



SPE 77597

## Identification of Successful Practices in Hydraulic Fracturing Using Intelligent Data Mining Tools; Application to the Codell Formation in the DJ –Basin

Mohaghegh, S.D., Popa, A., Gaskari, R., Ameri, S., West Virginia University, and Wolhart, S. L., Pinnacle Technologies

Copyright 2002, Society of Petroleum Engineers Inc.

This paper was prepared for presentation at the SPE Annual Technical Conference and Exhibition held in San Antonio, Texas, 29 September–2 October 2002.

This paper was selected for presentation by an SPE Program Committee following review of information contained in an abstract submitted by the author(s). Contents of the paper, as presented, have not been reviewed by the Society of Petroleum Engineers and are subject to correction by the author(s). The material, as presented, does not necessarily reflect any position of the Society of Petroleum Engineers, its officers, or members. Papers presented at SPE meetings are subject to publication review by Editorial Committees of the Society of Petroleum Engineers. Electronic reproduction, distribution, or storage of any part of this paper for commercial purposes without the written consent of the Society of Petroleum Engineers is prohibited. Permission to reproduce in print is restricted to an abstract of not more than 300 words; illustrations may not be copied. The abstract must contain conspicuous acknowledgment of where and by whom the paper was presented. Write Librarian, SPE, P.O. Box 833836, Richardson, TX 75083-3836, U.S.A., fax 01-972-952-9435.

### Abstract

In a detail data mining study about 150 wells that have been completed in the Codell formation, DJ Basin, have been analyzed to identify successful practices in hydraulic fracturing. The Codell formation is a low permeability sandstone within the Wattenburg field in the DJ Basin of Colorado. Since 1997 over 1500 Codell wells have been restimulated. As part of a Gas Research Institute restimulation project 150 wells were studied to optimize candidate selection and identify successful practices.

Hydraulic fracturing is an economic way of increasing gas well productivity. Hydraulic fracturing is routinely performed on many gas wells in fields that contain hundreds of wells. During the process of hydraulically fracturing gas wells over many years, companies usually record the relevant data on methods and materials in a database. These databases usually include general information such as date of the job, Service Company performing the job, fluid type and amount, proppant type and amount, and pump rate. Sometimes more detail information may be available such as breakers, additives, amount of nitrogen, and ISIP to name a few.

These data are usually of little use in complex 3-D hydraulic fracture simulators. These models require additional and more detailed information. On the other hand, the collected data contain valuable information that can be processed using virtual intelligence tools. The process covered in this paper takes the above-mentioned data and couples it with general information from each well (things like latitude, longitude and elevation), any information available from log analysis and production data. The conclusion of the analysis is a set of

successful practices that has been implemented in a particular field and recommendations on how to precede with further hydraulic fracture jobs.

In this paper the results of applying this process to about 150 Codell wells during the GRI sponsored project is presented. This process provides an important step toward constructing a comprehensive set of methods and processes for data mining, knowledge discovery, and data-knowledge fusion from data sets in oil and gas industry.

### Introduction

Patina Oil and Gas has been very active in the DJ basin in recent years. They have been one of the most active operators in the United States in identifying and restimulating tight gas sand wells. Patina has over 3,400 producing wells in the basin, and has restimulated over 230 Niobrara/Codell completions so far. Furthermore, it is estimated that the results they are achieving in terms of incremental recoveries are up to 60% better than other operators.

Studies and analysis such as the one being presented in this paper has the potential to help operators like Patina Oil & Gas to increase their chance of success even to a higher percentage. It also has the potential to help other operators in increasing their chances of success in DJ Basin or any other locations throughout the North America. This study is probably one of the most comprehensive analyses of its kind ever to be performed on a set of wells in the United States.

In this technical paper the authors' intention is to introduce this new and novel methodology in its entirety and present as much of the results as the page limitations of this paper allows. Please note that due to the comprehensive nature of this methodology many of the topics cannot be discussed in much detail. It is our intention to introduce these topics in much more detail in series of upcoming technical papers.

### Methodology

The process of "Successful Practices Identification" using state-of-the-art data mining, knowledge discovery and data-knowledge fusion techniques<sup>1</sup> includes the following five steps. In order to comprehensively cover the theoretical

background of each step involved in this process, a separate paper may be needed for each. Some of the ideas have been introduced in the past and are referred to in the references. Detail on other topics will be the subject of future papers. In this article the goal is to provide a view of the methodology as a whole, and therefore, authors will simply introduce and give a brief explanation of each of the topics to clarify their role in the process.

### **Step One: Data Quality Control**

The process starts with a thorough quality control of the data set. During this process the outliers and their nature (are they really valid data elements or are they the result of human error either in measurement or in recording the values?) as well as missing data are identified, and using advanced intelligent techniques,<sup>2,3,4</sup> the data set is repaired. It is important to note that the repair of the data set at this stage of the analysis is aimed at rescuing the remaining data elements in a particular record (a data record here is referred to a row of the data matrix) that includes many features (features are the columns of the data matrix). The goal is not to “magically” find the piece of the missing data or substitute the outlier with the correct value. The goal is merely to put the best possible value in place of the missing data or the outlier that would allow the analysis to continue without losing the information content that exists in the rest of the features in a particular data record. Moreover, a new and novel methodology has been developed in order to identify and eliminate erroneous data records from the data set. These techniques, verification of their accuracy and how they are implemented are subjects of a future paper.

### **Step Two: Fuzzy Combinatorial Analysis**

The second step of the process is a complete “Fuzzy Combinatorial Analysis – FCA” that examines each feature in the data set in order to identify its influence on the process outcome<sup>5</sup>. The process outcome is a feature (usually a production indicator such as cumulative production, 5 year cum., 10 year cum., best 12 months of production, etc.) in the data set that is designated to identify the success of the practices in a field. For example if 5 year cumulative gas production is selected to be the process outcome, then a high 5 year cum. would indicate performance of good practices for that particular well. During that “Fuzzy Combinatorial Analysis” each feature’s influence is examined on the process outcome both individually and in combination with other features. This is due to the fact that influence of a particular feature (say a fracturing fluid) on the process outcome may be altered once it is combined with the effects of other features (say specific additives) that are present in the process. Therefore it is important to perform the analysis in a combinatorial fashion (hence the name combinatorial analysis) in order to reveal the true influence of the features present in the data set on the process outcome.

A note of caution is in order here. Many commercial, off-the-shelf neural network software applications claim to identify

the influence of features on the output once a neural network model is build for a data set, and many practitioners in our industry have been using them as the true influence of parameters on the output. These products simply use the summation of the weights connected to a particular input neuron in order to achieve this. Authors believe that this is a gross simplification of a complex problem, and does not provide an accurate account of the influence of each feature and therefore should not be used as such. This method simply is an artifact of the modeling process and in the event of changing the architecture or learning algorithms of the modeling process the influence of features may be altered.

### **Step Three: Intelligent Production Data Analysis**

The third step in the “Successful Practices Identification” methodology is a process called “Intelligent Production Data Analysis – IPDA.” The word “intelligent” in the above phrase refers to the use of intelligent systems techniques in the production data analysis process. During the IPDA process production data is used to identify a series of “Production Indicators” that would represent the state of the production from a particular field in time and space. The time and space representation is aimed at capturing the depletion and pressure decline in the field as new wells are drilled and put into production at different rates. The dynamic nature of this analysis (simultaneous analysis of the data in four dimensions x, y, z, and t) allows the user to identify the sweet spots as well as bad (unproductive) spots in a field. Such analysis would prove quite valuable during field development strategies that include infill drilling programs and candidate selection for stimulation, restimulation and workovers.

### **Step Four: Neural Model Building**

Next step (Step 4) in the process calls for building a predictive neural model based on the available data. This step has been covered in detail in the several prior papers and will not be repeated here<sup>4,6-9</sup>.

