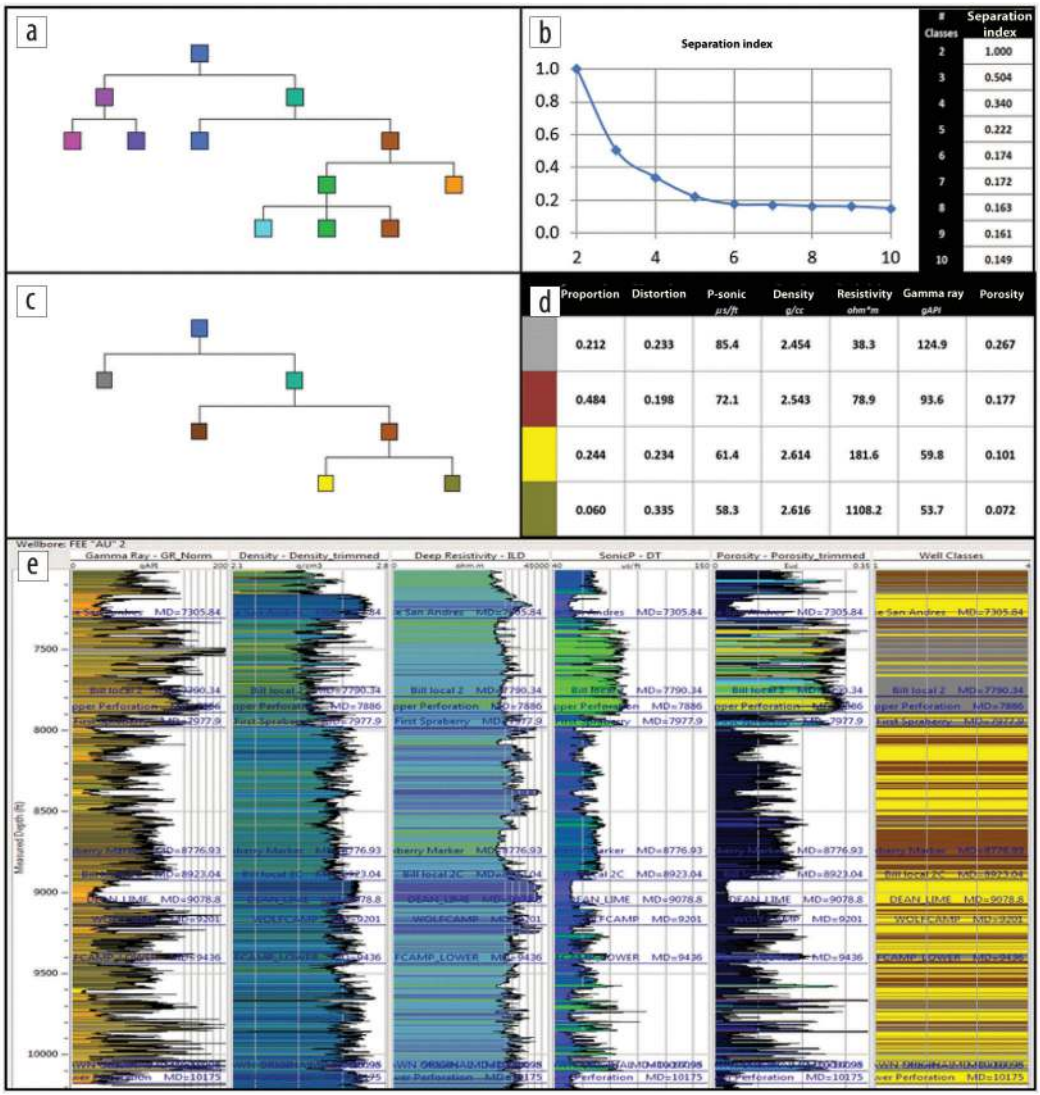

Figure 2. Results of hierarchical classification for well logs. (a) Initial run of hierarchical classification allowing four children and minimum class proportion of 5%, yielding seven final classes. (b) *K*-means scree plot for number of classes 2–10. (c) Collapsing hierarchy back to four primary classes from seven in (a). (d) Table of centroid locations for final four classes. (e) Application of hierarchical classification model to the FEE "AU" 2 well (API 42-135-39513).

each child node must contain at least 5% of the total sample count. In this case, four splits yielding seven final classes were made. The scree plot in Figure 2b is useful for understanding how the variance in the data is explained as the number of classes grows. A useful measure for variance is the separation index (y-axis in the scree plot). It is simply the ratio of the minimum intercentroid distance to the maximum intercentroid distance. For a two-class model, the minimum and maximum are the same, so the separation index is 1.

Often, one might wish to select the number of classes in which the scree plot begins to flatten. Five classes are a good selection from the plot, but four were selected as most geologically meaningful for the objectives of this study: three primary classes and a fourth minor class in the Wolfcamp (low-density, high-resistivity zone primarily found within the Spraberry). The splits were "lumped back" to the final four (Figure 2c), and the proportion, normalized distortion, and location of each class centroid is given in Figure 2d.

The application of this model to the Fee "AU" 2 well (API 42-135-39513), one of seven wells containing all five log types, is shown in the log plot of Figure 2e. Better coverage is desired for quantitative analyses so another classification algorithm is required, and a classification scheme is now available. This allows supervised classification to be employed; many methods exist to do this, and discriminant analysis was chosen. The sonic log is the most limiting log type, so it is the first to be removed. Next, the resistivity log is removed, and the sonic log is added back. Finally, both resistivity and sonic logs are removed. In each of these steps, all of the available samples were used (total count of 16,580). The error tables and confusion matrices for all three models are given in Figure 3. Application to the appropriate wells and using a log alias for the models in the order described now yields 67 wells with lithotype logs.

