Application of Statistical Analysis to Optimize Reservoir Performance

R. FORTH, B. SLEVINSKY
Petro-Canada
D. LEE, L. FEDENCZUK
University of Calgary

Abstract

In order to properly manage the depletion of a petroleum reservoir one must understand its physical characteristics and their relationship to production performance. Often the challenge is to interpret an extensive amount of available information. This paper presents an approach to this common petroleum engineering problem using statistical analysis.

Statistical analysis can provide an improved understanding of any reservoir. Beginning with the compilation of a comprehensive database, intrinsic relationships among physical parameters and production are characterized. Mapping of statistically processed data allows field-wide interpretation and visual comparison with the geological model. Probabilities of production success can be mathematically determined and used to focus optimization efforts on areas of the reservoir with favourable characteristics.

The techniques provide an unbiased appraisal of the available data that can lead to re-examination of assumptions about the underlying mechanisms governing production behaviour. Technical hypotheses can be tested for consistency by determining if expected correlations are present in the data.

A recent evaluation of the performance of a heavy oil waterflood at Golden Lake, Saskatchewan, is discussed to illustrate the application of the method. Modification of well completion practices and the reservoir depletion strategy have resulted from the study, and favourable areas for infill drilling have been identified.

Introduction

The initial challenge when analysing the performance of a petroleum reservoir is often to marshal and interpret an overwhelming amount of raw data. Technological advances in reservoir characterization and data processing have provided access to an abundant supply of numerical data. Coincidentally, the importance of understanding reservoir heterogeneity to maximize depletion efficiency is increasingly recognized. Complex variation of fluid and rock properties within reservoirs is the norm, demanding more detailed description and sophisticated technical analysis. Techniques for processing large databases and distilling critical knowledge are therefore of growing interest to the petroleum engineer.

The data overload problem may be addressed by arbitrarily discarding information or by working with averaged values. A compromise between these shortcuts and utilization of all of the data available is usually necessary; however, the quality of technical analysis may be sacrificed by over simplification. A method of extracting the important information concealed in an extensive dataset is required. Determination of the significant factors controlling production success will usually suggest the proper course to optimize the field. Statistical analysis techniques are well suited to achieving these objectives.

Statistical analysis of data is best regarded as a supplement to, not a replacement for, the engineering analysis typically undertaken to solve technical problems. As an integrated component of a study, it can generate ideas, provide direction, and confirm that the data actually supports the theoretical conclusions. The value added by statistical analysis comes from ensuring no important trends hidden in the data are overlooked, and from the improved level of confidence in the results.

Petro-Canada initiated a comprehensive technical review of the Golden Lake heavy oil field in 1994. The scope of the work included standard waterflood analysis, laboratory experiments, field tests, and numerical simulation in conjunction with the statistical analysis discussed in this paper. The overall objective was to determine why the field did not respond as expected to a pattern waterflood initiated in 1992.

Heavy oil production is technically complex, involving in this case significant variation in oil viscosity throughout the pool, foamy oil effects, production of the unconsolidated Waseca sand matrix, and apparent high permeability water conduits between wells. Over two hundred production and injection wells were directly involved in the flood patterns. Some wells exhibited improved oil rates, some suffered from premature water breakthrough, while others showed no response at all. The overall response was poorer than that of analogous fields where waterflooding was successful.

The problem was regarded as difficult since the physical production mechanisms were poorly understood and an abundance of data and conflicting evidence were present. Statistical analysis was employed in an attempt to isolate the significant parameters controlling the observed field behaviour. It was hoped that insights into the processes occurring in the reservoir would be gained by investigating the relationship of observed production performance to the geological, petrophysical, and operational data available. A secondary objective was to establish distributions and ranges for various parameters to conduct appropriate sensitivities with analytical calculations and numerical simulation.

Methodology

The general approach described in this paper is applicable to a wide range of problems. The data employed will be specific to the circumstance, thus the primary focus of the following discussion is the method. Table 1 summarizes the main steps in the process.

TABLE 1: Outline of statistical analysis process.

