SPE-184050-MS

Multivariate Statistical Analysis for Resource Estimation in Unconventional Plays Application to Eagle Ford Shales

Saurabh Sinha, Deepak Devegowda, University of Oklahoma; Bhabesh Deka, Reliance Holdings USA Inc.

This paper was prepared for presentation at the SPE Eastern Regional Meeting held in Canton, Ohio, USA, 13-15 September 2016.

Abstract

Unconventional resource plays have shaken the energy industry in last decade changing the energy import – export equation especially in United States. US shale plays now contribute towards 40% of the total natural gas production (U.S. Energy Information Administration, 2012)With highly heterogeneous formations and Multi-Fractured Horizontal Wells (MFHW), resource estimation in these plays are still debatable. From sophisticated numerical simulation models to much simpler hyperbolic decline curves, none can be used with confidence to estimate resources in these shale plays.

In this paper we have described the application of multivariate statistical techniques to generate type curves (TC's) for the Arp's decline analysis. Our study includes more than 1500 wells from different fluid type's (dry gas, lean gas, rich gas and light oil) in eagle ford shale (EFS).

Data such as well logs, completion, production and geology is included in the analysis as input variables. The analysis demonstrates the application of multivariate statistical techniques such as principal component analysis (PCA), k means clustering, self-organizing maps (SOM) and multivariate regression (MVR).

Our analysis show that univariate statistics is insufficient for analyzing the data due to significant amount of heterogeneity in shale reservoirs. Using univariate statistics impact of different variables such as well spacing, condensate gas ratio (CGR), perforated lateral lengths, zipper fracture etc. cannot be accounted in a holistic way. Multivariate analysis can quantify the impact of these variables and can be used as a good predictive tool for determination of Arp's parameters.

Introduction

Arp's decline curves (Arps, 1944) are used as the primary method of reserves accounting by most of the operators across United States. These are preferred over other conventional methods due to the following reasons:

- Low permeability of shale requires massive hydraulic fracturing (MHF) for production. Hence, the contribution from the reservoir beyond the fracture tips is negligible (Ambrose et al., 2011). This scenario is different from connected conventional reservoirs where connectivity between the different facies requires full field scale simulation.

- Number of wells in conventional scenario are limited hence making numerical simulation and other conventional methods (well testing, material balance etc.) feasible and reliable. These methods are not applicable in shale reservoirs and the reservoir behavior is mainly volumetric. This makes decline curve analysis (DCA) a preferred method to forecast reserves.

A detailed discussion on different types of declines is out of scope of this work and only Arp's hyperbolic declines are discussed. Arp's hyperbolic equation is given in the general form as:

$$q = q_i \frac{1}{(1 + bD_i\, t)^{1/b}} \tag{1}$$

Where, q_i is the initial production (IP) rate (volume/time), b is Arp's hyperbolic decline constant (0<b<1) and D_i is Arp's initial decline rate (1/time). In conventional Arp's analysis b factor is limited to a maximum value of 1. Exceeding this value makes the reserves unbound(Lee, Texas, & Sidle, 2010), and it is usually handled by operators through a Securities and Exchange Comission (SEC) defined values for terminal declines.

In a typical type (TC) curve analysis setting, the wells are grouped into various clusters depending on multiple factors:

a. Geological setting (such as presence of major fault systems, petrophysical properties gradation, thermal maturity levels etc.)
b. Reservoir parameters (reservoir fluid type, reservoir pressure, compositional gradient etc.)
c. Geomechanical parameters (local stress regimes, zipper fracture effect, compaction, dilation etc.)
d. Completion design (proppant volumes, perforated lateral length, surface pumping rates etc.)
e. Well spacing, re-fracturing and other special operations

The process of grouping the wells is mainly the "best engineering judgement" of the asset reservoir engineering team. From the local knowledge of the area supported by the various datasets (well logs, micro seismic, daily production, tracers, well interference studies etc.) forecasts are given on yearly basis.

This method although useful, depends mainly on the skill of the interpreter. With limited dataset and multiple variables to consider, it could break down and the TC's would have to be revised frequently. This directly affects the reserves reporting and the valuation of the assets. Most of the small to mid-scale oil and gas business in United States operates on reserves based lending (RBL) scheme. Which is nothing but a process of taking loan from banks against reserves as mortgage to develop the asset.

This process makes the TC generation at the center of all unconventional business. Even in the process of univariate TC grouping, the asset team tries to group the well together looking for factors that correlate best with the Arp's decline parameters (IP, D, and B). Subcategories are then created within these groups to normalize the dataset to account for lateral length variation and other factors which correlate strongly with any of the Arp's decline parameters.

Major decisions such as changing the completion for the whole field (or a specific area of the field), well spacing's, single well PAD drilling vs. multi well PAD drilling are also taken based on these results. However, due to the vast amount of heterogeneity present in the shale reservoir, it very difficult to account for every parameter that affects the Arp's decline inputs in univariate statistics.

One such example is demonstrated in a correlation matrix in Figure 1. Different Arp's variables depend on more than one factor. This makes grouping the data and analysis extremely difficult with univariate statistics. In this study, we have tried to address this issue using a multivariate statistical approach.

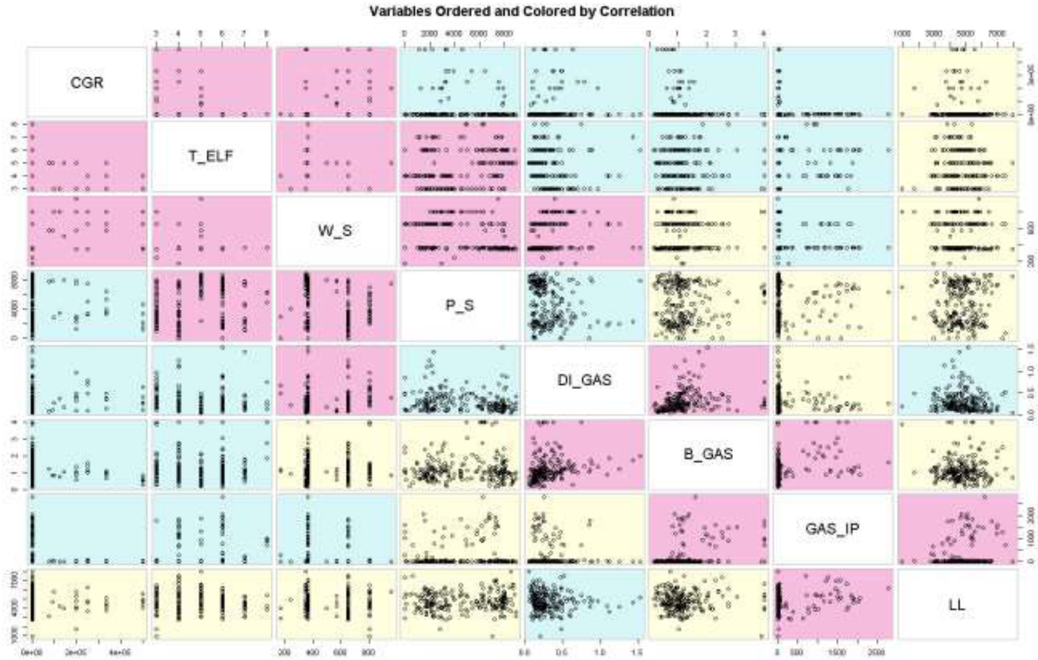


Figure 1—Correlation matrix of decline parameters (IP, D and B) with different input parameters in area of interest. The strongest correlation is in pink and weakest in yellow. Definitions of input variables are summarized in Table 2.

