geoconvention
Calgary • Canada • May 13-17 2019

A Data Driven Method For Sweet Spot Identification In Shale Plays Using Well Log Data

Bianca Zadrozny, Matthias Kormaksson, and Marcos R. Vieira.
IBM Research - Brazil

Summary

In recent years, interest in shale plays has grown substantially due to horizontal drilling and hydraulic fracturing techniques. Special interest is in shale plays previously exhausted with vertical wells that are believed to still have high potential if developed with horizontal wells. However, with drilling costs at an all-time high, choosing the right locations for new wells is a crucial issue. Therefore, identifying so called "sweet spots" with high potential for oil and gas is of great importance for oil companies worldwide.

Well log data from millions of wells drilled using conventional techniques since the industry's inception is available and generally not used. In this paper, we propose a data analytical solution that 1) automatically extracts simple features from complex and high-dimensional well log curves arising from vertical wells using functional Principal Component Analysis (fPCA), and 2) builds models that predict sweet spots in shale plays by correlating extracted features with production data from horizontal wells. Our solution builds predictive models for production directly using previous production data and petrophysical well logs alone, thus circumventing time consuming and expensive geological analysis.

In this paper we present a method involving 3 analytical components: Functional Principal Components Analysis (FPCA), Kriging, and Multiple Linear Regression. Since both kriging and multiple linear regression are well-established methods in Geostatistics we simply point to the references Cressie, N. (1993), for kriging, and Cohen, J. et al. (2003), for multiple linear regression. Using petrophysical well logs and geological core, the shale properties can be evaluated around the boreholes of drilled vertical wells as with reference to Passey, Q.R. et al. (1990).

The methodology was applied to well log data from 2020 vertical wells and production data from 702 horizontal wells in a single field. In predicting whether a given horizontal well has production above a given high-production threshold, we were able to achieve an accuracy of 90% for gas wells, and 71% for oil wells.

Theory

Regression in conjunction with interpolation (Regression-Kriging) is a well-known approach that can be used to correlate well log curves with production data. However, this method is only applicable once the set of (one-dimensional) predictors and the corresponding regression model have been defined. Summary statistics such as means, maximum or minimum peak heights are obvious candidates but are too simple to capture all the relevant features from the well logs. Instead, we extract a set of one-dimensional features from each of the well log curves, facilitating simple 2D interpolations as opposed to more difficult 3D interpolations. This is particularly advantageous in situations where seismic data are not available and 3D interpolations are challenging. Finally, by regressing previous production data from horizontal wells on the extracted features we show that it is possible to predict production at new

locations directly. We have implemented our method using the R statistical package. We tested it using well log data from 2020 vertical wells and production data from 702 horizontal wells in a single field. For gas we get an accuracy of 90% at predicting whether a given horizontal well has production above a given high-production threshold, while for oil, we get an accuracy of 71%.

The main novelty of our method is the systematic extraction of one-dimensional predictors using a statistically robust method called functional Principal Components Analysis (fPCA). To the best of our knowledge this is the first time fPCA is applied to well log curves in the context of oil and gas exploration.

Geological and petrophysical studies have shown that identification of shale plays sweet spots involves finding locations with the right combination of thickness, TOC (total organic carbon), maturity, porosity, high gas-in-place and fracability (Liu, Y. (2013)). Using petrophysical well logs and geological core analysis (see for example the DlogR method of Passey, Q.R. et al. (1990)), these shale properties can be evaluated around the boreholes of drilled vertical wells. Then, traditional geostatistical methods like the ones in Deutsch, C.V. (2002), can be applied to interpolate the properties from a given borehole to a new proposed drilling site.

The main drawback of such a workflow is that geological core analysis (such as in Passey, Q.R. et al. (1990)) is very time consuming and expensive as it entails one-at-a-time analysis of each well by a geological expert. Secondly, propagation of the derived properties onto new borehole locations is not trivial as the interpolation is not just two-dimensional along the plane but also along the depth of the shale play. In presence of seismic data, which possess high lateral resolution, efficient 3D-interpolation algorithms, such as SVM, do exist (see, for example, Liu, Y and Sacchi, M.D. (2002)). However, in absence of seismic data, which is often the case, the quality of such 3D interpolation decreases, in particular when boreholes are sparse with large distances between boreholes. The final drawback is that once shale properties have been derived at new locations, the geological expert will need to spend additional time interpreting and analyzing the results.

