Well performance predictions from geologic, petrophysical and completions-related parameters using generative topographic mapping: A field case study
Atish Roy*, Tao Zhao, University of Oklahoma; Vikram Jayaram, Global Geophysical Services, Inc.; and Deepak Devegowda, University of Oklahoma

Summary

This paper explores and discusses the use of generative topographic mapping (GTM) in estimating estimated ultimate recovery (EUR) from geologic, petrophysical, and completion parameters, and further distinguishing poorly performing wells from the high productivity wells and simultaneously to quantify the ranges of the explanatory reservoir and completion related parameters that dictate well performance. Using an application on a field dataset from a prominent shale gas play with over a hundred horizontal wells, we demonstrate the advantage of GTM when employed in predictive mode with non-Gaussian datasets.

Introduction

Analyzing large datasets with several explanatory variables can often be challenging. The complexities arise due to the non-Gaussian distribution of the variables that comprise the dataset, strong and weak interactions between the explanatory variables that may not be apparent from simple univariate cross-plots and the need for accurate or reasonable predictive capabilities. Additionally, the underlying functional form of the relationship between the various variables may not be known and may necessitate a trial-and-error approach to uncover these relationships. An example of such a dataset is related to well performance analysis as a function of several explanatory variables related to reservoir properties and completion design parameters. Considering the rapid pace of development in unconventional plays where the cost of a single horizontal well with a multiple stage hydraulic fracture treatment can be in the vicinity of a few million dollars, it is extremely important and crucial to determine the key variables influencing well performance in order to minimize the capital outlays associated with reservoir development activity.

In this paper explore the use of a recently developed machine learning technique known as generative topographic mapping (GTM) (Bishop, 1998) which is a more probabilistic approach of self-organizing maps (SOM) and overcomes the shortcomings of SOM and helps in data classification. We use GTM to classify wells based on their performance and identify the key petrophysical, geologic and completion-related parameters that explain the variability in the observed well performances. The dataset comprises of 137 horizontal wells from a prominent shale play in the continental US.

Roy et al. (2013) implemented GTM to handle much larger input seismic attribute volumes to perform multi-attribute seismic facies classification of a carbonate conglomerate oil field in the Veracruz Basin of southern Mexico. Our workflow in this paper is focused on applying the GTM technique to generate classification of the horizontal well completion parameters. We also exploit the fact that GTM is deeply rooted in probability theory and can be utilized in predictive mode. We verify the classification scheme using GTM to a set of wells that were not utilized for the classification and demonstrate the power and utility of this approach.

Data Description and Methodology

The dataset employed in our study comprises EUR data for 137 horizontal wells from a prominent shale play in the US. The relative locations of the wells and their respective EUR values are provided in Figure 1. The EUR values are scaled from a value of 0 to 1. The dataset includes 14 geologic and completion-related parameters for each of the wells that are likely to impact the EUR (Table 1).

Although from an engineering standpoint, some of these variables may have only a limited influence on well performance or may be related (for example, total volume of proppant and total clean volume of sand), in this work we apply several clustering techniques without utilizing any priori engineering judgment to the dataset. Prior to clustering these wells according to their EUR’s, all explanatory variables are normalized by z-score algorithm to remove any bias in any of the variables. We then use GTM to calculate and project the posterior probabilities of the data in a 2D lower dimensional latent space. The weighted sum of the EUR for each well is calculated to form the most likely EUR map. Then we validate our approach by estimating EUR’s from the GTM property map and comparing the predicted values against the true EUR’s for a set of 8 wells that constitute our validation dataset. Redundant parameters were then removed through parameter sensitivity analysis to only consider the more sensitive parameters impacting well EURs. The summarized workflow is illustrated in Figure 2.
Well performance predictions using generative topographic mapping

Figure 1: Well locations for the unconventional shale play with an area of approximately 1000 km². The wells are color-coded according to their corresponding EUR values which are scaled between 0 and 1.

<table>
<thead>
<tr>
<th>Parameters used to correlate with EUR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Total clean volume of sand</td>
<td>8 Total perforations</td>
</tr>
<tr>
<td>2 Total proppant volume</td>
<td>9 Contour permeability</td>
</tr>
<tr>
<td>3 Total non-100 mesh sand</td>
<td>10 Number of perforation clusters</td>
</tr>
<tr>
<td>4 Average proppant concentration</td>
<td>11 Average treating rate</td>
</tr>
<tr>
<td>5 Cluster spacing</td>
<td>12 Thickness of the formation</td>
</tr>
<tr>
<td>6 Number of hydraulic fracture stages</td>
<td>13 Porosity</td>
</tr>
<tr>
<td>7 Total 100 Mesh sand</td>
<td>14 Total perforation length</td>
</tr>
</tbody>
</table>

Table 1. Geologic and completion-related parameters used to estimate EUR.

