
API BUL D14
Second Edition
April 30, 1984

STATISTICAL ANALYSIS OF CRUDE OIL RECOVERY AND RECOVERY EFFICIENCY

OFFICIAL PUBLICATION



REG. U.S. PATENT OFFICE

AMERICAN PETROLEUM INSTITUTE

Washington, D.C.



Issued by

AMERICAN PETROLEUM INSTITUTE

Production Department

211 North Ervay, Suite 1700

Dallas TX 75201

Copyright © 1984 by American Petroleum Institute

Users of this publication should become familiar with its scope, content, and inherent limitations.
This publication is intended to supplement rather than replace individual engineering judgment.

Documents

Document Sharing Hub

STATISTICAL ANALYSIS OF CRUDE OIL RECOVERY AND RECOVERY EFFICIENCY

TABLE OF CONTENTS

	Page
Foreword	4
SECTION I	
Summary and Conclusions	5
SECTION II	
Review of <i>API Bulletin D14</i> , October 1967 Data and Correlations	7
A. The Weighted <i>Bulletin D14</i> (1967) Correlations	7
B. Analysis of the Unweighted <i>Bulletin D14</i> (1967) Data	8
C. Correlation of Recovery Efficiency	9
SECTION III	
Fitting the Updated Data Sample to a <i>Bulletin D14</i> (1967) Format.....	10
SECTION IV	
Search for a New Multi-Variable Regression Format	12
SECTION V	
Correlation Based on Geographical Province	13
SECTION VI	
Secondary Recovery	17
A. Introduction	17
B. Pattern Waterflooding.....	17
C. Edge Water Injection	17
D. Gas Injection Into Gas Cap.....	18
SECTION VII	
Possible Reasons for Failure to Develop Adequate Empirical Correlations	19
APPENDIX A	
Some Statistical Principles	20
A.1 Introduction	20
A.2 The Principle of Least Squares.....	20
A.3 Standard Error or Variance Estimate.....	20
A.4 The Correlation Coefficient	25
A.5 Illustrative Examples	25
A.6 Rejection of Data Outliers	29
APPENDIX B	
Development of Alternative Multi Variable Correlations by Stepwise Linear Regression.....	30
B.1 Introduction	30
B.2 Selection of Samples	30
B.3 Stepwise Regression Procedure	30
B.4 Analysis of Solution Gas Drive	31
B.5 Analysis of Water Drive	35

TABLE OF CONTENTS

(Continued)

APPENDIX C	
Illustrations Depicting Recovered Oil Versus Original Oil-in-Place	39
APPENDIX D	
Nomenclature	44
D.1 Engineering Symbols and Definitions	45
D.2 Probability and Statistical Symbols and Definitions	46
APPENDIX E	
References	47
ILLUSTRATIONS	
Fig. II-1: Solution Gas Reservoirs Unweighted <i>Bulletin D14</i> Samples	9
Fig. V-1: Texas-Sandstone Natural Water Drive	14
Fig. V-2: Histograms of Reservoir Parameters	16
Fig. VI-1: Histograms of Secondary Recovery Reservoir Parameters	18
Fig. A-1: Probability Distribution Plot: Percentage Distribution of Residuals, Texas-Sandstone Solution Gas Reservoirs	21
Fig. A 2: Distribution of Residuals, Texas-Sandstone Solution Gas Reservoirs	22
Fig. A-3:	23
Fig. A-4:	23
Fig. A 5:	24
Fig. A-6:	24
Fig. A 7:	24
Fig. A-8: Data and Regression Equation for Illustrative Example, "Data Set 1"	26
Fig. A-9: Data and Regression Equation for Illustrative Example, "Data Set 2"	27
Fig. C-1: Primary and Primary Plus Secondary Recovery, California- Sandstones Pattern Waterflood	39
Fig. C 2: Primary and Primary Plus Secondary Recovery, Louisiana- Sandstones Pattern Waterflood	40
Fig. C-3: Primary and Primary Plus Secondary Recovery, Oklahoma- Sandstones Pattern Waterflood	40
Fig. C-4: Primary and Primary Plus Secondary Recovery, Texas- Sandstones Pattern Waterflood	41
Fig. C-5: Primary and Primary Plus Secondary Recovery, Wyoming- Sandstones Pattern Waterflood	41
Fig. C-6: Primary and Primary Plus Secondary Recovery, All Carbonates Pattern Waterflood	42
Fig. C-7: Primary and Primary Plus Secondary Recovery, Texas- Sandstones Gas Cap Injection	42
Fig. C-8: Primary and Primary Plus Secondary Recovery, Louisiana- Sandstones Edge Water Injection	43
Fig. C-9: Primary and Primary Plus Secondary Recovery, Texas- Sandstones Edge Water Injection	43

STATISTICAL ANALYSIS OF CRUDE OIL RECOVERY AND RECOVERY EFFICIENCY

Foreword

a. This bulletin was prepared by the API Subcommittee on Recovery Efficiency working under the administration of the American Petroleum Institute Production Department's Executive Committee on Drilling and Production Practice.

b. *API Bulletin D14: A Statistical Study of Recovery Efficiency*, First Edition, October 1967, presented conclusions and equations which evolved from a study undertaken by the API Subcommittee on Recovery Efficiency in 1956. The Subcommittee's assignment then was to study reservoir recovery processes based on actual performance data from producing fields rather than on theoretical or laboratory data. The objective of the current study was to expand the data base of the original study and study oil recovery based on actual field performance data.

c. This Second Edition supersedes and replaces the First Edition, October 1967.

d. Every effort has been made by API to assure the accuracy and reliability of data contained in this report. However, the Institute makes no representation, warranty, or guarantee in connection with the report contents and hereby expressly disclaims any liability or responsibility for loss or damage resulting from use or application hereunder, for any violation of Federal, state, or municipal regulation with which the contents may conflict, or for the infringement of any patent resulting from the use of this publication.