1. **Assembly of Database**
   - Review of Available Data
   - Selection of Variables to Include
2. **Database Quality Check**
   - Statistical Measures, Frequency Distributions
   - Removal of Bad Data, Replacement of Missing Data
3. **Univariate Analysis**
   - Variable Pair Correlations
4. **Multivariate Analysis**
   - Linear Regression Models
   - Determination of Significant Variables
   - Calculation of Studentized Residuals, Grouping Wells
5. **Discriminant Analysis**
   - Development of Discriminant Functions
   - Determination of Probabilities of Good Performance
   - Combination of Probabilities
6. **Data Contouring**
   - Areal Performance Visualization
   - Geological Model Cross Reference and Checking
7. **Non-linear Regression**
   - Determine Optimal Values of Independent Variables
8. **Interpretation of Results**
   - Explanation of Trends Using Physical Principles
   - Hypothesis Testing

Database Preparation

The first step was to select the parameters to include in the database. Careful initial consideration of parameter selection is warranted to ensure that all of the relevant data is gathered at the outset. Involvement of all project engineering staff and field operations personnel at this stage is recommended. The process of reviewing the available information and debating how various parameters are related to production performance can be beneficial in itself, as various avenues of thought on the subject are explored.

The database assembly work may account for the majority of the project time. Generally, it is advisable to cast a wide net and collect as much data as possible. Variables that could reasonably impact on the problem should not be excluded on the basis of opinion about what is significant. The objective is to find out, not pre-judge, what the data can tell about reservoir behaviour. Highgrading variables by statistical means occurs later in the process.

An extensive database was assembled on Golden Lake using a spreadsheet program. Over 140 variables were assigned to the 372 wells included in the study. Table 2 contains a partial listing of the variables that were selected. The data can be classified into two broad categories: (1) well completion configuration and production history parameters that change with time, and (2) geological, petrophysical, and geometrical parameters considered constant during the field life.

TABLE 2: Partial list of variables.

- Well Location Coordinates
- Well Status
- Completion Date
- Position in Waterflood Pattern (corner, edge)
- Proximity to Bottom Water
- Primary Cumulative Oil
- Primary Cumulative Watercut
- Average Primary Oil Rate
- Average Primary Water Rate
- Recovery Factor
- Net Pay (by flow unit)
- Average Porosity (by flow unit)
- Water Saturation (by flow unit)
- Dead Oil Viscosity
- Formation Top
- Past History Code (previous injector, fireflood well, observation well, etc.)
- Perforation Charge Size and Density
- Workover Count
- Waterflood Cumulative Oil
- Apparent Breakthrough Time

The production history data was taken at monthly intervals during three periods of the life of the field: primary production, waterflood initiation to breakthrough, and post-breakthrough. The production data was cross referenced by date with well perforation configuration for wells which had produced from multiple zones during their history in order to properly determine cumulative zone volumes. Average oil rate, water rate, and watercut values were calculated by well for each period.

A detailed geological model prepared in conjunction with the technical study subdivided the Waseca sand into four flow units, designated Upper Waseca, and Lower Waseca 0, 1, and 2, as illustrated in Figure 1. Average reservoir parameters were assigned to individual flow units at each well.

Univariate Analysis

The compiled database was checked for completeness and quality to promote confidence in subsequent results. Minimum, maximum and average values and standard deviations were calculated for each variable. Frequency distributions were generated as a basis for removal of bad observations (outliers having identifiable causes). Correlations among pairs of variables were tested to identify any strong relationships; this was done for the periods prior to and during the waterflood to ascertain the impact of the flood on average field performance. Spurious values within the production data known to result from accounting practices rather than field performance were removed.

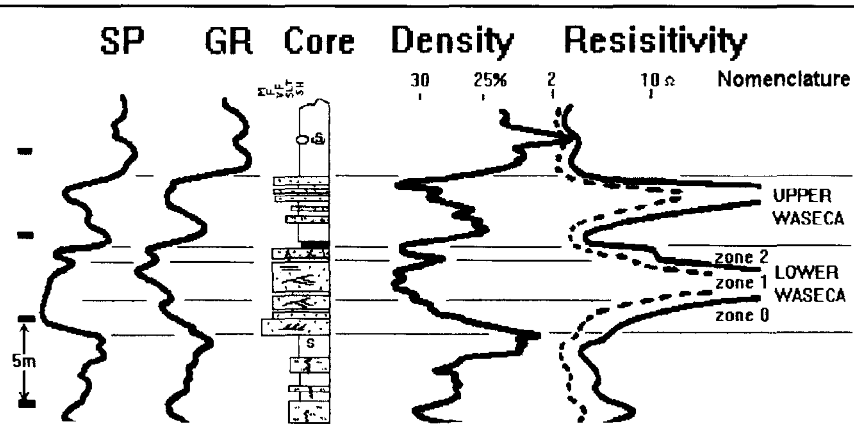


FIGURE 1: Golden Lake's Waseca formation.