Reviewing the errors in each of the tables reveals that not having a sonic log still permits acceptable accuracy in all lithotypes. As we will see later, this is most likely because neutron porosity and velocity are highly correlated. The lack of a sonic log is compensated by having a porosity log. However, the lack of a deep resistivity log does cause a significant increase in error rate, especially in the yellow lithotype. Fortunately, the availability of resistivity logs is much higher than the availability of sonic logs. The confusion matrices permit us to see what is predicted when the model value is wrong. For example, as highlighted by the red box in Figure 3, in the case of removing only the sonic log, while the model got 3228 predictions correct, there were 292 wrong predictions of the yellow lithotype: seven were the olive or tan lithotype, 284 were brown, and one was gray.

The two distributions shown in Figure 4 illustrate the importance of segregating porosity by lithotype. The distribution in Figure 4a contains nearly 48,000 neutron porosity samples over

		No sonic				No resistivity				No sonic or resistivity			
Lithotype	Total	Right	Percent	Wrong	Percent	Right	Percent	Wrong	Percent	Right	Percent	Wrong	Percent
	3520	3228	91.7%	292	8.3%	2485	70.6%	1035	29.4%	2577	73.2%	943	26.8%
	149	124	83.2%	25	16.8%	134	89.9%	15	10.1%	111	74.5%	38	25.5%
	7606	7250	95.3%	356	4.7%	7038	92.5%	568	7.5%	6858	90.2%	748	9.8%
	5305	4634	87.4%	671	12.6%	4621	87.1%	684	12.9%	4576	86.3%	729	13.7%

	3228	2	142	30	2485	0	360	0	2577	0	540	0
	7	124	15	22	1	134	23	27	3	111	16	16
	284	1	7250	619	997	6	7038	657	885	5	6858	713
	1	22	199	4634	37	9	185	4621	55	33	192	4576

Figure 3. Error tables and confusion matrices for QDA modeling of well lithotypes given the elimination of P-sonic only, deep resistivity only, and both sonic and resistivity.

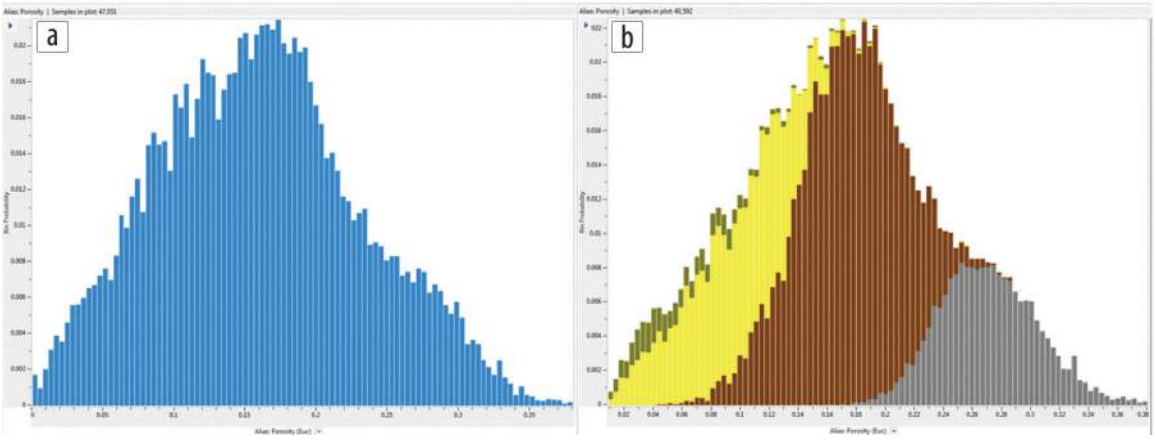

Figure 4. Distribution of porosity: (a) no segregation by facies and (b) segregated by facies (nonoverlapping stacked distributions).

	Acoustic impedance	Brittleness	Bulk modulus	Density	Lambda rho	P-wave modulus	Poisson's ratio	Shear impedance	Shear modulus	V_P/V_S ratio	Velocity P	Velocity S	Young's modulus
Acoustic impedance	1.000	0.617	0.902	0.643	0.749	0.990	0.206	0.886	0.883	0.206	0.955	0.870	0.937
Brittleness	0.617	1.000	0.289	0.555	0.026	0.573	-0.576	0.895	0.898	-0.576	0.510	0.890	0.830
Bulk modulus	0.902	0.289	1.000	0.411	0.953	0.935	0.549	0.630	0.641	0.549	0.953	0.649	0.734
Density	0.643	0.555	0.411	1.000	0.279	0.549	-0.082	0.707	0.642	-0.082	0.433	0.556	0.638
Lambda rho	0.749	0.026	0.953	0.279	1.000	0.793	0.759	0.400	0.411	0.759	0.831	0.420	0.522
P-wave modulus	0.990	0.573	0.935	0.549	0.793	1.000	0.261	0.848	0.857	0.261	0.986	0.858	0.918
Poisson's ratio	0.206	-0.576	0.549	-0.082	0.759	0.261	1.000	-0.212	-0.208	1.000	0.324	-0.199	-0.086
Shear impedance	0.886	0.895	0.630	0.707	0.400	0.848	-0.212	1.000	0.995	-0.212	0.785	0.977	0.984
Shear modulus	0.883	0.898	0.641	0.642	0.411	0.857	-0.208	0.995	1.000	-0.208	0.807	0.993	0.989
V_P/V_S ratio	0.206	-0.576	0.549	-0.082	0.759	0.261	1.000	-0.212	-0.208	1.000	0.324	-0.199	-0.086
Velocity P	0.955	0.510	0.953	0.433	0.831	0.986	0.324	0.785	0.807	0.324	1.000	0.824	0.874
Velocity S	0.870	0.890	0.649	0.556	0.420	0.858	-0.199	0.977	0.993	-0.199	0.824	1.000	0.983
Young's modulus	0.937	0.830	0.734	0.638	0.522	0.918	-0.086	0.984	0.989	-0.086	0.874	0.983	1.000