e. Contents of this report are not intended to inhibit developing technology or improvements in equipment or improved operational procedures. The report is not a substitute for qualified engineering analysis and judgment to fit specific operations.

f. Members of the Subcommittee on Recovery Efficiency who devoted a substantial amount of time and effort in preparing this report are as follows:

T. M. Doscher, *Chairman*, The Doschers Group Inc., Los Angeles, Calif.
G. G. Cook, Marathon Oil Co., Findlay, Ohio
Carl Grubeman, Sun Expl. & Prod. Co., Dallas, Tex.
J. C. Joers, Conoco Inc., Houston, Tex.
R. L. Jewett, Pennzoil Co., Houston, Tex.
B. C. McRee, Petromac Inc., Houston, Tex.
R. T. Miller, Shell Oil Co., Houston, Tex.
R. P. Murray, Cities Service Co., Tulsa, Okla.
A. Omelchuck, Marathon Oil Co., Findlay, Ohio
L. E. Roberts, ARCO Oil and Gas Co., Dallas, Tex.
M. B. Roy, Mobil Oil Corp., Oklahoma City, Okla.
D. N. Sawyer, ARCO Oil and Gas Co., Dallas, Tex.
A. T. Sayre, Jr., Union Oil Co. of Calif., Los Angeles, Calif.
A. M. Skov, Sohio Petroleum Co., San Francisco, Calif.
W. J. Sleeper, DeGoyler and MacNaughton, Dallas, Tex.
A. E. Smith (deceased), Consultant, Dallas, Tex.
A. S. Trube, Getty Oil Co., Houston, Tex.
V. V. Valleroy, Exxon Production Research Co., Houston, Tex.
W. R. Voss, Chevron USA Inc., San Francisco, Calif.

— 0 —

Requests for permission to reproduce all or any part of the material published herein should be addressed to the Director, Production Department, American Petroleum Institute, 211 North Ervau, Suite 1700, Dallas, Texas 75201.



Documents

Document Sharing Hub

STATISTICAL ANALYSIS OF CRUDE OIL RECOVERY AND RECOVERY EFFICIENCY

SECTION I SUMMARY AND CONCLUSIONS

The API Subcommittee on Recovery Efficiency undertook a follow-up study to assess and, if possible, to extend the correlations of *API Bulletin D14: A Statistical Study of Recovery Efficiency*, October 1967. The extension aimed to gather a larger sample of crude oil reservoir data than was available in the 1967 *Bulletin D14* study. The Subcommittee assignment in both the 1967 study and the 1984 study (roughly six times as large) was to study oil recovery based on actual field performance of producing fields rather than on theory or laboratory data, and to develop an empirical correlation for the prediction of recovery efficiency.

In the current study no statistically valid correlation has been found between oil recovery and definable reservoir parameters. A direct correlation of recovery efficiency yields an equation which fits the data very poorly (r^2 values of the order of 0.1 or less). This was found to hold true for both the *Bulletin D14* data (1967) and the expanded data set of North American reservoirs gathered in the 1984 study. The correlations achieved in the earlier work (1967) relate recoverable oil to original oil-in-place, the latter modified somewhat by the inclusion of additional reservoir parameters. Equations for recovery efficiency were obtained by dividing the recoverable oil equations by original oil-in-place. However, the resulting equations for recovery efficiency do not have the same statistical significance as do the original equations for recoverable oil.

Correlations of the same format as those developed in *Bulletin D14* (1967), when applied to the expanded data base of the 1984 study, are even poorer in predicting recoverable oil than those of the original study (1967). By breaking the total data sample down into subsets based on geographical area (state), lithology (sandstone or carbonate), and producing mechanism (solution gas drive or water drive), the quality of the correlations is improved. However, the improvement is still inadequate for the resulting correlations to be of value as tools for predicting the recoverable oil from any one reservoir.

The poor quality of the correlations is illustrated by the wide range of values — 50% to 75% — which must be cited to be certain that the recoverable oil for a single reservoir has a 95% probability of being included within the range of predicted values. For a 50% probability, this range would drop to 15% to 25%.

The average recoverable oil of a large group of reservoirs having the same value of the independent variable(s) can be predicted well by the correlations of *Bulletin D14* (1967) and the 1984 study. However, the existence of a large group of reservoirs of more or less identical parameters is not expected; therefore, the correlations will provide very minimal use for such predictions.

Some of the 1984 reported studies of the Subcommittee did disclose significant empirical data on recoverable oil (and calculated recovery efficiency) which has not been available before. The following average calculated recovery efficiencies for reservoirs under various producing drives and in various producing areas will be of interest.

TABLE I-1
PRIMARY RECOVERY EFFICIENCIES

Production Mechanism	Lithology	State	Average Primary Recovery Efficiency, % At Average Value of OOIP*
Solution Gas Drive	Sandstones	California	22
		Louisiana	27
		Oklahoma	19
		Texas 7C, 8, 10	15
		Texas 1-7B, 9	31
		West Virginia	21
		Wyoming	25
Solution Gas Drive	Carbonates	All	18
Natural Water Drive	Sandstones	California	36
		Louisiana	60
		Texas	54
		Wyoming	44
Natural Water Drive	Carbonates	All	44

*OOIP = Original oil-in-place, refer to Appendix D, Nomenclature.



Documents

Document Sharing Hub

TABLE I-2
SECONDARY RECOVERY EFFICIENCIES

Secondary Recovery Method	Lithology	State	Primary Plus Secondary Recovery Efficiency at Average OOIP*, percent	Ratio of Secondary to Primary Recovery Efficiency at Average OOIP*, percent
Pattern Waterflood	Sandstone	California	35	0.33
		Louisiana	51	0.40
		Oklahoma	28	0.62
		Texas	38	0.50
		Wyoming	45	0.89
Pattern Waterflood	Carbonates	Texas	32	1.05
Edge Water Injection	Sandstone	Louisiana	55	0.33
		Texas	56	0.64
Gas Cap Injection	Sandstone	California	44	0.48
		Texas	43	0.23

*OOIP = Original oil-in-place, refer to Appendix D, Nomenclature.