TABLE 3: Well grouping by studentized residuals.

Well Classification	Studentized Residuals
Group N - negative	<= -1
Group O - neutral	> -1, < +1
Group P - positive	>= +1

Multivariate Analysis

The multivariate analysis locates correlations among several variables that are not necessarily evident from univariate analysis[1].

Linear regression models were developed to determine if reasonable predictions could be made for key dependent variables as functions of independent geological and petrophysical parameters. Five dependent variables were chosen for modelling: oil rate, water rate, watercut, recovery percentage, and waterflood breakthrough time. The general form of a linear regression model is shown in Equation (1).

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 \quad \text{(1)}$$

where; y is the dependent variable
$x_1 \ldots x_n$ are the independent variables
$a_0 \ldots a_n$ are the regression coefficients

Development of a linear regression model containing only the significant independent variables (those with large regression coefficients) is the goal of the multivariate analysis procedure. The independent variables contained in the final models were selected with a stepwise regression procedure. In some cases an R-squared procedure was used to confirm the validity of the elimination process. The final models were developed by the iterative elimination of insignificant variables, and standardized estimates were then used to evaluate the relative importance of the regressor coefficients. Standardized estimates of coefficients are obtained by performing a regression on standardized data having zero mean and unit standard deviation[1]. Thus the magnitudes of the resultant coefficients are not affected by the ranges of the original variables and reflect the influence of each independent variable on the modelled dependent variable.

As is common in petroleum engineering projects, data was not available for every variable for all of the wells. For example, core data was not always present. Thus, a sparse data matrix resulted. Initially, multivariate analysis was attempted on the original dataset without replacement of missing values. This eliminates wells with missing data from the process of selecting the most significant variables from the set of all variables. Including fewer wells degrades the confidence in the results, and makes the models less stable. To circumvent the problem, several stages of stepwise regression were performed on smaller sub-sets of variables. The significant ones identified from these sub-sets were potential candidates for the complete set necessary to explain the observed variability in the modelled parameters.

The models generated by this technique consisted of three to seven independent variables and were based on approximately fifty wells. Although these models accounted for a considerable portion of the variability, they were unstable when any observations were added or deleted. In order to support multivariate analysis, it was therefore necessary to fill in the missing values.

Missing data was replaced using a grid interpolation process available in the SAS/GRAPH® module. Each variable was gridded on a 100 m square pattern overlaid on the field map. The missing grid point values were then added by linear interpolation, but kriging techniques can be used, if preferred. Missing well data was then assigned based on the value of the variable at the nearest grid point.

Ideally, values for all of the variables would be present for each well, and the process of grid interpolation would be unnecessary. Although the data replacement process is not without controversy, it allowed the inclusion of 160 wells in further model development and made possible the identification of significant parameters using linear regression models. The importance of thorough data collection during the lifetime of a field is highlighted when statistical methods are applied.

Another complication with the data was distinguishing missing data from true zero values. The geological flow units were discontinuous throughout the field, requiring input of true zero values for reservoir parameters for flow units not present at a given well.

Tests were applied to ascertain how well the models fit the data. The general linear additive models, of the form shown in Equation (1), accounted for only a portion of the total variance (R-squared values of 0.1 to 0.3) and represented a poor fit. Selection of significant sub-sets of variables, those with regression coefficients not close to zero, necessary to predict dependent variables, was still possible, however.

The next step was to generate studentized residuals for the five models. Residuals are the difference between the model predicted values and the observed values for the dependent variable. Studentized residuals are produced by dividing the residuals by the standard error of the residuals.

The studentized residuals were used to classify observations (wells) into three categories as summarized in Table 3. Those with negative residuals greater than one formed group N, those with positive residuals greater than one comprised group P, and those in between were group O. Linear regression models were then developed for these groups. The fit of these models to the data was considerably better (R-squared values of 0.60-0.80). On the basis of selected cutoff values for the dependent variables, the wells in group O were reclassified into groups N and P when required.

Discriminant Analysis

Discriminant analysis is a technique used to classify observations into groups based on selected parameters. The factors common to the wells contained in the above groups were identified. Of particular interest was finding the characteristics associated with wells that tended to be good performers, that is, those with high oil rates, low watercut, and high recovery factors.