Regression in conjunction with interpolation (Regression-Kriging) is a well-known analytical approach in Geostatistics (see, for example, Hengl, T. et al (2007)) that can be used to correlate well log curves with production data. However, this method is only applicable once the set of (one-dimensional) predictors and the corresponding regression model have been well defined. In the literature there is no clear systematic way to extract such predictors from high-dimensional and complex well log curves. Summary statistics such as means, maximum or minimum peak heights are obvious candidates but are too simple to capture all the relevant features from the well log curves. Therefore, there is a need for a systematic and principled framework for extracting meaningful features from well logs that correlate well with production data.

With the numbers of available well log data in the tens of thousands or even hundreds of thousands in some fields, the need for an automatic, efficient, and robust method for identifying sweet spots in shale plays is very high. The current geological methods are too time consuming and the analytical methods available lack a principled framework for extracting simple yet meaningful features from complex and high-dimensional well log curves. The method presented in this paper aims at meeting the above need using state of the art statistical techniques.

Our method builds predictive models for production directly using previous production data and petrophysical well logs alone, thus circumventing time consuming and expensive geological core analysis. Our method extracts simple one-dimensional features from each of the well log curves, thus facilitating simple 2D interpolations as opposed to more difficult 3D interpolations. This is particularly advantageous in situations where seismic data are not available and 3D interpolations are challenging. Finally, by regressing previous production data from horizontal wells on the extracted features we can

predict production at new locations directly. This has a clear advantage over predicting only the shale properties that may or may not correlate well with production and then later need to be analyzed by geological experts.

One of the main novelties of our method is the systematic extraction of one-dimensional predictors using a statistically robust method called “functional Principal Components Analysis (fPCA)”. To the best of our knowledge this is the first time fPCA is applied to well log curves in the context of oil and gas exploration.

Figure 1 illustrates the proposed workflow for identifying sweet spots in a given shale play. The shale play consists of several vertical wells with petrophysical well logs along with production data from horizontal wells. Note that the horizontal wells are not located at the same coordinates as the vertical wells. The analytical workflow consists of a model building phase and a prediction phase.

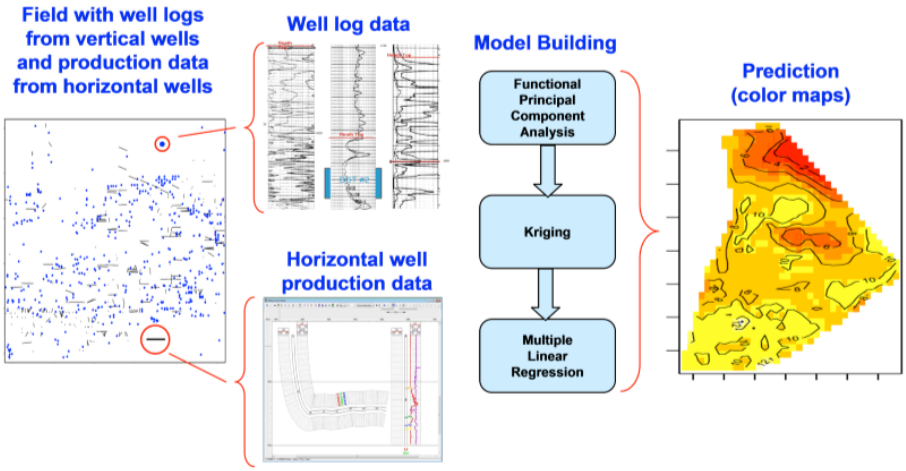

Figure 1: On the left we see an image of a shale play with vertical wells along with horizontal wells. At each vertical well we have several well logs (displayed middle left, upper panel) that consist of petrophysical measurements along the depths of the wells. At each of the horizontal wells (displayed middle left, lower panel) we have production data available (oil and gas production). These data are used to build predictive models through the following three steps: 1) Functional Principal Component Analysis, 2) Kriging, and 3) Multiple Linear Regression.

