API BUL D14
First Edition
October 1967

A STATISTICAL STUDY OF RECOVERY EFFICIENCY

AMERICAN PETROLEUM INSTITUTE
New York, N. Y. 10020

Issued by

AMERICAN PETROLEUM INSTITUTE
Division of Production
Dallas, Texas 75201

A STATISTICAL STUDY OF RECOVERY EFFICIENCY

A Report By
The API Subcommittee on Recovery Efficiency

Chairman
J. J. Arps
Petroleum Consultant
Dallas, Texas

Current Membership of the Subcommittee

F. H. Brinkman
Esso Production Research Co.
Houston, Texas

*Prof. Folkert Brons
University of Texas
Austin, Texas

R. W. Buchwald
Sun Oil Company
Dallas, Texas

R. C. Craze
Esso Production Research Co.
Houston, Texas

J. N. Dew
Continental Oil Co.
Ponca City, Oklahoma

E. F. Egan
Texaco Inc.
Houston, Texas

*L. F. Elkins
Sohio Petroleum Co.
Oklahoma City, Oklahoma

T. M. Geffen
Pan American Petroleum Corp.
Tulsa, Oklahoma

R. K. Guthrie
R. K. Guthrie & Assoc.
San Antonio, Texas

G. L. Hancock
Gulf Oil Corp.
Houston, Texas

G. E. Hendrickson
Cities Service Oil Co.
Bartlesville, Oklahoma

T. S. Hutchinson
Atlantic Richfield Co.
Dallas, Texas

Prof. H. T. Kennedy
Texas A&M University
College Station, Texas

L. C. McColloch
Sun Oil Co.
Dallas, Texas

*V. T. McGhee
Phillips Petroleum Co.
Bartlesville, Oklahoma

J. M. Miller
Standard Oil Co. of Calif.,
Western Operations, Inc.
San Francisco, California

W. J. Rogers
Sinclair Oil & Gas Co.
Tulsa, Oklahoma

A. E. Smith
Petroleum Consultant
Dallas, Texas

*A. F. van Everdingen
DeGolyer & MacNaughton
Dallas, Texas

*O. E. Van Meter, Jr.
Mobil Oil Corp.
Shreveport, Louisiana

J. W. Wilson
Union Oil Co. of California
Brea, California

*Task Force Chairmen

TABLE OF CONTENTS

		Page
I	SUMMARY AND CONCLUSIONS	1
II	INTRODUCTION	4
III	PRELIMINARY ANALYSIS OF CASE HISTORIES	5
	A. Number and Distribution of Case Histories by Groups	5
	B. Range of Basic Parameters	6
	C. Rating and Weighting of Case Histories	9
	D. Predominant Drive Mechanisms	9
	E. Recovery Efficiencies of the Various Mechanisms	9
	F. Reservoir Rock	13
	G. Well Spacing	13
IV	CORRELATION METHODS	13
	A. Recovery Parameters or Dependent Variables	14
	B. Reservoir Parameters and Independent Variables	16
	C. Regression Equations	16
V	EARLY CORRELATION EFFORTS	17
	A. Primary Recovery from Water Drive Reservoirs	17
	B. Primary Recovery from Solution Gas Drive Reservoirs	18
	C. Secondary Recovery by Fluid Injection	18
	D. Inter-Correlation of P. V. T. Data	18
	E. Inter-Correlation of Rock Parameters	19
	F. Overall Results	19
VI	FINAL CORRELATIONS	19
	A. Results	19
	1. Water Drive	19
	2. Solution Gas Drive	20
	B. Numerical Examples	20
	C. Discussion of Results	22
VII	NOMENCLATURE	26
VIII	REFERENCES	27
APPENDIX: REGRESSION ANALYSIS AND TESTS OF SIGNIFICANCE		28

Figures

Fig. 1—Comparison of Observed and Computed Values of Recovery Factors in Barrels per Acre Foot from 70 Water Drive Reservoirs 3

Fig. 2—Comparison of Observed and Computed Values of Recovery Factors in Barrels per Acre Foot from 80 Solution Gas Drive Reservoirs below Bubble Point 4

Fig. 3—Cumulative Frequency Distribution of Permeabilities for Water Drive and Solution Gas Drive Cases—Sandstone Reservoirs 10

Fig. 4—Cumulative Frequency Distribution of Permeabilities for Water Drive and Solution Gas Drive Cases—Limestone Reservoirs 11

Fig. 5—Cumulative Frequency Distribution of Recovery Efficiencies for Four Drive Mechanisms 12

Fig. 6—Cumulative Frequency Distribution of Recovery Factors for Four Drive Mechanisms 14

Fig. 7—Cumulative Frequency Distribution of Residual Oil Saturation for Water Drive and Residual Gas Saturation for Solution Gas Drive Reservoirs 15

Fig. 8—Correlation of Observed and Computed Dependent Variables 29

Fig. 9—"*student*" *t* Distribution 31

Fig. 10—Significant Coefficient of Correlation 32

Fig. 11—Significant Increase in Coefficient of Correlation 32

Tables

Table 1—Distribution of Case Histories by Reservoir Mechanism and Rock Type 5

Table 2—Range of Parameters for Water Drive Reservoirs 6

Table 3—Range of Parameters for Solution Gas Drive Reservoirs without Supplemental Drives 7

Table 4—Range of Parameters for Solution Gas Drive Reservoirs with Supplemental Drives 7

Table 5—Range of Parameters for Gas Cap Drive Reservoirs 8

Table 6—Range of Parameters for Gravity Drainage Reservoirs 8

Table 7—Median Values of Recovery Efficiency by Recovery Mechanisms 12

Table 8—Summary of Exponents of Terms in Regression Equations, Recovery Factor for Water Drive Reservoirs 20

Table 9—Summary of Exponents of Terms in Regression Equations, Recovery Factor for Solution Gas Drive Reservoirs 21

Table 10—Effect of the Addition of Various Terms in the Regression Equations on Computed Recovery Factors 21

Table 11—Effect of Variation in Independent Variables on Recovery Factor from Water Drive Reservoirs 24

Table 12—Effect of Variation in Independent Variables on Recovery Factor from Solution Gas Drive Reservoirs 25

A Statistical Study of Recovery Efficiency

J. J. ARPS, FOLKERT BRONS, A. F. VAN EVERDINGEN, R. W. BUCHWALD, AND A. E. SMITH

I — SUMMARY AND CONCLUSIONS

The equations and conclusions presented in this report have grown out of a study initiated by the American Petroleum Institute's Subcommittee on Recovery Efficiency in 1956. The Subcommittee's assignment then and now was to study reservoir recovery processes based on actual performance of producing fields rather than on theory or laboratory data.

The study began as an effort to apply machine methods of analysis to updated data from the 103 reservoirs investigated by Craze and Buckley in 1945,[1] using multiple correlation analysis procedures described by Guthrie and Greenberger in 1955.[2] This initial effort was later expanded to include collection and analysis of reservoir and recovery data on additional reservoirs.

Availability of advanced computers made it possible to apply more sophisticated techniques to arrive at statistically significant correlations between indicated recovery factors and the reservoir parameters.

The subcommittee collected data on 312 reservoirs which were weighted 1, 2, and 3 according to their reliability. This report deals with:

a. Oil reservoirs under a *water drive mechanism*, where invading bottom or edge water is the dominant displacing medium; and

b. Oil reservoirs under a *solution gas drive mechanism* (also referred to as depletion or internal gas drive) where the expansion of gas liberated from solution is the sole source of expulsion of the oil.

Not covered in this report are combinations of water drive and solution gas drive, gas cap drive, or gravity drainage which together comprised 30 to 40 percent of the case histories. Another 10 percent of all cases received could not be used for correlation purposes because of lack of adequate performance control or because pertinent data were missing.

This left about one-half of the case histories for use in the correlation work. The subcommittee found 70 case histories to be of acceptable accuracy for the water drive study, all from sand and sandstone reservoirs. The analysis of the solution gas drive mechanism was based on 80 case histories, 67 from sand and sandstone and 13 from carbonate reservoirs.

During the Subcommittee's early efforts the parameters which appeared to control the recovery process were determined by regression analysis and the computer-derived correlations were then studied to determine the possible groupings of these parameters which, from a reservoir engineering standpoint, would be meaningful. During the final correlation effort a series of regression equations was developed for each of the two main mechanisms relating the Recovery Factor *(BAF)* in barrels per acre foot as the dependent variable with physically meaningful groups of parameters as independent variables. The danger of "over fitting" the data was avoided by the requirement that addition of each variable must show a statistically significant improvement in the coefficient of multiple correlation. The resulting best equations were:

For Water Drive Reservoirs (Sands and Sandstones)

$$BAF = (4259) \cdot \left\{\frac{\phi(1-S_w)}{B_{oi}}\right\}^{+1.0422} \cdot \left(\frac{k\mu_{wi}}{\mu_{oi}}\right)^{+0.0770} \cdot \left(S_w\right)^{-0.1903} \cdot \left(\frac{p_i}{p_a}\right)^{-0.2159} \text{ barrels per acre foot} \quad (1)$$

For units used, see VII — NOMENCLATURE.*

The coefficient of multiple correlation r for this equation is 0.958. This coefficient indicates the "goodness of fit" and means that the regression was successful in removing $(0.958)^2$ or 91.8 percent of the original variance in the data. The standard error of estimate of this equation is 17.6 percent. After dividing Equation (1) by 77.58 times the oil-in-place term $\frac{\phi(1-S_w)}{B_{oi}}$, the corresponding equation for the Recovery Efficiency (in percent of initial stock tank oil in place) was found to be:

$$RE = (54.898) \cdot \left\{\frac{\phi(1-S_w)}{B_{oi}}\right\}^{+0.0422} \cdot \left(\frac{k\mu_{wi}}{\mu_{oi}}\right)^{+0.0770} \cdot \left(S_w\right)^{-0.1903} \cdot \left(\frac{p_i}{p_a}\right)^{-0.2159} \text{ percent} \quad (2)$$

[1]See p. 27 for references.

*Symbols used are those adopted as standard by Society of Petroleum Engineers of AIME.

For Solution Gas Drive Reservoirs (Sands, Sandstones, and Carbonate Rocks)

The best equation for the Recovery Factor below the bubble point was found to be:

$$BAF = (3244) \cdot \left\{\frac{\phi(1-S_w)}{B_{ob}}\right\}^{+1.1611} \cdot \left(\frac{k}{\mu_{ob}}\right)^{+0.0979} \cdot \left(S_w\right)^{+0.3722} \cdot \left(\frac{p_b}{p_a}\right)^{+0.1741} \quad \text{barrels per acre foot} \quad (3)$$

The coefficient of multiple correlation r for this equation is 0.932 and the standard error of estimate is 22.9 percent. In applying this correlation to under-saturated oil reservoirs, (bubble point pressure p_b below initial pressure p_i) the additional recovery due to expansion of the reservoir and its fluids over the pressure difference $(p_i - p_b)$ should be added to the computed values. After dividing Equation (3) by 77.58 times the oil-in-place term $\frac{\phi(1-S_w)}{B_{ob}}$, the Recovery Efficiency (in percent of stock-tank oil in place at bubble point) is obtained:

$$RE = (41.815) \cdot \left\{\frac{\phi(1-S_w)}{B_{ob}}\right\}^{+0.1611} \cdot \left(\frac{k}{\mu_{ob}}\right)^{+0.0979} \cdot \left(S_w\right)^{+0.3722} \cdot \left(\frac{p_b}{p_a}\right)^{+0.1741} \quad \text{percent} \quad (4)$$

A comparison of the observed Recovery Factors and those computed by means of Equations (1) and (3) is shown on Fig. 1 and 2. The relative weighting of the data points is indicated by the size of the circles.