Methodology

For analysis in this work, a total of 1514 horizontal multi fractured wells from Dewitt County, TX are chosen. Eagle ford shale (EFS) in this county offers a significant number of wells with substantial well history. Also, a band of fluid types (dry gas to light oil) is available in this county. This allows to have multiple clusters of wells to be grouped into meaningful type curves. Figure 2 elaborates our area of interest (AOI) and the wells chosen for analysis.

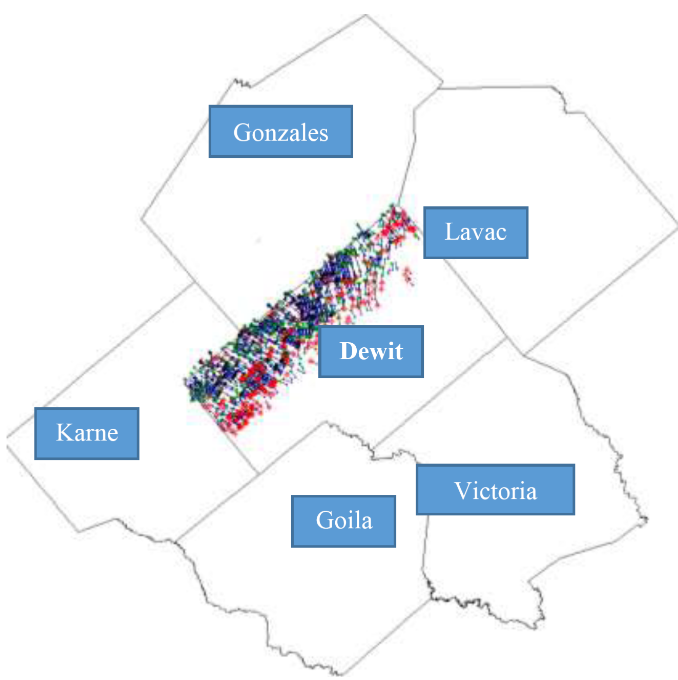


Figure 2—Area of interest for analysis (Dewitt County, TX). Overlay includes selected wells and surrounding counties.

Overall production trends (both oil and gas are summarized) in Figure 3 for all the wells. For all datasets in this analysis, a public information database (drillinginfo®) is used.

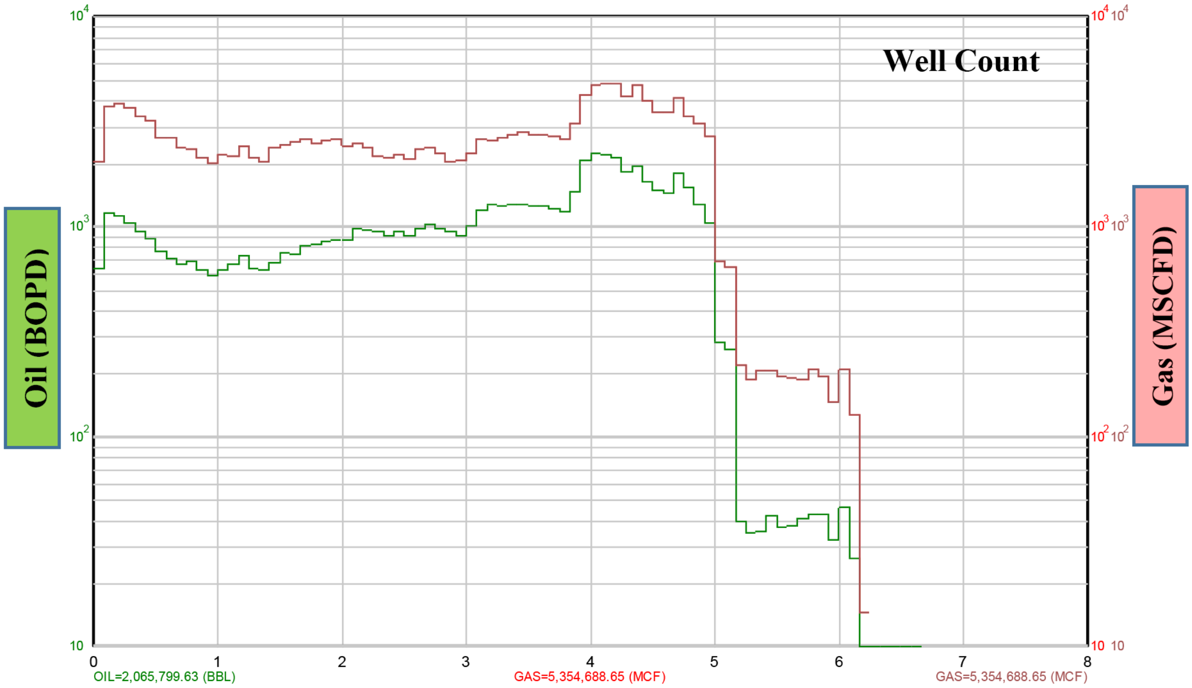


Figure 3—Production trend (all wells). Data normalized to maximum rate

Clustering of wells

First the wells are clustered into different type curve groups for analysis. The clustering of data is based on following parameters:

a. ***Flowing pressures***

The flowing pressures recorded at stabilized flow rates (IP tests) is a good indicative of actual reservoir pressure. In case of wells which are flowing through the casing, casing pressure is taken. For wells with tubing flow, tubing pressure is used. This way a flowing pressure variation across the AOI is mapped. As the dataset is taken from a public database, a detailed geology is not available. In case a detailed geological information is available this data can be further fine-tuned by geology (for example pressure variations across a major fault mapped on seismic, or an area of the reservoir where reservoir seal is not intact and hence unusually low pressures).

b. ***Shut in pressures***

Shut in pressures similar can be utilized similar to the flowing pressure. Although like conventional analysis, it cannot be utilized to derive any meaningful parameters like permeability or skin (due to supercharging (Chang, Hammond, & Pop, 2005) effect in shale reservoir), it can further bolster the confidence in pressure profile across the reservoir. An initial reservoir pressure can also be taken if the data is available for significant number of wells. As post fracture reports for limited number of wells is available, extended shut in pressures are taken as a proxy for initial reservoir pressure.

c. ***Condensate gas ratio (CGR)***

CGR (BBL /MMSCF) is an important parameter that influences the flow potential in the reservoir. CGR values can not only define the fluid type in the reservoir but combined with reservoir pressures is a strong indicator of reservoir drive mechanism and hence the expected ultimate recoveries (EUR's).

However, what CGR value to consider is very important for the analysis. A typical CGR trend in condensate well is shown in figure 4. It is clear from the plot that the CGR trend gradually decreases and stabilize at a constant value.

It can be due to multiple reasons including but not limited to: flow stabilization (from linear to boundary dominated), compositional gradient within the reservoir or gradual fracture tip closure (initial production from upper EFS (high CGR) + lower EFS gradually moving to lower CGR of lower EFS only).

Hence, using cumulative CGR's or initial test CGR's to define long term production and hence the EUR's is not a wise choice. In this study all CGR values used are the stable CGR's. For wells in which CGR has yet not stabilized, CGR values are borrowed from the adjacent wells for best approximation of stable CGR's.

d. ***Flow potential (productivity index)***

To analyze the flow potential of the wells, use of cumulative volumes has been avoided in this study. Cumulative volumes not normalized by time and pressures are a poor choice for flow potential. As a detail pressure history for the wells is not available on drillinginfo®, hence the IP test values are used.

The wells are first classified into oil or gas wells on the basis of CGR. Wells with CGR<400 are considered as gas wells and wells with CGR> 400 are considered as oil wells. Fine subdivisions (dry gas, lean gas, rich condensate, light oil etc.) are not done at this stage.