The Subcommittee, with completion of this study, concludes that a reliable statistical correlation cannot be achieved for the prediction of recovery and/or efficiency for individual reservoirs based on the readily definable and available reservoir parameters. An important factor, reservoir heterogeneity, cannot be readily defined, certainly not by the assignment of a mere numerical value. It is believed that not being able to include heterogeneity and possibly other factors as independent parameters contributes to the inability to develop valid statistical correlations.

The accuracy of definable parameters such as porosity and initial water saturation is limited by the quality of measurement techniques available at the time of discovery and development. Such measurements have been subject to constant appraisal and improvement, but there is no simple way to correct older measurements. The Subcommittee did not attempt such corrections nor did they request respondents to do so. The impact of older, less accurate measurement techniques on such a study as this can be appreciated when it is recognized that approximately 75% of anticipated ultimate oil production from presently known U.S. fields will be derived from fields discovered prior to 1951.

The recoverable oil to be expected from a given reservoir can be estimated with reasonable accuracy only after a thorough engineering study of that reservoir is undertaken. Such estimates of recoverable oil are improved by matching the performance of a reservoir model to the reservoir's production history and then using the matched model to predict ultimate recovery.

Although statistical correlations of recovery efficiency based on readily definable reservoir and fluid parameters fall short of being able to predict accurately the recovery of an individual reservoir, the calculated average recoveries in a single geological trend are significant. Comparison of these calculated averages shows substantial recovery efficiency differences between reservoirs with different indigenous drive mechanisms, with the application of different secondary recovery techniques, and in different producing areas.

The Subcommittee on Recovery Efficiency cautions against continued use of the correlations from API Bulletin D14: A Statistical Study of Recovery Efficiency, October 1967, to predict recovery or recovery efficiency for any one reservoir. Further, to avoid any undue significance being attached to the correlations developed in the current study, only those results required to substantiate the expressed conclusions are cited in this report.

SECTION II

REVIEW OF API BULLETIN D14, OCTOBER 1967 DATA AND CORRELATIONS

A. THE WEIGHTED BULLETIN D14 (1967) CORRELATIONS.

The original work (1967) of the Subcommittee on Recovery Efficiency was based on 312 questionnaires supplied for the most part by the companies who were represented on the Subcommittee. Only about one-half of the completed questionnaires had sufficient data and pertained to reservoirs which were being produced by a single dominant mechanism. Seventy of these reservoirs had "acceptable accuracy" for the water drive correlation, all from sandstones. Eighty were used for the solution gas drive correlation, of which sixty-seven were sandstones and thirteen were carbonates. The selected reservoirs were then weighted 1, 2, and 3 according to the Subcommittee members' estimates of the quality of the data. A reservoir deemed to have good data was counted three times, and a reservoir with relatively poor data was counted only once.

Correlations were developed for recoverable oil (not recovery efficiency) and were as follows:

For solution gas drive reservoirs:

$$RO = 3244 \left[\frac{\phi (1-S_w)}{B_{ob}} \right]^{1.1611} \left[\frac{k}{\mu_{ob}} \right]^{0.0979} [S_w]^{0.3722} \left[\frac{P_b}{P_a} \right]^{0.174} \quad (1)$$

For water drive reservoirs:

$$RO = 4259 \left[\frac{\phi (1-S_w)}{B_{oi}} \right]^{1.0422} \left[\frac{k \mu_{wi}}{\mu_{oi}} \right]^{0.0770} [S_w]^{-0.1903} \left[\frac{P_i}{P_a} \right]^{0.2159} \quad (2)$$

Note that the last three terms in equations (1) and (2) have relatively little impact on the recoverable oil, RO, as indicated by the small exponents.

The correlation coefficients of the regression equations were reported as 0.9317 and 0.9575, and the standard errors of estimate, i.e., the range within which the probability of a correct estimate is 68%, were 22.0% and 17.6% for the solution gas drive and water drive reservoirs, respectively.

For a (ln y) versus (ln x) correlation, as used in *Bulletin D14* (1967) the parameters are truly (ln y) and (ln x), respectively, not y and x. Hence, the confidence limits calculated from the statistical analysis are now defined as the range of (ln y) bounded by (ln y ± ts). Thus, for the *Bulletin D14* (1967) solution gas drive and water drive samples, respectively, the confidence limits* for a 95% confidence level at the average values of recovered oil are (ln 112 ± 0.40) and (ln 576 ± 0.32).

When these confidence limits are "translated" to actual values of recovered oil, then the confidence limits are 75 to 167 and 418 to 793 barrels per acre foot.** On a percentage basis, the confidence limits are roughly ±40% and ±32% of the predicted values in the case of the solution gas drive and water drive, respectively.

*The confidence limits for a predicted value of a single value of an independent parameter is given by the following expression:

$$C.L. y_{i, a\%, df} = \pm (t) \cdot (s_{y \cdot x}) \left[1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]^{0.5} \quad (3)$$

For examples of the nature and size in this work, the following equation is a good approximation over most of the range of the independent parameter:

$$C.L. y_{i, a\%, df} = \pm (t) \cdot (s_{y \cdot x}) \quad (4)$$

There is only an a% chance that the true value of recoverable oil (the dependent parameter) will be outside these calculated confidence limits. The confidence level is (1 - a)%.

**The skewness of the confidence limits around the predicted value of the log of the recovered oil is caused by the fact that the antilog of (ln y ± t · s_{y·x}) is not equal to [y the antilog of (ln y ± t · s_{y·x})].