### **Step Five: Successful Practices Analysis**

Once a representative neural model is successfully trained, calibrated and verified, the process of “Successful Practices Identification” using state-of-the-art data mining, knowledge discovery and data-knowledge fusion techniques is concluded with a three stage analysis. The three stage analysis combines the neural model with Monte Carlo simulation, genetic algorithms search and optimization routines<sup>3</sup>, and fuzzy set theory<sup>2</sup> to identify the successful practices on a single well basis, on groups of wells basis, and on a field wide basis.

During the single well analysis, each well is thoroughly analyzed in order to identify the sensitivity of that particular well to different operational conditions. This analysis can identify the distance of actual practices on that well from the successful practices that could have been performed on that well. The larger the distance the higher would be the potential

of a good response from the well to future activities such as stimulation, restimulation and workover.

During the analysis of successful practices on groups of wells, two different kinds of grouping (clustering) procedures are used. First, the grouping of wells in the field is performed on the basis of their location (geology), using the information provided during the step three analyses. Then another set of grouping of the wells are performed based on specific attributes identified during the step two analyses. This grouping is done using either hard K-Mean clustering and/or fuzzy C-Mean clustering procedures, dictated by the data set. Upon clustering of the wells successful practices for each cluster is identified using the neural model in conjunction with genetic algorithms. As a new well is introduced to the process for analysis, it is first identified to which group or cluster it belongs and then accordingly, the successful practices for that particular group are recommended for the new well.

The last stage is the identification of the field-wide successful practices. In this stage all the wells in the field are treated as one group and the successful practices are identified for the entire field operation.

## Results and Discussions

The original database for stimulation of Codell wells in the DJ basin needed considerable quality control in order to remove erroneous records including records that did not belong to that database and had to be removed. It is interesting to note that several data records were identified by our data quality control procedure as suspect. These findings were communicated with the engineers in the operating company. After further investigation the engineer at the operating company confirmed that some of the records identified should not have been included in the database and their inclusion was the result of an oversight. This was a confirmation of the accuracy of the new and novel methodology that has been developed for the data quality control procedure.

Once the quality control of the data set was completed, Fuzzy Combinatorial Analysis (FCA) was performed. The analysis was performed for the combination of up to five features. Table 1 shows the result of FCA. As shown in this table the most influential parameter is “Flow Back Volume” followed by “Codell Gas-Ft” which is a reservoir quality parameter identifying the product of gas saturation and the net pay for the Codell formation. The next two influential parameters are related to the water quality used for the restimulation frac jobs. These are water bicarbonate content and the water peak viscosity.

The parameters in Table 1 can be divided into six different categories. These categories are Location, Reservoir Quality, Original Completion, Original Frac Treatment, Second Frac Treatment, and Production. Using the “Fuzzy Combinatorial Analysis” (FCA) average scores for each of the categories, one can identify the importance of each category on the

process outcome. It should be noted that the process outcome in this analysis is the “Actual Post-Restimulation Peak Production” which will only be referred to as Actual Peak. Figure 1 shows the influence of each category on the Actual Peak. From this figure it is clear that Reservoir Quality followed by Well Location (which is an indication of the geology) are clearly the most influential categories in the DJ basin. It is also interesting to note that original frac treatment plays a more important role than the restimulation treatment in identifying the success of a restimulation process. Please note that in this figure the lower the FCA value (shorter bars) the higher the influence of the parameter, or category.

Details on two of the parameters from Table 1 are presented in Figures 2 and 3. Each figure provides three graphs. These graphs include the conventional regression analysis, the conventional frequency distribution and finally the fuzzy combinatorial analysis curve for each of the parameters. Figure 2 represents the details on the most influential parameter – Flow Back Volume. It shows a well behaved frequency distribution of the values of this parameter in the data set that depicts a normal distribution slightly skewed to the left toward the lower volumes, slightly peaking around 225 bbls. The actual peak of the frequency distribution is around 500 bbls. The regression plot shows that no significant outliers can be immediately identified. Please note that the regression plot provides a  $R^2$  of about 0.065 that mainly identifies no correlation.

Next, looking at the FCA plot one can see a smooth and continuous curve that simply means that the values of this parameter in the dataset are behaving normally. There are three regions in this plot that need to be identified. First is the region in blue shade (Region B from 380 to 650 bbls). This is the region that has provided the best consistent results in this process. The area indicated by the shade of green (Region A from 80 to 240 bbls) is the region with the least desirable outcome. The rest of the plot with no shades (or white shades) is consisted of several regions of this parameter that seem to show sensitivity to other parameters in the data set. This simply means that in Flow Back Volumes between 250 – 380 bbls and those above 650 bbls, importance and role of other parameters determines the degree of success of the process. In other words, regions in shades of white are the regions of high sensitivity to other parameters. In order to identify the process outcome in these areas one need to build a neural model of the process and study each well separately. This is performed during the first stage of Step five. Figures 3 is a similar graphs for Peak Viscosity.

Much more can be discussed and shown as the result of the “Fuzzy Combinatorial Analysis” but to keep this paper short the detail of the findings of this analysis will be the subject of a future paper.

Upon completion of the first two steps the stage is set for the third step or the “Intelligent Production Data Analysis –

IPDA.” In this analysis the goal is to assist the operating company in their future field development strategies by identifying the locations of the field that would be a prime candidate for further investment (infill drilling, stimulation, restimulation or workovers) as well as identifying the portions of the field in which no new investments are recommended. Please keep in mind that all this is done based on the successful practices of the operating company in the past. Figure 4 shows the identified regions in this field. Region identified by number 1 is the sweet spot. New investment in this part of the field is recommended. The conclusion of this analysis is that new investment in this part of the field (infill drilling, stimulation, restimulation or workovers) has a high probability of providing high returns. Region 2, as designated in this graph would be the next best region in the field. Region 3 in the figure would be the region in which new investment is not recommended based on this analysis. Many times identification of bad spots in a field is just as important as the good ones. Table 2 shows the characteristics of each region. It can be seen that regional segmentation is supported by “Post-Restimulation Actual Peak Production,” and the reservoir quality identifier which is the product of gas saturation and Codell net pay. Values of both these parameters reduce as we go from one region to the next.

The fourth step in this analysis is the neural network model building. In order to facilitate a thorough analysis of the wells in DJ basin it was decided that two neural models should be constructed, trained, calibrated and verified. The first model only uses the parameters prior to the restimulation process (Pre-Restimulation Neural Model) and the second model uses parameters related to the restimulation process (Post-Restimulation Neural Model). This is an effort in order to study the possibility of identification of key issues that affect the ultimate process outcome (Post-Restimulation Actual Peak.)

In this paper the results and analysis of the first neural network (Pre-Restimulation Neural Model) will be discussed. The results of the second network may be presented in a future paper. The Pre-Restimulation neural model was developed using eight inputs and one output (Post-Restimulation Actual Peak). Following parameters were used during the neural model building as the input to the neural network.

1. Well Location – Latitude
2. Well Location – Longitude
3. Reservoir Quality – Codell porosity-net pay product
4. Reservoir Quality – Nibrara porosity-net pay product
5. Completion – Number of perforations in Codell
6. Original Stimulation – Amount of sand used as proppant (Mlbs)
7. Original Stimulation – Amount of fluid used (Mgal)
8. Result of cluster Analysis – Entropy

Figure 5 through 7 show the results of training, calibration and verification of the Pre-Restimulation neural model. Figure 5

shows the results of the training data set. The training data set has a  $R^2$  of 0.783 and a correlation coefficient of 0.901. Figure 6 shows the results of the calibration data set. The calibration data set has a  $R^2$  of 0.821 and a correlation coefficient of 0.907. Figure 7 shows the results of the verification data set. The verification data set has a  $R^2$  of 0.516 and a correlation coefficient of 0.809.

Last step in this analysis is the “Successful Practices Identification.” As was mentioned in the previous section the “Successful Practices Identification” is performed in three stages. During the stage one the analysis is performed on a single well basis. Please note that only the results of the “Successful Practices Identification” for the first neural network (Pre-Restimulation Neural Model) are being presented here and the concentration will be on three parameters, one completion parameter and two parameters from the original stimulation performed on each well. These parameters are the number of perforations in the Codell as the completion parameter, amount of sand (proppant) in thousands of pounds, and amount of fluid in thousands of gallons as the original stimulation parameters. The goal is to identify if the effect of the original completion and stimulation on the restimulation outcome.