Once, the wells are classified into oil or gas wells the major phase stable rate during the test is taken as flow rate from the well. This is then divided by drawdown (Stabilized Shut-in pressure – Flowing tubing head pressure). This tubing head pressure (THP) based PI's are then taken for all the wells. Out of 1514 wells, IP tests are reported in only 945 wells. For other wells, PI's are borrowed from immediate neighboring wells.

e. ***Separator oil and gas gravities***

Oil and gas gravities are indicative of the fluid type similar to CGR. Separator oil and gas gravities are included in the clustering process.

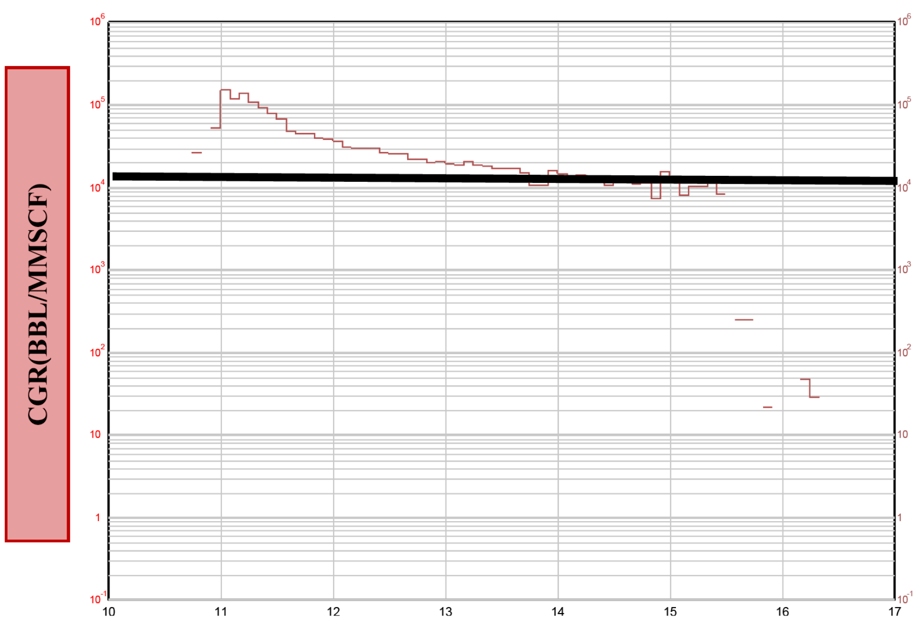


Figure 4—CGR trend over well life.

We then clustered the wells applying Unsupervised Clustering Techniques. In this case, we used k-means and self-organizing maps (SOM). These unsupervised learning algorithms are used for exploratory data analysis to find hidden patterns or grouping in data.

The objective in k-means is to define for each cluster a centroid. Then, each data point belonging to the set is compared and associated with the nearest centroid. The centroids are re-calculated and change their location step by step until no more changes are necessary. On the other hand, self-organizing maps (SOM) work as a data visualization technique that reduces the dimensions of a data set through the use of self-organizing network. The map units are allowed to change themselves by learning to become more like to the samples from the data they are compare with.

In k-means, we determine an estimate of the appropriate number of clusters required using a built-in algorithm in an inhouse developed code. For this, we calculate the sum of squares between each cluster initially assuming different numbers of clusters, and looking for an "elbow-point". Figure 5, exhibits the results obtained from the algorithm. The change in the slope starts around 5 clusters. But the ideal number may be anywhere between 5 and 9 clusters. For the purpose of defining the groups of wells a number of 5 clusters was selected.

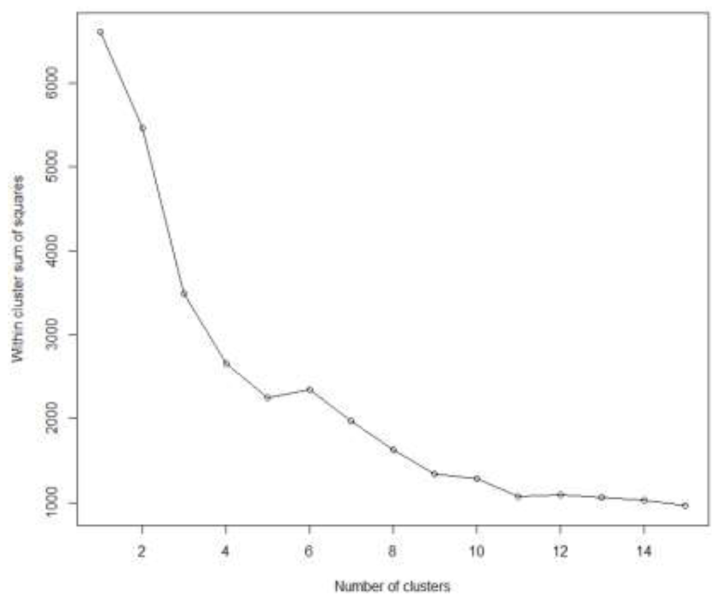


Figure 5—Number of clusters from k-means

After defining the number of clusters, we determined the data distribution in each cluster specifying an output of 5 groups. The cluster results are plotted in 2 dimensions using the first principal components, as shown in Figure 6.The principal components selected explain around 64% of the variability, which means that they are a good representation of all the parameters selected.

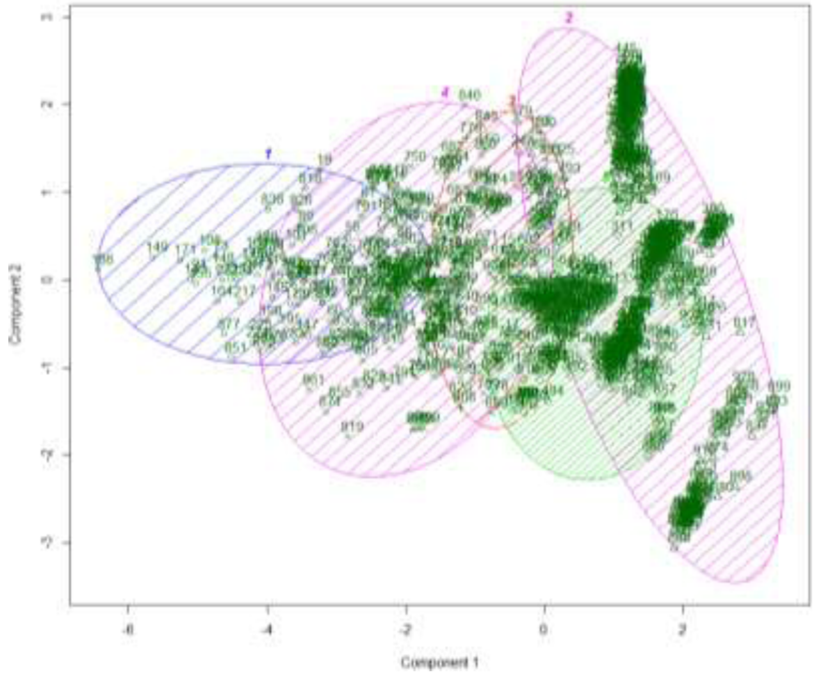
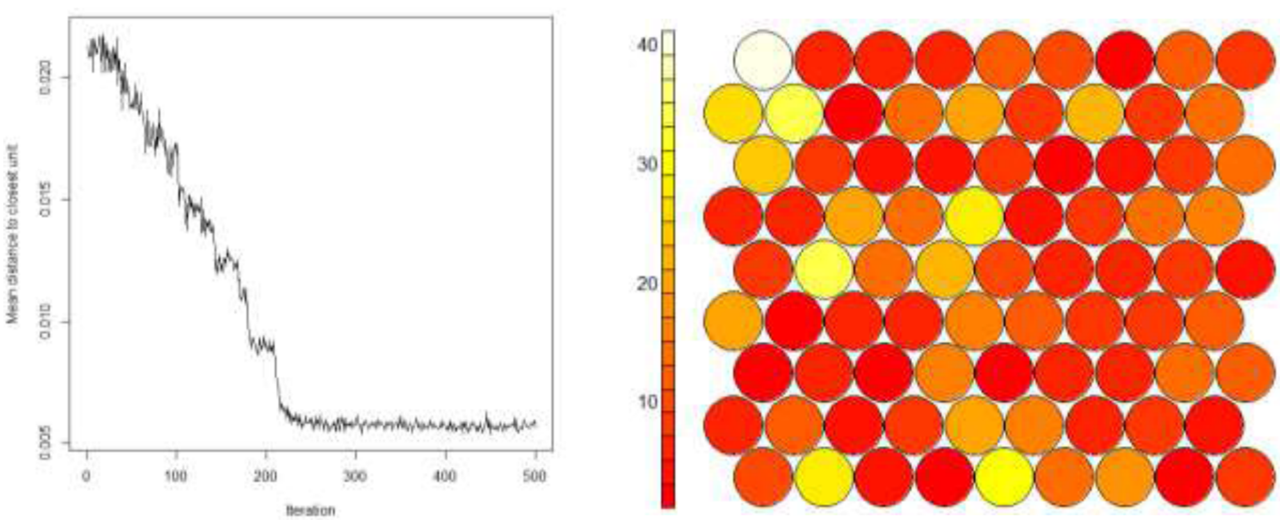