The *Bulletin D14* (1967) recoverable oil correlations do a far better job in predicting the average value of a large group of reservoirs that have more or less the same value of the independent parameter(s). The 1967 correlations predict confidence limits for such averaged performance of 5% and 4% for solution gas and water drive reservoirs, respectively. Unfortunately, such predictions are not often called for because of the lack of occurrence of a large number of reservoirs having more or less identical values of oil-in-place.

The *Bulletin D14* (1967) correlation shows an extremely strong dependence of recoverable oil on the original oil-in-place. Other reservoir and fluid parameters were included in the correlation since calculations showed a statistically significant increase in the correlation coefficient when using the "F" test for multiple factor regressions. The dependence of the recovery factors on fractional power functions of the additional independent reservoir parameters suggests that, at the most, the additional parameters serve as correction factors, e.g., a three-fold change in the value of (k/μ_{ob}) causes only a 10% change in the recovery factor.

B. ANALYSIS OF THE UNWEIGHTED *BULLETIN D14* (1967) DATA

The *Bulletin D14* (1967) study employed a ln-ln relationship for the correlation, viz., $\ln y = \ln a + b (\ln x)$. This is identical to the so-called power curve fit, $y = ax^b$. The use of the ln-ln relation forces the correlation through the origin. Although there is no need when developing an empirical correlation to heed physical observations, the well established fact that crude oil will not flow below some residual saturation is overlooked when the correlation is forced through the origin. A correlation in keeping with this fact would intersect the abscissa at a positive value, i.e., no recoverable oil if the original oil-in-place is present at or below the residual saturation.

The ln-ln relation was chosen because the residuals (actual ln of the recoverable oil less the predicted ln of the recoverable oil) showed less dependence on the independent parameter of the correlation (ln of OOIP) than that shown by the residuals of a linear correlation.

During the course of the 1984 study, it was noted that the residuals for the linear correlation for the *Bulletin D14* (1967) samples were not highly dependent on the independent variable when the data for individual reservoirs were not weighted for quality. There was then little basis for choosing the power correlation over the linear one.

Using unweighted data the 80 solution gas reservoirs of the 1967 study gave the following correlations:

Linear Correlation:

$$RO = 0.290 (OOIP) - 50.0 \quad (5)$$

for which: $r^2 = 0.65$, $r = 0.81$, $s_{y-x} = 55.6$, $RO_{avg} = 164$, and $OOIP_{avg} = 789$.

Power Curve Correlation:

$$RO = 0.0426 (OOIP)^{1.24} \quad (6)$$

for which: $r^2 = 0.76$, $r = 0.87$, $s_{y-x} = 0.29$, $(\ln RO)_{avg} = \ln 140$, and $(\ln OOIP)_{avg} = \ln 687$.

It is to be noted that r is higher for the power curve fit. The confidence limit for the predicted values of recoverable oil using the power curve fit is 79%, while for the linear correlation it is 68% of the predicted values.

The high value of the correlation coefficient which is achieved in the ln-ln correlation results from the compression in the spread of the values of the parameters when employing their logarithms. " r " is only a measure of the goodness of fit of the data to the regression line, and is not a direct measure of the precision of a prediction. (Curve fitting can be enhanced by correlating the logarithms of logarithms, etc., but this does not enhance the ability of the correlation to be a prediction tool.)

The relationship of the linear correlation for the unweighted *Bulletin D14* (1967) solution gas drive reservoirs to a power curve fit is shown in Figure II-1. The points represent the unweighted reservoir data. It can be seen why there is little superficial reason to choose the linear or power curve fit in preference to the other.

It is evident that the subjective weighting of the data employed in developing the *Bulletin D14* (1967) correlations resulted in a seemingly high statistical reliability. The weighting also enhanced the apparent ability of the regression equations to be used for predictive purposes. Even so, using the 95% confidence level as a yardstick, the *Bulletin D14* (1967) regression equations cannot predict recoverable oil from an individual reservoir in the edited sample to within about 50% of the true value.

C. CORRELATION OF RECOVERY EFFICIENCY

Bulletin D14 (1967) did not attempt to directly correlate recovery efficiency (recoverable oil/original oil-in-place). An equation for recovery efficiency was presented in *Bulletin D14* (1967) which is the result of dividing both sides of the correlation between recovered oil and original oil-in-place by the original oil-in-place.

If a direct correlation of the unweighted, calculated recovery efficiencies of the solution gas drive reservoirs of *Bulletin D14* (1967) against original oil-in-place is made, the following equations are obtained:

Linear Correlation:

$$R.E. = 0.00915 (OOIP) + 14.7 \quad (7)$$

for which: $r^2 = 0.11$, $r = 0.34$, $s_{y \cdot x} = 6.34$, and $R.E._{avg} = 21.4$

Power Curve Correlation:

$$R.E. = 4.26 (OOIP)^{0.240} \quad (8)$$

for which: $r^2 = 0.093$, $r = 0.31$, $s_{y \cdot x} = 0.29$, $(\ln R.E.)_{avg} = \ln 20.4$

Immediately, it is apparent the correlation coefficients for recovery efficiency are drastically smaller than those obtained for the correlation of recoverable oil. This must be expected as a result of the reduction in the slope of the correlation.

Application of the "t-test" to evaluate the significance of the regression equation's slope being different from zero shows that there is only a one in four chance of being wrong by assuming the slope of the correlation equation is in fact zero. This level of chance suggests that there is not a significant difference between the value of 0.00915 for the slope and zero, viz., that there is no significant correlation between recovery efficiency and original oil-in-place.