Production prediction can then be performed on a grid across the whole field, thus facilitating color map visualization of where sweet spots are located

The three main components of the model building phase are

1. Functional Principal Component Analysis: Used to extract one-dimensional features (principal component scores) from the petrophysical well log curves. The number K of principal components to be extracted can be chosen according to how much proportion of the variation in the curves is explained by the first K components.
2. Kriging: Spatial interpolation of the extracted features at the vertical wells onto the (center) coordinates of the horizontal wells.
3. Multiple Linear Regression: By regressing production at the horizontal wells onto the interpolated principal component scores we build predictive models that can be used to predict production at new sites.

Once the model building is completed, prediction can be performed at any given location in the field. First, the extracted principal components from step 1 above are interpolated (using the kriging from step

2) onto a new location. Then the estimated regression model from step 3 is used to predict the production. This prediction can be performed at any individual location or on a grid across the whole field. The latter can be visualized with color maps that provide the user with a visual representation of where exactly the sweet spots are located.

Functional Principal Component Analysis

The petrophysical well log curves (in this paper: Sonic DT, log(Resistivity), log(Gamma Rays)) consist of discrete observations x_j observed at (high resolution) discrete depths d_j , j = 1,...,p, where p is the number of depth units observed. Since the number of depth units can be in the thousands and since the scatterplots (d_j, x_j) demonstrate smooth (non-noisy) functional relationships, these data can be represented as functions. This facilitates the use of Functional Data Analysis, Ramsey, J. and Silverman, B.W. (2005), which can be used to extract (functional) principal components from the curves. The functions (for a particular petrophysical property) are represented using linear expansion

$$x(d) = \sum_{k=1}^{K} c_k \varphi_k(d)$$

where $\varphi_k(\cdot)$ represent known basis functions (e.g., Fourier Bases, B-spline bases). The coefficients c_k are calculated as part of the preprocessing of the data. This leads to an efficient computational representation of the discrete data as functional objects, where the data for each well have been reduced from thousands of measurements to only K << p coefficients c_k and known basis functions (φ_k). This can readily be achieved using the "fda" R-package publicly available on the R project website: *http://cran.r-project.org/*.

Assume now that the well logs have all been converted to functional objects $x_1(\cdot),\ldots, x_n(\cdot)$, where n denotes the number of wells in the field. Functional Principal Component Analysis (fPCA) involves the following steps:

- Find principal component weight function $\xi_1(\cdot)$ for which the principal component scores

$$\rho_{1i} = \int \xi_1(t) x_1(t) dt$$

  Maximize $\sum_i \rho_{1i}^2$ subject to

$$\int \xi_1^2(t) dt = 1$$

- Next, compute weight function $\xi_2(\cdot)$ and principal component scores maximizing $\sum_i \rho_{2i}^2$, subject to $\int \xi_2^2(t) dt = 1$ and additionally $\int \xi_1(t)\xi_2(t) dt = 0$.

- Continue calculating orthogonal weight functions until desired number of principal component scores have been calculated

The computation of functional principal component scores can be performed using the "fda" R-package.

Method

In Figure 2 we see a map of the shale play, demonstrating geographical coordinates of the underlying wells. The shale play consists of several vertical wells with petrophysical well logs along with horizontal wells containing production data. Note that the horizontal wells are not located at the same coordinates as the vertical wells. The grey and blue dots represent conventional vertical wells where petrophysical well log data are available. The grey dots represent vertical wells that we exclude from our analysis because they do not contain the desired petrophysical data, which are: Sonic DT, log(Resistivity), and

log(Gamma Ray). The blue dots represent vertical wells that have all these data available and are subsequently used in all analyses. The choice of these three particular petrophysical properties is inspired by the DlogR method, see e.g. Q.R. et al. (1990). The green and red dots represent the start and end locations, respectively, of the horizontal wells.

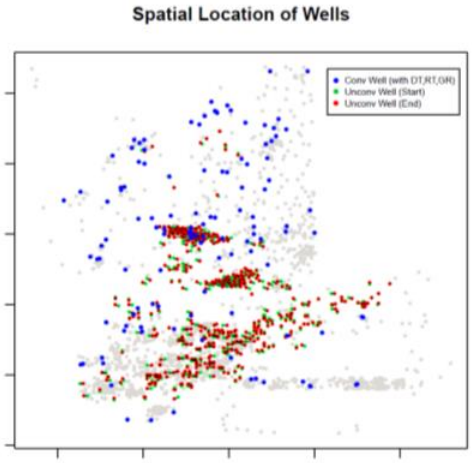

Figure 2: Shale Play demonstrating geographical location [x: latitude vs. y: longitude] of (a) conventional vertical wells in grey (all) and blue (those that have DT, RT, and GR well log data), (b) unconventional horizontal wells in red/green.