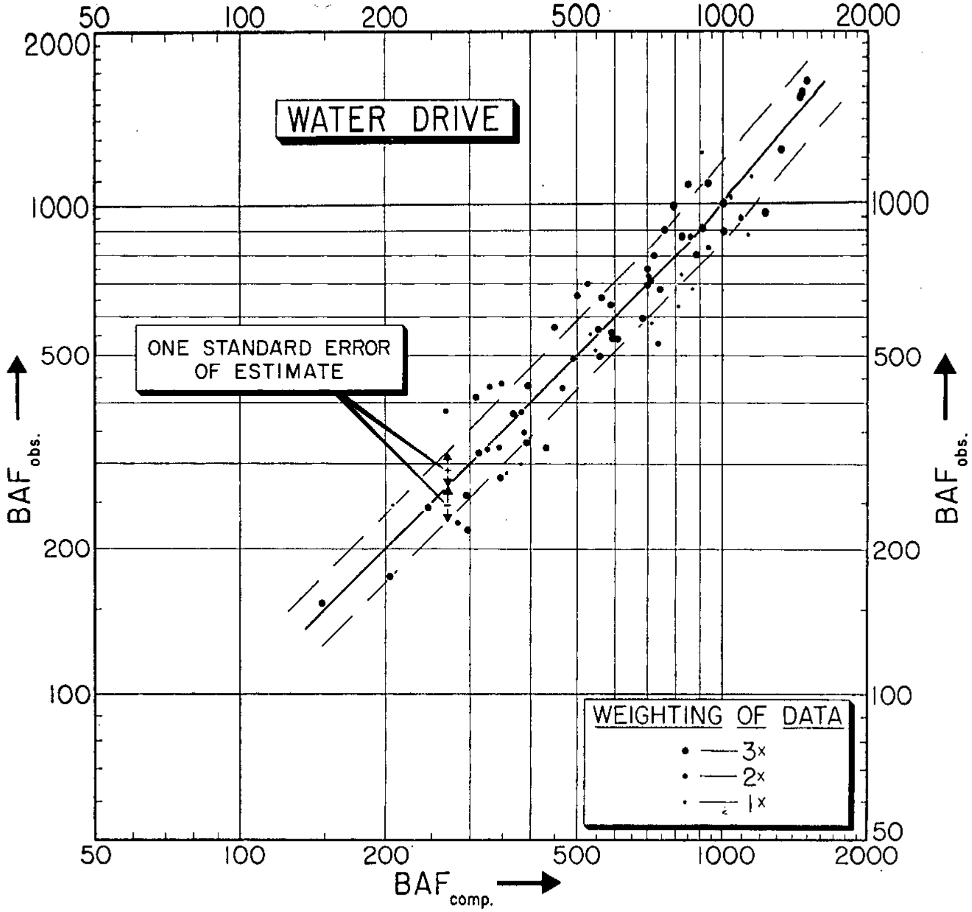
It may be noted that in Equations (1) and (3) for the Recovery Factor, in barrels per acre foot, four of the significant independent variables are not only similar but their exponents are also of the same order of magnitude (although of opposite sign in the case of the water saturation and pressure ratio). These findings may point to a fruitful field for further research.

Independent variables other than those listed in these equations may have a significant effect on recovery, but the number of case histories of acceptable accuracy was insufficient to support inclusion of additional parameters under the requirements of statistical validity. Continuation of this study with a larger number of case histories could well lead to the addition of other parameters to the regression equations.

The Subcommittee wishes to emphasize that these correlations represent a statistically valid relationship between empirical data and that computed results should never be considered as an absolute answer. The equations should be applied with particular caution to cases where the basic parameters fall outside the range of data as shown by Tables 2 and 3. They will yield the most probable Recovery Factor; but there is, as in all statistical correlations, a chance that the actual recovery will still fall outside the range shown on Fig. 1 and 2. It should be emphasized that the accuracy of estimates based on these equations will not be better than the reliability of the input data. Therefore, collection of accurate information on the reservoir rock and fluids early in the life of a field remains essential if reliable reserve estimates are desired. In applying volumetric factors to reservoirs which are known to be lenticular, allowance should be made for the volume of the reservoir actually connected to the producing wells. Relatively few of the reservoirs used in this study were in the category requiring hydraulic fracturing to make production at commercial rates possible. Caution should be exercised, therefore, in applying these equations to such tight reservoirs. Also, caution is warranted in applying these equations to water drive reservoirs produced in a way which takes advantage of "trapped gas" effects.

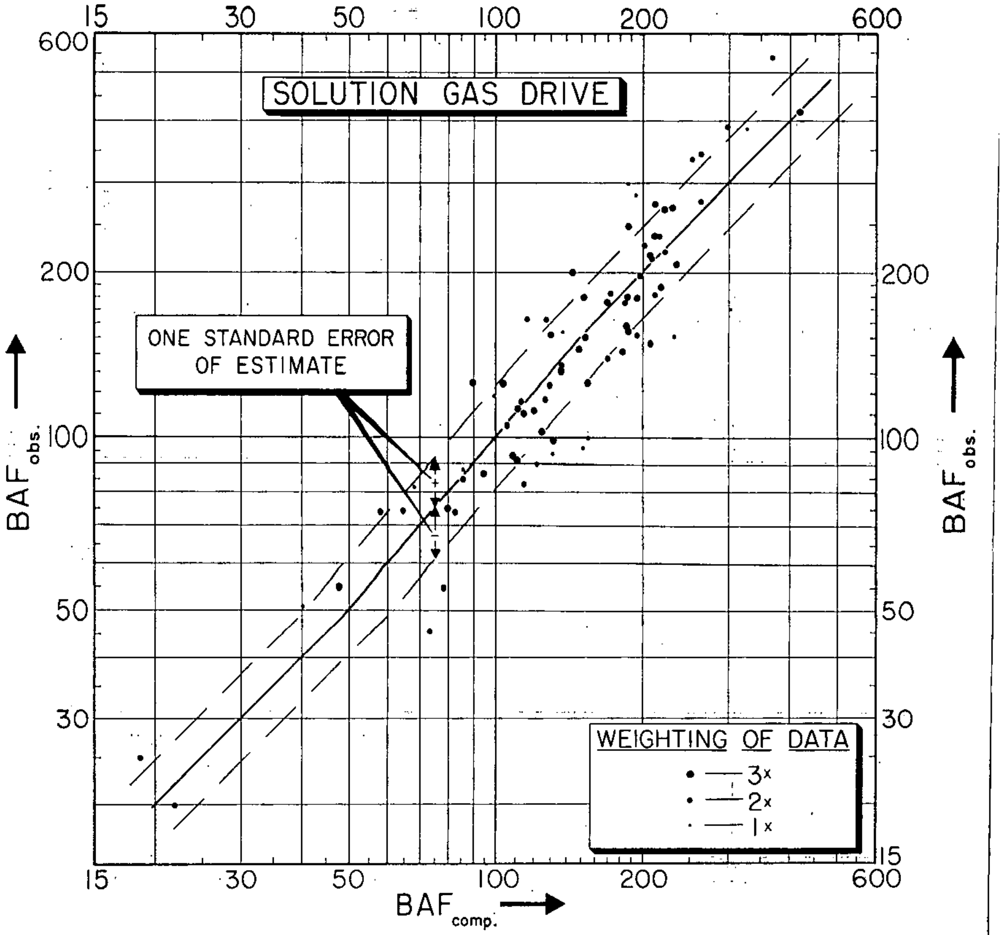
Because of the relatively small number of case histories available for the study of each mechanism, the statistical error in these equations is still rather large, and could be improved upon. Nevertheless, the Subcommittee feels that the results are so encouraging that continuation of this work is recommended. In order to reduce the standard error of estimate and be able to obtain statistically meaningful correlations with additional parameters, it is also recommended that an attempt be made to secure at least ten times more case histories for each mechanism than the number available for this study. In this respect, primary recovery data from fieldwide units studied extensively as fluid injection projects are particularly desirable. For future studies the questionnaire should be revised and simplified to include only parameters which are likely to be physically significant in the recovery process.

Considerable computer work during the early correlation efforts was contributed by British-American Oil Producing Company, Jersey Production Research Company, Standard Oil Company of California, The Atlantic Refining Company, Texas A & M University, DeGolyer & MacNaughton, Sun Oil Company, and Gulf Oil Corporation. The computer work on the correlations for the 1967 water drive study was performed by the firm of D. R. McCord in Dallas. The 1967 computer work on the solution gas drive cases and all final computer runs were contributed by Sun Oil Company in Dallas under the direction of Mr. R. W. Buchwald. The Subcommittee gratefully acknowledges these contributions.



$$\ln BAF_{comp.} = \ln 4259 + 1.0422 \ln \frac{\phi(1-S_w)}{B_{oi}} + .0770 \ln \frac{k\mu_{wi}}{\mu_{oi}} - .1903 \ln S_w - .2159 \ln \frac{p_i}{p_a}$$

Fig. 1—Comparison of Observed and Computed Values of Recovery Factors in Barrels per Acre Foot from 70 Water Drive Reservoirs
Coefficient of Multiple Correlation, r = 0.958
Standard Error of Estimate S_y = 0.162 or 17.6 percent



$$\ln BAF_{comp.} = \ln 3244 + 1.1611 \ln \frac{\phi(1-S_w)}{B_{ob}} + .0979 \ln \frac{k}{\mu_{ob}} + .3722 \ln S_w + .1741 \ln \frac{p_b}{p_a}$$

Fig. 2 — Comparison of Observed and Computed Values of Recovery Factors in Barrels per Acre Foot from 80 Solution Gas Drive Reservoirs below Bubble Point
Coefficient of Multiple Correlation, $r = 0.932$
Standard Error of Estimate, $S_y = 0.207$ or 22.9 percent

II — INTRODUCTION

In 1945 Craze and Buckley[1] published their now classic investigation on the effect of well spacing on ultimate recovery efficiency using data from 103 reservoirs. In 1955 Guthrie and Greenberger[2] used the same data to illustrate the use of multiple correlation analysis in obtaining the best relation between the recovery efficiency of water drive reservoirs and various combinations of parameters which were thought to control it.

During 1956 a group of engineers consisting of Messrs. R. J. Dobson, J. H. Sullivan, and S. C. Oliphant recommended the formation of an API Subcommittee on Recovery Efficiency. It was approved by the Executive Committee on Drilling and Production Practice in June of the same year. The assignment to the Subcommittee, as indicated in Mr. A. W. Thompson's report to the Board of Directors in the 1956 *Proceedings*, was "to evaluate the relative importance of the various types of data which are used or can be used in the estimation of reserves."

As a first step, the Subcommittee under its energetic Chairman, Mr. E. Kraus of The Atlantic Refining Company, designed a comprehensive questionnaire covering over 100 items of pertinent information on each reservoir.