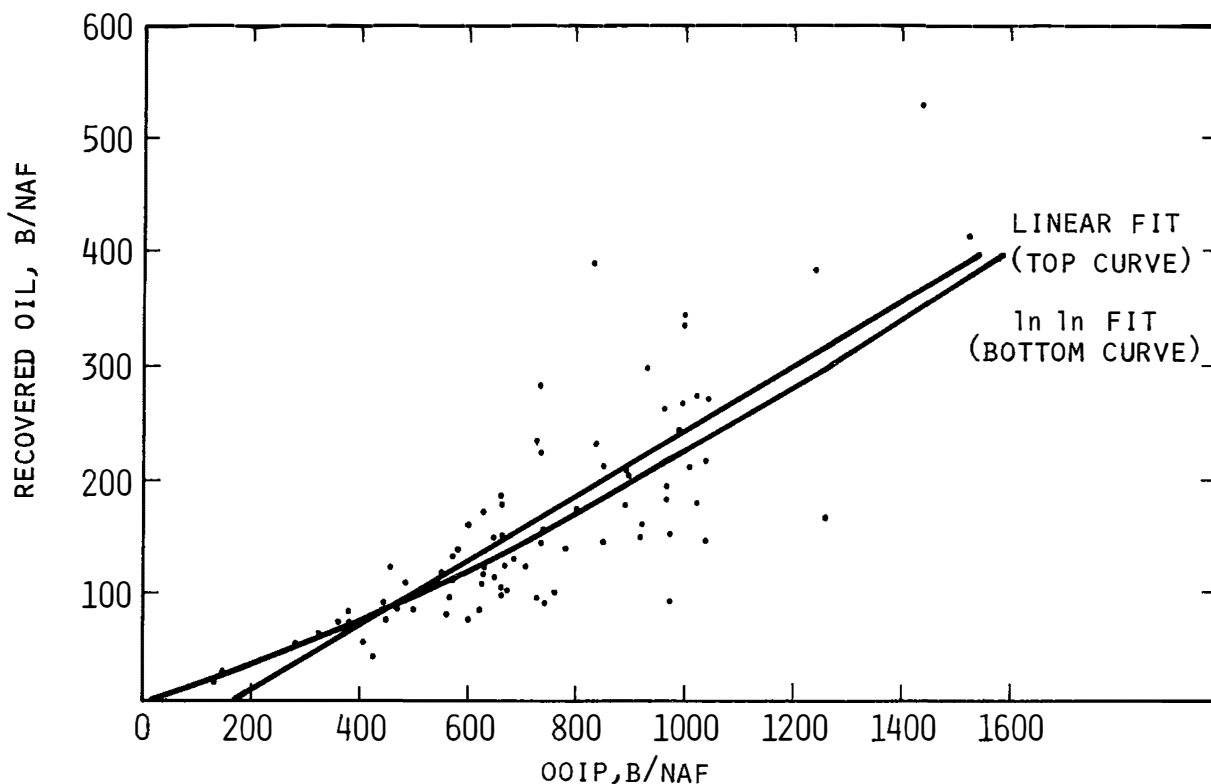


FIGURE II-1
SOLUTION GAS RESERVOIRS
UNWEIGHTED *BULLETIN D14* SAMPLES

SECTION III

FITTING THE UPDATED DATA SAMPLE TO A *BULLETIN D14* (1967) FORMAT

The updating work of the Subcommittee on Recovery Efficiency sought to improve the correlations of *Bulletin D14* (1967). The goals of the updating Subcommittee work can be described as follows:

1. Test the *API Bulletin D14* (1967) correlations on a new, larger data set than that which was available in the 1967 study.
2. Improve, if possible, the reliability of the *Bulletin D14* (1967) correlations by calculating new coefficients for the independent parameters.
3. Investigate the feasibility of alternate and possibly more useful correlations.

The reservoir data were solicited by members of the Subcommittee on an informal basis from appropriate sources within their own company, and through professional contacts from companies not represented on the Subcommittee. Data questionnaires were submitted in confidence to API. When requested, the confidentiality extended to withholding the name of the operator from Subcommittee members themselves. Some 20 companies identified themselves on the returned questionnaires. Each of five companies contributed 15%, 13%, 10%, 8%, and 7% of the questionnaires; a total of 53%. Confidential contributions comprise 12% of the sample, and the remaining 35% of the questionnaires were contributed by 15 other companies.

A total of 675 questionnaires were submitted, of which 620 reservoirs in the United States and Canada could be verified to have a dominant drive mechanism of solution gas or natural water drive and also had a well-defined lithological character, sandstone or carbonate. The remaining 55 samples were dropped because the data on primary recovery were lacking, or the primary drive was by gravity drainage or the lithology was ill-defined.

There were 376 reservoirs produced by solution gas drive and 244 by natural water drive. The geographical distribution of the reservoirs is shown in Table III-1.

TABLE III-1
GEOGRAPHICAL DISTRIBUTION OF RESERVOIRS

Solution Gas Drive			Natural Water Drives		
Sandstones:	Texas	90	Sandstones:	Texas	102
	California	51		Louisiana	42
	Oklahoma	41		Wyoming	27
	Wyoming	23		California	9
	Louisiana	17		Others	17
	West Virginia	13		Total Sandstones	197
	Others	41			
	Total Sandstones	276			
Carbonates:	Texas	61	Carbonates:	Texas	33
	Others	39		Others	14
	Total Carbonates	100		Total Carbonates	47
Total Solution Gas Drive Reservoirs		376	Total Water Drive Reservoirs		244

The Subcommittee first set out to rate the questionnaires following the procedures used in developing the *Bulletin D14* (1967) correlations. As a first try to test the *Bulletin D14* (1967) correlations, a set of sandstone reservoirs having a rating of 2 or 3 (adequate or good, respectively) and for which the principal drive mechanism was said to account for at least 75% of the reservoir energy was chosen from the new data as a pilot sample. A total of 116 solution gas drive sandstone reservoirs (approximately 40% of the total so categorized) fell into this select sample. The regression equation derived for this pilot set is as follows:

$$RO = 6533 \left[\frac{\phi(1-S_w)}{B_{ob}} \right]^{1.312} \left[\frac{k}{\mu_{ob}} \right]^{0.0816} \left[S_w \right]^{0.463} \left[\frac{P_b - P_a}{P_b} \right]^{0.249}$$

The coefficient of correlation is 0.74, and the standard percentage error of estimate is 70%. More importantly, however, $s_{y,x}$ is 0.27, and in order to predict the recoverable oil with a confidence of $\pm 95\%$, the range of the prediction is $\pm 70\%$ of the true value. Despite a fairly good fit, this correlation is of little value for reliably predicting recoverable oil from a crude oil reservoir.