This analysis can guide the operator and the Service Company on what has worked in the past on a particular well and what has failed to work. It can provide a valuable roadmap for design and implementation of restimulation procedures for such a well. Furthermore, coupling the results from this stage with the similar study on the same well using the Post-Restimulation Neural Model can provide much more insight into the future practices that should be implemented on this or similar wells.

The analysis on each well is performed and presented in two ways in order to allow the operator and the Service Company to take full advantage of the process. First the analysis is performed and presented in the form of a series of two-dimensional graphs. In these Cartesian graphs each of the selected parameters is plotted against the process outcome. The 2-D graph displays the response of an average restimulation procedure as a function of different values of the parameter being studied. In each plot the dependent variable (parameter being studied) is swept throughout its range from the minimum to the maximum values present in the dataset. Then the the dependent variable, being the process outcome (post-restimulation actual peak in Barrels of Oil Equivalent – BOE) is calculated and plotted as a function of different values of dependent variable. This is done by dividing the range of the independent variable by 100 and running the model 100 times – once per each value of the independent variable. The results of the model are then plotted in a 2-D graph.

Figure 8 shows this analysis performed for well “Bohlender 8-5”. There are three plots in this figure for three independent

variables, namely, original 20/40 sand, which is the amount of the proppant (Mlbs) pumped during the original stimulation job, original fluid, which is the amount of fluid (Mgal) pumped during the original stimulation job, and finally total number of perforations in Codell.

As it can be seen in this figure, in the case of the original fluid a maximum can be clearly seen at about 1,364 BOE, while in the other two cases a minimum value can be observed for each. The minimum value in the case of proppant is about 970 BOE and in the case of number of Codell perforations is about 1100 BOE. Since the post-restimulation actual peak for this well is about 1,344 BOE, the conclusion might be that in the case of this particular well the amount of the original fluid used during the stimulation was at its optimum or close to its optimum value. On the other hand by increasing the amount of proppant or the number of perforations, one might be able to increase the restimulation performance. But it has to be noted that there is not much one can do about the independent variables being studied in this model, since we cannot go back and change the original stimulation job's specifications. But are there lessons that can be learned from this for the future wells that will be drilled in this area? May be we find out that once a well has been damaged (or stimulated in a way that it would hamper its future restimulation performance) it would be an indication to stay away from such wells (for restimulation purposes) and invest on other wells.

The straight black lines in each of the plots in Figure 8 show the actual value of the independent variable that was used in the original completion and stimulation. This shows that although much could not have been done in the case of amount of fluid, but there definitely was room for improvement as far as the amount of proppant and the number of perforations was concerned.

The second part of the single well analysis includes three-dimensional sensitivity analysis. The three-dimensional sensitivity analysis is, in nature, very much similar to the 2-D analysis. During the three-dimensional sensitivity analysis two independent variables are selected. Then as the value of one of the variables are changed in sequence from its minimum to its maximum value in ten steps, at each step the value of the second variable is swept for the entire range. The result is a three-dimensional graph that shows the sensitivity of the post-restimulation actual peak in Barrels of Oil Equivalent to both variables simultaneously. Figure 9 shows three of these plots for the well "Bohlender 8-5" examining original sand, original fluid and number of perforations in Codell.

Last but not least for the single well analysis is the Monte Carlo simulation. During the Monte Carlo simulation analysis for each well, first the independent variables are selected. Then the probability distribution for each parameter is identified along with the corresponding parameters. Finally the model is executed several times in order to generate a

distribution of the process outcome that is the post-restimulation actual peak in this case.

Table 4 shows the characteristics of the Monte Carlo simulation that was performed for a particular well. These are the same three variables that were selected and studied for well "Bohlender 8-5". Uniform probability distribution was selected for all three variables. The Monte Carlo simulation was performed for 5000 runs. The results for well "Rumsey 10-4" are shown in Figures 10 and 11. In Figure 10 the frequency distribution is shown. As you can see the frequency distribution demonstrates a sort of bi-modal characteristics around 1,195 BOE and 2,095 BOE.

Figure 11 shows the results of the Monte Carlo simulation for the same well in the form of cumulative probability distribution. The actual value of the post-restimulation actual peak for this well was 2,100 BOE. This means that there is not much that could have been done for this well to enhance its post-restimulation actual peak above and beyond the achieved value. The green line in Figures 10 and 11 identifies the actual value of the post-restimulation actual peak for this well.

Stage two of the "Successful Practices Identification" performs the study on groups of wells. The groups are identified in two fashions. First using the geological separation as performed in step three (Figure 4.) In such cases a new well will be subject to successful practices recommendation based on where it is located in the field. The second way of grouping the wells is using either hard K-Mean cluster analysis or fuzzy C-Mean cluster analysis. The cluster analysis is performed based on the information provided in the step two (Fuzzy Combinatorial Analysis) of this methodology. Cluster analysis by nature does not provide a unique solution. In order to identify the best possible clusters (groups of wells) a novel method was developed to couple the cluster analysis routines with a robust search and optimization routine in order to converge to the best clustering of the wells based on the identified features. The objective function of the search and optimization algorithm was chosen to be the minimum average entropy associated with each cluster as well as with the entire data set. The entropy is defined as an indication of amount of information content (as oppose to noise) in a data record.

Once the grouping and the clustering of the wells are completed, a genetic algorithm routine, using the developed neural network as the objective function, will identify the successful practices for each of the parameters involved in the process. The results are presented in the form of characteristics tables identifying the general behavior of the optimized frac jobs for each group.

In stage three of the "Successful Practices Identification" the same procedure (as explained in the above paragraph) is performed for all the wells in the field, with results being presented to the user in the form of a characteristics table.

Table 5 and 6 are examples of such “Successful Practices” characteristic tables. The next and final step would be to deduce rules of thumb or recommendations based upon these characteristics tables. Following are examples of such conclusions from field-wide “Successful Practices” conclusions for amount of proppant and peak viscosity.

#### Amount of Proppant:

The dominant trend is Linear Descending – about 83% of the wells. Four more cases present the same trend but are defined by a second order polynomial. Ten cases present the opposite trend. The dominant trend for amount of proppant (Table 5) would imply that high Actual Peak would be achieved if low amounts of sand were used in treatment. The trend indicates that placing more sand in the fracture will not bring significant increase in Actual Peak.

#### Peak Viscosity:

The dominant trend is Half-Parabolic Descending – about 53% of the wells. However, another 31 cases present the same Descending trend but are defined by a second order polynomial. This would bring the same type of trend to approximately 95%. The dominant trend implies that high Actual Peak is achieved for low viscosity of the frac fluids. The trend indicates that low viscosity fluids are preferred to higher viscosity fluids. This somehow agrees with the trends identified above in the amount of proppant analysis. Indeed, for low proppant concentrations there is no need for high viscosity fluids. Furthermore, high viscosity fluids would produce high friction when pumped at high rate.

In conclusion of this section it is important to note that in the case of field-wide “Successful Practices Identification” (as well as the analysis for groups of wells but at a lesser extent) the law of averages will play an important role. Therefore, it is recommended that all three stages of analysis, namely single well analysis, groups of wells analysis, and field-wide analysis should be performed and the final recommendations and decision making process should use the results of all three stages simultaneously. Moreover, once the relevant graphs from the stage one and the characteristics tables from stages two and three are generated, it is suggested that the final recommendations, rules of thumb and future action items should be decided upon by a team of experts in the particular field (hydraulic fracturing in this case) together with engineers from the operating company that are intimately familiar with the intricacies of the particular field being studied.