	Density	V_P/V_S Ratio	Velocity P
Density	1.000	-0.082	0.433
V_P/V_S ratio	-0.082	1.000	0.324
Velocity P	0.433	0.324	1.000

Abs Correlation	Color
0.900 - 1.000	
0.600 - 0.899	
0.000 - 0.599	

Figure 5. Original pair-wise crosscorrelations of all elastic inversion attributes are shown in the upper table. In the lower left table are the final pair-wise crosscorrelations of the remaining three attributes deemed most informative after multicollinearity analysis at a multiple correlation level of 0.8. Lower right shows color legend of correlation ranges.

the Dean to Strawn interval from the vertical well control. Subdividing the samples into the four categories as defined from classification modeling and then replotting the porosity distribution but coloring by lithotype is shown in Figure 4b. Note that the ranges and shapes of each lithotype are different from considering all the porosity as a single population. Any statistical modeling of porosity from other well log or seismic attributes should include the lithotype categorization.

Seismic attributes and horizon interpretation

Interpretation of the seismic amplitude volume for the Base San Andres and for the Strawn is relatively straightforward because both events are marked by strong acoustic impedance contrasts. Although there is an acoustic impedance contrast at the top of the First Spraberry, it does not manifest itself as a reliable and spatially consistent amplitude event; however, it is relatively obvious in the acoustic impedance volume. The Dean, Upper Wolfcamp, and Lower Wolfcamp events are much more problematic in both the amplitude and acoustic impedance volumes. Is it possible that one of the other volumes or some combination of all of them would assist in interpreting these more spatially variable geologic events? If so, given that the volumes are interrelated by three primary variables — Lamé coefficients (lambda and mu) and density (rho) — which volumes do we use to provide the greatest amount of information with the least amount of data?

To address these questions, a multicollinearity analysis based on forward selection (also known as step up or stepwise) (Whitley et al., 2000) was employed to determine the most informative attributes. The correlation table in Figure 5 contains many off-diagonal crosscorrelation coefficients above 0.9 (shown in green). These same attribute combinations also have rank crosscorrelations above 0.9 (not shown). Therefore, collinearity among multiple attributes does exist. Setting the rejection level on the multiple crosscorrelation to 0.8 returns three attributes as most informative: P-velocity, density, and V_P/V_S ratio. It is interesting that the analysis found three attributes given the inherent interdependencies on the three coefficients previously mentioned.

The integration of the three most informative attribute volumes was accomplished using *k*-means classification, a very easy-to-use algorithm. The only purpose for this classification step is to provide a volume to aid in the interpretation of the horizons not evident on the amplitude volume or on any one of the three elastic attribute volumes. The only required parameter for *k*-means classification is the number of output classes. Because four classes were retained in the DGSOT classification for the well logs, it was decided to use the same number. As hoped, the *k*-means model, when applied, produces a very geologically intriguing volume for interpretation. Figure 6 shows an inline and crossline of the classification results after the application of an additional smoothing step using a simple 3D mix (triangular weighting) to accentuate spatial continuity of class patterns. The color definition of the four classes was determined by assigning the colors from the well log classification using the same ranking of class proportions. This final volume is extremely helpful, along with the original amplitude data (corendered on the crossline in Figure 6 for illustration), in interpreting the Wolfcamp events (also shown in Figure 6). The interpreted time horizons and the well depth picks are used in defining a 3D time-depth model (i.e., velocity model) for converting the seismic attribute volumes to depth. These depth-converted attributes, extracted along all wellbores, both vertical and horizontal, are then used in more rigorous conditioning by the well lithotype and porosity logs for porosity estimation from the seismic attributes.

Lithotype-dependent porosity estimation from seismic attributes

The key to the quantitative analyses from this point forward in the workflow is the ability to "extract" from the seismic attribute volumes along wellbores, both vertical and horizontal. This is made possible by the precise and detailed time-depth conversions from the velocity model constructed from the interpreted time horizons and corresponding depth picks in the wells. The extracted logs from the volumes are sampled at 2.5 ft increments. This sampling rate is between the well logs (0.5 ft) and the seismic depth samples (5 ft). QDA proved a quick but predictive, tool for modeling well lithotypes from the three seismic attributes retained from the multicollinearity analysis. Other supervised classification tools such as random forest, support vector machine, and neural networks might also prove to be good classifiers, and future work will be to investigate them. QDA gave predictions exceeding 60% accuracy for three of the four classes. The exception is understandably the class that is predominantly defined by high resistivity (olive or tan). Fortunately, this lithotype is rare in the Wolfcamp. Figure 7 shows the three input volumes and the resulting facies model after application of the QDA model.

The next step is to define a lithotype-dependent porosity model from the three seismic attributes using a nonlinear regression based on the ACE algorithm. This algorithm generates optimal forward transformations for each independent variable and an inverse transform for the sum of transformed independent variable values. Figure 8 shows the final transforms for P-velocity and for V_P/V_S ratio. Density was found to provide no significant contribution to porosity for each lithotype when P-velocity and V_P/V_S ratio were included. Thus, it was removed from the modeling. In all, 225,000 samples were used in the modeling, and the final prediction correlation is 0.491 with a rank correlation of 0.504. The actual versus predicted results are shown in Figure 8a.