Figure 6—Cluster plot using principal components.

As a clustering alternative we train a Self-Organizing Map (SOM). The SOM is a map of size of 9x9 with a hexagonal topology, the number of times the complete data set is presented in the network is 500. The learning rate (indicating the amount of change) declines from 0.05 to 0.01 over the number of updates. The variables are standardized with the mean and standard deviation for the model construction.

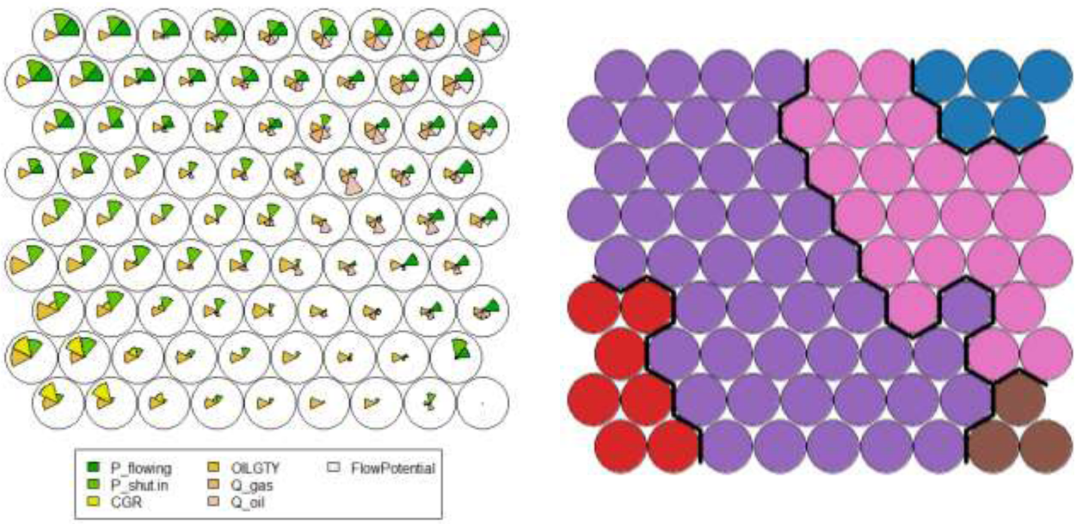
Figure 7a. shows the training progress of the map over time (iterations). The y-axis shows the distance from each node's weights to the samples represented by that node. This distance should reach a minimum plateau; in this case the plateau is reached around 200 iterations. Figure 7b shows the counts plot, the color in each circle shows the number of times the unit was a best matching unit, for example, a unit colored in yellow was best matching unit for around 30 counts or times. The counts plot measures the map quality, ideally the sample distribution should be uniform. For this study a network size of 9X9 was selected, which is adequate for the data set size.



(a) (b)

Figure 7—SOM (a) Training progress, (b) Counts plot

To interpret the map obtained we use the codes plot shown Figure 8a. Which is a fan diagram representation for each parameter weight in the nodes. From the plot we can see patterns in the distribution. The data points with high flowing and shut-in pressures are located in the left upper corner, while the data points with high oil density are located in the left lower corner.



(a) (b)

Figure 8—SOM (a) Codes plot, (b) Clusters obtained from hierarchical clustering on the SOM nodes.

Selecting the boundaries for each cluster is a subjective task that can be based on the visualization of the plots shown above. For these data set, a number of 5 clusters were selected. Then, we use hierarchical clustering from the Kohonen package for the classification. The resultant clusters are shown below in Figure 8b.

The final selection of Clustering method is based on following criterion:

- Predictability
- Data distribution

Both methods give similar predictability (~70%). However, with k means the classification suits the well parameters analyzed and is more efficient than the hierarchical clustering used in SOM. Hence k means clustering is selected for clustering the data. However, the maps generated with SOM, give an idea of the data distribution and the relations between the parameters

The clustered wells or TC groups are shown in the Figure 9. On the basis of clustering, a total of 5 type curve areas are defined.

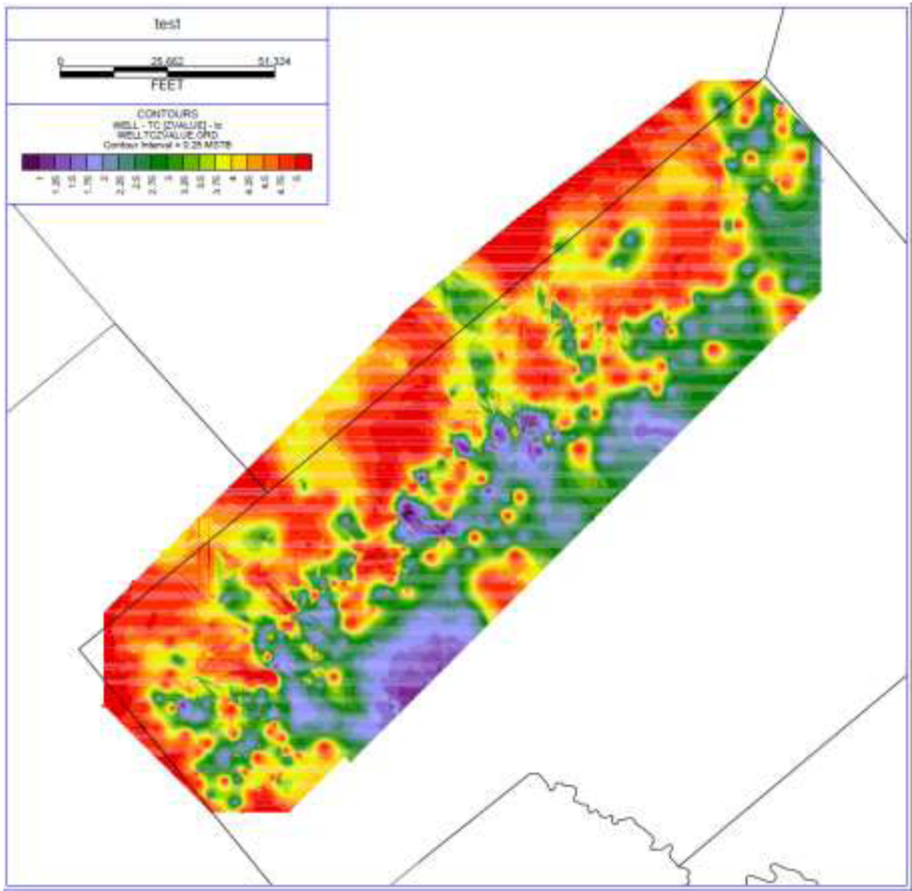


Figure 9—Type curve classification (1-5). TC contours overlay with county outline.