It was intended to use electronic computers and statistical techniques, to determine from the information obtained to what extent various factors and parameters appear to affect ultimate recovery from oil reservoirs. The industry would thus be provided with recovery factors based on actual performance of producing oil fields instead of theory or laboratory studies.

Toward the end of 1959 the Subcommittee had received in excess of 300 case histories on oil reservoirs in the United States and abroad, and by the middle of 1960 the time-consuming job of checking the individual questionnaires was well on its way.

With Mr. Kraus' retirement in October 1960, the Subcommittee lost its inspiring Chairman who had provided much capable guidance through the organizational and data-gathering phases of the assignment.

Mr. J. J. Arps took over as Chairman and directed the Subcommittee through the correlation and computational work to the completion of the present report. He divided the work by geographical grouping of the membership between: *1*, a Houston task force (Chairman, Prof. Folkert Brons, University of Texas, succeeding Mr. E. K. Schluntz of Shell Development Company) to handle the correlations for water drive reservoirs; *2*, a Dallas task force (Chairman, Mr. A. F. van Everdingen, DeGolyer and MacNaughton, succeeding its original chairman, Mr. E. G. Trostel), to study possible correlations for solution gas drive reservoirs; *3*, an Oklahoma City-Tulsa task force (Chairman, Mr. L. F. Elkins, Sohio Petroleum Company) to study the recoveries from reservoirs under fluid injection; *4*, a Bartlesville-Ponca City task force (Chairman, Mr. V. T. McGhee, Phillips Petroleum Company) to assist in checking parameters of the reservoir fluids; and *5*, a California task force (Chairman, Mr. O. E. Van Meter, Jr., Mobil Oil Corporation) to assist in checking the parameters of the reservoir rocks.

The present report is the result of the combined effort of the entire Subcommittee membership represented by these different task forces.

III — PRELIMINARY ANALYSIS OF CASE HISTORIES

A. Number and Distribution of Case Histories by Groups

By the end of 1960 the number of questionnaires received by the Subcommittee had risen to 312, reporting on many different drive mechanisms and rock types. The number was considered somewhat small for statistical validity, nevertheless it was decided to proceed with the correlation work rather than attempt to secure a larger sample. The tedious and detailed work of reviewing and checking for obvious errors, duplication, or missing data and rating the questionnaire data was essentially finished by the end of 1963. For solution gas drive reservoirs this included correction of all recovery data to bubble point conditions; i.e., the recovery factors used in correlation studies were corrected by the Bartlesville-Ponca City task force under the direction of Mr. V. T. McGhee for the amount of oil recovered between initial reservoir pressure and bubble point pressure. Complete card decks and printouts of the factual data were then made available to the entire membership. Later additional adjustments to the data became necessary in some cases where more complete and conclusive production data had become available.

Table 1 shows a distribution of the original 312 case histories according to their predominant reservoir mechanism and rock types. The supplemental drive mechanisms listed in the third line in Table 1 were edge or bottom water drive or gravity. In a few cases there was some gas or water injection.

Table 1

Distribution of Case Histories by Reservoir Mechanism and Rock Type

Predominant Reservoir Mechanism	Reservoir Rock Type: Sand and Sandstone	Reservoir Rock Type: Limestone, Dolomite, Other	Total
Water Drive	72	39	111
Solution Gas Drive (without supplemental drive)	77	21	98
Solution Gas Drive (with supplemental drive)	60	21	81
Gas Cap Drive	11	3	14
Gravity Drainage	6	2	8
Total	226	86	312

B. Range of Basic Parameters

The range of the various parameters for each of the predominant mechanisms is listed on Tables 2 through 6. On these tables the independent variables are subdivided into three groups to define: *a*, the reservoir rock; *b*, the fluids in the reservoir; and *c*, the environment under which these fluids are being produced. For each parameter the minimum, median, and maximum values are listed; and, when sufficient cases were available, separate data are shown for sands and sandstones and for limestone, dolomite, and other.

The variables used as recovery parameters in the correlation studies are: barrels per acre foot *(BAF)* and recovery efficiency *(RE)* for all cases; residual oil saturation (S_{or}) for water drive reservoirs; residual gas saturation (S_{gr}) for solution gas drive reservoirs. The minimum, median, and maximum values are shown for each separate parameter; the values shown do not necessarily belong to the same case history.

Many of the contributors requested that all data submitted be treated as confidential. For that reason identification of field and reservoir names and their geographic locations were deleted in the card decks and printouts. Also, of course, no tabulation of the actual basic data could be included with this report.

Table 2

Range of Parameters of Reservoirs with Water Drive as Predominant Drive Mechanism

Parameters		Sand and Sandstone			Limestone, Dolomite and Other		
		Minimum	Median	Maximum	Minimum	Median	Maximum
Rock							
k	; darcys	0.011	0.568	4.000	0.010	0.127	1.600
ϕ	; fraction	0.111	0.256	0.350	0.022	0.154	0.300
S_w	; fraction	0.052	0.250	0.470	0.033	0.180	0.500
Fluids							
g_o	; deg API	15.5	35.3	50	15	37	54
μ_{oi}	; cp	0.2	1.0	500	0.2	0.7	142
μ_w	; cp	0.24	0.46	0.95	—	—	—
B_{oif}	; ratio	0.997	1.238	2.950	—	—	—
B_{obf}	; ratio	1.008	1.259	2.950	1.110	1.321	1.933
B_{oaf}	; ratio	1.004	1.223	1.970	—	—	—
$\rho_w - \rho_o$	; g/cc	0.054	0.241	0.490	—	—	—
Environment							
h_n	; ft	6.5	17.5	160	9	50.2	185
α	; deg	0–5	0–5	15–45	—	—	—
D_{bs}	; ft	1,400	6,260	12,400	2,210	6,790	13,100
T	; deg F	84	163	270	90	182	226
p_i	; psig	450	2,775	6,788	700	3,200	5,668
p_b	; psig	52	1,815	5,400	30	1,805	3,821
p_a	; psig	100	1,970	5,010	—	—	—
Ultimate Recovery							
BAF	; bbl/AF	155	571	1,641	6	172	1,422
RE	; percent	27.8	51.1	86.7	6.3	43.6	80.5
S_{or}	; fraction	0.114	0.327	0.635	0.247	0.421	0.908

Table 3

Range of Parameters of Reservoirs with Solution Gas Drive as Predominant Drive Mechanism without Supplemental Drives

Parameters		Sand and Sandstone			Limestone, Dolomite and Other		
		Minimum	Median	Maximum	Minimum	Median	Maximum
Rock							
k	; darcys	0.006	0.051	0.940	0.001	0.016	0.252
ϕ	; fraction	0.115	0.188	0.299	0.042	0.135	0.200
S_w	; fraction	0.150	0.300	0.500	0.163	0.250	0.350
Fluids							
g_o	; deg API	20	35	49	32	40	50.2
μ_{ob}	; cp	0.3	0.8	6	0.2	0.4	1.5
R_{sb}	; cf/bbl	60	565	1,680	302	640	1,867
B_{obd}	; ratio	1.050	1.310	1.900	1.200	1.346	2.067
B_{obf}	; ratio	1.050	1.297	1.740	1.200	1.402	2.350
B_{oad}	; ratio	1.000	1.090	1.400	1.060	1.120	1.420
Environment							
h_n	; ft	3.4	32.2	772	3.9	27	425
α	; deg	0–5	5–15	>45	0–5	0–5	5–15
D_{bs}	; ft	1,500	5,380	11,500	3,100	6,300	10,500
T	; deg F	79	150	260	107	174	209
p_b	; psig	639	1,750	4,403	1,280	2,383	3,578
p_a	; psig	10	150	1,000	50	200	1,300
Ultimate Recovery							
BAF	; bbl/AF	47	154	534	20	88	187
RE	; percent	9.5	21.3	46.0	15.5	17.6	20.7
S_{gr}	; fraction	0.130	0.229	0.382	0.169	0.267	0.447

Table 4

Range of Parameters of Reservoirs with Solution Gas Drive as Predominant Drive Mechanism and with Supplemental Drives

Parameters		Sand and Sandstone			Limestone, Dolomite and Other		
		Minimum	Median	Maximum	Minimum	Median	Maximum
Rock							
k	; darcys	0.010	0.216	2.500	0.002	0.019	0.867
ϕ	; fraction	0.120	0.210	0.359	0.033	0.133	0.248
S_w	; fraction	0.100	0.310	0.579	0.035	0.250	0.600
Fluids							
g_o	; deg API	15.5	36	46	22	38	46
μ_{ob}	; cp	0.4	0.9	20	0.3	0.8	2
R_{sb}	; cf/bbl	10	390	1,010	60	615	1,325
B_{obd}	; ratio	1.010	1.230	1.580	1.050	1.328	1.682
B_{obf}	; ratio	1.015	1.230	1.845	1.050	1.310	1.680
B_{oad}	; ratio	1.000	1.050	1.220	1.020	1.110	1.500
Environment							
h_n	; ft	4	30	714	8	31	154
α	; deg	0–5	0–5	>45	0–5	0–5	5–15
D_{bs}	; ft	300	4,237	10,280	2,800	6,000	10,530
T	; deg F	77	146	260	88	128	225
p_b	; psig	5	1,360	4,275	530	1,830	2,935
p_a	; psig	5	100	800	40	200	1,550
Ultimate Recovery							
BAF	; bbl/AF	109	227	820	32	120	464
RE	; percent	13.1	28.4	57.9	9.0	21.8	48.1
S_{gr}	; fraction	0.077	0.255	0.435	0.112	0.260	0.426

Table 5

Range of Parameters of Reservoirs with Gas Cap Drive as Predominant Drive Mechanism

All Rock Types Combined

Parameters		Minimum	Median	Maximum
Rock				
k	; darcys	0.047	0.600	1.966
ϕ	; fraction	0.086	0.225	0.358
S_w	; fraction	0.150	0.262	0.430
Fluids				
g_o	; deg API	34	40	43
μ_{ob}	; cp	0.3	0.6	2.3
R_{sb}	; cf/bbl	226	703	1,335
B_{obd}	; ratio	1.116	1.374	1.675
B_{obf}	; ratio	1.116	1.350	1.631
B_{oad}	; ratio	1.040	1.159	1.490
Environment				
h_n	; ft	7	15	35
a	; deg	0–5	0–5	15–45
D_{bs}	; ft	3,300	5,500	7,675
T	; deg F	108	175	200
p_b	; psig	854	2,213	3,583
p_a	; psig	0	500	2,900
Ultimate Recovery				
BAF	; bbl/AF	68	289	864
RE	; percent	15.8	32.5	67.0
S_{gr}	; fraction	0.223	0.271	0.517

Table 6

Range of Parameters of Reservoirs with Gravity Drainage as Predominant Drive Mechanism

Sand and Sandstone Only

Parameters		Minimum	Median	Maximum
Rock				
k	; darcys	0.305	1.285	2.000
ϕ	; fraction	0.194	0.329	0.350
S_w	; fraction	0.030	0.293	0.400
Fluids				
g_o	; deg API	15	22.5	38.5
μ_{ob}	; cp	0.7	4	8
R_{sbd}	; cf/bbl	96	200	735
B_{obd}	; ratio	1.070	1.106	1.383
B_{obf}	; ratio	1.070	1.106	1.340
B_{oad}	; ratio	1.030	1.040	1.100
Environment				
h_n	; ft	40	71	200
a	; deg	0–5	5–15	15–45
D_{bs}	; ft	1,170	2,400	6,500
T	; deg F	100	100	132
p_b	; psig	497	1,044	2,670
p_a	; psig	0	20	50
Ultimate Recovery				
BAF	; bbl/AF	250	696	1,124
RE	; percent	16	57.2	63.8
S_{gr}	; fraction	0.151	0.377	0.654

C. Rating and Weighting of Case Histories

The Subcommittee members reviewed and rated all questionnaires with respect to reliability of the volumetric data and degree of performance control on which estimates of recovery were based. From these two ratings the reviewers furnished a combined rating of each questionnaire as either "good," "adequate," or "poor." On this basis 47 percent of the questionnaires were rated as good, 34 percent as adequate and the remaining 19 percent as poor. In later regression analysis these ratings were "weighted" as follows:

Combined Rating of Volumetric Data and Performance Control	*Weight*
Good	3
Adequate	2
Poor	1

In the final regression equations the weighted data consistently showed better correlations than the unweighted. Consequently, the correlations presented in this report are based on data weighted as shown.