Note that the transforms found are petrophysically sound. Examination of the inverse transform to porosity (Figure 8b) shows that for each lithotype, as the sum of transformed P-velocity and transformed V_P/V_S ratio increases, the prediction of porosity increases. The transform of V_P/V_S ratio (Figure 8c) is linear for each lithotype, albeit with different slopes, meaning that as V_P/V_S ratio increases, its contribution to porosity also increases. This is expected (Castagna et al., 1985). Also, the transforms for P-velocity indicate that as velocity increases, its contribution to porosity decreases (Figure 8d). Again, this well-known relationship (Wyllie et al., 1956) is expected, and dependence upon lithotype is evident.

Using this regression model with the seismic lithotype, P-velocity and V_P/V_S ratio volumes yield the porosity volume in Figure 8e. The resulting porosity trace for each well in the study area can be extracted. These traces for the vertical wells are displayed in

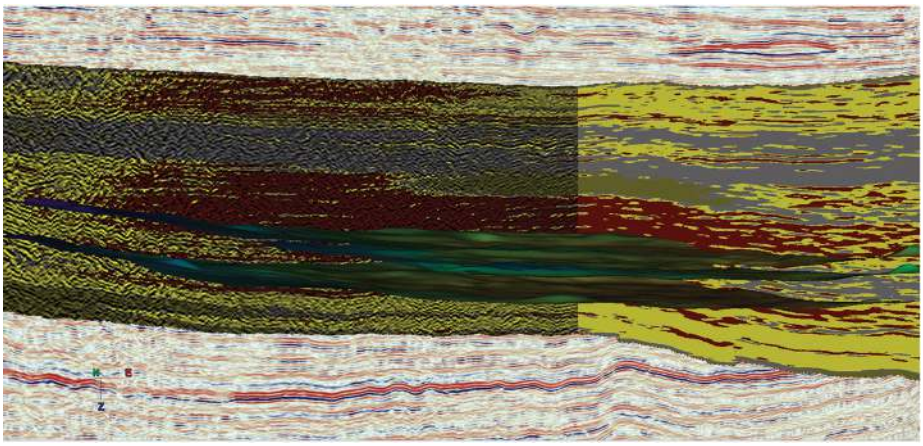

Figure 6. Unsupervised classification volume aided the interpretation of three spatially discontinuous horizons corresponding to three of the well log interpreted picks (Dean, top Wolfcamp, top Lower Wolfcamp). The crossline, shown on the left, is bump map corendered with the seismic amplitude to help guide the interpretation as well.

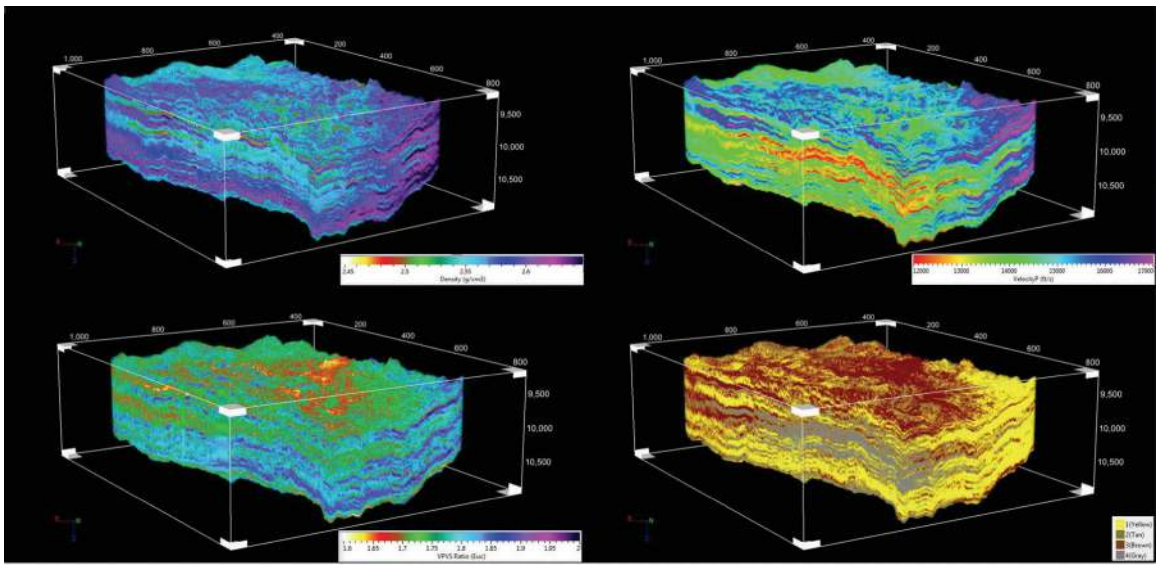

Figure 7. Application of the QDA model using the three elastic property volumes and the well lithotype logs results in the seismic-based lithotype volume shown in the lower right.

Figure 8f. The crosscorrelation of these extracted logs over the Dean to Strawn interval with the actual neutron logs for the same wells is 0.536 with a rank correlation of 0.538.

Porosity estimation using well and seismically derived porosity logs

Now that porosity logs derived from seismic data have been obtained, the question remains as to whether this information improves the prediction of porosity for horizontal wells. To answer this question, it is necessary to run parallel estimation analyses with and without the seismically derived porosity on the vertical wells where actual porosity logs exist. Ordinary kriging and collocated cokriging, respectively, are the two regression methods used to perform these parallel analyses.