Decline curve analysis (DCA)

Within these type curve areas, now correlations for IP, decline (D) and b factor (B) has to be defined. Hence, these parameters are chosen for regression. (Lee et al., 2010) has described the applicability of decline analysis. One of the most important criterion for applicability of decline curve analysis (DCA) is well should be in boundary dominated flow (BDF). BDF in this whole work is referred to fracture interference based on a horizontal well with transverse fractures (Figure 10).

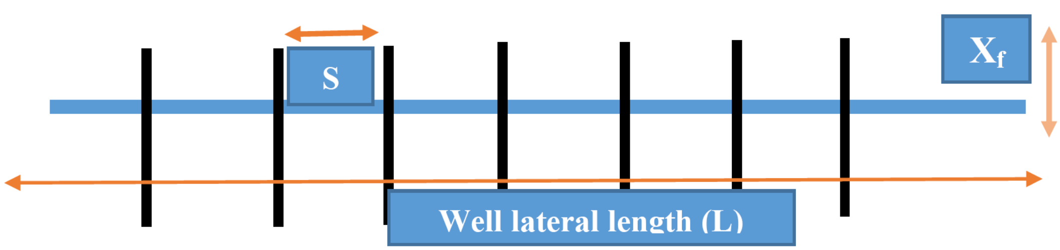


Figure 10—Well geometry assumed in the analysis. Horizontal well with lateral length (L), fracture half length (Xf) and fracture spacing S.

With a limited dataset, there is a limitation on methods which can be applied to identify the wells in BDF. In this study a combination of Fetkovich type curves (rate –time) and reciprocal-rate cumulative plot((Wattenbarger, Robert A; El-Banbi, Ahmed H; Villegas, Mauricio E; Maggard, 1998)) are used. Reciprocal rates are plotted against cumulative volumes instead of conventional square root time plots to reduce noise in dataset.

These are elaborated in Figure 11and Figure 12 figure 18. Fetkovich TC (Fetkovich, 1980) do not require pressure data and are ideal in this situation. Reciprocal –rate cumulative curves are used in conjunction with the type curves. All 1514 wells are tested for flow regime identification.

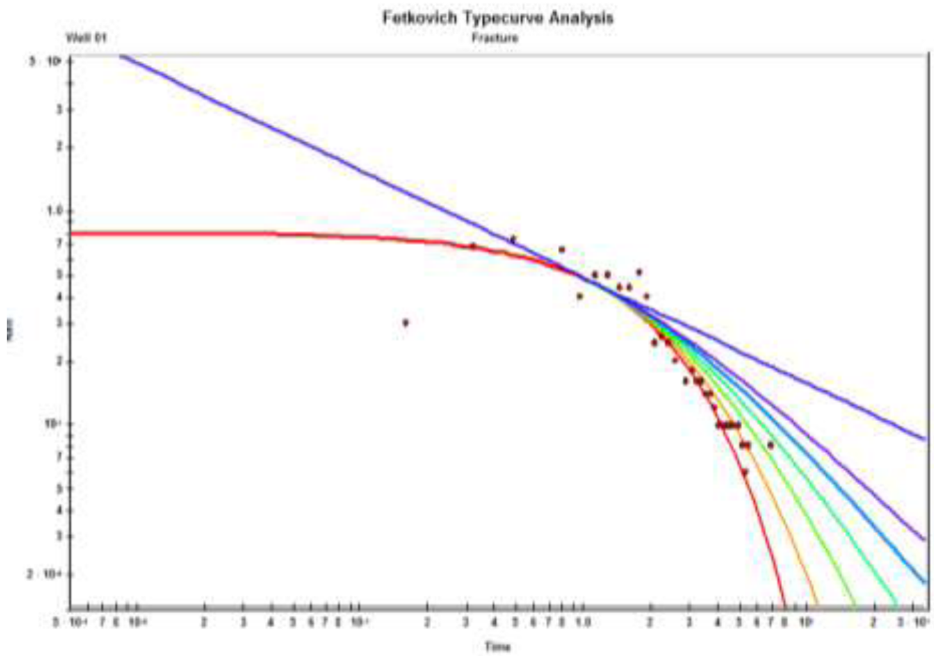


Figure 11—Fetkovich rate-time type curve. Well in BDF.

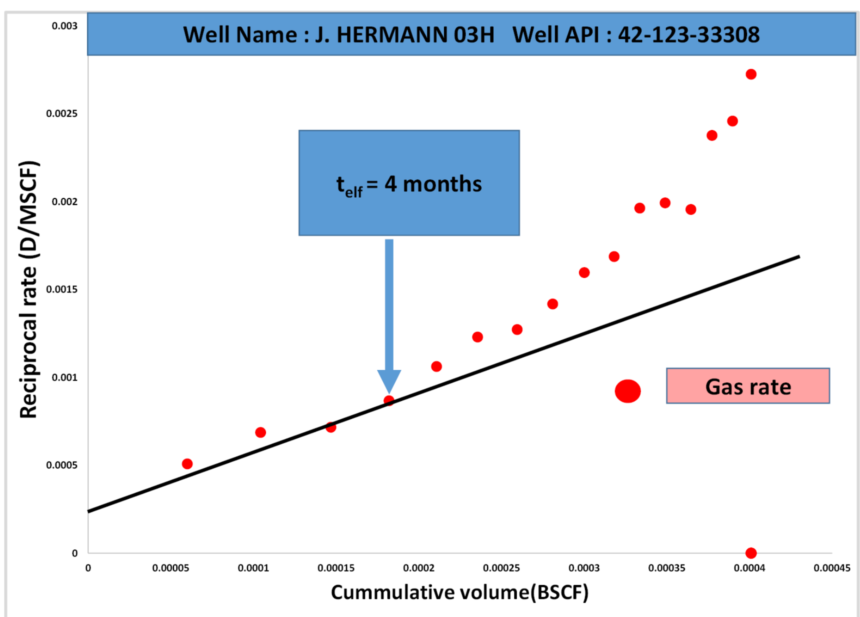


Figure 12—Fig diagnostic plot (reciprocal rate vs. cumulative volumes) for the major producing phase (downtime adjusted). Well in BDF. Time end of linear flow(telf) is 4 months for this well.