D. Predominant Drive Mechanisms

In 81 case histories (25.9 percent) solution gas drive was listed as the predominant mechanism supplemented by water drive, gravity and, in a few cases, partial gas or water injection. Since the exact contribution of these supplemental mechanisms was not known and the recovery factors of water drive and solution gas drive differ considerably, no correlations were attempted for this group. Also, the number of gravity drainage cases (8) and gas cap drive cases (14) was considered too small for statistical treatment. As a result only the recovery characteristics of water drive cases and unsupplemented solution gas drive cases were investigated in detail. No differentiation was made between those cases showing edge or bottom water as the dominant displacing medium and the analysis of the solution gas drive cases was restricted to those where the expansion of gas liberated from solution was considered the sole source of expulsion of oil.

It is realized that classification into water drive or solution gas drive used is still open to criticism. This classification has been adopted even though it is nearly impossible to conceive of any solution gas drive reservoir in which gravity would not be operative once free gas exists, nor of a water drive reservoir where free gas coming out of solution does not help recovery.

Sorting the questionnaires by predominant recovery mechanisms revealed some interesting relationships. For example, the two curves on the log probability charts of Fig. 3 and 4 may indicate that the probability of natural water influx into a sandstone or limestone reservoir increases as the permeability increases. The top curve on each of these charts relates the frequency distribution of the number of cases with water drive as a predominant mechanism to the logarithm of their average permeability, while the bottom curve shows the same distribution for those cases where no water influx is reported either as a predominant or supplemental mechanism. The distribution of this last curve also includes those cases which show gas cap drive or gravity as the predominant mechanism, without supplemental water influx. It may be noted from the curves on Fig. 3 and 4 that the permeabilities for the sandstone reservoirs in both cases exceed severalfold those for the limestone reservoirs.

Combination of the data from these two curves may be useful, when information other than permeability is lacking, in estimating the probability of a natural water drive occurring. This, in turn, may influence the choice of recovery factor for a given case between the extremes of water drive and solution gas drive.

E. Recovery Efficiencies of the Various Mechanisms

Recovery efficiency has been expressed as a percentage of the oil initially in place in the reservoir. In solution gas drive cases, corrections were applied to account for the volume of oil obtained between initial pressure and bubble point pressure.

Percentage figures can be studied in different ways, but for the comparison here the median value of each group was selected; e.g., in a group of 37 reservoirs the median value would be the recovery efficiency of reservoir #19, had the reservoirs been arranged in either ascending or descending order of efficiency.

The median values found for the various predominant drive mechanisms and rock types are shown on Table 7.

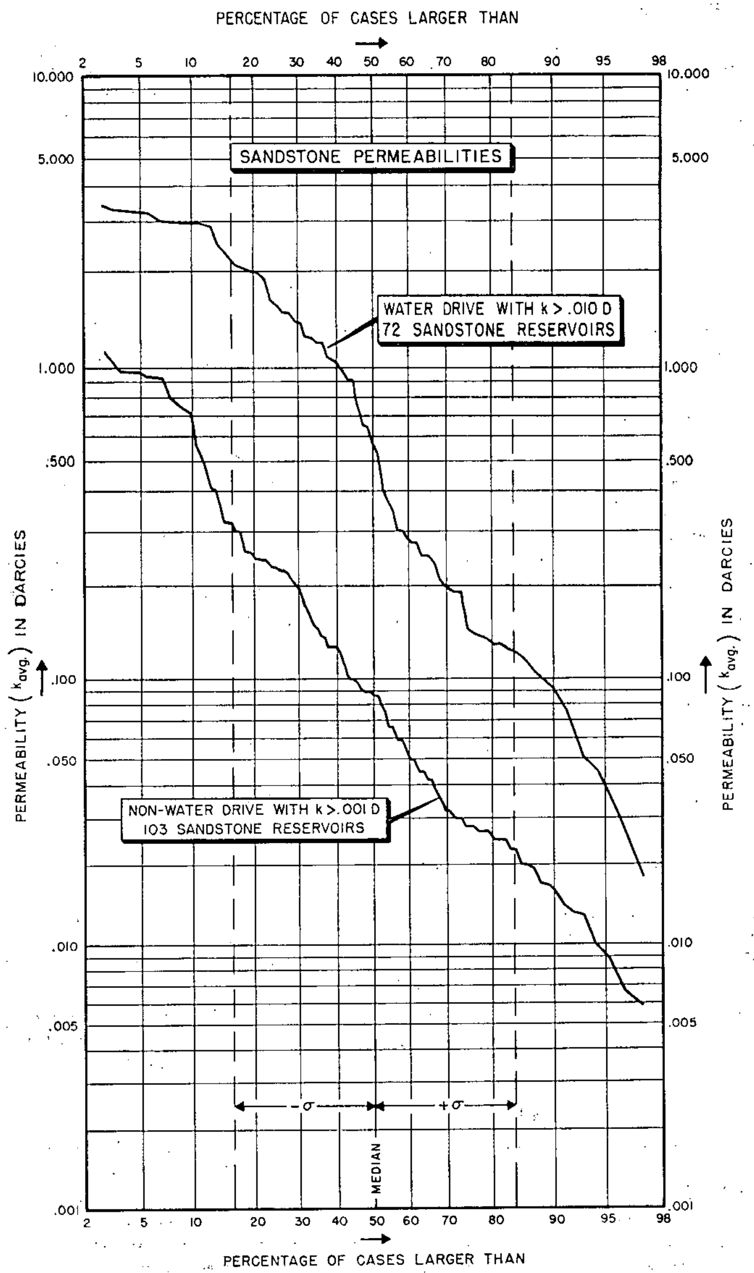


Fig. 3 — Cumulative Frequency Distribution of Permeabilities for Water Drive and Solution Gas Drive Cases — Sandstone Reservoirs

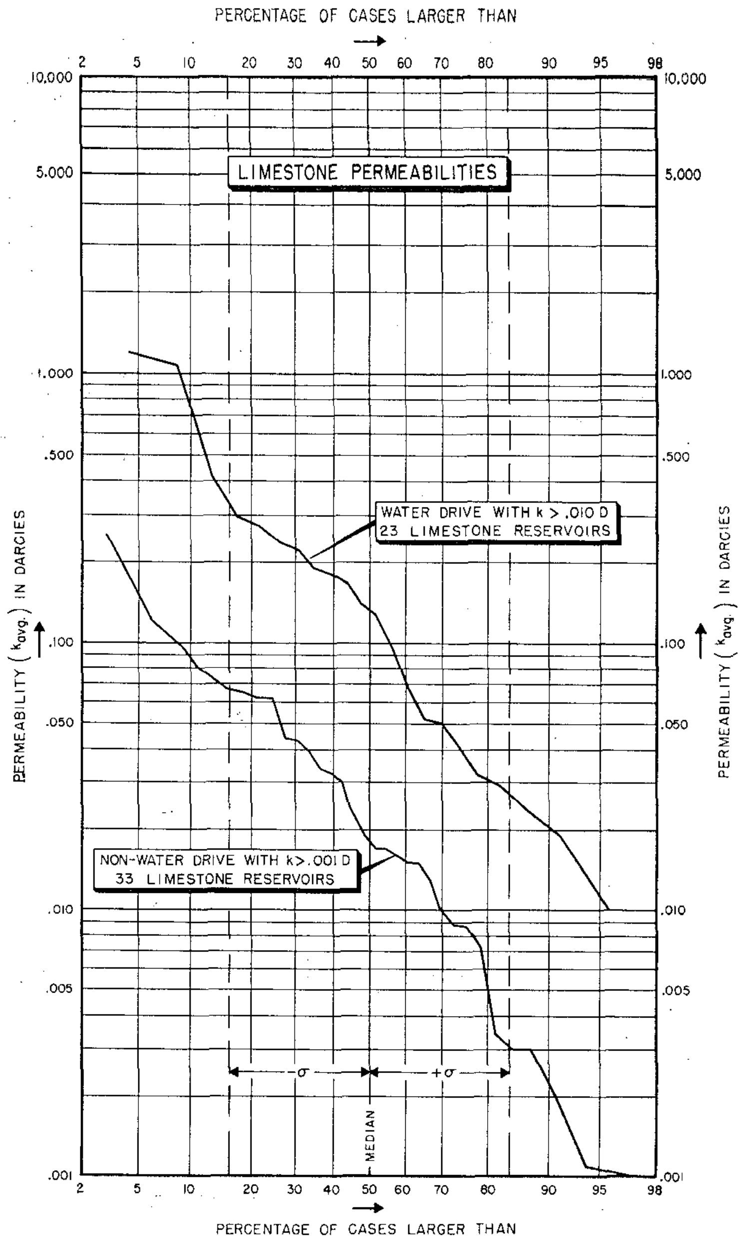


Fig. 4 — Cumulative Frequency Distribution of Permeabilities for Water Drive and Solution Gas Drive Cases — Limestone Reservoirs

Table 7

Median Values of Recovery Efficiency by Recovery Mechanisms

Predominant Recovery Mechanism	Whether Supplemented	Rock Type*	Median Value of Recovery Efficiency (*RE*), Percent
Water Drive	—	SS	51.1
Water Drive	—	LS	43.6
Gas Cap Drive	Yes	SS & LS	32.5
Solution Gas Drive	Yes	SS	28.4
Solution Gas Drive	Yes	LS	21.8
Solution Gas Drive	No	SS	21.3
Solution Gas Drive	No	LS	17.6

*SS = Sand and Sandstone. LS = Limestone, Dolomite and other.

Table 7 shows: *a*, that dissolved gas as an expulsive force is the least efficient recovery mechanism; and *b*, that it works somewhat better in reservoir rock with intergranular porosity, such as sand and sandstone, than in reservoirs with the intermediate type of porosity (carbonate rocks).[3] The table further underscores the much higher efficiency of water drive. This mechanism also seems to work better in sands and sandstones than in carbonate rocks. The number of gravity cases (8) was too small to permit a well-founded comparison. However, under ideal conditions the recovery efficiency of gravity drainage appears to rank with the best water drive cases.