Discriminant analysis employs a learning data set to develop discriminant functions and a testing data set to compare the predictions of the functions to actual observations. The DISCRIM procedure[2] in the SAS/STAT® module allows the use of the same data set for both functions using the cross-validate option. Alternatively, observations from one of the well groups could be used as the learning set, and the resultant function tested against observations in the other groups.

Since many of the variables involved were not found to be normally distributed, a non-parametric discriminant procedure had to be used. Discriminant functions were developed to classify wells as either good or bad performers. Stepwise discrimination was initially applied to find the most significant parameters supporting this discrimination. The probability of membership in the group of good performers was next determined.

A mapping program was used throughout the project to generate contour maps of variables to assist in visualization of areal distributions. Figures 2 through 5 illustrate the type of maps generated for the Golden Lake South Waseca pool. The calculated probability of obtaining good oil rates, for example, is displayed in Figure 3.

A further step was taken in combining the probability maps obtained from separate discriminant functions through the use of Boolean operators[3] (union, difference, complement and intersect). Since the discriminant functions are based on several variables the process of generating the combined probability maps is equivalent to simultaneously overlaying contour maps of the significant variables on a light table, and identifying areas with consistently favourable values. Figure 4 shows the combined probability of good oil rates, low watercuts, and high recovery percentages. The map of combined probability of production success can be used to focus optimization efforts and choose infill drilling locations.

Valuable synergies were obtained by virtue of conducting the statistical analysis in parallel with the geological study of Golden

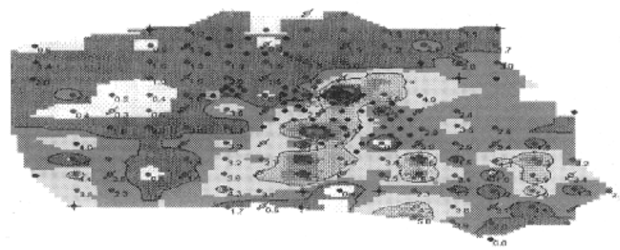


FIGURE 2: Golden Lake's averaged produced oil rate 1991, m^3/day.

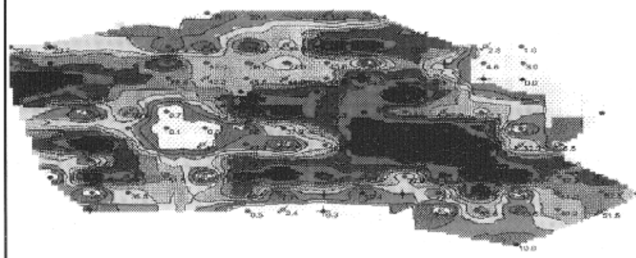


FIGURE 3: Golden Lake's relative probability for producing good oil rates.

Lake. Standard geological flow unit maps from seismic, log, and core data were compared with contour maps of variables generated from the statistical analysis, and visually recognizable patterns of similarity were discovered. Attempts to relate areas exhibiting certain production characteristics such as poor oil rates to geological features, such as tight streaks, post depositional channels, flow unit discontinuities, and erosional features, were generally successful.

Intuitively, variations in well production behaviour should arise from a fundamental relationship between reservoir performance and geological heterogeneity. Thus, contour maps of various production parameters seem to reveal underlying areal geological heterogeneity on the scale corresponding with the well spacing. The ability to visualize the physical and performance characteristics in this manner added significantly to the technical understanding of the reservoir. For example, areas of good waterflood performance were visually similar to areas of good primary production leading immediately to the useful understanding that poor primary performance does not indicate untapped potential exploitable by waterflood.

The graphical comparison of the production behaviour to the geological model contributed to the refinement of the latter. Closing the loop between the geology and the production performance in this fashion provided a sense of closure and an improved level of confidence in the final product.

Non-linear Regression

An interesting prediction of the linear regression model for oil rate was that perforation density and charge size are governing factors. A closer investigation of this relationship was undertaken using a second order non-linear regression model of the general form illustrated in Equation (2).

$$y = a_0 + a_1 x_1 + a_2 x_1^2 \qquad (2)$$

where, y is the dependent variable
x_1 is the selected independent variable
$a_0 \ldots a_2$ are the regression coefficients

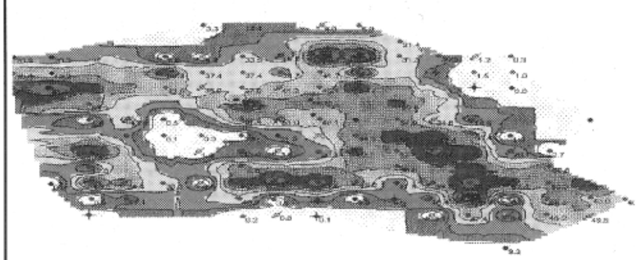


FIGURE 4: Golden Lake's relative probability for producing good oil rates, water cuts and per cent recoveries.