### Conclusions

A new and novel methodology for comprehensive “Successful Practices Identification” of the oil and gas operations has been developed. This methodology uses state-of-the-art data mining, knowledge discovery and data-knowledge fusion techniques inspired by virtual intelligence paradigms. The new methodology was applied to hydraulic fracturing program of about 150 wells in the Codell formation in DJ Basin, Rock

Mountain region. This methodology has the potential to open new doors in comprehensive and detail field studies as an alternative to field wide simulation and modeling that sometimes can be prohibitive due to high cost and lack of proper information and data. If enough information (data) is available and cost is not an issue, field-wide simulation and modeling would provide valuable detail that can prove to be crucial in large scale operations. Even in such cases, studies like the one presented in this paper can be performed on a relatively low cost and can provide much needed insight at a short time to the engineers and field managers.

### Acknowledgement

The authors wish to thank Dr. Rober Siegfried and Mr. Iraj Salehi of Gas Technology Institute, and Mr. David Hill of Ticora Geosciences for their help and support throughout the life of this project. We also wish to thank Mr. Donald Shaw for his help and insightful observations during this project. We would like to extend our gratitude toward Patina Oil & Gas Corporation for providing the data set.

**References:**

1. Mohaghegh, S., Gaskari, R., Popa, A., Ameri, S., and Wolhart, S.: "Identifying Best Practices in Hydraulic Fracturing Using Virtual Intelligence Techniques", *SPE 72385, Proceedings*, 2001 SPE Eastern Regional Conference and Exhibition, October 17-19, North Canton, Ohio.
2. Mohaghegh, S.: "Virtual Intelligence Applications in Petroleum Engineering: Part 3 – Fuzzy Logic." *Journal of Petroleum Technology, Distinguished Author Series*, November 2000, pp 82-87.
3. Mohaghegh, S.: "Virtual Intelligence Applications in Petroleum Engineering: Part 2 – Evolutionary Computing." *Journal of Petroleum Technology, Distinguished Author Series*, October 2000, pp 40-46.
4. Mohaghegh, S.: "Virtual Intelligence Applications in Petroleum Engineering: Part 1 – Artificial Neural Networks." *Journal of Petroleum Technology, Distinguished Author Series*, September 2000, pp 64-73.
5. Mohaghegh, S., Mohamad, K., Popa, A. S. and Ameri, S.: "Performance Drivers in Restimulation of Gas Storage Wells." *SPE Reservoir Evaluation & Engineering Journal*, December 2001, pp 536-543.
6. Mohaghegh, S., Goddard, C., Popa, A., Ameri, S., and Bhuiyan, M.: " Reservoir Characterization Through Synthetic Logs", *SPE 65675, Proceedings*, 2000 SPE Eastern Regional Conference and Exhibition, October 17-19, Morgantown, West Virginia.
7. Mohaghegh, S., Reeves, S., and Hill D.: " Development of an Intelligent Systems Approach to Restimulation Candidate Selection", *SPE 59767, Proceedings*, SPE Gas Technology Symposium, Calgary, Alberta, April 2000.
8. Mohaghegh, S., Koperna, G., Popa, A. S. and Hill, D. G.: "Reducing the Cost of Field-Scale Log Analysis Using Virtual Intelligence Techniques", *SPE 57454, Proceedings*, 1999 SPE Eastern Regional Conference and Exhibition, October 21-22, Charleston, West Virginia.
9. Mohaghegh, S., Popa, A. S. and Ameri, S.: "Intelligent Systems Can Design Optimum Fracturing Jobs", *SPE 57433, Proceedings*, 1999 SPE Eastern Regional Conference and Exhibition, October 21-22, Charleston, West Virginia.

Rank	Feature	FCA Value	Rank	Feature	FCA Value
1	Flowback Volbbl	0	22	Frac Type	2.2848
2	CO -Phi-H	0.5811	23	No-CO-Perfs	2.303
3	Bicarbonate ppm	0.6666	24	Chloride ppm	2.3298
4	Peak Visc	0.7486	25	NI- Perfed-H	2.3302
5	Lat	0.7734	26	Water pHLab	2.3665
6	Orig20/40 Sand-Mlbs	0.9214	27	Pre-Refrac Mcfd	2.3956
7	Long	1.1	28	Cum MMcf	2.4009
8	Refrac Date	1.1934	29	Water Source	2.4018
9	ViscShear 100-30Min	1.3324	30	Iron ppm	2.4351
10	TotHardness ppm	1.518	31	MGAL	2.496
11	Calcium ppm	1.6692	32	TotalPerfs	2.5045
12	AvgRate BPM	1.7415	33	Sulfate ppm	2.5164
13	Est-Ult- GOR	1.7706	34	New Perfs	2.552
14	No-NI -Perfs	1.7863	35	Sodium ppm	2.6039
15	AvgPsi	1.8438	36	Magnesium ppm	2.6108
16	ViscShear 100-5Min	1.9401	37	ViscShear 100-0Min	2.6649
17	Top CO Perf	1.9819	38	Pre- FracISDP	2.7127
18	TDSolid ppm	2.0084	39	TestedPH	2.8066
19	MMcf	2.0777	40	Post- FracISDP	2.8256
20	Orig Fluid-Mgal	2.0855	41	Mlb20-40	2.8907
21	DOFP	2.2451	42	Communication	2.9554

**Table 1. Ranking of all the parameters in the dataset from most to least influential parameter. The lower the FCA values the higher the rank.**

	Region 1	Region 2	Region 3
Longitude - Low	-104.854	-104.704	-104.61
Longitude - High	-104.666	-104.579	-104.496
Latitude - Low	40.33	40.27	40.197
Latitude - High	40.42	40.369	40.315
<b>Actual Peak (BOE)</b>	<b>2,234</b>	<b>1,792</b>	<b>1,235</b>
<b>Res. Quality (Gas-ft)</b>	<b>2.58</b>	<b>2.47</b>	<b>2.25</b>
No Wells Included	58	50	9

**Table 2. Statistics on the selected regions during the IPDA.**

	Training	Calibration	Verification
Rsquare	0.783	0.821	0.516
Correlation Coefficient	0.901	0.907	0.809

**Table 3. Statistics of the Pre-Restimulation neural network model.**

Variable	Distribution	Minimum	Maximum
No. of Perforation in Codel (Number)	Uniform	4	80
Original 20-40 Sand pumped (Mlbs)	Uniform	85.5	600
Original Fluid Pumped (Mgal)	Uniform	44.5	147.6

**Table 4. Characteristics of the variables used in the Monte Carlo simulations.**

PARAMETER BEING STUDIED: 20-40 Sand (Mlb)												
No. of Wells	Percent of Total	Curves Shape Characteristics					Curves Statistics					
		Curve Type	Shape Type	Order	Trend	Convexity		Min.	Avg.	Max.	Entry Value	End Value
73	83%	Line	Line	ax+b	Descending		Min	292	292	504	504	292
							Max	2,855	2,927	2,927	2,927	2,855
							Average	1,566	1,714	1,862	1,862	1,566
							Stand. Dev.	675		619	619	675
							Variance	455,463		382,917	382,917	455,463
6	7%	Line	Line	ax+b	Ascending		Min	986	986	1,046	986	1,046
							Max	2,281	2,346	2,346	2,281	2,346
							Average	1,763	1,821	1,878	1,763	1,878
							Stand. Dev.	481		493	481	493
							Variance	231,429		242,924	231,429	242,924
4	5%	Polynomial	U-Turn	X <sup>2</sup> -Type	Ascending/Descending	Concave	Min	1,895	1,895	1,920	1,895	1,920
							Max	2,984	2,990	2,990	2,986	2,984
							Average	2,346	2,357	2,369	2,348	2,362
							Stand. Dev.	401		394	401	394
							Variance	160,761		155,206	160,592	155,254
4	5%	Polynomial	Half-Parabolic	X <sup>2</sup> -Type	Descending	Concave	Min	2,142	2,142	2,205	2,205	2,142
							Max	2,586	2,621	2,621	2,621	2,586
							Average	2,431	2,451	2,472	2,472	2,431
							Stand. Dev.	172		160	160	172
							Variance	29,618		25,718	25,718	29,618

Table 5. Characteristics table for "Amount of Proppant" in field-wide study (Stage Three) of the "Best Practices Identification."