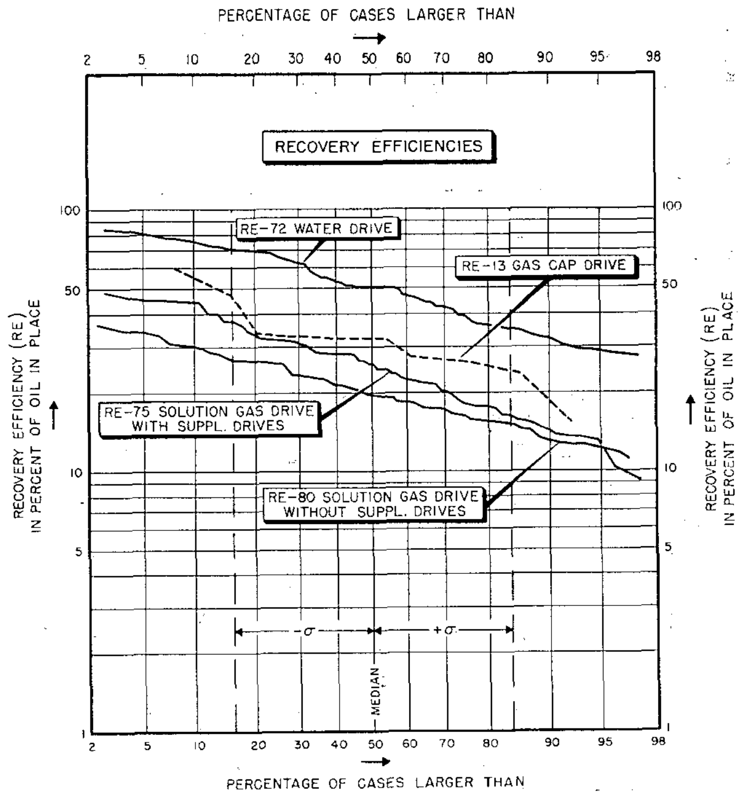


Fig. 5 — Cumulative Distribution of Recovery Efficiencies for Four Drive Mechanisms

The distribution of recovery efficiency values for the four main drive mechanisms is illustrated on the log. probability chart of Fig. 5. The extreme values outside of the 2 to 98 percent range are not shown.

It may be noted on Fig. 5 that:

a. Within the range of one standard deviation on either side of the median value, the curves for the four mechanisms follow the typical straightline relationships of a lognormal distribution.

b. The various trends are essentially parallel except for divergence in the trend for supplemented solution gas drive cases.

c. The highest recovery efficiency under water drive conditions reaches 84 percent of the oil in place, or about 3 times the lowest recovery shown for this mechanism. Two-thirds of the water drive cases fall in the 35 to 71 percent range with 51.1 percent as the median value.

d. In general, water drive seems to be around 2.6 times more efficient than unsupplemented solution gas drive.

e. The highest recovery efficiency under an unsupplemented solution gas drive mechanism reaches 37 percent, or about 3 times the lowest recovery shown for this mechanism. Two-thirds of the cases fall in the 15 to 27 percent range, with a median value of 19.5 percent.

f. The distribution curves for the supplemented and unsupplemented solution gas drive mechanisms appear to merge at the low end of the range and show a diverging trend toward the high end of the range. This clearly indicates that the effect of supplementation is much more pronounced when the reservoir characteristics themselves become more favorable. At the median value the improvement due to such supplementation by partial water drive, gas cap drive, or gravity is about one-third.

g. For those cases where a gas cap drive is reported as the predominant mechanism the average recovery efficiency improves by about two-thirds over that of solution gas drive alone.

F. Reservoir Rock

For complete statistical analysis it would have been desirable to separate the case histories according to the type of drive mechanism, with further subdivisions according to the type of reservoir rock. However, because the number of case histories was limited, separate correlations could not be made for the different rock types.

Of the 80 cases with acceptable accuracy in the solution gas group, which were used by the Dallas task force in their final correlations, 67 were obtained from sand and sandstone reservoirs and 13 from carbonate rock reservoirs.

Most of the case histories of water drives in limestone reservoirs came from the Middle East where limited depletion made it diffcult to confirm recovery estimates by performance. The remaining water drive case histories covering carbonate rock reservoirs lacked adequate supporting data. Therefore, the Houston task force restricted its study of water drive reservoirs to the 70 sand and sandstone cases which exhibited acceptable accuracy.

G. Well Spacing

The inclusion of well spacing as an independent variable affecting recovery received considerable attention in the study of the data. It is believed that the choice of well spacing normally is strongly affected by economic factors, such as ultimate recovery and well cost. This was confirmed by an analysis of the data obtained from the available questionnaires on the domestic fields, which showed a consistent increase of both the average-ultimate recovery per well and the average well spacing with increasing well depth.

In view of this observation it may be expected that in those cases which, for reasons other than well spacing, exhibit a recovery factor appreciably higher or lower than average, a bias toward closer or wider well spacing may also be present. Because of this the Subcommittee felt that the inclusion of well spacing as an independent variable in the analysis of recovery efficiency could lead to erroneous conclusions, and it was, therefore, omitted from the regression analysis.

IV — CORRELATION METHODS

In an analytical correlation study variables are separated into dependent variables which are known and independent variables which are thought to affect them. The connection between them is called a regression equation. This section will discuss in succession *a*, dependent variables; *b*, independent variables; and *c*, the various types of regression equations used.

A. Recovery Parameters or Dependent Variables

Several recovery parameters were considered as dependent variables.

a. Recovery Factor (BAF)

The commonly used Recovery Factor in barrels per acre foot of net pay became the first choice. In the study of solution gas drive reservoirs the observed Recovery Factors were corrected by the Bartlesville-Ponca City task force for expansion above the bubble point. The distribution of Recovery Factor values for various drive mechanisms shows essentially straight-line relationships on the log probability chart of Fig. 6, paralleling each other.

b. Recovery Efficiency (RE)

Recovery Efficiency in percent of stock tank oil initially in place was the second choice. Because of the logarithmic format selected for the regression equation, Recovery Efficiency is related to Recovery Factor so that essentially the same equation expresses the correlation for both. Distribution of the Recovery Efficiency values for different drive mechanisms also shows a parallel straight-line relationship on the log probability chart of Fig. 5. It may be noted that the respective curves show a much lesser slope, and therefore smaller variance, than the Recovery Factor curves of Fig. 6.

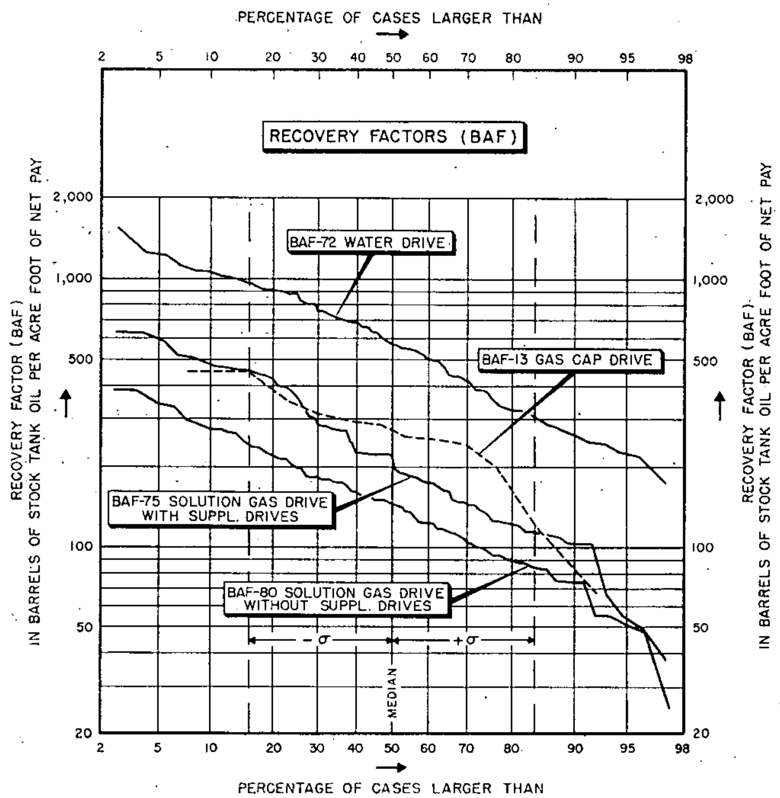


Fig. 6 — Cumulative Distribution of Recovery Factors for Four Drive Mechanisms

c. Residual Saturations (S_{or} *and* S_{gr})

The actual physical displacement mechanism in the reservoir was considered best represented by Residual Oil Saturation and Residual Gas Saturation. This opinion is not new because both were used as recovery parameters in the 1945 Craze-Buckley study of the effect of well spacing on recovery.[1] The Guthrie-Greenberger study of 1955[2] used Recovery Efficiency in regression analysis of data on the water drive reservoirs reported in the Craze-Buckley study. Residual Oil Saturation in water drive reservoirs at abandonment time is expressed as a fraction of the total pore space. Distribution of Residual Oil Saturation values computed for the water drive cases in this study is shown as a dashed curve on the log probability chart of Fig. 7.

Residual Gas Saturation of solution gas drive reservoirs at abandonment time is also expressed as a fraction of total pore space. The distribution of Residual Gas Saturation values for solution gas cases is shown as a solid curve on the log probability chart of Fig. 7.

Both Residual Oil Saturation and Residual Gas Saturation are computed indirectly from Recovery Factor and the parameters of rock and fluids. Many attempts were made to find direct correlations between Residual Oil Saturation and Residual Gas Saturation and the various independent variables. However, the results were generally unsatisfactory since the coefficients of multiple correlation were consistently lower than those for the recovery factors.

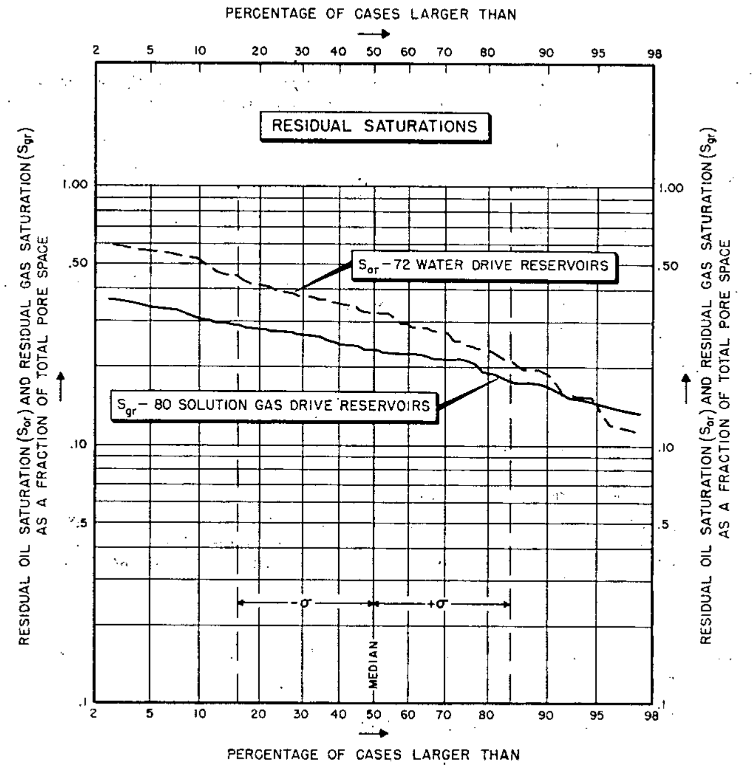


Fig. 7 — Cumulative Frequency Distribution of Residual Oil Saturation for Water Drive and Residual Gas Saturation for Solution Gas Drive Reservoirs

B. Reservoir Parameters and Independent Variables

In the preliminary correlations, practically all of the available reservoir and fluid characteristics (parameters) were used as independent variables. From the results it was possible to pinpoint those parameters which seemed consistently to affect the ultimate recovery. The Subcommittee was fortunate in having outstanding reservoir engineering talent among its membership, which proved most helpful in selecting and grouping these parameters. The independent variables finally selected were:

a. For the Water Drive Study

1. $\frac{\phi(1-S_w)}{B_{oi}}$, oil in place (initially).
2. $\frac{k\mu_{wi}}{\mu_{oi}}$, mobility ratio (under initial reservoir conditions).