Functional Principal Component Analysis

Let *H* and *V* denote the set of horizontal and vertical wells respectively. Let *P* denote the set of petrophysical well log properties. The well log data comprise of the petrophysical curves, f_{ij}(*depth*), at vertical wells i ∈ *V* and petrophysical property *j* ∈ *P*. In this paper *P* denotes the set {Sonic DT, log(Resistivity), log(Gamma Ray)}. In Figure 3 we see a plot of the observed petrophysical curves as functions of depth. For each of the three petrophysical properties we calculate the functional principal component scores (See section 2) x_{ijk} , for *k* =1,…, K_j, where K_j can either be chosen in advance or selected in such a way that the explained variance exceeds certain thresholds. Let $K = \sum_{j\in \mathrm{P}} K_j$ and for ease of notation we omit the *j* subscript and write the complete set of scores (across all petrophysical properties) as x_{i1},..., x_{iK}. Note that the principal component scores are calculated at all vertical wells *i* ∈ *V*.

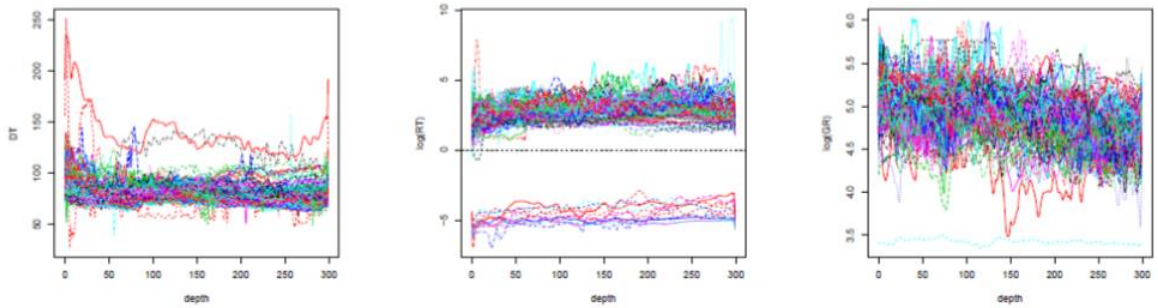

Figure 3: Plot of the petrophysical well log curves as observed at the vertical wells.
Left: Sonic DT, *Middle*: log(Resistivity), *Right*: log(Gamma Ray).
Each color represents different wells to facilitate visualization.

Kriging

We now take the above principal component scores, as calculated at vertical wells, and perform kriging to obtain interpolated scores at the horizontal wells. More specifically, for each horizontal well *i*'∈ *H* and principal component score *k* we calculate the weighted average

$$\hat{x}_{i'k} = \sum_{i \in V} weight\ (dist\ _{i,i'}) \cdot\ x_{i,k}\ ,$$

where $dist\ _{i,i'}$ denotes the distance between vertical well $i \in V$ and horizontal well $i' \in H$. The kriging weights are obtained by specifying the covariance function parametrically (e.g. Matérn, exponential, spherical) and using the Maximum Likelihood criteria to estimate the corresponding parameters.

Multiple Linear Regression

Let $y_{i'}$ denote the production (e.g. 6-month oil/gas production) at horizontal well $i' \in H$. We now regress the interpolated principal component scores $\hat{x}_{i'1}, \ldots, \hat{x}_{i'K}$ on $y_{i'}$ through the multiple linear regression model:

$$y_{i'} = \beta_0 + \sum_{k=1}^{K} \hat{x}_{i'k}\beta_k + \varepsilon_{i'},$$

where $\varepsilon_{i'}$ is independent and normally distributed with mean zero and constant variance. We fit the above model using Least Squares and obtain parameter estimates $\hat{\beta}$.