PARAMETER BEING STUDIED: Peak Viscosity												
No. of Wells	Percent of Total	Curves Shape Characteristics					Curves Statistics					
		Curve Type	Shape Type	Order	Trend	Convexity		Min.	Avg.	Max.	Entry Value	End Value
53	60%	Polynomial	Half-Parabolic	X <sup>2</sup> -Type	Descending	Concave	Min	354	354	1,062	1,062	354
							Max	1,743	3,078	3,078	3,078	1,743
							Average	1,040	1,746	2,451	2,451	1,040
							Stand. Dev.	285		350	350	285
							Variance	81,320		122,668	122,668	81,320
31	35%	Line	Line	ax+b	Descending		Min	307	307	1,906	1,906	307
							Max	1,605	2,990	2,990	2,990	1,605
							Average	793	1,704	2,615	2,615	793
							Stand. Dev.	249		224	224	249
							Variance	61,836		49,957	49,957	61,836
3	3%	Polynomial	U-Turn	X <sup>3</sup> -Type	Ascending-Descending	Concave	Min	1,523	1,523	2,432	2,377	1,523
							Max	2,119	2,530	2,530	2,530	2,119
							Average	1,804	2,147	2,490	2,471	1,804
							Stand. Dev.	245		42	67	245
							Variance	59,828		1,746	4,554	59,828

Table 6. Characteristics table for "Peak Viscosity" in field-wide study (Stage Three) of the "Best Practices Identification."

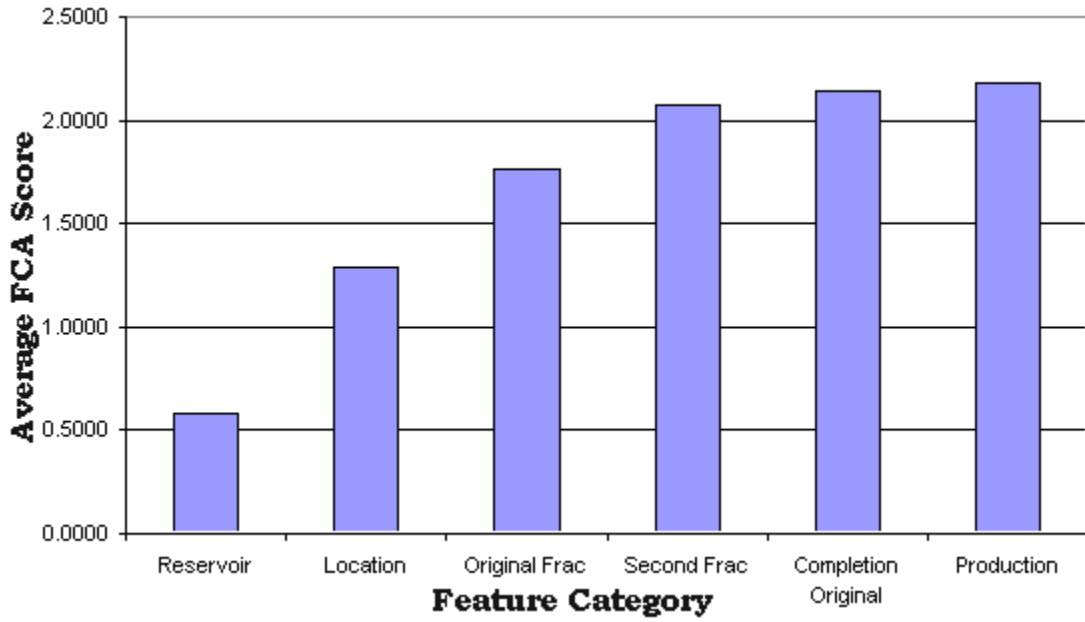


Figure 1. Influence of different categories on the post-restimulation actual peak.

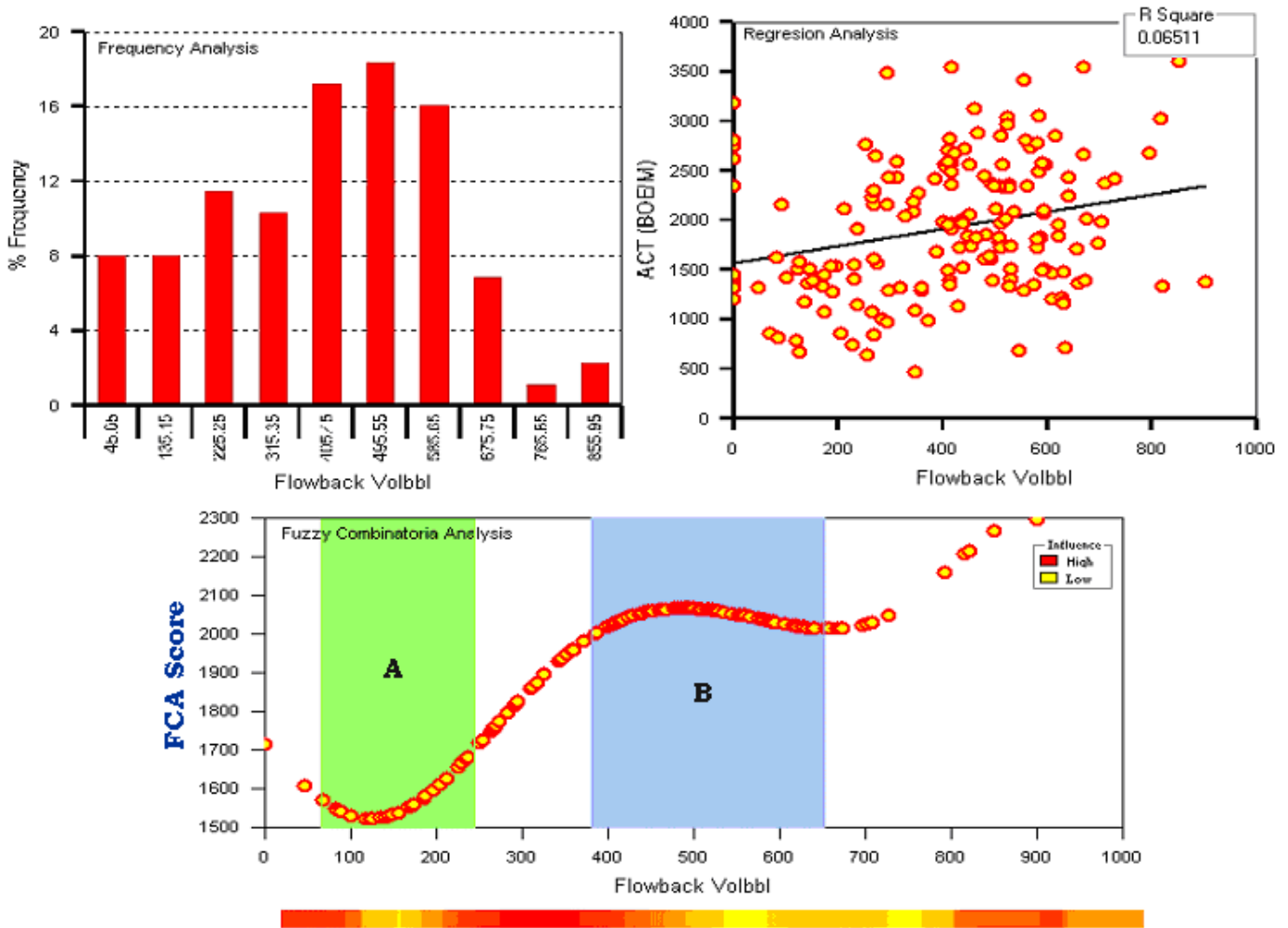


Figure 2. Conventional statistics and FCA results for “FlowBack Volume”, Ranked #1.