A total of 688 wells are found to be in BDF. These wells are then fitted with hyperbolic decline curves (Figure 13). The wells are fitted with a portion of decline which best represents the reservoir characteristics. Following criterion is applied on the decline curves before attempting a hyperbolic decline:

- Portion of decline curve in which the decline is well developed (free of choke management and line pressure issues)
- If there are operational changes (coke change, adjacent well fracture shut-in) the well-developed decline trend before or after the operation is taken
- If there is a refracture, loading or artificial lift signature on the production profile, the portion of curve with anomaly is not considered in analysis

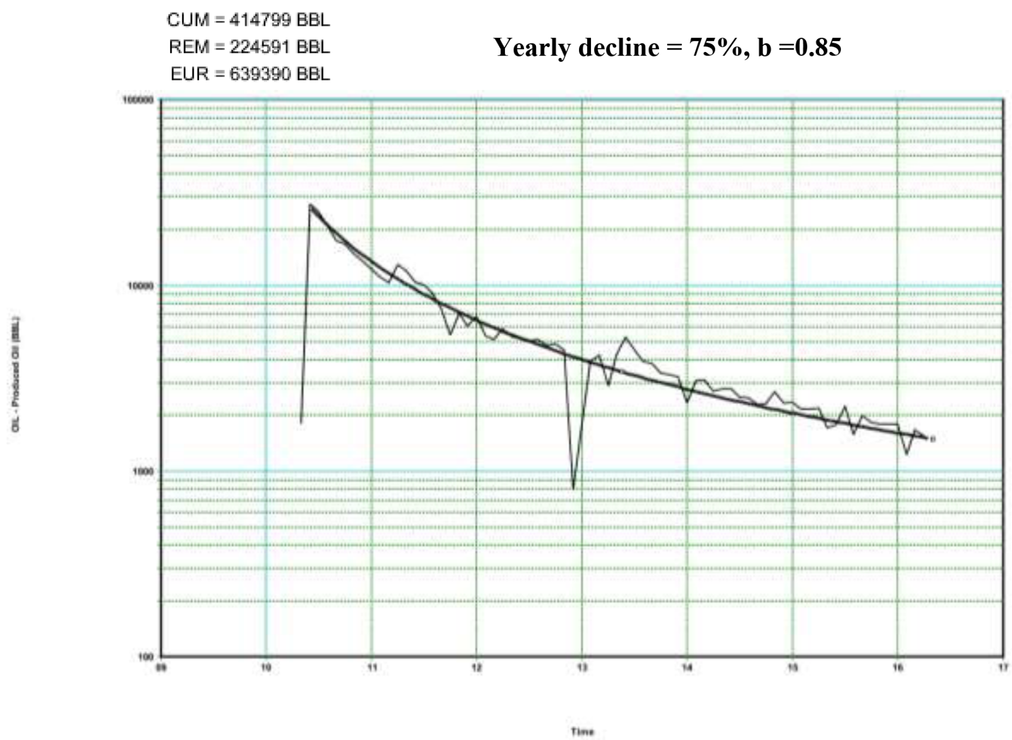


Figure 13—Example decline analysis (rate-time) for oil phase for an example well

This is an important step as the purpose is to develop generic TC's. This includes modeling declines which are best representative of the reservoir decline behavior and not the latest decline trend. Hence, the process is completely different from yearly reserves reporting in which well wise forecasts are generated for proved developed producing (PDP) wells. All wells are fitted with both oil and gas declines (if both phases are available)

Example distribution of decline and b factors in the AOI are summarized in Figure 14 and figure 21. Reservoir decline and b factors are reservoir properties and hence can be evaluated from decline curves. IP's however has a strong correlation on the mode of well operation (casing vs tubing flow, choke sizes, downhole separators etc.). For this purpose, the IP's are taken from the well IP tests. Once reconciled, IP, decline and b factor (for both oil and gas phase) are taken as response variables for next step.

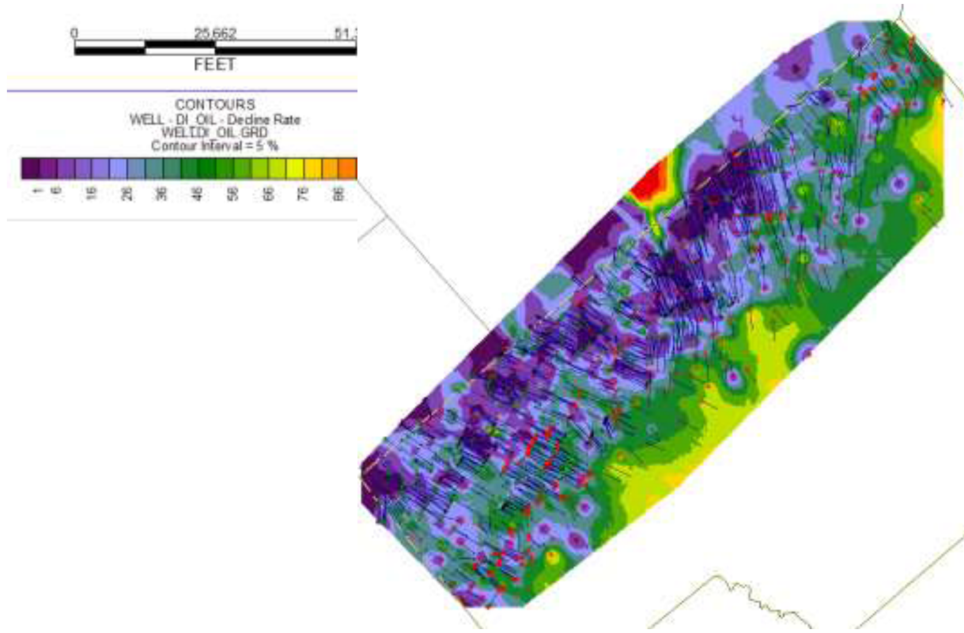


Figure 14—Oil monthly decline distribution in AOI. Contours are overlain with wells and county outline

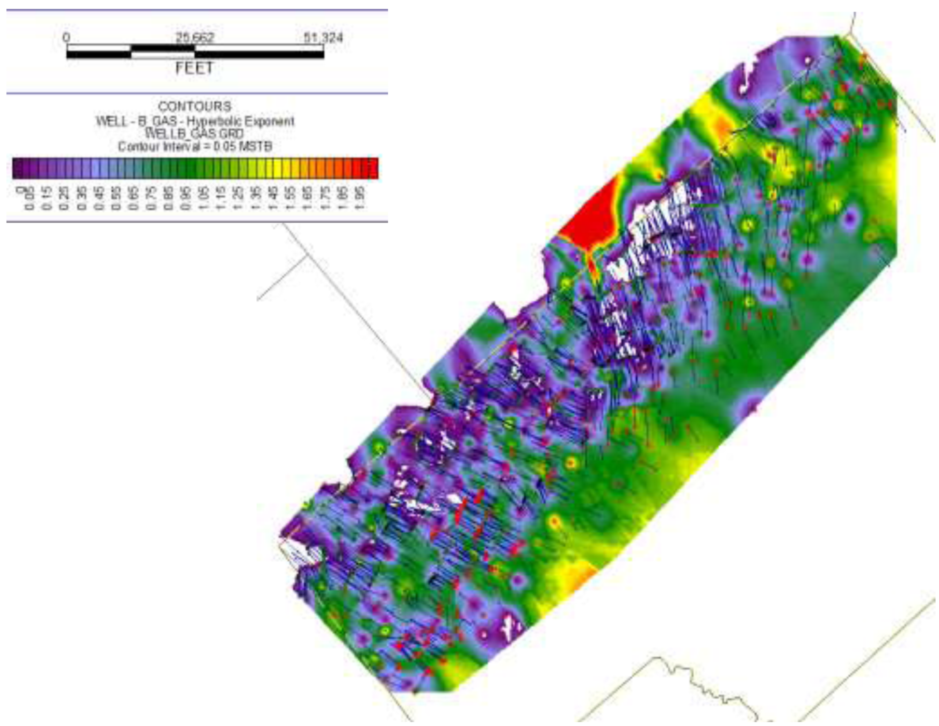


Figure 15—Gas b factor distribution in area of interest

Multiple Regression

The multiple regression analysis is performed only in half of the wells considered in this study, and the remaining wells are used for results validation. The input variables for the analysis are listed below:

Petrophysical properties

Seven pilot well log are used in the analysis to delineate a trend of effective porosity, water saturation and pay. Lower + upper EFS thickness is considered for the analysis. As all the wells are horizontal multi-fracture wells, mapping of fracture growth is required to assign the exact pay zone. In absence of relevant dataset, the whole EFS thickness is considered as payzone.