It is interesting to note that on this new sample the coefficients are significantly different than were obtained for the *Bulletin D14* (1967) sample for solution gas drive reservoirs. Specifically, the multiplicative coefficient at 6533 is higher by a factor of 2 than the value of 3244 found earlier, and the exponents are different by ratios of 1.13, 0.83, 1.24, and 1.43, respectively, for the 1984 data.

Several attempts were made to regroup, delete, and add parameters; searching for a regression equation that would have greater significance when using the "new data set" (1984) than was obtained with the *Bulletin D14* (1967) format, but one which nevertheless conserved the spirit of *Bulletin D14* (1967). This could not be achieved.

Omitting the S_w term, for example, decreased the correlation coefficient to 0.71, and $s_{y,x}$ was increased slightly to 0.28 making the 95% confidence range $\pm 74\%$ of the predicted value. This indicated that S_w had some but little significance compared to that of the other reservoir parameters all together.

Suspecting that California reservoirs might be different in their production performance, because of very different geological parameters, than those of any other geographical region in the United States, the California reservoirs were eliminated from the sample. The square root of the remaining variance, $s_{y,x}$, was reduced to 0.26, only slightly affecting the reliability of the correlation.

A correlation based on the original oil-in-place, alone, gave the following equation:

$$RO = 2788 \left[\frac{\phi (1-S_w)}{B_{ob}} \right]^{1.813} \quad (10)$$

The correlation coefficient for this regression is 0.69, and $s_{y,x}$ is 0.29. When these statistical measures of the quality of the correlations, which omit selected parameters or selected groups of reservoirs, are compared to those of the correlation which include all the *Bulletin D14* (1967) parameters, it is found that the differences are slight. Although the differences in the correlations may have some statistical significance they do not result in any heightened predictive qualities of the regression equations.

The water drive pilot sample, withdrawn in a similar manner to that described above for the selection of the solution gas drive pilot sample, also gave poor results. It is inferior to the quality of the correlation for the weighted *Bulletin D14* (1967) sample, but comparable to the quality of the correlation for the unweighted *Bulletin D14* (1967) sample developed in the 1984 study.

It was concluded at this point of the study that the *Bulletin D14* (1967) correlations are valid only for that data sample and for the quality weighting factors which were utilized. Further, that correlations of recoverable oil as a function of oil-in-place, when samples are not weighted, are poor for predicting recoverable oil from new reservoirs. It was also found that calculated values of recovery efficiency do not appear to be dependent, in a statistically significant sense, upon reservoir and fluid parameters for the samples utilized in these studies. It is likely that simple errors in data, due in part to necessarily incomplete knowledge of reservoir heterogeneity and continuity, mask the effect of measurable physical parameters which are known to influence the recovery of oil.

SECTION IV

SEARCH FOR A NEW MULTI-VARIABLE REGRESSION FORMAT

The work already described suggested that original oil-in-place is the only parameter that has a statistically significant relationship to recoverable oil. The Subcommittee chose to continue to search for a regression relationship which might show that other parameters also had a significant effect on recoverable oil and recovery efficiency.

A multi-variable, stepwise, linear regression program, part of the *UCLA Biomedical Computer Program (BMD)*, which had been developed since *Bulletin D14* (1967) was released, was employed. This program successively seeks to add independent variables which represent reservoir data (as stored in the computer in logarithmic and/or algebraic form) in a stepwise manner so as to progressively increase the quality of the fit of the regression equation to the data. It measures the increase in the quality of the fit by the reduction in the sum of the squares of the deviations of measured values from the values of the dependent variable calculated from the correlation. The procedures employed in the use of this program to seek an improved correlation are detailed in Appendix B.

This program was used to develop correlations on a selected group of 112 solution gas drive reservoirs, which had a full complement of data. All the correlations that were developed showed the overwhelming influence of the original oil-in-place, either calculated at the bubble point or as cited by the operator at initial conditions. In many trials, the program chose a reservoir parameter more than once (in different groupings), and in some cases specified the inclusion of a parameter in a sense opposite to intuition, empirical observations, or fluid flow theory.

Further, none of these correlations were significantly better than a *Bulletin D14* (1967) format regression fitted to the 112 sample subset.

A similar attempt was made to develop an improved correlation for a subset of the water drive reservoirs. This sample was of a very high quality in that all the data necessary for the correlations were explicitly available in all the completed questionnaires. The original oil-in-place in the reservoir was the dominant term, although not quite at the same level of significance as in the case of the solution gas drive subset. Other parameters again had far less significance and would appear and reappear depending upon the type of correlation the program was instructed to develop. The correlations were only marginally better than a *Bulletin D14* (1967) format fit of the data. Severe loss in reliability of the correlation occurred in a trial increase in the sample size.

The results of the application of the *UCLA BMD* program confirmed that an adequate predictive correlation for the recovery of crude oil as a function of reservoir and crude oil parameters could not be achieved for the new sample of reservoirs. Further, it was again apparent in all the correlations that were developed by the program that the dominant independent parameter is the original oil-in-place in the reservoir.