   Permeability k was tested in three different ways against the Recovery Factor (BAF): as arithmetic average permeability, as median permeability, and as log mean of permeability. It was found that the strongest correlation was obtained with the arithmetic average permeability so this parameter was used throughout.
3. S_w, interstitial water saturation.
4. $\frac{p_i}{p_a}$, pressure ratio.
5. $\rho_w - \rho_o$, density difference between reservoir water and oil.
6. $\sqrt{\frac{k}{\phi}}$, "hydraulic radius" function.[4]

   In correlation work with the logarithmic model the regression process independently determines the appropriate exponent for each parameter and the k/ϕ variable was, therefore, introduced without its square root sign.

b. For the Solution Gas Drive Study

1. $\frac{\phi(1-S_w)}{B_{ob}}$, oil in place (at bubble point).
2. $\frac{k}{\mu_{ob}}$, mobility ratio (at bubble point).

   Actual data on gas viscosities were reported for only a limited number of cases. Since the range in viscosity values of reservoir gas is usually rather narrow, it was decided to retain only the oil viscosity at the bubble point in the mobility ratio for the solution gas drive reservoirs.
3. S_w, interstitial water saturation.
4. R_s, dissolved gas-oil ratio.
5. $\frac{p_b}{p_a}$, pressure ratio.
6. $\sqrt{\frac{k}{\phi}}$, "hydraulic radius" function.[4]

In the correlation work the square root sign was dropped.

Other independent parameters may prove to have a measureable effect on the Recovery Efficiency. In particular those related to relative permeability of the rock, variance of the permeability distribution, and interfacial tension would be of interest for further study. Such a study must, however, await the collection of more detailed data.

C. Regression Equations

Several types of regression equations were considered in the early correlation work with the following three main types actually being used.

a. Logarithmic Model

$$\ln y = a + b \ln x_2 + c \ln x_3 + \ldots \qquad (5)$$

$$y = (e)^a \cdot (x_2)^b \cdot (x_3)^c \ldots \qquad (6)$$

The nomenclature of these symbols is set out separately in Sec. VII.

This logarithmic model was selected as most applicable for the final correlations because:

1. Both the dependent variables and most of the independent variables indicate a good lognormal sampling distribution (see Fig. 3, 4, 5, 6, and 7).
2. The indicated Recovery Factor *(BAF)* showed a strong and almost linear correlation with the oil-in-place term so that after dividing by this factor and changing the constant *(a)* essentially the same correlation applies to Recovery Factor and Recovery Efficiency.
3. The effect of any specific independent variable having only single terms would be unidirectional.

b. Algebraic Model

$$y = a + b \cdot x_2^{\alpha} + c \cdot x_3^{\beta} + \ldots \quad (7)$$

The algebraic model was used largely in some of the early correlation work, but was dropped in the final correlation work in favor of the more suitable logarithmic model.

c. Logistic Model

$$y = \frac{C_1}{C_2 + e^{-(a + b \ln x_2 + c \ln x_3 + \ldots)}} \quad (8)$$

This model was used extensively during the early correlation efforts because it has the rather unusual advantage of allowing preset limits in the correlation relationship. For example, the range of the dependent variable would be between a minimum value of zero, when the term $(a + b \ln x_2 + c \ln x_3 + \ldots)$ reached minus infinity and a maximum of $\frac{C_1}{C_2}$ when this term reached plus infinity. In this manner the correlation could be prevented from assuming physically impossible values. However, the logistic model consistently yielded lower correlation coefficients than those obtained with the other models, in addition to being somewhat unwieldy. It was, therefore, dropped in favor of the logarithmic model.

V — EARLY CORRELATION EFFORTS

Complete sets of data printouts on all 312 case histories were prepared by Mr. R. K. Guthrie and distributed to members of the Subcommittee at a meeting in Dallas in August 1962. At that meeting, preferred recovery yardsticks and the best mathematical correlation models were also agreed upon. Correlation work was then assigned to each of the Subcommittee's five task forces. Results of these early efforts are summarized here.

A. Primary Recovery from Water Drive Reservoirs

1. In August 1962, Mr. J. C. Burke, British-American Oil Producing Company, presented a correlation of unweighted data for 52 sandstone reservoirs with water drive as the predominant drive mechanism. Using the "logistic model", Mr. Burke developed five regression equations with recovery efficiency as the dependent variable. His work demonstrated that the coefficient of multiple correlation improved steadily as additional independent variables were included in the regression equation. His best equation for recovery efficiency included permeability, viscosity ratio, and interstitial water saturation as independent variables and yielded a multiple correlation coefficient of 0.691.

2. In February 1964, Mr. R. E. Carlile, representing Prof. H. T. Kennedy of Texas A&M University, presented correlations used for his graduate Ph.D. thesis, "Recovery, Recovery Fraction, and Residual Oil Saturation Correlations for Sandstone Water Drive Reservoirs." Mr. Carlile, with data for 57 sandstone water drive reservoirs, developed 29 equations with recovery factor, recovery efficiency, and residual oil saturation as dependent variables. Independent variables were investigated as follows: Rock parameters of permeability, porosity, and water saturation; fluid parameters of oil-water viscosity ratio, solution gas-oil ratio, formation volume factor, and API gravity of the oil; and environmental parameters of net pay thickness, reservoir temperature, depth, and bubble point pressure. Mr. Carlile's best equation was obtained by relating residual oil saturation to viscosity ratio, porosity, bubble point pressure, reservoir temperature, and solution gas-oil ratio.

3. In February 1966, Mr. G. L. Hancock, Jr., Gulf Oil Corporation, presented correlations developed by Messrs. W. R. Cook and D. G. Strittmater of Gulf Research and Development Company, using 71 sandstone reservoirs with water drive as the predominant drive mechanism. The Gulf group developed two sets of 14 regression equations; one with unweighted data and a second one with weighted data. Algebraic, logistic, and, for the first time, logarithmic models were used. Dependent variables correlated were: recovery factor, recovery efficiency, and residual oil saturation. Independent variables which proved to be significant were: in the rock parameter category, permeability, porosity, and interstitial water saturation; in the fluid parameter group, oil-water viscosity ratio, solution gas-oil ratio, and initial formation volume factor; and in the enviromental category, initial pressure, and bubble point pressure. The highest coefficients of multiple correlation obtained were: for recovery factor, 0.886; for recovery efficiency, 0.708; and for residual oil saturation, 0.715. Gulf's correlation study showed the superiority of the logarithmic model and focussed attention upon the combination of parameters into groupings significant from a reservoir engineering standpoint. The most important groupings were the "oil-in-place term", and the "mobility ratio".

B. Primary Recovery from Solution Gas Drive Reservoirs

1. In May 1963, Mr. W. Sturm, under the able direction of Mr. E. G. Trostel, DeGolyer and MacNaughton, reported on his correlations for data from 95 sandstone and limestone reservoirs with solution gas drive as the predominant drive mechanism. Using a logistic-type model with 13 independent variables, his best correlation coefficient for recovery efficiency below the bubble point reached 0.756, while his best correlation for residual gas saturation showed a maximum of 0.802.

2. In February 1964, Messrs. J. K. Latimer, Jr. and W. C. Sturdivant, Jr. with Sun Oil Company in Dallas, presented their best correlations on 183 solution gas drive reservoirs, including those with and without other supplemental mechanisms. The algebraic model was used for 130 case histories with recovery factor and residual oil saturation as dependent variables, and 161 cases with recovery efficiency as the dependent variable. Independent variables which proved to be significant were: for rock parameters, permeability, porosity, and water saturation; for fluid parameters, API gravity, mobility ratio, and oil compressibility at the bubble point; and for enviromental parameters, net pay thickness, bubble point pressure, reservoir temperature, productive area, and depth subsea. The best coefficients of multiple correlation obtained were: for recovery factor 0.918, for recovery efficiency 0.517, and for residual oil saturation 0.973.

3. Also in February 1964, Mr. F. H. Hunter, with Jersey Production Research, presented an equation for recovery efficiency below the bubble point for 68 sandstone solution gas drive reservoirs without supplemental mechanisms. Independent variables were mobility ratio, porosity, interstitial water saturation, depth, ratio of well drainage radius squared to oil pay thickness, and ratio of solution gas-oil ratio to formation volume factor at the bubble point. Mr. Hunter reported that his equation provided a calculated recovery efficiency of within 25 percent for over two-thirds of the reservoirs and within 10 percent for one-third of the reservoirs.

4. During 1964, Mr. R. C. LeRoy, with The Atlantic Refining Company, prepared and presented to the Dallas Task Force a large number of correlations on solution gas drive reservoirs. Dependent variables correlated were: recovery factor, recovery efficiency, and residual gas saturation. As independent variables he used rock parameters of permeability, porosity water saturation, and critical gas saturation. For fluid parameters he used solution gas-oil ratio and API gravity of the oil. As significant parameters to describe environment he used reservoir temperature, bubble point pressure, and ratio between bubble point and adbandonment pressures. In his algebraic equations, different functions of these individual parameters were used, such as their linear, square root, square, and logarithmic form. The best coefficients of multiple correlation obtained in this work for various recovery parameters and for different groupings of solution gas drive case histories are enumerated below.

Solution Gas Drive Reservoirs Correlated	Rock Type	Highest Correlation Coefficient for: *BAF*	*RE*	S_{gr}
70 without supplemental drive	SS	0.885	0.719	0.860
55 with supplemental drive	SS	0.885	0.702	0.762
125 with and without supplemental drive	SS	0.892	0.709	0.794
21 with and without supplemental drive	LS	—	0.865	—
13 with supplemental drive	LS	—	0.884	—
39 with k_g/k_o data	—	0.958	0.894	—

Both the algebraic and the logistic mathematical models were used in the foregoing work.

In an effort to improve correlations further several additional runs were made, using inter-related functions of the independent parameters (x) of the type $a(x^{\alpha} + C_1)^{\beta}$ and $b \ln (cx^{\alpha} + C_2)$; but the additional complexity did not yield any further improvement in the correlation coefficients.