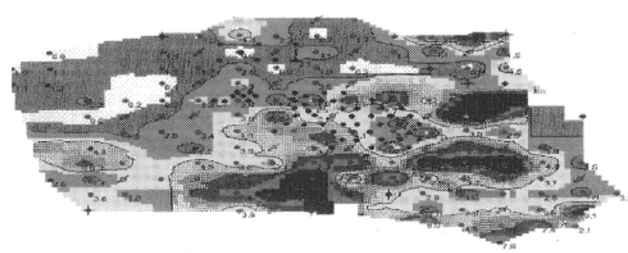


FIGURE 5: Golden Lake's averaged produced oil rate 1994, m^3/day.

The results indicate that there is an optimum charge size and that the perforation density should be maximized to realize the full production potential of a well. This result can be interpreted within the context of near wellbore production mechanisms postulated for heavy oil[4] in unconsolidated reservoirs. Viscous heavy oils are known to transport unconsolidated formation sand to the well, and the grains can form bridges across the perforations. Heavy oil wells often exhibit unstable production and tend to sand in periodically. Both phenomena may be related to the random collapse of the sand bridges at the perforations. The stability of the sand bridges has been related to perforation size, among other factors. In addition, higher sand production has been linked to better oil production in unconsolidated heavy oil reservoirs. Sand production should be facilitated by increased perforation density.

Physical processes associated with heavy oil production are the topic of current industry research. The Golden Lake study illustrates the use of statistical analysis for practical determination of optimal values for controllable parameters, such as perforation density and charge size, that cannot presently be calculated analytically. The model predicts a statistical improvement in oil production rates of up to 30% for an optimal completion relative to the average. Completion practices have been modified for infill wells; however, to date, the number of new completions is insufficient to verify the predicted improvements.

General Interpretation of Results

The statistical analysis procedure produces a set of results requiring interpretation. The interpretation is interesting for the petroleum engineer, as it involves attempting to relate the statistical models back to engineering principles. Application of engineering judgment and practical experience is essential to ensure that meaningful conclusions are drawn and the pitfalls associated with deduction from statistics are avoided.

The following aspects of the results should be examined:

- Which variables are present in the models. These are the key factors influencing the dependent variables.

TABLE 4: Significant variables in regression models.

Oil Rate	Water Rate	Watercut	Recovery %	Breakthrough Time
Lower 2 Zone Net Pay	(Upper Formation Top)	Well Location	(Upper Waseca Porosity)	(Lower 2 Zone Sw)
(Lower 2 Zone Sw)	(Well Location)	(Perf Density)	Upper Waseca Sw	(Upper Waseca Sw)
Well Location	(Perf Charge Size)	Lower 1 Zone Porosity	(Perf Charge Size)	(Upper Formation Top)
Perf Density	(Lower 0 Zone Porosity)	Lower 1 Zone Sw	(Oil Viscosity)	
Lower 1 Zone Net Pay				

Note: Bracketed values indicate negative correlation—dependent variable increases when independent variable decreases.

- Which variables are most important, based on the magnitude of the standardized regression coefficients.
- The sign of the regression coefficients. Positive signs indicate positive correlation, and vice versa. For example, oil production rates would normally be positively correlated to permeability, and negatively correlated to watercut.
- Which variables, if any, are present in all of the models.
- Which variables are significant in both the regression models and discriminant functions.
- Which variables allow distinction between classes of wells.
- Which variables are controllable by the operator.