Figure 2. The flowchart for EUR prediction scheme utilizing GTM.

Application

We wish to determine which combination of completion process and reservoir properties are correlated with high and low expected ultimate recovery (EUR). We performed 2 case studies with GTM; the first case included all of the 14 variables, while the second case worked with a reduced subset of the 14 variables where some redundant parameters expected to have a minimal influence to the EUR prediction were removed.

For case 1, all the 14 horizontal well variables are normalized using a z-score algorithm to minimize any bias due to units of measure. The 2D latent space is uniformly sampled using 144 points and forms a square grid of 12 x 12 points. These 144 latent space variables are mapped into the N=14 dimensional data-space using 16 Gaussian basis functions with equal variance. After GTM training the manifold in the data space will be stretched in regions of low-data density and compressed in the region of high data density. With the trained GTM model parameters, the posterior probabilities of the data-vectors are calculated using Bayes’ theorem. We then use these posterior probabilities and project them onto the 2D latent space to form a mean distribution map (Figure 3). After the final iteration the mean projections are more distributed in the latent space with less overlap. Finally projected points are colored by the scaled EUR values (Figure 4). Once trained we validate the clustering using the estimated EURs from the GTM property map and compare with the true EURs for the 8 wells not used in training (Figure 5a and 5b).

Figure 3. The mean posterior distribution map of the “responsibilities” of the data (case 1) in the 2D latent space. The PDF of a representative well vector projected onto the latent space showing the mean projection as a weighted average of the posterior distribution values and will in general fall in between neighboring values of $u_k$. (a) Initial and (e) final distribution of the posterior mean projections of all 137 well data onto the latent space after 200 iterations. Note that after 200 iterations the mean projections are more distributed in the latent space with less overlap.
Well performance predictions using generative topographic mapping

Figure 4. EUR prediction through GTM modeling. (a) The posterior probability of the data-vector from the \(n\)th well. (b) The EUR for the \(n\)th well, \(E_n\), is multiplied with the posterior projection values onto 2D latent space in (a). The result gives an EUR map for 1 well. (c) Then, we can formulate a weighted sum of the EUR at each grid point \(k\) in the latent space for all the wells (given by Equation 1) and form the EUR “map” over the latent space.

Figure 5. (a) EUR property map for case 1 (b) Plot showing the predicted EUR from the EUR property map in (a) vs. the true EUR for case 1. (c) EUR property map for case 2 (d) Plot showing the predicted EUR from the EUR property map in (c) vs. the true EUR for case 2.

Some of the wells from the mean posterior probability distribution map (Figure 3b) are analyzed as highlighted in Figure 6a. Three wells having high EURs (dark red dots) and three wells with low EURs (dark blue dots) are considered for analysis of the well parameters. The averages of the normalized well parameters for high and low EUR wells are plotted in Figure 7b. The average well parameters from wells with high EURs (in red) are having different characteristics compared to the wells with low EURs (in blue). The wells with good EURs possess the following characteristics: a higher proppant volume, higher sand volume, less cluster spacing, higher number of fracture stages, perforation clusters and a higher porosity, whereas the wells with poor EURs tend to have the opposite characteristics.

By analyzing Figure 6, some of the variables such as total perforations, average treating rate, formation thickness and perforation length are removed (highlighted in gray). The underlying reason is that these variables appear not to contribute much towards well productivity in either the good performing wells or the poorly performing ones. Although formation thickness is likely to contribute to well productivity, Figure 6 essentially illustrates that the strongest controls on well productivity for the good performing wells are related to the completions. For the wells characterized by low EUR’s, the role of formation thickness is significant however. This approach reduces the number of well variables to ten thereby decreasing the data dimensionality. Similar to the preceding GTM analysis the mean posterior projections are plotted in the 2D latent space after the final iteration and show less overlap in the final projection (Figure 7).