Prediction

Assume we now wish to drill a new horizontal well $i_0 \in H$ at a location whose spatial coordinates are known. We then proceed to calculate the interpolated principal component scores at i_0 through $\hat{x}_{i'k} = \sum_{i \in V} weight\ (dist\ _{i,i'}) \cdot\ x_{i,k}$,as described in the Kriging section. We then use the parameter estimates obtained from our Multiple Linear Regression model to form the predicted production at the new horizontal well:

$$\hat{y}_{i_0} = \hat{\beta}_0 + \sum_{k=1}^{K} \hat{x}_{i_0 k}\hat{\beta}_k.$$

Data Processing

For our experiments we employed a dataset consisting of 2,020 vertical wells containing up to 26 petrophysical well log properties (e.g., Calipur, Bulk Density, Deep Resistivity, Neutron Porosity, Sonic DT, Resistivity, Gamma Ray, etc.) within a region of 40,000 square miles. The resolution of our raw well log data was .50ft and for each well log we observed measurements down to a maximum depth of 11,000ft. We also had access to another source of data containing the start and end depths of the shale formation of interest for each well log. Since our analysis involved only the relevant parts of the wells pertaining to the shale play we had to filter the well logs accordingly. The average thickness of the formation was approximately 250ft. We further restricted our analysis only to those wells that contained all of the following three indicators of carbon matter content: Sonic DT, log(Resistivity) and log(Gamma Rays) [see Q.R. et al. (1990)]. We see scatterplots of these three petrophysical properties as functions of depth (starting from beginning of shale formation) in Figure 3.

Once we had pre-processed the well log data in the above manner, we next determined the functional representation of each curve through the basis function expansion given in (1). We used Fourier basis functions to represent the underlying curves and the number of basis functions was chosen to be $N_d - 2$ where N_d represents the number of depth measurements (within the shale formation) for each curve. The above representation was calculated using the fda R-package. The resulting functional data was used in all experimental analyses. In particular, they were used to calculate the functional principal components as described in section 2. The production data consisted of 702 horizontal wells and we defined the oil/gas production at each well as the total production during the first 6 months of production.

Results

A prediction analysis was performed on a shale play consisting of 2020 vertical wells and production data from 702 horizontal wells. We only focused on those vertical wells that had all three desired petrophysical properties: Sonic DT, log(Resistivity), and log(Gamma Ray).

Sweet Spot Identification

At the vertical wells we extracted for each petrophysical property the 10 first principal component scores, resulting in a total of 30 features. We then interpolated these 30 features onto the coordinates of the 702 horizontal wells using (2). Finally, we regressed the 30 features on 6-month oil and gas production at the 702 horizontal wells using (3). We used backward stepwise regression (and the AIC model selection criteria) to reduce the initial number of 30 features and achieve a more parsimonious model. We then fit the resulting linear model to obtain least squares parameter estimates. Finally, using (4) we performed prediction at unobserved locations (i.e. locations for potential new horizontal wells) across a grid inside the convex hull of our data. In Figure 4 we see color maps of these predicted oil (left) and gas (right) productions. These color maps can help reservoir engineers identify sweet spots, i.e. locations with high predicted production.