The procedure is to remove the neutron porosity log from a given vertical well and then krige an estimated porosity log from surrounding vertical wells that have a neutron porosity log. This process is repeated for all 740 vertical wells. The process is then repeated using the seismically derived porosity log as soft data when a porosity log is estimated using collocated cokriging. In both cases, an octant search for three wells per octant with a maximum search distance of 12,000 ft was used. Further, the variogram model was isotropic with a range of 6000 ft. This range was obtained by modeling the experimental variable on porosity feet over the Wolfcamp interval. This distance also corresponds well with the maximum well-spacing distance for vertical wells where there was any indication of spatial influence on production. No anisotropy was strongly evident, so the variogram was kept isotropic.

One issue that was considered in these estimation analyses deals with well spacing. It is a reasonable assumption that when conditioning well locations are closer to the estimated well location, the influence of the wells might dominate the seismic-based information. In other words, how small is the well spacing before the seismically derived porosity data are no longer significantly contributing? To address this issue, the minimum search distance was varied from 600 to 3000 ft for both estimations, with and without the seismic-based porosity. The correlation coefficient used in the collocated cokriging is 0.450. This is slightly smaller than both the regression modeling results (0.491) and the correlation of the application of the model (0.536). It was determined by cross validating a range of coefficients for each minimum search distance, and the coefficients with the smallest mean absolute error were selected. Nicely, these coefficients clustered around 0.450. Because collocated cokriging is performed in Gaussian space, the correlation could be different than in the original data space.

Figure 9 displays the porosity estimation correlation coefficients for well-only kriging versus well plus seismic-derived porosity soft data. At a minimum search distance of 600 ft, the

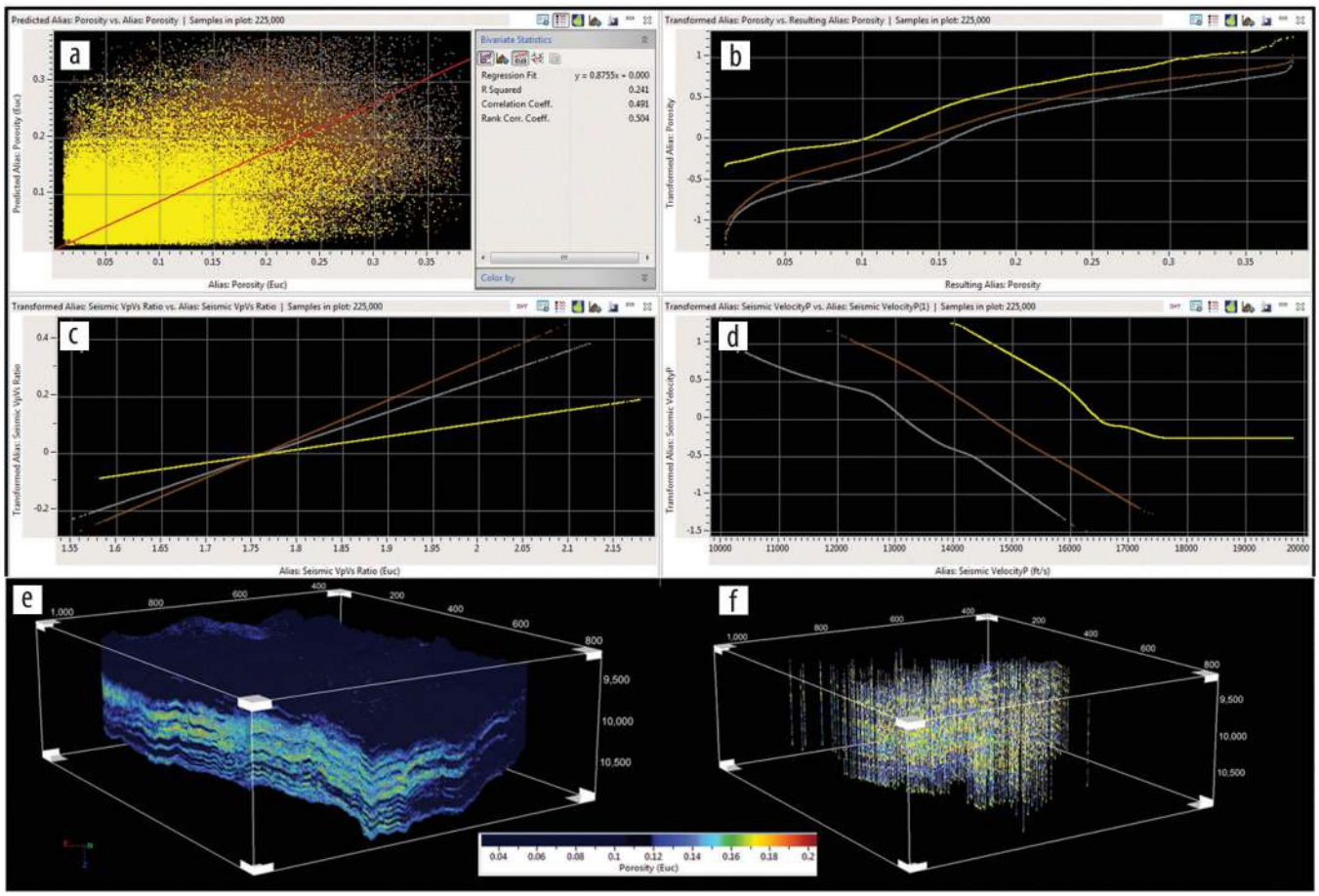

Figure 8. The nonlinear regression model with the prediction crossplot, (a) the transformations for the response variable (neutron porosity), (b) and the independent variables ([c] V_P/V_S ratio and [d] P-velocity). The transformations are colored according to the corresponding lithotype. The (e) porosity volume below the model is the result of its application to elastic volumes and the seismic lithotype volume. (f) Also shown are the extractions along the vertical wellbores.

results are almost identical between well-only and including the seismic-derived porosities: correlations of 0.814 versus 0.818 and error standard deviations of 0.0316 and 0.0312, respectively. As the minimum search distance increases, the separation of the two curves for both correlation and error standard deviation is seen with the collocated cokriging producing better estimation fits than well-only kriging. The greatest difference occurs at a 1500 ft minimum search distance where the contribution of the seismic-derived porosity is a reduction in error standard deviation of nearly 0.2 porosity units. Curiously though, they tend to converge at a minimum search distance of 3000 ft.