Figure 6. (a) Average parameter values for 3 wells belonging to the class of high (red) and low (blue) EUR values are considered for parameter sensitivity analysis. (b) 14 normalized average well parameters for each of the wells are plotted forming two sets of averaged data-vectors. Note that most of the well parameters differ substantially for the two cases. The well parameters highlighted in gray do not show considerable differences between both classes of wells and are removed for further analysis.

However to quantitatively compare Case 1 with all well variables included in the analysis with Case 2 using a reduced subset of variables, we predict the EUR using both the results for a set of validation wells. The full distribution of the posterior probabilities \(R_{nk}\) are crucial to our application of the GTM for EUR predictions. The workflow is as follows: The EUR, \(E_n\) is known for each well. We also have the posterior probabilities of the \(n\)th well data-vector.
projected on the 2D latent space as shown in Figure 4a. From this we get the EUR map for 1 well (Figure 4b). This then provides us a weighted sum of the EUR, \( q_k \) at each grid point \( k \) in the latent space for all the \( N \) wells given by:

\[
q_k = \sum_{n=1}^{N} \frac{R_{nk} E_n}{N} \tag{1}
\]

This will give us a most likely EUR map from all the wells over the latent space (Figure 4c). We first should analyze this EUR map to qualitatively determine if there is significant correlation between the latent space and the physical property. In addition, we can also use this new \( q_k \) to predict the EUR at another location in the latent space \( E_m \), which is not present in the training data set. This is done by calculating the posterior probability \( V_{km} \) of the new \( m \)th data-vector and multiplying this posterior probability value \( V_{km} \) at each grid point \( k \) with the weighted sum of the EUR values \( q_k \). Thus, for predicting the EURs \( E_m \) we use,

\[
E_m = \sum_{k=1}^{K} V_{km} q_k, \tag{2}
\]

where the indices \( m \) representing a validating well.

**Figure 7.** The mean posterior distribution map of the “responsibilities” of the data with reduced set of parameters (case 2) in the 2D Latent space plot color-coded by the scaled EURs. (a) Initial and (b) final distribution of the posterior mean projections of all 137 well data onto the latent space after 200 iterations. Note that after 200 iterations the mean projections are more distributed in the latent space with less overlap.

For the training dataset, the EUR’s derived from the latent space variables are shown in Figures 5a and 5c for Case 1 and Case 2 respectively. The posterior probabilities for the validation set of 8 wells for Case 1 and Case 2 are calculated and are multiplied with their corresponding weighted EUR values to give the plot of the predicted EURs from the GTM model versus the true given EURs as shown in Figures 5b and 5d. Both plots demonstrate a reasonably satisfactory linear trend with similar least-square fits. One key implication of the results presented in Figure 5 is that in spite of using redundant variables in the training, the final analysis is very close to an analysis based on choosing a set of variables using some engineering priori judgments. Consequently, by virtue of using the concept of ‘responsibilities’, the GTM analysis essentially ignores explanatory variables that do not contribute towards the output and can therefore be utilized in a ‘black-box’ approach by the practicing engineer dealing with similar datasets.

**Conclusions**

While there is significant uncertainty related on the method of EUR evaluation, EUR is perhaps the most common means of comparing well value by a given operator. The classification of EUR from production data is a challenging problem for two main reasons. Firstly, the high dimension of the input data (more the number of well parameters poses a high dimensionality of the input data). Second, the parameters are often coupled, with the total amount of proppant being correlated to the number of perforations, with no clear accepted normalization technique. We find that GTM has the advantage that it defines probability densities for the dataset, which can be used to prioritize alternative reservoir targets or completion techniques.

In this paper we fit a deformed 2-D manifold to 14 well measurements that reside in 14-dimensional space. The 145 data vectors projected onto this are represented by a grid of Gaussian probability functions. Using 2D conventional crossplotting tools, the interpreter interactively defines clusters in the 2D latent space that are subsequently correlated to production, or problems in completion. Furthermore, using a distance metric defined by Bhattacharya, we are able to predict EUR from the 14 well measurements. GTM Prediction of EUR at the eight validation wells resulted in an \( R^2 \approx 0.92 \). In spite of this excellent prediction, GTM (and any other regression or clustering technique) implicitly assumes that the provided well measurements represent the variability in production. In our example, there are few if any natural fractures. In other reservoirs, we would wish to include some type of fracture measurement, either from image logs or from surface seismic data.

**Acknowledgement**

The authors thank their institutions for supporting this work.