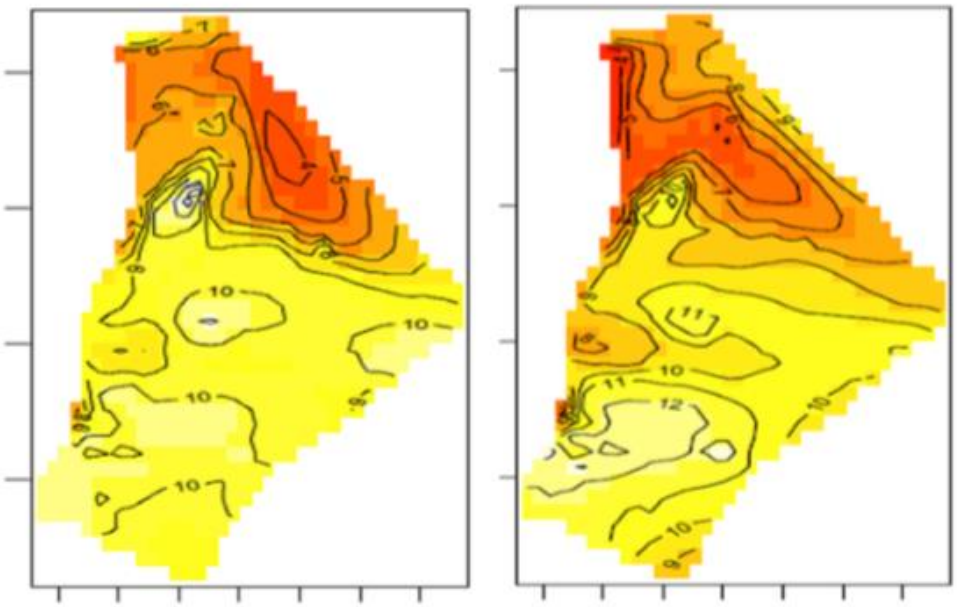

Figure 4. Color maps of predicted (6-month) production at a grid of locations within the convex hull of the data. Left panel shows predicted oil production, Right panel shows predicted gas production.

Cross-Validation

In the oil and gas industry a particular well location is called a sweet spot if its production exceeds a given threshold. Therefore, in order to get a sense of how well our method would work in practice we performed a leave-one-out cross validation on the above data. As before we extracted 30 principal component scores and interpolated them onto the 702 horizontal well locations. For each horizontal well $i_0 \in H$ we then 1) estimated the regression model using all remaining wells, i.e. $V \setminus \{ i_0\}$, and then 2) predicted the production at i_0 using the obtained parameter estimates. In Figure 5 we see scatterplots of the predicted versus observed log(production) for oil (left) and gas (right). We declare wells with predicted production above a given threshold “sweet spots” and then compare to actual sweet spots obtained by thresholding the observed production values. To assess the performance we then calculated the accuracy for a given “sweet-spot” threshold. To be more specific the accuracy is defined as (TP+TN)/(# of wells), where TP=“True Positives” represents the number of correctly specified sweet spots and TN=“True Negatives” represents the number of correctly specified non-“sweet spots”.

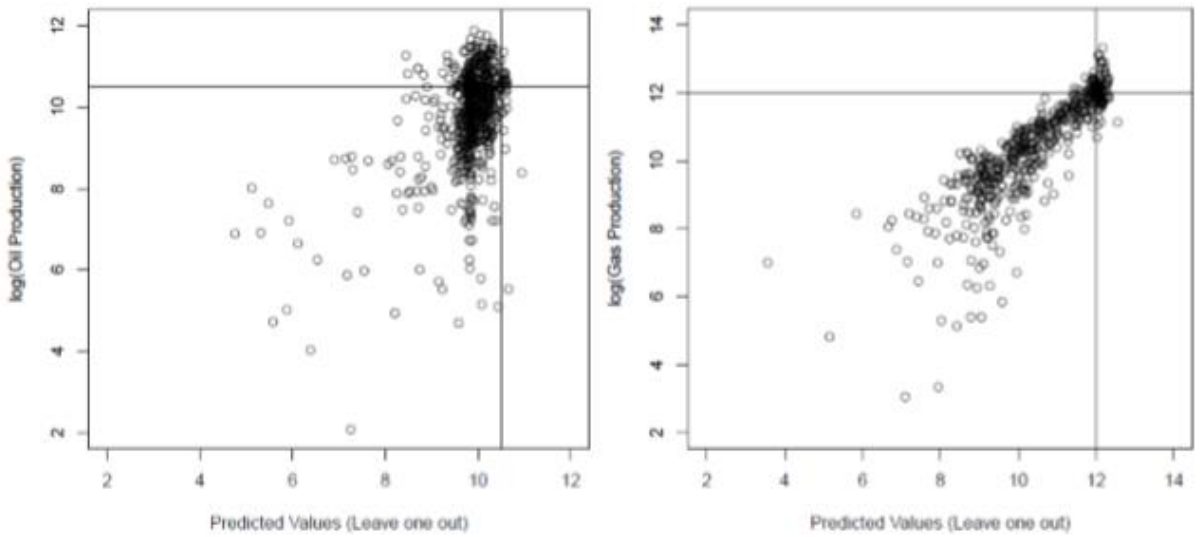

Figure 5. Scatterplots of predicted versus observed log(production) using leave-one-out cross validation. Left plot shows results for oil while the right plot shows those for gas.

Figure 6 shows the Accuracy as a function of the sweet spot threshold for oil prediction (left) and gas prediction (right). For a sweet spot threshold of 40,000 barrels we obtained an accuracy of 0.71 for oil predictions and 0.9 for gas predictions.