Application to horizontal wells and production modeling

It has been shown that seismic data provide a better prediction of porosity as compared with only using existing porosity logs. With that said, how does this translate to a better prediction of expected oil production for these wells? To answer that question, it is necessary to develop a model of production as a function of porosity. Moreover, the model should also consider variables related to drilling and completion, because two wells with the same porosity profiles may have different production profiles as a consequence of those processes in how they were drilled and completed.

As was done for the vertical wells, log traces were extracted from the seismic porosity volume for each of the horizontal wellbores. These traces were used then as soft data for estimating porosity logs by collocated cokriging using the vertical well porosity logs as hard data. A second estimated porosity log was created by kriging using only the well logs. The same variogram and correlation coefficient as specified in the previous analyses was used. No restriction was placed on minimum search distance. Porosity feet was then computed from the resulting estimated porosity logs over the horizontal section of the wellbore. This scalar for each well was included in the production regression modeling.

Cumulative oil production in the first six months was selected as the response variable. Six months was considered long enough to filter out early production variations in the first couple of months

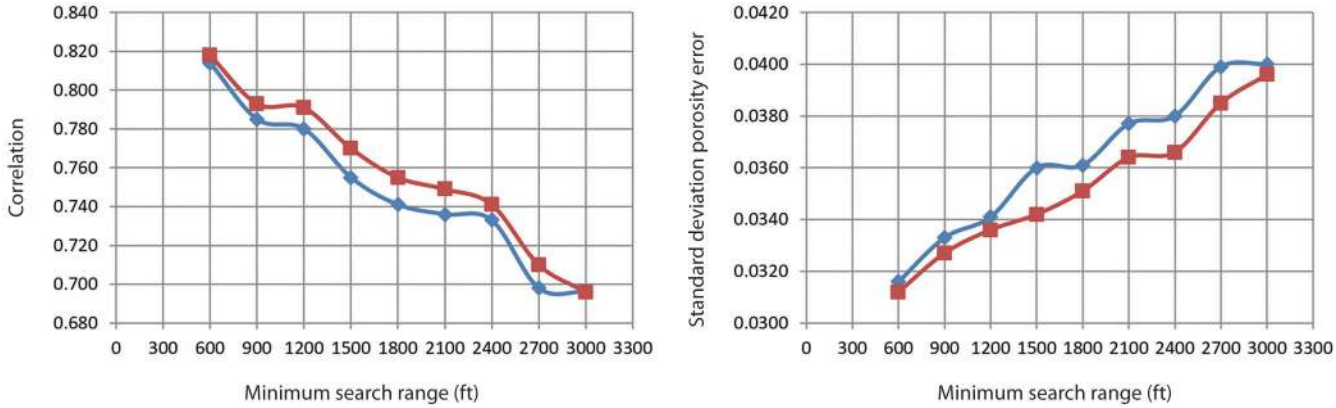

Figure 9. Plots of minimum search range versus estimation correlation and error standard deviation for each of the ordinary kriging with well logs only (blue) and collocated cokriging of the well logs with the seismically derived porosity as soft (red).