The interpretation continues with the development of explanations for the statistical results on the basis of multi-diciplinary knowledge of the field and accepted scientific principles. Several situations can arise during this phase of the analysis:

- The causal link is obvious, as in the correspondence of higher oil rates with better permeability. Nevertheless, areas of the field with low rates despite high permeability may be present—a discrepancy requiring consideration and explanation.
- An expected correlation is not present. Prior to embarking on a detailed statistical analysis there will likely be pre-existing ideas about causes of the observed reservoir behaviour. Consequently, certain correlations will be expected in the data. For instance, water production might be attributed to edge water influx, in which case a correlation of watercut with proximity to known edge water is expected. The absence of the correlation would provide information as significant to the interpretation as the identification of a strong correlation, refuting or confirming the original hypothesis respectively.
- The significant factor or factors among many possible causes for a given observation is identified. For example, oil rates may be strongly correlated to the characteristics of one geological flow unit, but not the others. This result may indicate the layer responsible for contributing most of the production.
- The data suggests a behaviour which is not easily explained or contradictory to accepted engineering principles. The data can force a re-thinking of the production mechanism that may ultimately yield useful new knowledge. Assuming the integrity of the data leading to the anomalous conclusion is confirmed, hypotheses to explain the results should be developed and field tested. In Golden Lake, the data indicates that oil rates may not increase as wells are pumped off. This is contrary to both the fluid flow equations and intuition. In this case, it is speculated that the oil production may be hindered by dramatic viscosity increase occurring at very low pressures, as the well is pumped faster to lower the fluid level down to the pump intake. The suspected viscosity increase as the oil foams has been observed in the laboratory. The data shows a negative correlation between pump efficiency and stroke rate, suggesting that foam is increasingly present at the pump when a well is sped up to a lower fluid level. No testing has been undertaken to validate this hypothesis to date, but the statistical process indicates the need for such a test.
- A false deduction is made from valid data. An illustration is the dependence of cumulative recovery on the time a well has been on production. If the database did not include production time, new wells or wells with low cumulative production time and low recovery could be erroneously classified as having poor recovery due to poor reservoir quality.

Interpretation for Golden Lake

As mentioned previously, models were created to determine the primary factors influencing oil rate, water rate, watercut, recovery percentage, and breakthrough time. Table 4 summarizes the significant variables in each model. The sign and magnitude of the standardized coefficients indicate the direction of influence and relative importance of each independent variable on the five dependent variables.

Oil Rate

The most important factors here are the pay thickness and water saturation in the Lower Waseca 2 flow unit, the east-west location in the field, the net pay thickness in the Lower Waseca 1 flow unit, and the perforation charge density. The net pay in flow units 1 and 2 may be interpreted to provide most of the oil production from the Waseca; this is consistent with the flow unit permeabilities observed. That the water saturation in flow unit 2 is inversely correlated to oil rate is interesting since this factor was not significant to water rate or watercut. The relationship may be related to lower oil reserves where water saturations are higher. The wells toward the southeast end of the field tend to have higher oil rates, a reflection of both better reservoir quality, and less pressure depletion corresponding to the more recent development of the south portion of the field. The impact of perforation density was discussed in the non-linear regression section.

Water Rate

This is affected primarily by the structural elevation of the Upper Waseca zone, the east-west location in the field, the perforation charge size, and the porosity in the Lower Waseca 0 flow unit. The structural connection with water rate makes sense since there is bottom water located around the edge of the pool. The field location influence may relate to the development timeframe or to the presence of perched water and edge water on the east side of the field. The porosity in the Lower Waseca 0 zone may be a factor in water rate because the bottom flow unit provides a path for the bottom water to move up structure; the water has slightly greater density than the oil. The inverse relationship of perforation charge size to water rate is not well understood, but may be associated with higher oil rates being produced with larger perforations, or evolution of perforating techniques as the field was developed.

Watercut

This is strongly influenced by location in the field and perforation density, which may be explained in the same way as the rela-

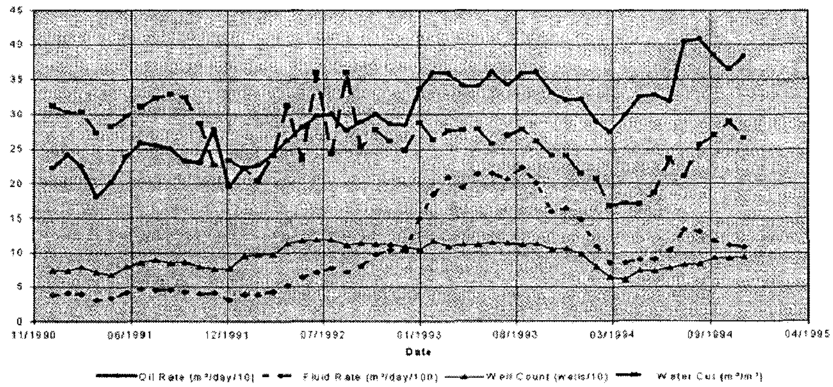


FIGURE 6: Golden Lake's production history.

tionships to water rate. Additionally, the porosity and water saturation of the Lower Waseca 2 flow unit are proportional to watercut. This suggests that connate water is mobilized in this zone along with the oil which may explain why wells that are high on structure and not adjacent to known bottom water exhibit increasing watercuts during primary production. Proximity to bottom water relates to higher watercuts, perhaps because water mobilized in the formation generally provides a conduit for bottom water to encroach.