Perforated lateral length

The difference between top perforation and bottom perforation from completion reports is used for analysis. In this analysis a simplistic model is assumed (horizontal well with multiple fractures, Figure 10).

Permeability / cluster spacing proxy

As there is no permeability log available, a different approach to include the permeability is taken. The time end of linear flow as described by (El-Banbi & Wattenbarger, 1995)is direct indicative of fracture interference. The time end of linear flow (t_{elf}) depends on fracture spacing (figure 18) and the permeability of the reservoir. Hence, t_{elf} is included as a variable in the analysis. This takes into account both the fracture spacing (a completion parameter) and permeability (reservoir parameter). In instances, where pressure data is available to conduct a full RTA, these parameters can be taken separately.

Initial reservoir pressure

Instead of taking initial reservoir pressure from post fracture gradients, shut in pressures from the IP tests are taken. As the post fracture gradient is available from very few wells, a proxy in terms of extended shut in pressures is expected to have more reliability due to more data points.

Well spacing

Well spacing is included for all wells on a well to well basis as a parameter. This is expected to take into account the effect of well interference (if any) in response variables (Arp's parameters).

Condensate gas ratio

Stable CGR is taken from the production data as an input described before.

First, a principal component analysis (PCA) is used to reduce the number of variables. It was observed during the quality check process that seven pilot wells are not enough to model the scale of area considered in this study. Hence, geological parameters are removed from analysis. With 6 parameters, principal components were derived using an in-house developed code. Based on the scree plot obtained from PCA (Figure 16), we decide how many principal components are enough to explain the variability between the parameters. Usually three principal components (PC's) are enough to explain more than 90% variability in data. Hence, in this case three PC's are considered for regression instead of 7 original variables.

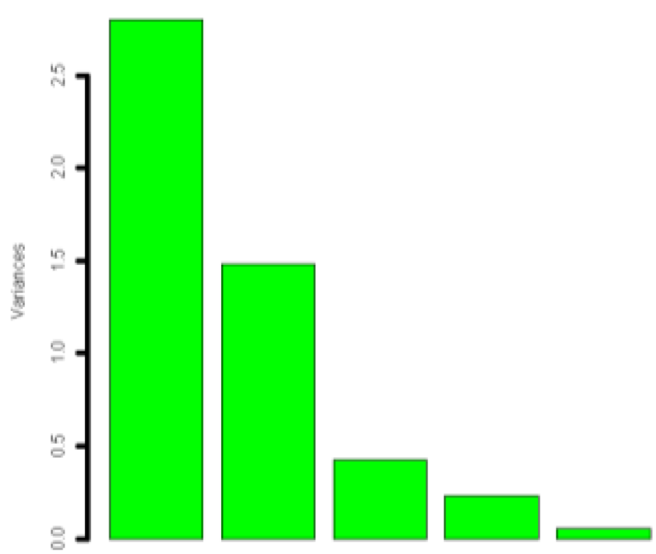


Figure 16—Scree plots for PCA performed on gas IP in TC area 5

Then, a multiple regression is run on these three principal components (PC's). A general additive model is used for the analysis. The regression is tested with a null hypothesis (p value) test. Principal components with p values greater than 0.05 (PC's>0.05) are rejected and considered as poor predictors of response variables (IP, decline and b). R^2values are then reported for each case (every TC group, every variable). The regression model derived from these values (half the dataset) is then tested to predict the decline parameters and hence the EUR's of remaining wells. The original EUR's (from actual decline fit), predicted EUR's and the errors are documented.

Results

Table 1 summarizes the variable name and their definition. Table 2 summarizes the R^2 values for IP, D and b factors for different TC areas.

Table 1—Variables used for principal component analysis and their definition.

Variable	Definition
T_ELF	Time end of linear flow
W_S	Well spacing
P_S	Extended shut-in pressure
LL	Perforated lateral length
CGR	condensate gas ratio

Table 2—R2 values obtained in different type curve area.

TC Area	Parameter	R^2 VALUE
1	Gas IP	0.75
	Gas D	0.62
	Gas B	0.70
	Oil IP	0.77
	Oil D	0.75
	Oil B	0.72
2(gas only)	Gas IP	0.55
	Gas D	0.63
	Gas B	0.62
	Oil IP	NA
	Oil D	NA
	Oil B	NA
3(gas only)	Gas IP	0.67
	Gas D	0.71
	Gas B	0.68
	Oil IP	NA
	Oil D	NA
	Oil B	NA

(continued on next page.)

Table 2—(continued).

TC Area	Parameter	R^2 VALUE
4	Gas IP	0.77
	Gas D	0.85
	Gas B	0.87
	Oil IP	0.77
	Oil D	0.72
	Oil B	0.79
5(gas only)	Gas IP	0.80
	Gas D	0.85
	Gas B	0.88
	Oil IP	NA
	Oil D	NA
	Oil B	NA

Figure 17 and Figure 19 are EUR maps with initial actual well decline fits and EUR's by 50-50 approach (half the wells are considered for generating type curves and rest half are predicted) respectively. All TC regions show low p values, however TC area 2 and 3 show low R^2values. This is a classic case of low P, low R^2 and the direct interpretation of this case is the insufficient number of variables chosen for analysis. Also, the number of wells in these TC's are significantly less than in TC area 1, 4 and 5 (Figure 9).

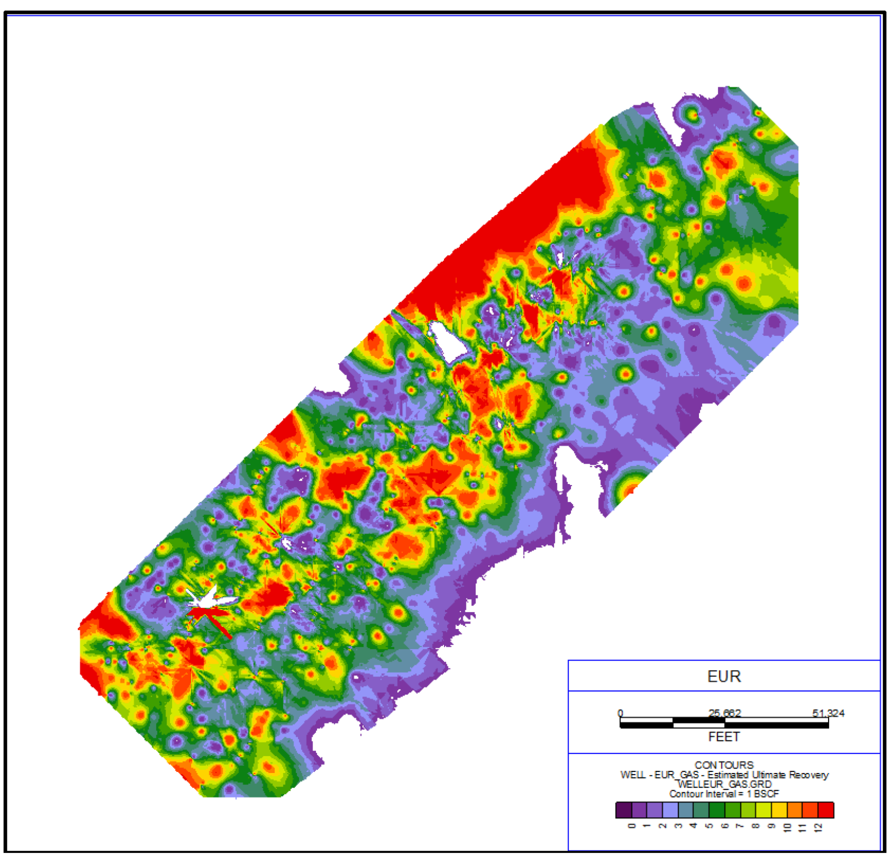


Figure 17—Original gas EUR (BSCF), from decline fit of all wells in BDF.