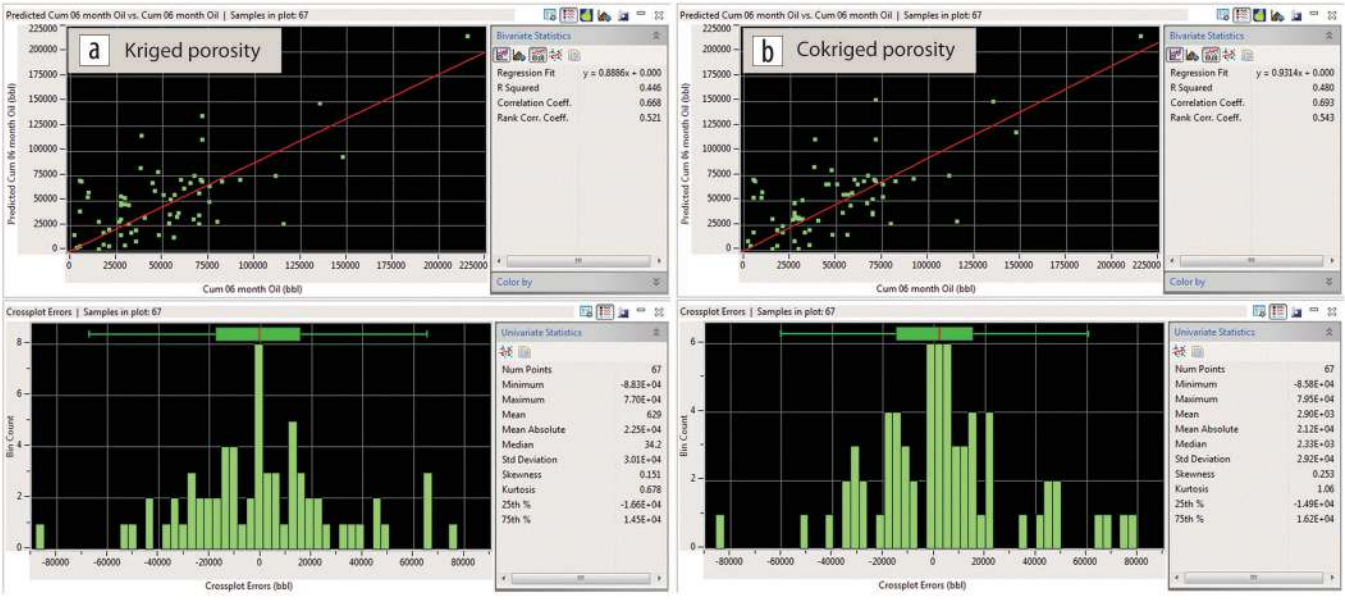

Figure 10. Comparison of prediction of six-month cumulative oil using porosity feet determined from the estimated porosity logs obtained by (a) kriging well porosity logs only and (b) cokriging well porosity logs and using seismically derived porosity as soft data. The upper plots show the prediction crossplots, and the lower plots show the corresponding distribution and statistics of the prediction errors.

as compared with three-month cumulative production. Going to 12 months reduced the number of total wells because many of the horizontal wells haven't been on production that long. Correlation between six months and 12 months is 0.940. Selecting six-month production gives 67 horizontal wells in the Wolfcamp interval within the study area.

Because the sample size of 67 is relatively small, it was decided to limit the final number of independent variables in the model to avoid possible overfitting. In both linear (principal component regression) and nonlinear regression (ACE), three variables consistently ranked as most significant: porosity feet, proppant concentration, and proppant per foot. In all cases, porosity feet, whether from kriging or from collocated cokriging, was the most significant. The ACE models were selected for subsequent use, although all transformations were pseudolinear and monotonic, because they gave higher estimation fits. All the models show that as porosity feet increases, production increases. Similar transforms were found for proppant concentration and proppant per foot. Interestingly, using porosity feet from the estimated porosity logs calculated with the seismic-derived porosity produced slightly better predictions of six-month cumulative oil than porosity feet from the well-only kriged estimated porosity (0.693 versus 0.668, respectively). The final predictions for each are given in the upper plots of Figure 10.

Previously, it was shown that seismic data can improve the prediction of porosity in the interwell space over using just the existing porosity well control. Here it has been shown that this improvement in porosity prediction translates to an improvement in predicting six-month cumulative oil. The mean absolute error of prediction for the 67 wells using the seismic-derived porosity is 21,200 barrels versus 22,500 barrels using only the well porosity logs. The distribution and statistics of the respective errors are given in the lower plots of Figure 10.

Workflow summary

The workflow followed in this study is outlined in graphical form in Figure 11. At least five different multivariate statistical modeling algorithms were employed. *K*-means classification was used to construct a pseudolithotype volume to aid in interpreting time horizons within the Dean and Wolfcamp intervals. This