Recovery Percentage

This is positively correlated to the water saturation and inversely related to the porosity in the Upper Waseca zone. The correlation may be due to the sporadic completion of the zone—where it contains a significant oil volume, the overall recovery from the entire section appears lower. As expected, the recovery percentage is inversely related to the dead oil viscosity which varies significantly throughout the field and is well correlated to structure. Perforation charge size is inversely related to recovery percentage for unknown reasons.

Breakthrough Time

This was established from changes in the watercut trend of each producing well following initiation of water injection. Breakthrough time is inversely correlated to structural elevation, and the water saturation in the Upper Waseca and Lower Waseca 2 zones. The correlation indicates that more water initially present results in quicker breakthrough. This is consistent with the hypothesis that primary production mobilizes water in place. The correlation to elevation is partially due to the quick breakthrough in the structurally high centre of the north pool which was previously waterflooded (from 1971 to 1981).

Conclusions about the Golden Lake field were derived from a multi-faceted technical study that included geological modelling, field testing, laboratory experiments, and numerical simulation in addition to the statistical analysis. The information from the statistical work was used as a cross-check and catalyst for further analysis; thus it served to both support and direct the other work. The statistical findings were tested in the numerical reservoir simulation model to confirm that it performed in a manner consistent with the data. The statistical distributions of variables were used to select appropriate ranges for simulation sensitivities. This linkage to the data provided some validation of the model, and narrowed the options considered for obtaining a history match.

It is unlikely that any single component of the technical study would have provided the understanding obtained from the synergetic combination.

In summary, the main conclusions are:

- The statistical analysis augmented the technical study of Golden Lake by providing insights into information contained in the field data that are not apparent from unaided inspection.
- The significant variables impacting on key performance variables (oil rate, water rate, watercut, recovery factor, and breakthrough time) have been identified, as shown in Table 4.
- The waterflood breakthrough mechanism was water mobilized by primary production and prior waterflood projects in the field. Water breakthrough was related to a free gas saturation created by primary pressure depletion, as shown by pattern similarities in the contour maps of water production during the flood and primary oil production. Overall flood performance was impacted negatively by viscosity variation which is strongly correlated to structural elevation within the pool and to the degree of pressure depletion which correlates to geographic location of wells within the pool. These conclusions do not arise from the statistical analysis alone, but from the integrated technical study. Taken together, these results were sufficient to make it clear that the waterflood could not succeed or meet its performance expectations, thus it was terminated.
- Completion of wells could be optimized by using 30 to 34 gram perforation charges and high shot density.
- Selection of favourable areas to conduct further development work is possible utilizing the probability of success maps and the conclusions from the statistical models. Completion of the Upper Waseca zone in some wells may provide additional reserves, and infill drilling if economically justified can now be targeted within the field.

The Golden Lake Waseca waterflood was terminated based on the results of the technical study and the improved understanding of the production mechanism. The statistical analysis contributed significantly to the new understanding by forcing a congruence with the that returning the field to primary depletion evidence of the raw data. Figure 6 illustrates the positive impact on oil volumes over the period from March 1994 to January 1995. Substantial operating cost reductions were also achieved. The field is presently being optimized on primary production, and may be a candidate for application of future EOR technologies or infill drilling.

Limitations and Cautions

Appropriate use of statistical analysis for resolving petroleum engineering problems requires awareness of its limitations. Valid concerns may exist about the quality and completeness of available data. However, used with appropriate caution, the techniques are not only useful, but may be critical to the solution of problems involving large quantities of data.

The capability of the statistical process can be limited by the incompleteness of the database, and error/noise in the data that will always be present to a degree. Dynamic changes in the production parameters are lost due to taking averages over time. Practical aspects of data collection may have an influence; for example, production volumes may be allocated back to individual wells based on infrequent capability tests. Factors such as regulatory limits on production or facility constraints may obscure true reservoir performance capability. The geological and petrophysical parameters are necessarily derived from minute samples of the reservoir that are assumed to be widely representative. The analysis results reflect both reservoir behaviour and the process of collecting and reducing data. These considerations should be kept in mind.