C. Secondary Recovery by Fluid Injection

In February 1964, Mr. F. H. Hunter, with Jersey Production Research, presented a correlation for recovery efficiency of secondary recovery by water injection based on data from 17 sandstone reservoirs. Mr. Hunter also presented a correlation for recovery efficiency of secondary recovery by gas injection based on data from 15 sandstone reservoirs. Because of the limited availability of relative permeability data, such data were not included in his correlations.

Mr. Lincoln Elkins, Sohio Petroleum, Chairman of the Oklahoma City — Tulsa Task Force, recommended, and the Subcommittee agreed, that no further effort should be made to correlate fluid injection recoveries against available parameters because of the small number of available cases.

D. Inter-Correlation of P.V.T. Data

Under the chairmanship of Mr. V. T. McGhee, with Phillips Petroleum Company, the Bartlesville — Ponca City Task Force rendered valuable service to the other task forces in solution of problems relating to P.V.T. data. In 1963, this group provided the Dallas Task Force with a correlation of oil compressibility and oil formation volume factors so that correlations for residual gas saturations at abandonment and recovery factors below the bubble point for solution gas drive reservoirs could be developed. In 1967 the P.V.T. group provided the Houston Task Force with reservoir water and oil density values for water drive reservoirs so that the density difference between water and oil could be tested as an independent variable in the regression equations.

E. Inter-Correlation of Rock Parameters

In February 1964, Mr. J. M. Miller, with Standard Oil Company of California, reported that his group had investigated relative permeability relationships and had attempted correlations with other rock properties. They had develepod interesting distributions of k_g/k_o data by geographical areas and by geological age; and also correlations between air and water permeabilities. A general correlation equation for residual gas saturation was also attempted with five independent parameters, including relative permeability and permeability variance. This equation indicated a correlation coefficient of 0.738.

F. Overall Results

Early correlation efforts, as summarized here, provided the necessary background for development of final correlations presented in this report. In particular, these earlier results proved helpful in:

a. Selecting the best dependent recovery variables for the final equations;

b. Selecting the best mathematical model for the final equations;

c. Indicating the independent parameters which proved to be persistent and dominant in the computer-derived regression equations so that their physical significance and relative importance could be studied further from a reservoir engineering viewpoint; and

d. Determining, by observing coefficients and exponents, which significant parameters could be combined into physically meaningful groupings so that the final equations could be improved and simplified.

In addition to those recognized previously in this report, the Subcommittee expresses their thanks to all engineering and computer personnel who have so generously contributed their valuable time and effort to the ultimate goal of the Subcommittee.

VI — FINAL CORRELATIONS

A. Results

The final results of the regressions are shown on Tables 8 and 9 for water drive and solution gas drive, respectively. The values of the constants and the coefficients of the independent variables for each case are listed horizontally for each regression equation, together with the multiple correlation coefficient (r) and the standard error of estimate[5] (S_y) expressed as a percentage. These independent variables were added in succession until the improvement in the multiple correlation coefficient was no longer statistically significant. Whether to accept each improvement as significant was decided by the so-called "F-test" (see Appendix).

It was found in both cases that the probability was very small that addition of the first four independent variables was not statistically significant; in no case more than 3.0 percent and in most cases 0.02 percent or less. The probability that the fifth and sixth variables are not statistically significant increased sharply to around 9 and 50 percent, respectively. At the same time, the addition of these two variables showed only a minor improvement in the standard error of estimate. The Subcommittee decided, therefore, that an equation comprising only the first four variables — oil in place, mobility ratio, water saturation, and pressure ratio — be accepted as statistically significant.

1. Water Drive

Based on figures shown on Table 8, the following logarithmic regression equation for the Recovery Factor under a water drive mechanism with the four significant independent variables is recommended:

$$BAF = (4259) \cdot \left\{\frac{\phi(1-S_w)}{B_{oi}}\right\}^{+1.0422} \cdot \left\{\frac{k\mu_{wi}}{\mu_{oi}}\right\}^{+0.0770} \cdot \left\{S_w\right\}^{-0.1903} \cdot \left\{\frac{p_i}{p_a}\right\}^{-0.2159} \text{ barrels per acre foot} \quad (9)$$

, where $4259 = e^{+8.3569}$

The multiple correlation coefficient for this equation is 0.958 and the standard error of estimate is 17.6 percent. Fig. 1 shows observed Recovery Factors plotted against those computed by means of Equation (9). The two dashed lines on this chart indicate the range of values between the computed amount plus or minus the standard error of estimate of 17.6 percent. It should be recognized that this standard error of estimate strictly applies only at the mean of the distribution.

By dividing both sides of Equation (9) by 77.58 times the oil-in-place term, the following equation giving the Recovery Efficiency for a water drive reservoir is obtained:

$$RE = (54.898) \cdot \left\{ \frac{\phi(1-S_w)}{B_{oi}} \right\}^{+0.0422} \cdot \left(\frac{k\mu_{wi}}{\mu_{oi}} \right)^{+0.0770} \cdot \left(S_w \right)^{-0.1903} \cdot \left(\frac{p_i}{p_a} \right)^{-0.2159} \text{ percent} \quad (10)$$

2. Solution Gas Drive

Based on figures shown on Table 9, the following logarithmic regression equation for Recovery Factor under a solution gas drive mechanism below the bubble point with four significant independent variables is recommended:

$$BAF = (3244) \cdot \left\{ \frac{\phi(1-S_w)}{B_{ob}} \right\}^{+1.1611} \cdot \left(\frac{k}{\mu_{ob}} \right)^{+0.0979} \cdot \left(S_w \right)^{+0.3722} \cdot \left(\frac{p_b}{p_a} \right)^{+0.1741} \text{ barrels per acre foot} \quad (11)$$

, where $3244 = e^{+8.0845}$

The coefficient of multiple correlation for this equation is 0.932 and the standard error of estimate is 22.9 percent. Fig. 2 shows observed Recovery Factors plotted against those computed by means of Equation (11). The two dashed lines on this chart indicate the range of values between the computed amount plus or minus the standard error of estimate of 22.9 percent. It should be recognized that this standard error of estimate strictly applies only at the mean of the distribution.

By dividing both sides of Equation (11) by 77.58 times the oil-in-place term, the following equation giving the Recovery Efficiency below the bubble point in solution gas drive reservoirs is obtained:

$$RE = (41.815) \cdot \left\{ \frac{\phi(1-S_w)}{B_{ob}} \right\}^{+0.1611} \cdot \left(\frac{k}{\mu_{ob}} \right)^{+0.0979} \cdot \left(S_w \right)^{+0.3722} \cdot \left(\frac{p_b}{p_a} \right)^{+0.1741} \text{ percent} \quad (12)$$

B. Numerical Examples

To illustrate the effect of adding the six terms listed on page 16 in succession, Table 10 was prepared showing, for three typical cases in the low, average, and high range of values, both computed and observed recovery in barrels per acre foot.

Table 8

Summary of Exponents of Terms in Regression Equations

Recovery Factor *(BAF)* for Water Drive Reservoirs, Sands, and Sandstones — 70 Case Histories

(1) Constant	(2) Oil in Place $\frac{\phi(1-S_w)}{B_{oi}}$	(3) Mobility Ratio $\frac{k\mu_{wi}}{\mu_{oi}}$	(4) Water Saturation S_w	(5) Pressure Ratio $\frac{p_i}{p_a}$	(6) Hydraulic Radius $\frac{k}{\phi}$	(7) Density Difference $\rho_w - \rho_o$	(8) Coefficient of Multiple Correlation r	(9) Standard Error of Estimate, Percent S_y
9.0567	+ 1.3871	—	—	—	—	—	0.8650	32.2
8.7665	+ 1.1414	+ 0.1037	—	—	—	—	0.9268	23.2
8.1585	+ 0.9897	+ 0.1140	— 0.2280	—	—	—	0.9409	20.8
8.3569	+ 1.0422	+ 0.0770	— 0.1903	— 0.2159	—	—	0.9575	17.6
8.5079	+ 1.1027	+ 0.0940	— 0.2040	— 0.2014	— 0.0444	—	0.9594	17.4
8.6421	+ 1.1334	+ 0.0701	— 0.1959	— 0.1969	— 0.0243	— 0.0901	0.9597	17.4

The probability of improvement in the correlation due to the addition of the last parameter in the regression equation being insignificant is less than 0.02 percent when the variables representing oil in place, mobility ratio, water saturation, and pressure ratio are successively included. Inclusion of the hydraulic radius and density difference had a 9.2 and 48.0 percent chance, respectively, of being insignificant. As the standard error of estimate is not appreciably improved by adding these last two variables the Subcommittee accepted the following regression equation as significant:

$$BAF = e^{+8.3569} \cdot \left\{ \frac{\phi(1-S_w)}{B_{oi}} \right\}^{+1.0422} \cdot \left(\frac{k\mu_{wi}}{\mu_{oi}} \right)^{+0.0770} \cdot \left(S_w \right)^{-0.1903} \cdot \left(\frac{p_i}{p_a} \right)^{-0.2159}$$

Table 9

Summary of Exponents of Terms in Regressions Equations Recovery Factor (BAF) for Solution Gas Drive Reservoirs, Sands, Sandstones, Carbonates and Other — 80 Case Histories

(1) Constant	(2) Oil in Place $\frac{\phi(1-S_w)}{B_{ob}}$	(3) Water Saturation S_w	(4) Mobility Ratio $\frac{k}{\mu_{ob}}$	(5) Pressure Ratio $\frac{p_b}{p_a}$	(6) Solution Gas-Oil Ratio R_s	(7) Hydraulic Radius $\frac{k}{\phi}$	(8) Coefficient of Multiple Correlation r	(9) Standard Error of Estimate, Percent S_y
8.0360	+ 1.2693	—	—	—	—	—	0.8899	29.3
8.4156	+ 1.3049	+ 0.2399	—	—	—	—	0.8968	28.4
8.5200	+ 1.2216	+ 0.3128	+ 0.0776	—	—	—	0.9090	26.8
8.0845	+ 1.1611	+ 0.3722	+ 0.0979	+ 0.1741	—	—	0.9317	22.9
8.4007	+ 1.0619	+ 0.3586	+ 0.1162	+ 0.1723	— 0.0828	—	0.9345	22.5
8.5697	+ 1.0694	+ 0.3800	+ 0.0881	+ 0.1728	— 0.0567	— 0.0351	0.9349	22.5

The probability of the improvement in the correlation due to the addition of the last parameter in the regression equation being insignificant is not more than 3.0 percent when the variables representing oil in place, water saturation, mobility ratio, and pressure ratio are successively included. Inclusion of the solution gas-oil ratio and hydraulic radius has an 8.6 and 52.0 percent chance, respectively, of being insignificant. As the standard error of estimate is not appreciably improved by adding these last two variables the Subcommittee accepted the following regression equation as significant:

$$BAF = e^{+8.0845}\left\{\frac{\phi(1-S_w)}{B_{ob}}\right\}^{+1.1611}\left(\frac{k}{\mu_{ob}}\right)^{+0.0979}\left(S_w\right)^{+0.3722}\left(\frac{p_b}{p_a}\right)^{+0.1741}$$

Table 10

Effect of the Addition of Various Terms in the Regression Equations on Computed Recovery Factors

	Barrels per Acre Foot Computed: Low	Average	High
WATER DRIVE			
A constant plus oil-in-place term	301	714	1,135
Above terms plus mobility ratio	243	655	1,085
Above terms plus water saturation	225	581	1,231
Above terms plus pressure ratio	244	554	1,326
Above terms plus hydraulic radius	232	575	1,362
Above terms plus density difference	231	585	1,349
These figures must be compared to observed values of	245	570	1,250
SOLUTION GAS DRIVE			
A constant plus oil-in-place term	56	129	277
Above terms plus water saturation	53	133	244
Above terms plus mobility ratio	51	131	279
Above terms plus pressure ratio	48	119	262
Above terms plus gas-oil ratio	50	116	271
Above terms plus hydraulic radius	51	115	270
These figures must be compared to observed values of	55	112	276

The results obtained from the recommended regression Equations (9) and (11) are underlined in Table 10.