SECTION V

CORRELATIONS BASED ON GEOGRAPHICAL PROVINCE

The recovery statistics tabulated by the American Petroleum Institute in its annual publication, *Reserves of Crude Oil, Natural Gas Liquids, and Natural Gas in the United States and Canada*, indicate that average recovery efficiency varies significantly by state. The four chief producing states, Texas, Louisiana, California, and Oklahoma, accounting for over 70% of the anticipated ultimate recovery in the United States*, show significant variations in recovery efficiency. Certainly, some of this variation is due to the prevalence of different drive mechanisms in the reservoirs in the different states. However, there is much evidence to support the hypothesis that for the same mechanism the performance of reservoirs does vary as a result of the different subsurface geological and reservoir conditions that prevail in different geographical areas. Obviously, the regional geological and depositional processes also vary significantly within a state, particularly one as large as Texas.

The Subcommittee, therefore, chose to investigate the possibility that stronger correlations for recovery and recovery efficiency might be derived if the available reservoirs were broken down into subsets by geographical area, which for convenience was assumed to be that of the chief producing states. Such subdivision did result in somewhat higher quality correlations but they were insufficiently improved to be useful as predictive tools. Differences in average recovery among the geographical subsets were, of course, highlighted by the subdivision.

The Subcommittee chose not to reject any questionnaires based on subjective judgments of the "quality" of the data, although there certainly were superficial differences in the apparent quality of the data as submitted. Key data, such as permeability, oil viscosity, or pressures were missing from a number of the submissions, and some of the submissions had such seemingly exact data, e.g., a reported recovery efficiency of 50.00%, that the origin of the data could be questioned.

The procedure used by the Subcommittee is illustrated by reviewing the analysis for the subset, "Texas, sandstones, natural water drive." This 102 reservoir subset is the largest subset sample available.

The recoverable oil from one of these reservoirs was not cited. Obviously, it had to be discarded. A linear correlation was developed for the remaining 101 reservoirs, correlating recoverable oil against original oil-in-place.

$$RO = 0.565 (OOIP) - 73.0 \quad (11)$$

for which: $r^2 = 0.41$, $s_{y \cdot x} = 220$, $RO_{avg} = 629$, $OOIP_{avg} = 1242$, and the confidence limits at a 95% significance level are $\pm 70\%$.

This correlation is not an improvement over the *Bulletin D14* (1967) format correlation developed for the selected group of all water drive reservoirs. However, of the 101 samples, 11 did not have reported permeability and/or viscosity data. Should there be any merit to a correlation that would include these parameters, it was obvious that these 11 had to be rejected for sheer lack of appropriate data. A correlation of the remaining 90 reservoirs gave the following result:

$$RO = 0.548 (OOIP) - 67.4 \quad (12)$$

for which: $r^2 = 0.40$, $s_{y \cdot x} = 218$, $RO_{avg} = 616$, $OOIP_{avg} = 1248$, and the confidence limits at a 95% significance level are $\pm 71\%$.

Obviously, the 11 reservoirs fitted in with the remaining 90. Examination of the scatter plots of the data, Figure V-1, suggested that a number of the reservoirs with a large deviation from the correlation line had relatively low values of (k/μ) , permeability divided by viscosity. The Subcommittee, therefore, decided to reject those reservoirs which had a ratio of less than 110 mds/cp. All such reservoirs were rejected even though some plotted very close to the correlation line. The result was as follows:

$$RO = 0.513 (OOIP) - 31.3 \quad (13)$$

for which: $r^2 = 0.36$, $s_{y \cdot x} = 222$, $RO_{avg} = 609$, $OOIP_{avg} = 1247$, and the confidence limits at a significance level of 95% are $\pm 73\%$.

*Excluding the Prudhoe Bay Field on the Alaska North Slope.

Apparently, even these 12 reservoirs had enough good fitting values to offset the effect of reservoirs with significant departures from the correlation equation. There is a progressive flattening of the correlations as reservoirs are eliminated, but no substantial change in their value as predictive tools. They are all unacceptable for use by industry; a predicted value of recoverable oil could be cited only as falling within the range of approximately $\pm 70\%$ of the calculated value if one wanted to be correct in 95 of every 100 predictions. If one wanted to be correct only half the time, the range would still be approximately $\pm 25\%$ of the calculated value.

The correlation for recovery efficiency as a function of the original oil-in-place for the 101 reservoirs is as follows:

$$R.E. = 0.00886 (OOIP) + 38.9 \quad (14)$$

for which: $r^2 = 0.018$, $s_{y.x} = 17.3$, $R.E._{avg} = 49.9$, $OOIP_{avg} = 1242$, and the confidence limits at a significance level of 95% are $\pm 69\%$.

A "t-test" analysis to determine whether the slope of this correlation is significantly different than zero reveals that there is less than a 10% chance that the slope is different from zero, viz., the correlation is meaningless.