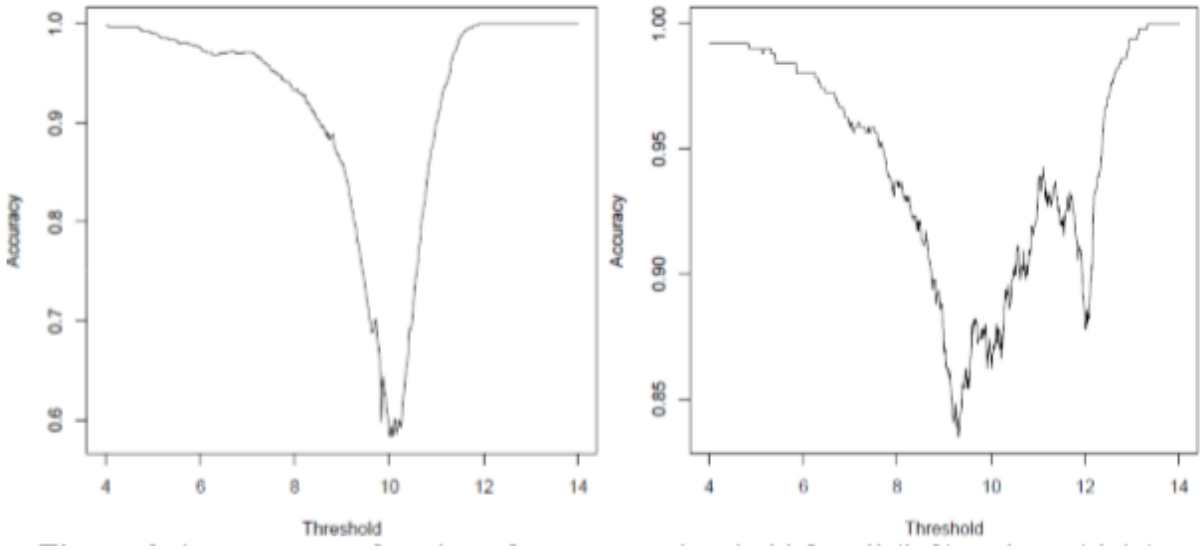

Figure 6. Accuracy as a function of sweet spot threshold for oil (left) and gas (right).

Conclusions

In this paper we proposed a data driven and automatic method for identifying sweet spots in shale plays using well log data from vertical wells and production data from horizontal wells. The main contribution is the systematic (and fully data-driven) extraction of one-dimensional features from the complex functional well logs. Instead of relying on ad-hoc summary statistics, chosen manually, we use a statistically robust method (fPCA) that automatically extracts relevant features from the functions. We demonstrated the method on a real data set from a shale play consisting of 2020 vertical wells and 702 horizontal wells and obtained promising results in terms of accuracy. Although we only explored three petrophysical properties (Sonic DT, Resistivity, and Gamma Ray) in this paper, theoretically the method may be applied to any number and any type of petrophysical well logs provided they are smooth and functional in nature.

References

Cressie, N. (1993). Statistics for Spatial Data, Wiley Series in Probability and Mathematical Statistics.

Cohen, J., Cohen, P., West, S.G. , and Aiken, L.S. (2003). "*Applied multiple regression/correlation analysis for the behavioral sciences", 2nd edition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Deutsch, C.V. (2002). "*Geostatistical Reservoir Modeling*." Oxford University Press.

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G. (2007). "*About regression-kriging: From equations to case studies."* Computers and Geosciences, v. 33, 1301-1315.

Liu, Y. (2013). "*Workflows for Sweet Spots Identification in Shale Plays Using Seismic Inversion and Well Logs*". GeoConvention 2013: Integration, Calgary, Alberta, Canada.

Liu, Y. and Sacchi, M.D. (2003). "*Propagation of borehole derived properties via a Support Vector Machine (SVM)*." CSEG Recorder, 28, 54-58.

Passey, Q.R., Creaney, S., Kulla, J.B., Moretti, F.J. and Stroud, J.D. (1990). "*A Practical Model for Organ Richness from Porosity and Resistivity Logs*." The American Association of Petroleum Geologists Bulletin, v. 74, No. 12.

Ramsey, J., and Silverman, B.W. (2005). "*Functional Data Analysis*", 2nd edition, Springer.