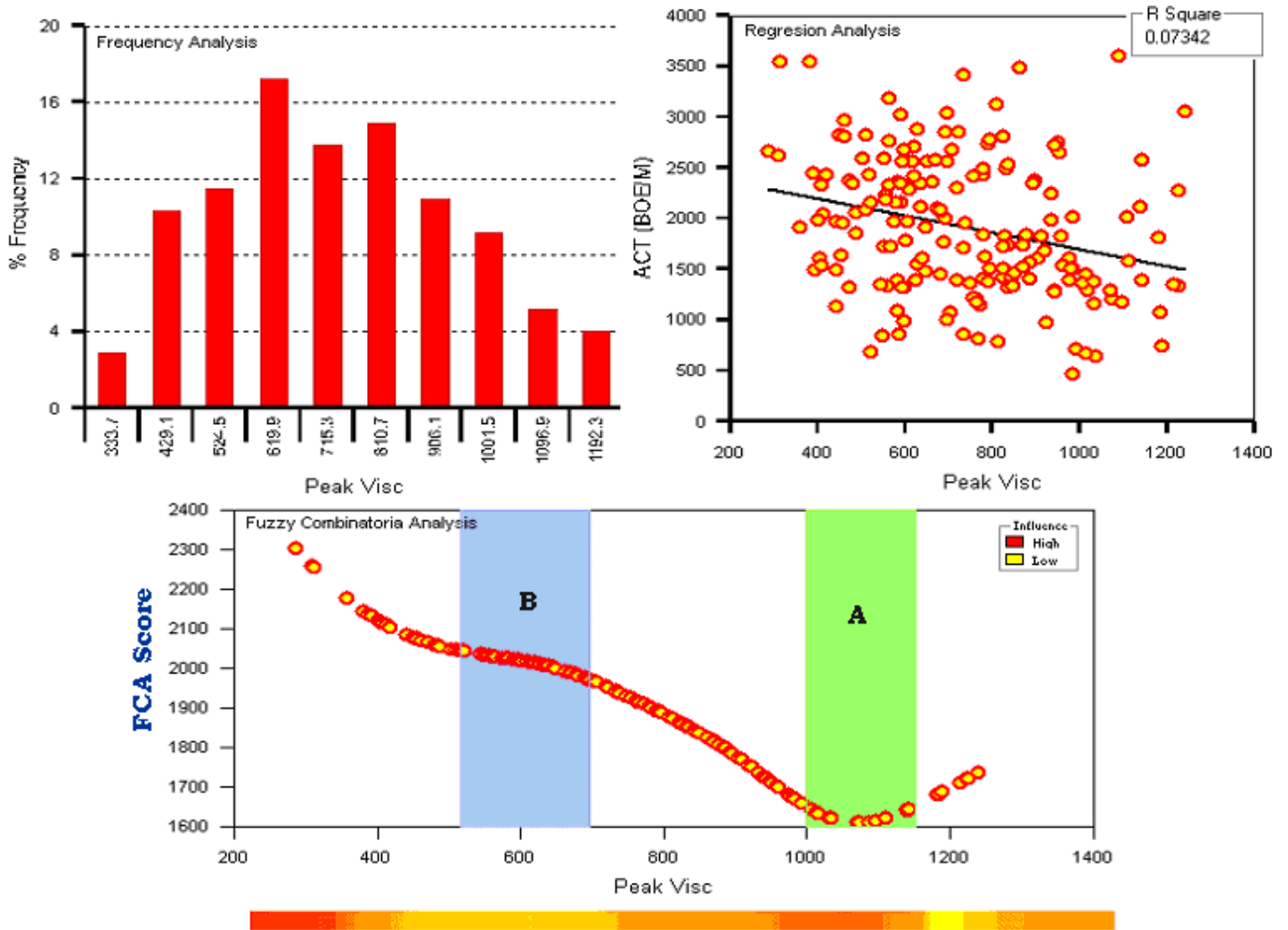


Figure 3. Conventional statistics and FCA results for "Peak Viscosity", Ranked #4.

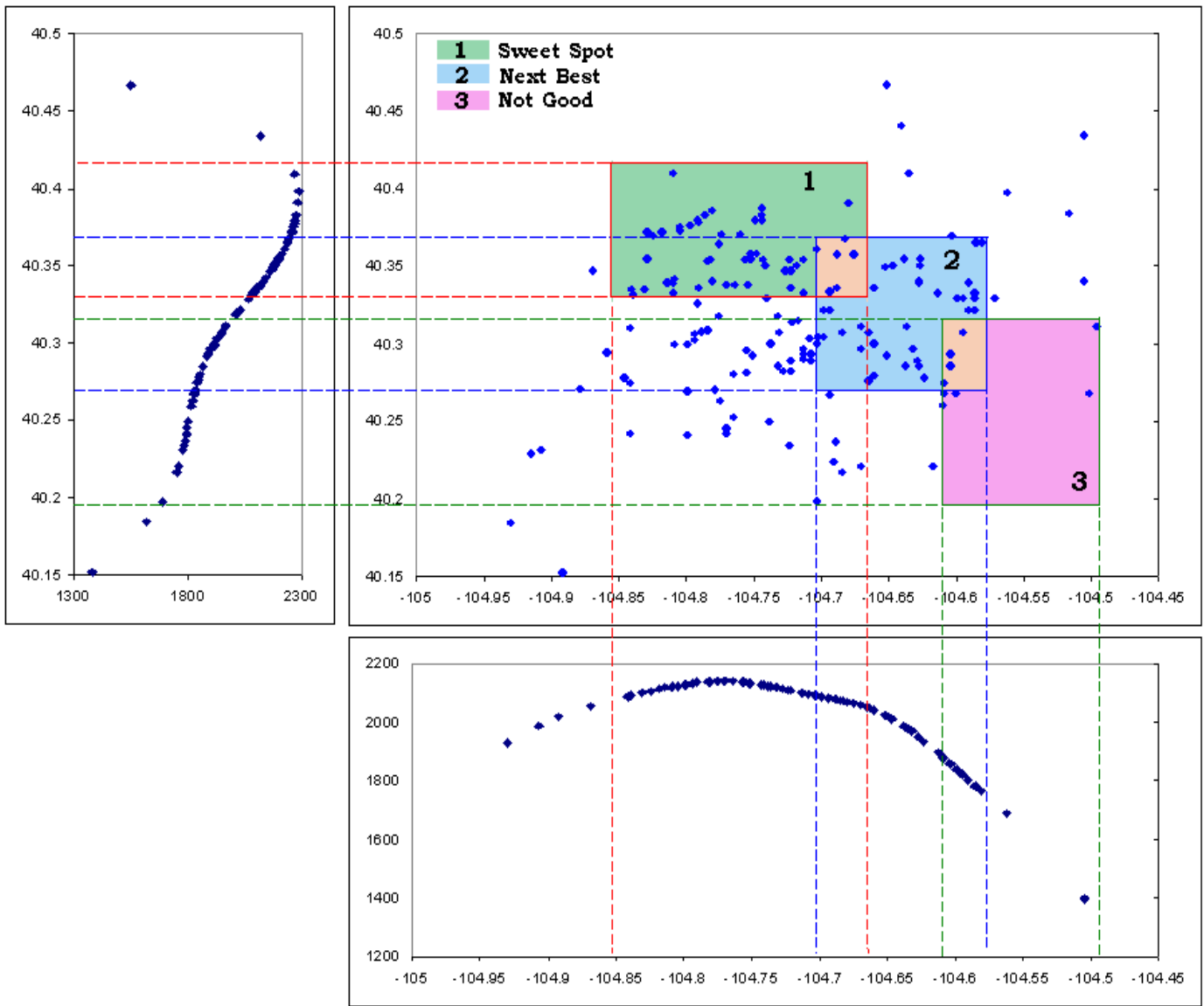


Figure 4. Graphical representation of the IPDA analysis.