Real reservoirs are typically heterogeneous on a much smaller scale than the inter-well distances. Contours of variables based on data at well spacing intervals do not capture inter-well variations in physical parameters.

A major concern is the attribution of trends in the data to the wrong causes. For this reason it is advisable to consider all possible factors that might result in observed correlation and devise tests to confirm the conclusions. The tests may be field trials or statistical tests on a modified or expanded set of data.

Relationships revealed by statistical procedures are valuable for obtaining a better general understanding of the behaviour of the reservoir but can only predict specific outcomes in the probabilistic sense.

Conclusions

- Statistical analysis techniques can complement conventional engineering analysis, and are increasingly worthwhile in addressing complex problems where large quantities of data are available.
- Examination of the data using statistical techniques can reveal relationships apparently influencing reservoir performance which might not otherwise be considered, and can indicate the significance of various controlling factors.
- In situations where the physical mechanisms that cause a given behaviour to occur are poorly understood, statistical results may suggest or refine avenues for investigation.
- The reasons for the poor performance of the Golden Lake waterflood are now understood. The difficult decision to terminate the project was taken with confidence based on a new technical understanding grounded on all of the available data. Through statistical analysis, the data was translated into useful information that supports the decision.
- The interaction between the applied sciences of engineering, geology and statistics proved to be invaluable in directing the analysis and interpreting the results.

Acknowledgements

The authors wish to thank Petro-Canada for permission to publish this paper. We would also like to acknowledge the contributions of John Dwyer, P.Geol., in supplying detailed geological parameters for the database. Also, thanks to the members of the Golden Lake Asset Team.

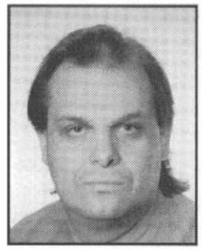
REFERENCES

1. J.C. DAVIS, Statistics and Data Analysis in Geology; *John Wiley & Sons, New York, ISBN 0-471-08079-9.*
2. SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1, The DISCRIM Procedure; *SAS Institute Inc., Cary, NC, ISBN 1-55544-376-1, pp. 677-772, 1990.*
3. L. FEDENCZUK, B. WYVILL, K. HOFFMANN, Scientific Visualization in Oil and Gas Exploration; *SAS User Group International Nineteen Annual Conference, Dallas, TX, April 1994, pp. 675-680 (invited paper).*
4. Mechanisms and Management of Sand Production; *Centre for Frontier Engineering Research, 1994.*

Provenance—Original Petroleum Society manuscript, **Application of Statistical Analysis to Optimize Reservoir Performance**, (95-32), first presented at the 46th Annual Technical Meeting, May 14-17, 1995, in Banff, Alberta. Abstract submitted for review October 25, 1994; editorial comments sent to the author(s) January 8, 1996; revised manuscript received February 16, 1996; paper approved for pre-press February 28, 1996; final approval March 22, 1996.

Authors' Biographies

Leon Fedenczuk holds his B.Sc. in electrical engineering, B.Sc. in computer science, and M.Sc. in physics. He is currently a Ph.D. candidate at the University of Calgary in the Department of Computer Science. Before starting his consulting career, he was involved in research in plasma chemistry, semiconductors, and computer applications in the petroleum industry. He is interested in scientific visualization of multidimensional data analysis through dimension reduction techniques. He is developing adaptive stochastic modelling methods for scientific visualization with emphasis on porous media.



Dale Lee received a Bachelor of Science degree from the Department of Chemical and Petroleum Engineering at the University of Calgary. He is currently completing a M.Eng. degree. Dale joined Petro-Canada in 1988, where his interests are in reservoir engineering.

Ron Forth received his Bachelor of Science in mechanical engineering from the University of Alberta in 1981. He has worked in various reservoir engineering assignments with Petro-Canada since 1982, and is currently a reservoir engineering specialist in the Oil Business Unit.

Bruce Slevinsky is senior advisor, reservoir engineering for Petro-Canada. Since joining the company in 1990, he has worked in the Technology Improvement Department providing consulting services to the business units and developing enhanced oil and gas recovery technologies and engineering software. Past experience includes 16 years of reservoir engineering for a major oil company and a Canadian consulting company. He received a B.Sc. in mechanical engineering (1971) and an M.Sc. in applied mechanics (1972) from the University of Alberta.