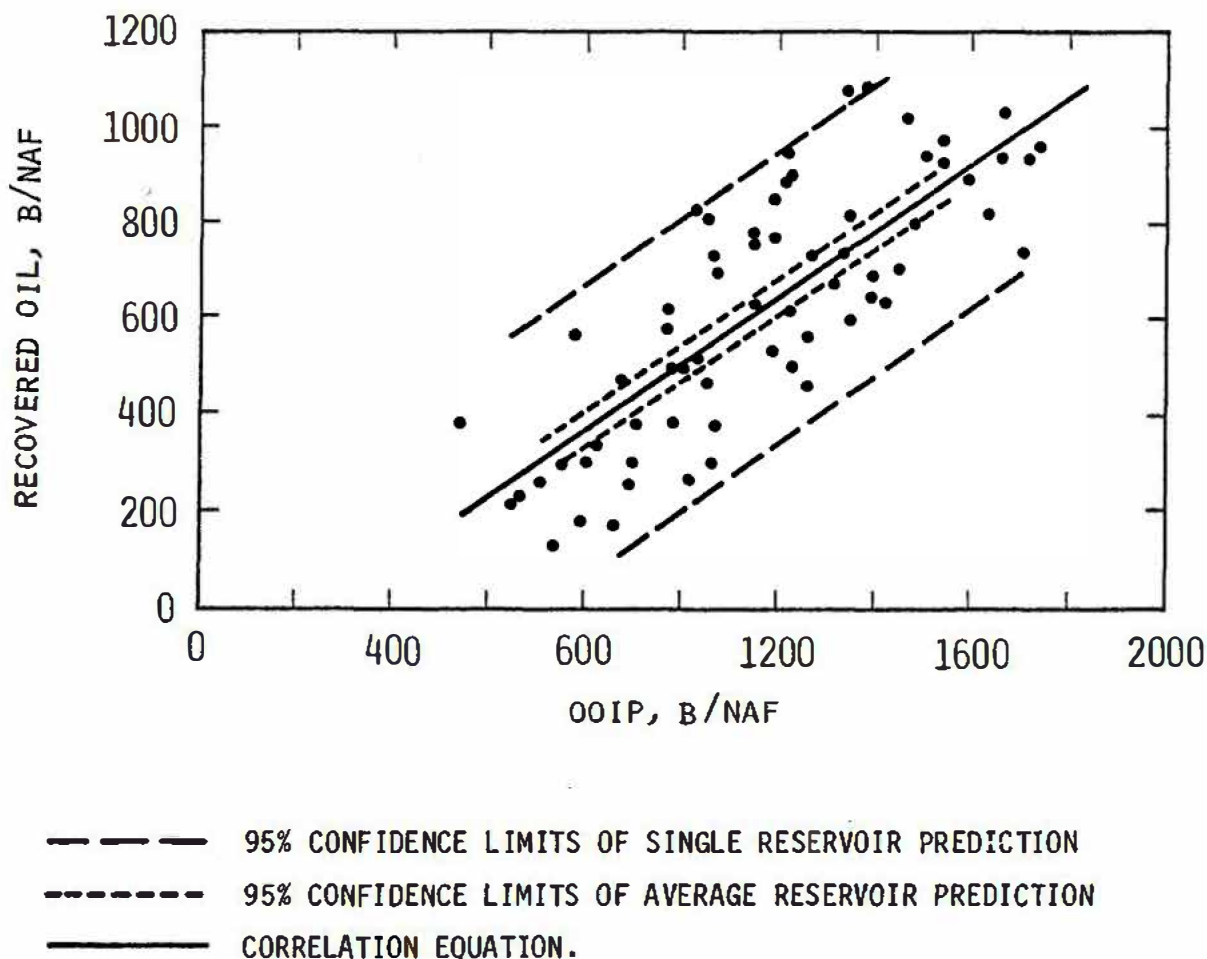


FIGURE V-1
TEXAS-SANDSTONE NATURAL WATER DRIVE



Documents

Document Sharing Hub

Finally, a correlation of recovery efficiency was attempted as a function of (k/μ) . For the 90 samples having reported data, the correlation is as follows:

$$R.E. = 51.7 - 1.71 (k/\mu) \quad (15)$$

for which: $r^2 = 0.04$, $s_{y.x} = 17.0$, $R.E._{avg} = 48.7$, and $(k/\mu)_{avg} = 1.7$.

Again, an analysis of the significance of the slope indicates that there is only a 20% chance that it is truly different than zero.

Note that the sign of the slope is negative, suggesting the physical absurdity that recovery efficiency decreases with increasing permeability and increases with increasing values of viscosity. This is further confirmation that the meaning of the slope is insignificant.

A correlation of recovery efficiency as a function of (k/μ) after excluding the 12 samples having a (k/μ) value less than 110 hardly changed the correlation and its significance.

A scatter plot of the recoverable oil versus the original oil-in-place for the 101 Texas-sandstone-water drive reservoirs is shown in Figure V-1, along with the calculated regression line, and the confidence limits of the latter.

Although the Subcommittee calculated numerous correlations for the various categories tabulated in Table V-1, these are not presented herein because of their complete lack of utility and significance. Table V-1 does show the average values of recoverable oil and original oil-in-place, and Figure V-2 presents histograms of the reservoir parameters. A magnetic tape file of the original data is available from the American Petroleum Institute to interested parties who may wish to attempt further statistical correlations.

TABLE V-1
PRIMARY RECOVERY OF RESERVOIR CATEGORIES STUDIED

	Number of Samples	OOIP, B/NAF		RO, B/NAF		Square Root of Variance
		Range	Average	Range	Average	
I. SOLUTION GAS DRIVEN						
A. Sandstone						
1. Texas, All	90	189 - 1620	731	16 - 655	159	61
1(a). Texas, 7C, 8, 10	24		578		86	39
1(b). Texas, 1-7B & 9	38		827		205	63
2. California	51	328 - 2146	1148	55 - 644	273	105
3. Louisiana	17	360 - 1540	1167	110 - 613	315	97
4. Oklahoma	41	386 - 1030	604	40 - 227	117	35
5. Wyoming	23	386 - 1135	692	82 - 361	176	21
6. West Virginia	13	546 - 882	678	100 - 250	143	22
7. Others	41		738		183	41
B. Carbonates						
1. All	100	59 - 1067	526	5 - 190	93	20
II. WATER DRIVEN						
A. Sandstones						
1. Texas	101	547 - 1801	1167	125 - 1094	633	150
2. California	9	508 - 1770	1020	169 - 568	367	99
3. Louisiana	42	535 - 1960	1203	219 - 1470	722	160
4. Wyoming	27		860		313	101
B. Carbonates						
1. Texas and others	47	85 - 705	265	29 - 370	117	28