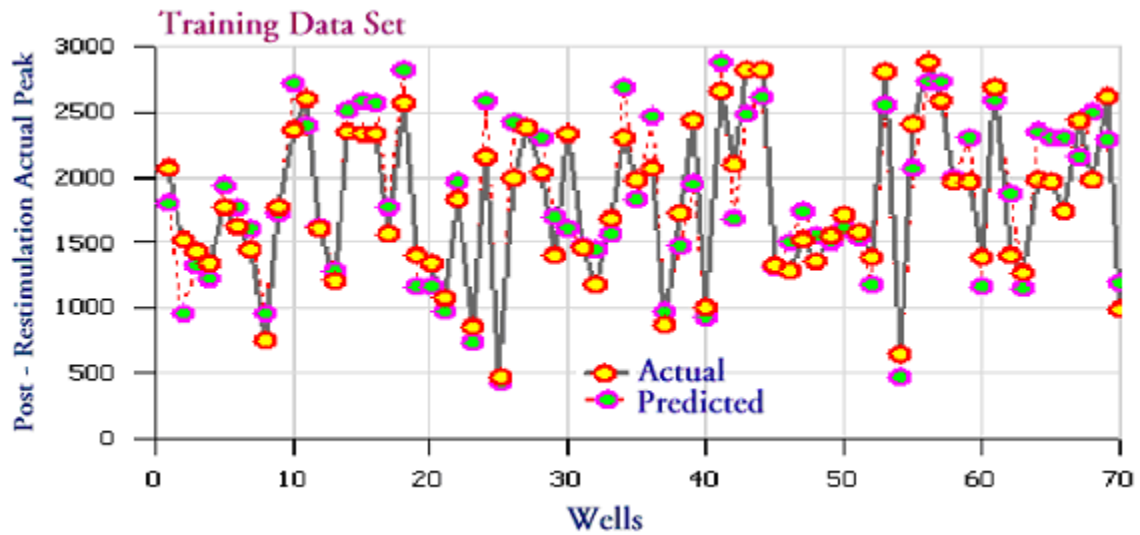


Figure 5. Actual and predicted Post-Restimulation Actual Peak for the 70 wells in the training dataset. The Pre-Restimulation neural model.

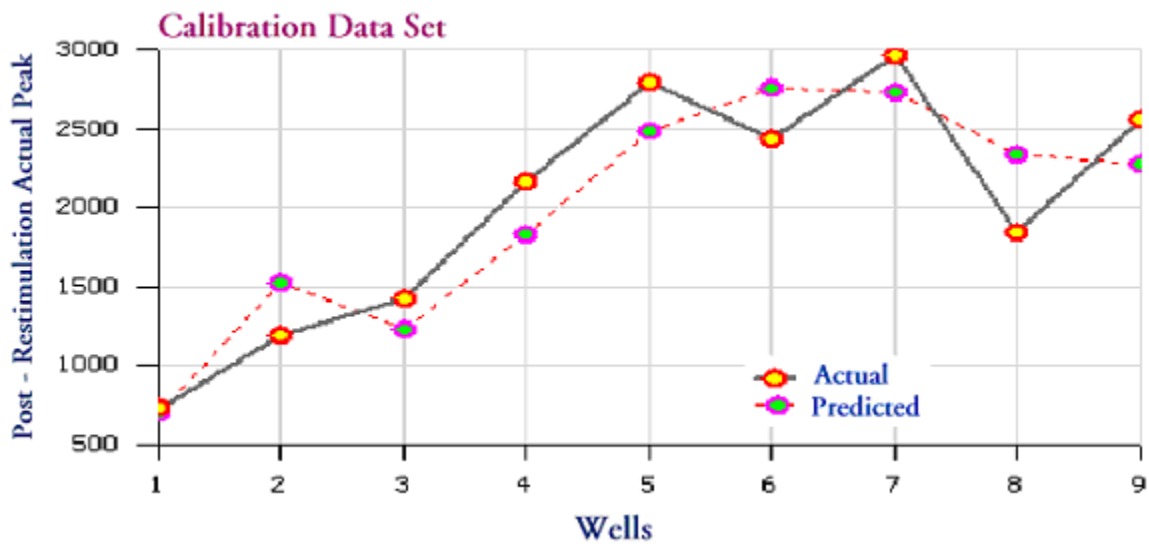


Figure 6. Actual and predicted Post-Restimulation Actual Peak for the 9 wells in the calibration dataset. The Pre-Restimulation neural model.

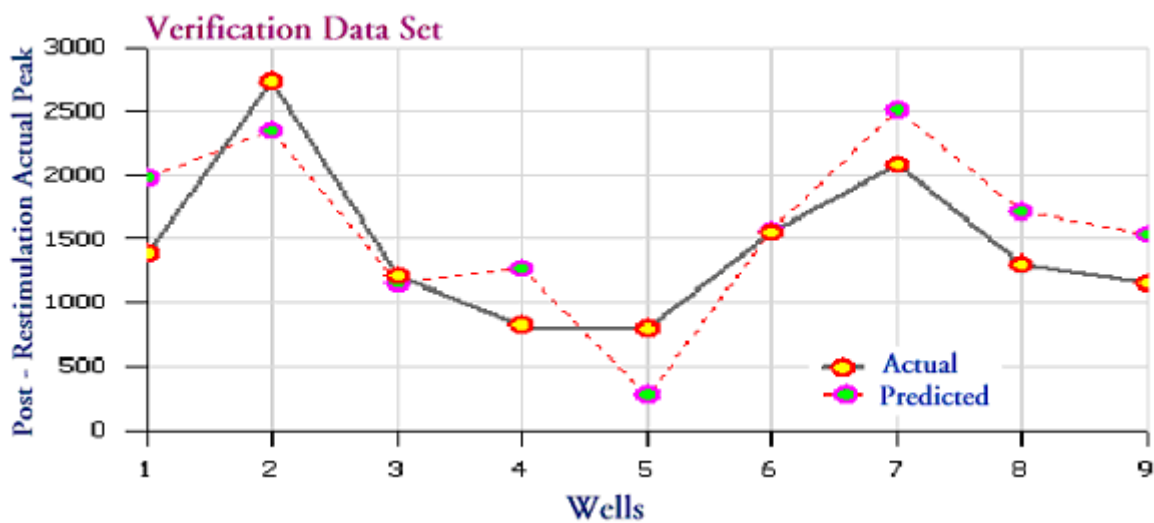


Figure 7. Actual and predicted Post-Restimulation Actual Peak for the 9 wells in the verification dataset. The Pre-Restimulation neural model.