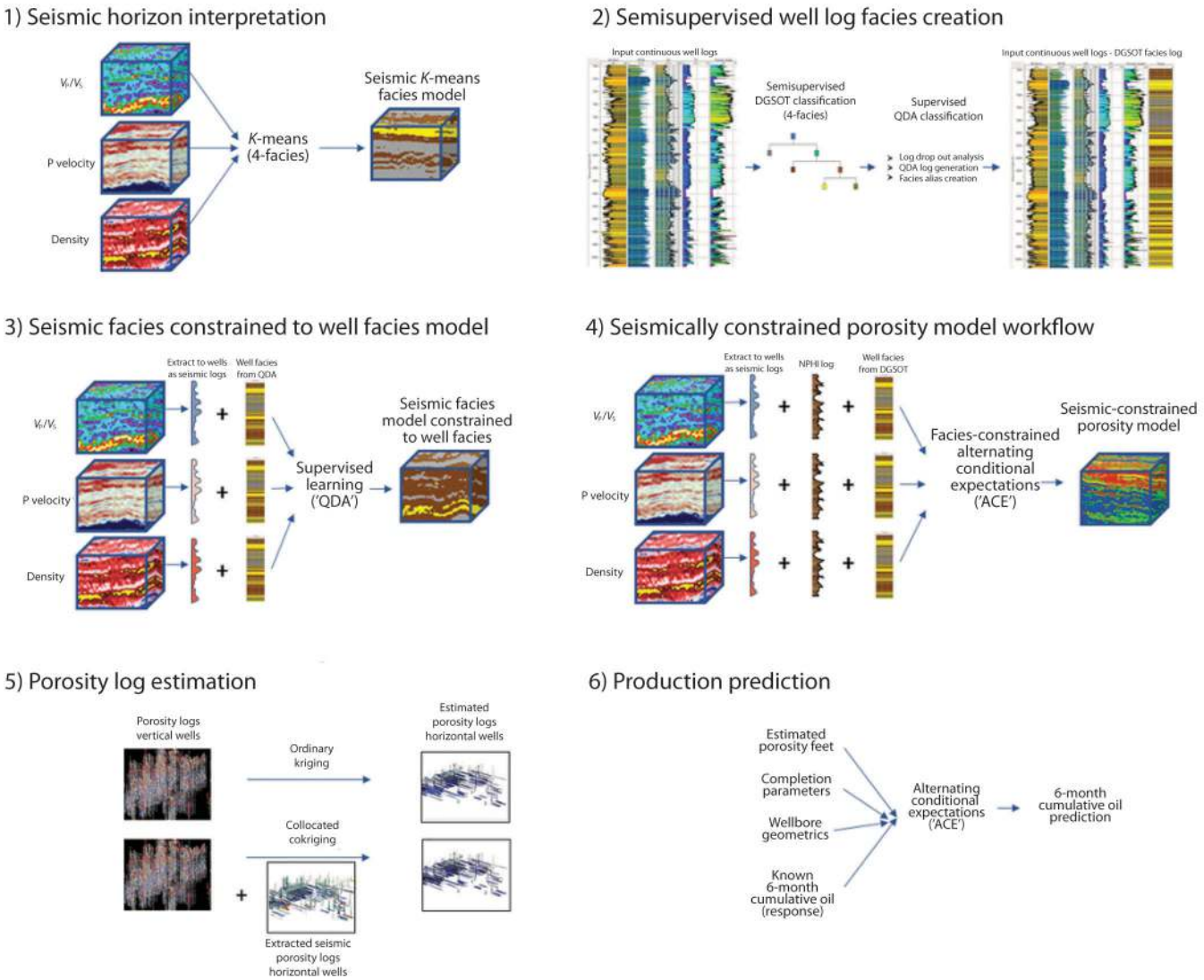

Figure 11. Graphical representation of the workflow presented in the paper of the various multivariate statistical analyses.

volume contains information from the multiple elastic inversion volumes (winnowed down in number by multicollinearity analysis). Lithotype logs were modeled using the DGSOT hierarchical classification approach. Because not all wells had the full suite of logs available, subsequent discriminant analysis (QDA) was applied on the DGSOT results by iteratively removing the missing log types. After depth converting the elastic attribute volumes, a QDA model was derived to convert these elastic volumes to a seismic lithotype volume. Finally, a lithotype-dependent nonlinear regression for porosity using ACE was modeled and applied to the seismic attributes and lithotype volumes. The result is a seismically derived porosity volume.

The next step is to estimate porosity logs on the horizontal wellbores using porosity logs from the vertical wells. Extracting the seismically derived porosity onto the horizontal wellbores gives "soft data" that can be used in collocated cokriging to compute porosity logs on the horizontal wellbores. It was shown that there is an improvement in estimation error over simply estimating porosity from the vertical well logs only using ordinary kriging. The last step in the workflow is to compute porosity feet over the lateral length and to use this scalar as an independent variable for predicting six-month cumulative oil. Additionally, completion and wellbore geometry parameters were analyzed for use in the final model. Again, the ACE algorithm was selected to perform this nonlinear regression.

Conclusions

This study set out to determine whether seismic-based information could be used to improve prediction of production in horizontal wellbores and, if so, whether this improvement could be quantified in barrels. The improvement in porosity prediction was shown to be well-spacing dependent ranging from very little at a 600 ft minimum search distance to nearly 0.2 porosity units at 1500 ft. Multiplying this error reduction over thousands of lateral feet can be substantial. The second part of the question regarding quantifying the improved porosity prediction into actual improvement in oil production prediction was answered as well. For six-month cumulative production, the average improved error in prediction was 1300 barrels per well. At US$50 per barrel, this equates to US$65,000. Typical all-in (permitting, acquisition, and processing) costs for seismic data in the Midland Basin are US$75–80 per acre. Presuming six horizontal wells per section, this translates to a seismic cost of about US$50,000 to get US$390,000 in improved predicted production — a very decent nominal return on investment. TLE

Acknowledgments

The authors thank Fasken Oil and Ranch for providing the seismic data used in this study and the right to publish our results. We also thank Drillinginfo for providing the well and production data. Drillinginfo's Transform software was used for the seismic and well interpretations, data analyses, and multivariate statistical modeling. Finally, we thank Patrick Rutty for his time and expertise in the review of this paper and for his helpful critiques.

Corresponding author: bill.bashore@drillinginfo.com

References

Baumgardner, R. W. Jr., H. S. Hamlin, and H. D. Rowe, 2014, High-resolution core studies of Wolfcamp/Leonard basinal facies, southern Midland Basin, Texas: Search and Discovery, 10607.

Breiman, L., and J. H. Friedman, 1985, Estimating optimal transformations for multiple regression and correlation: Journal of the American Statistical Association, **80**, no. 391, 580–598, https://doi.org/10.1080/01621459.1985.10478157.

Castagna, J. P., M. L. Batzle, and R. L. Eastwood, 1985, Relationship between compressional-wave and shear-wave velocities in clastic silicate rocks: Geophysics, **50**, no. 4, 571–581, https://doi.org/10.1190/1.1441933.

Hastie T., R. Tibshirani, and J. Friedman, 2008, The elements of statistical learning: Springer.

Isaaks, E. H., and R. M. Srivastava, 1989, An introduction to applied geostatistics: Oxford University Press.

Luo, F., L. Khan, F. Bastani, I.-L. Yen, and J. Zhou, 2004, A dynamically growing self-organizing tree (DGSOT) for hierarchical clustering gene expression profiles: Bioinformatics (Oxford, England), **20**, no. 16, 2605–2617, https://doi.org/10.1093/bioinformatics/bth292.

MacQueen, J. B., 1967, Some methods for classification and analysis of multivariate observations: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability: University of California Press, 281–297.

Whitley, D. C., M. G. Ford, and D. J. Livingstone, 2000, Unsupervised forward selection: A method for eliminating redundant variables: Journal of Chemical Information and Computer Sciences, **40**, no. 5, 1160–1168, https://doi.org/10.1021/ci000384c.

Wyllie, M. R., A. R. Gregory, and G. H. F. Gardner, 1956, Elastic wave velocities in heterogeneous and porous media: Geophysics, **21**, no. 1, 41–70, https://doi.org/10.1190/1.1438217.

Xu, W., T. Tran, M. Srivastava, and G. Journel, 1992, Integrating seismic data in reservoir modeling: The collocated cokriging alternative: Presented at SPE Annual Technical Conference and Exhibition.