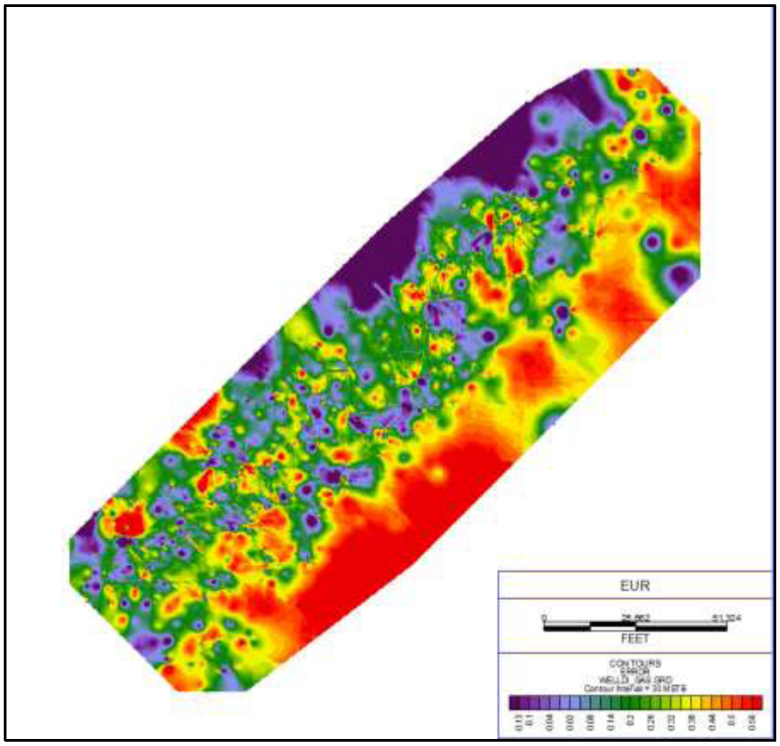


Figure 18—Error in calculation of gas EUR (predicted- actual)/ actual in percentage. High error areas has very few wells in TC area.

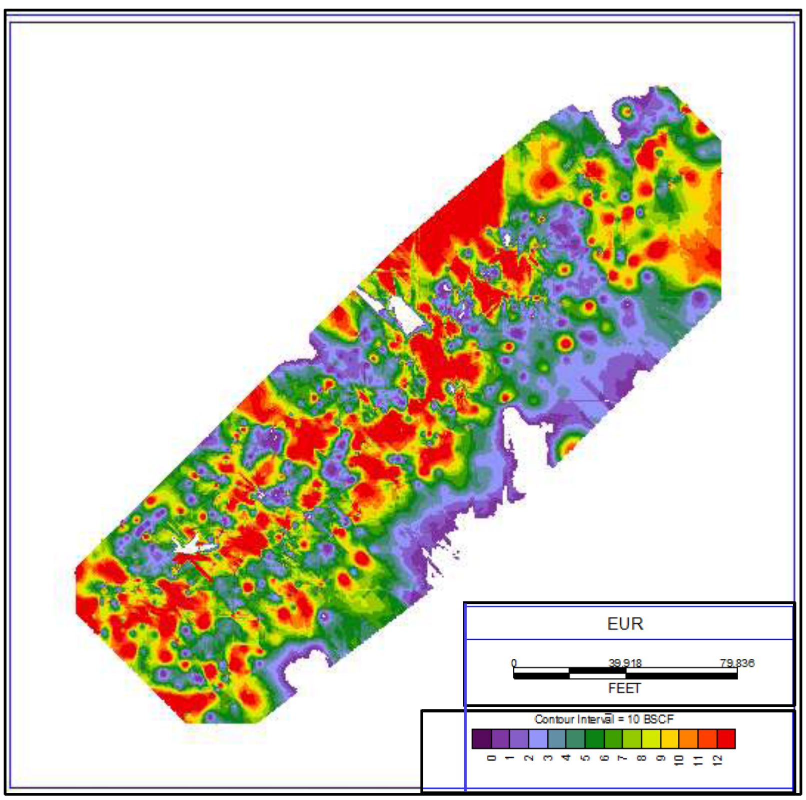


Figure 19—Predicted gas EUR (50% original and 50% predicted) in BSCF.

In our case, we have used a limited dataset available from public domain, however for operators with significant dataset, additional variables such as initial reservoir pressure (from multiple wells), petrophysical parameters from sizeable number of wells, rate transient analysis parameters, and completion parameters such as proppant volumes may increase the R^2 values significantly.

Example results of intial EUR's from decline fits, error from multiple regression (actual-predicted)/(actual) and predicted EUR's are summarized from figure 17-19.

Conclusions

Our analysis demonstrates the conventional method of deriving the TC's with univariate statistics is insufficient to account for variability in dataset. Using multivariate statistical methods can significantly improve the results and provide data trends with multiple variables. The key is to select the appropriate and sufficient parameters for analysis.

Once a type curves are generated, it can be used to predict EUR's with respect to multiple variables like well spacing, proppant volmes, zipper frac effect (from micro-sesmic mapping) or any other variables if included in input. This can provide a variable specific effect on EUR's from many variables. Simpler methods such as multiple regression is sufficient for generating TC's in our area. However, if optimal results are not obtained more complex methods such as neural networks may be used.

Nomenclature

q_i - initial production rate (volume/time)
b - Arp's hyperbolic decline constant ($0<b<1$)
D_i - Arp's initial decline rate (1/time)
t - time
EUR - Expected ultimate recovery
IP - Initial production

References

Ambrose, R. J., Energy, D., Clarkson, C. R., Youngblood, J., Adams, R., Nguyen, P., … Biseda, B. (2011). Life-Cycle Decline Curve Estimation for Tight / Shale Gas Reservoirs. *Society*, (2008), 607–621, SPE 140519. http://doi.org/10.2118/140519-MS

Arps, J. J. (1944). Decline Curve Analysis.pdf. *Aime*. http://doi.org/10.2118/945228-G

Chang, Y., Hammond, P., & Pop, J. (2005). When Should We Worry About Supercharging in Formation Pressure While Drilling Measurements? Proceedings of SPE/IADC Drilling Conference. http://doi.org/10.2118/92380-MS

El-Banbi, A. H., & Wattenbarger, R. A. (1995). Analysis of Linear Flow in Gas Well Production. The 1998 SPE Gas Technology Symposium, 15–18 March. http://doi.org/10.2118/39972-MS

Fetkovich, M. J. (1980). Decline curve analysis using type curves. *Journal of Petroleum Technology*. http://doi.org/10.2118/4629-PA

Lee, W. J., Texas, A., & Sidle, R. E. (2010). *SPE 130102 Gas Reserves Estimation in Resource Plays*, (February), 23–25.

U.S. Energy Information Administration. (2012). Annual Energy Outlook 2012. *U.S. Department of Energy*, (June).

Wattenbarger, Robert A; El-Banbi, Ahmed H; Villegas, Mauricio E; Maggard, J. B. (1998). Production Analysis of Linear flow Into Fractured Tight Gas Wells. SPE Rocky Mountain Regional/Low-Permeability Reservoirs Symposium and Exhibition. http://doi.org/10.2118/39931-MS