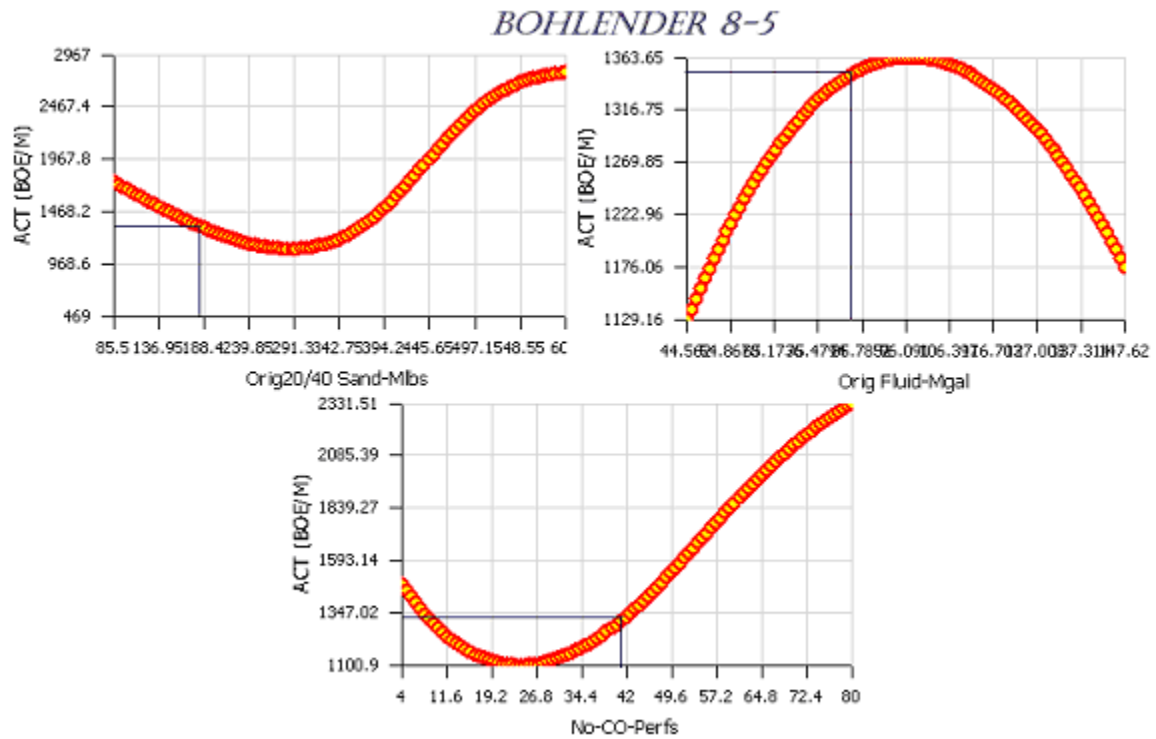


Figure 8. Sensitivity of the post-restimulation actual peak (after an average restimulation job) to different values of three parameters being studied, namely, amount of sand, amount of fluid, and number of Codell perforations for well Bohlender 8-5.

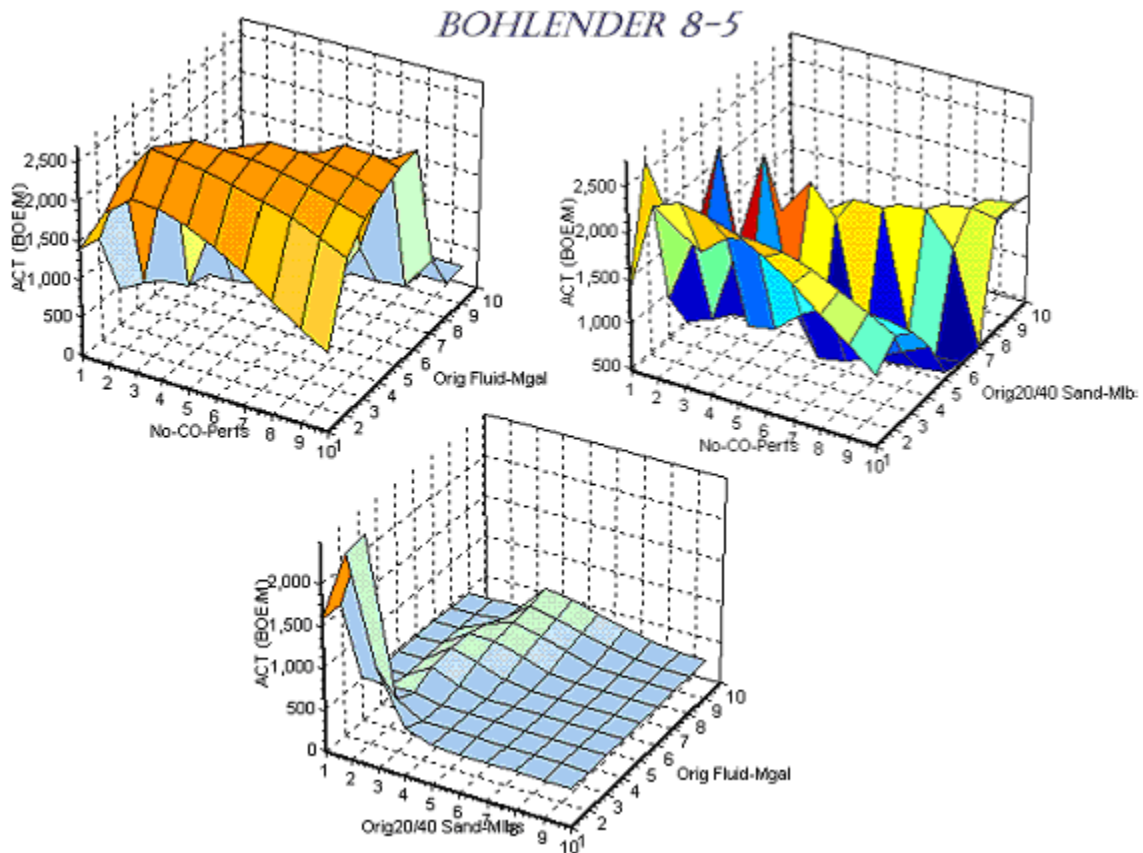


Figure 9. Sensitivity of the post-restimulation actual peak (after an average restimulation job) to different values of three parameters being studied (two variable at a time), namely, amount of sand, amount of fluid, and number of Codell perforations for well Bohlender 8-5.

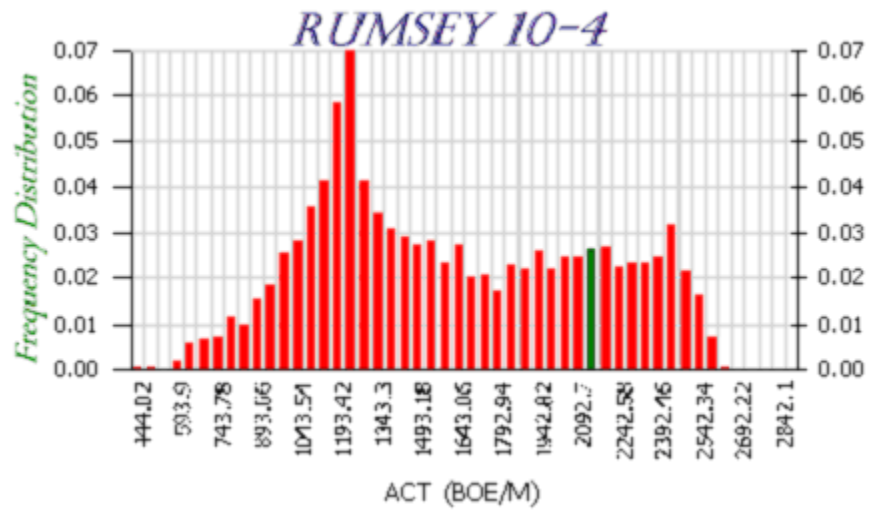


Figure 10. Results of Monte Carlo Simulation in the form of frequency distribution for well Rumsey 10-4.

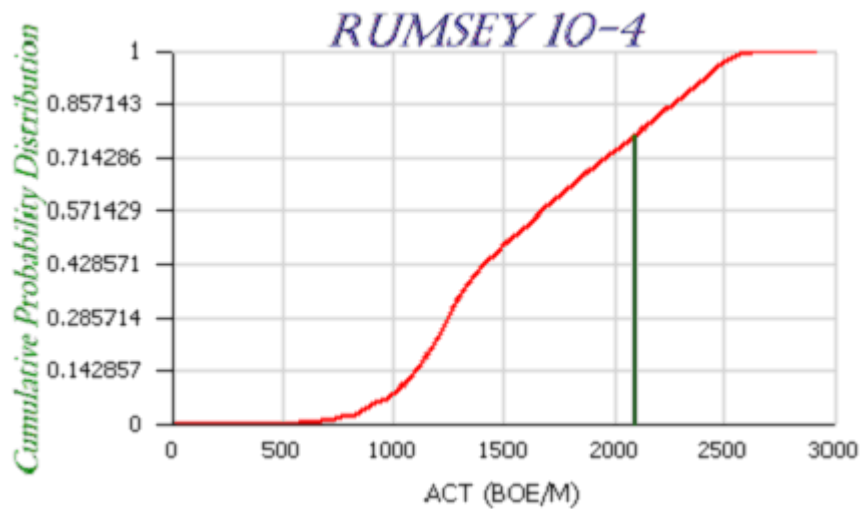


Figure 11. Results of Monte Carlo Simulation in the form of Cumulative probability distribution for well Rumsey 10-4.