It may be noted that, for the water drive mechanism, the inclusion of more terms results in a computed recovery nearer to the observed value. Adding terms in the solution gas drive mechanism appears to be less effective.

The following examples show, in detail, how the Recovery Factor may be computed for the forementioned two average cases by means of the preferred regression equations:

	Water Drive	Solution Gas Drive
Constant	4,259	3,244
Natural log of constant	8.3569	8.0845
Porosity	0.282	0.174
Oil-filled pore space	(1–0.350)	(1–0.340)
Shrinkage factor	1.100	1.402
Oil-in-place factor	0.16664	0.08191
Natural log of oil in place	—1.79194	—2.50212
Permeability	0.249	0.0199
Oil viscosity at bubble point	—	0.500
Oil viscosity at initial pressure	1.310	—
Water viscosity at initial pressure	0.540	—
Mobility factor	0.10264	0.03980
Natural log of mobility factor	—2.27652	—3.22389
Water saturation	0.350	0.340
Natural log of water saturation	—1.04982	—1.07881
Initital pressure	1,986	—
Bubble point pressure	—	3,660
Pressure at abandonment	800	580
Pressure ratio	2.48250	6.31034
Natural log of pressure ratio	0.90927	1.84219

The regression equation can now be written as follows:

Water Drive Case

$$\ln BAF = 8.3569 + 1.0422\,(-1.79194) + 0.0770\,(-2.27652) - 0.1903\,(-1.04982) - 0.2159\,(0.90927)$$

$$\ln BAF = 6.31752$$

$BAF = 554$ *computed vs. 570 observed.*

Solution Gas Drive Case

$$\ln BAF = 8.0845 + 1.1611\,(-2.05212) + 0.0979\,(-3.22389) + 0.3722\,(-1.07881) + 0.1741\,(1.84219)$$

$$\ln BAF = 4.78286$$

$BAF = 119$ *computed vs. 112 observed.*

The reader will recognize the figures outside the parentheses as the exponents appearing in Equations (9) and (11).

C. Discussion of Results

Equations (9) and (11) will yield the estimated values of the Recovery Factor for water drive and solution gas drive reservoirs, respectively. In the event the bubble point pressure of a solution gas drive reservoir, p_b, is less than its initial pressure, p_i, the amount of oil expected during the time the pressures drop from p_i to p_b must be added to the values computed from Equation (11). The variables appearing in these equations are arranged approximately in order of importance.

Oil-in-Place Term

In the accepted Equation (9) for water drive, the amount to be recovered, in barrels per acre foot, is equal to a constant, 4529, modified by the influence exerted by the succeeding four independent variables. A recovery proportional to the amount of oil in place needs no further explanation. The exponent of the oil-in-place factor, +1.0422, being slightly higher than unity, indicates that the recovery factor for water drive increases somewhat faster than the amount of oil in place.

In the accepted Equation (11) for solution gas drive reservoirs, the constant term, 3244, is about 25 percent less than the constant appearing in the water drive equation. It would be interesting to explore the reasons for the decrease; however, it is felt that an analysis in depth of both figures should be left until data on a greater number of sample reservoirs are available. The exponent of the oil-in-place variable, +1.1611, for the solution gas drive case, is somewhat larger than the one for water drive, which indicates that the recovery factor will increase more progressively with higher oil-in-place content.

The first two columns of Tables 11 and 12 list the oil content of the formation and the uncorrected recovery efficiency factor based thereon. To show the improvement in recovery efficiency when the oil content of the reservoir rock increases, the uncorrected recovery efficiency, RE', is given in percent of oil originally in place. This recovery efficiency for water drive reservoirs improves from 47.3 percent to 52.2 percent for a 10-fold increase in oil content. It increases from 23.8 percent to 34.4 percent for the same 10-fold increase in oil content of solution gas drive reservoirs.

Also shown on these tables is the effect of each of the three additional independent variables over the range of values observed in the reservoirs studied.

Mobility Term

The mobility term appears in the water drive case to the power +0.0770 and in the solution gas drive case to the power +0.0979. Tables 11 and 12 show that a maximum increase in the mobility term will approximately double the recovery from the solution gas drive case and increase the recovery from the water drive case by about 50 percent. The significance of this variable in the recovery process is well-recognized in the literature,[1] and the results of the present study confirm and give quantitative meaning to its effect.

Water Saturation Term

The interstitial water saturation parameter appears in the water drive equation to the power —0.1903 and in the solution gas drive case to the power +0.3722. From a reservoir engineering standpoint these opposite effects are as expected. Low water saturation in water drive reservoirs generally means a low surface area of the pores per unit of pore-space and, therefore, larger pores and a more efficient displacement process by water

In solution gas drive reservoirs low water saturation means more oil in place initially present in the pores. Apparently the void space created by shrinkage of this larger amount of oil upon pressure reduction causes gas to flow without moving much oil, thereby impairing the efficiency of the recovery process. These effects have been noted previously by several authors, including P. A. Dickey and R. B. Bossler in 1944[6] and M. Muskat and M. O. Taylor in 1946.[7]

According to the Tables 11 and 12, a 10-fold increase in the water saturation will decrease the contribution of the water saturation term in water drive reservoirs by about one-third. At the same time, such an increase will more than double the effect of the water saturation term in solution gas drive reservoirs.

Pressure Ratio Term

The pressure ratio can be considered as an indication of the effectiveness with which the available water drive mechanism is being utilized. A value close to one means that the operator is using all of the water drive. According to Table 11, an increase in the pressure ratio from 1 to 20 decreases the contribution of this term to the water drive recovery by almost 50 percent.

In solution gas drive reservoirs the expansion of the gas coming out of solution is the sole expulsive force, and it is to be expected that a maximum pressure drawdown in the reservoir is required to make full use of the available energy contained in the dissolved gas. Table 12 shows that an eventual drawdown of the reservoir pressure to 1/30th of the bubble point pressure will increase the contribution of this term to the recovery by around 75 percent.

In addition to the foregoing significant independent variables, the following terms were also investigated, but not included in the final correlations.

Gas in Solution Term for Solution Gas Drive Reservoirs

According to the data of Table 9 the amount of gas in solution appears to have only a small effect on the ultimate recovery from solution gas drive reservoirs. Apparently, the effect of the somewhat larger amount of free gas space created by a higher initial gas-oil ratio is just about offset by the effect of the correspondingly larger shrinkage of the oil in place, thus nullifying its effect on ultimate recovery in terms of stock-tank barrels. This confirms earlier findings, based on theoretical considerations by M. Muskat and M. O. Taylor in 1946[7] and J. J. Arps in 1955.[8]

Table 11

Effect of Variation in Independent Variables on Recovery Factor from Water Drive Reservoirs

Recovery Efficiency (RE) =

$$54.90\left\{\frac{\phi(1-S_w)}{B_{oi}}\right\}^{+0.0422} \times \left(\frac{k\mu_{wi}}{\mu_{oi}}\right)^{+0.0770} \times \left(S_w\right)^{-0.1903} \times \left(\frac{P_i}{P_a}\right)^{-0.2159} \text{percent}$$

OIL-IN-PLACE TERM		MOBILITY TERM		WATER SATURATION TERM		PRESSURE RATIO TERM	
If *OIP* equals:	Then the uncorrected recovery efficiency, *RE'*, is:	If mobility $\frac{k\mu_{wi}}{\mu_{oi}}$ equals:	Then the multiplier is:	If water saturation equals:	Then the multiplier is:	If P_i/P_a equals:	Then the multiplier is:
				0.05	1.77	1.0	1.00
3%	47.3%	0.01	0.701	0.10	1.55	1.5	0.92
5%	48.4%	0.03	0.763	0.15	1.43	2.0	0.86
10%	49.8%	0.10	0.838	0.20	1.36	3.0	0.79
15%	50.7%	0.30	0.911	0.30	1.26	5.0	0.71
20%	51.3%	1.00	1.000	0.40	1.19	10.0	0.61
30%	52.2%	3.00	1.088	0.50	1.14	20.0	0.52

Example:

$\phi = 0.282$), $S_w = 0.350$), $B_{oi} = 1.10$) $OIP = 16.7\%$, $RE' = 50.9\%$

$k = 0.249$ d), $\mu_{wi} = 0.540$ cp), $\mu_{oi} = 1.310$ cp) $\frac{k\mu_{wi}}{\mu_{oi}} = 0.103$; Multiplier = 0.839

$S_w = 0.350$; Multiplier = 1.22

$P_i = 1986\#$), $P_a = 800\#$) $\frac{P_i}{P_a} = 2.48$; Multiplier = 0.82

Recovery Factor (BAF) = 0.509 × 0.839 × 1.22 × 0.82 × 0.167 × 7758 = 554 Barrels per Acre Foot (vs. 570 observed)

Hydraulic Radius Term in Both Types of Mechanisms

According to the results shown on Tables 8 and 9, the effect of the "hydraulic radius" term in both equations appears to be small, probably because both permeability and porosity are already included in other variables and the addition of this term is, therefore, redundant.

Density-Difference Term in Water Drive Reservoirs

According to the data of Table 8, the effect of a density-difference term in the water drive equation appears to be small. It should also be pointed out that the addition of this term had a 48 percent probability of being insignificant. The Subcommittee feels that the behavior of this parameter in this correlation is probably not representative and may be deserving of further investigation when this study is continued.

General Observations

From the distribution pattern of the recovery efficiencies observed for the water drive and solution gas drive mechanisms on Fig. 5, it was noted in Sec. III E that for each of these mechanisms the recovery efficiency under maximum conditions is about three times higher than the percentage recoverable under minimum conditions. The foregoing analysis of the effect of the different variables on the recovery efficiency was extended to find out to what extent each variable contributes to this threefold range.

According to the data on Table 11, a maximum change in the oil-in-place variable in the water drive case may result in an improvement in recovery efficiency of 1.1 times. This improvement expresses essentially the effect of better reservoir rock. Such an improvement is usually accompanied by a simultaneous improvement in the other three

Table 12

Effect of Variation in Independent Variables on Recovery Factor from Solution Gas Drive Reservoirs

Recovery Efficiency (RE) =

$$41.82\left\{\frac{\phi(1-S_w)}{B_{ob}}\right\}^{+0.1611} \times \left(\frac{k}{\mu_{ob}}\right)^{+0.0979} \times \left(S_w\right)^{+0.3722} \times \left(\frac{P_b}{P_a}\right)^{+0.1741} \text{ percent}$$

OIL-IN-PLACE TERM		MOBILITY TERM		WATER SATURATION TERM		PRESSURE RATIO TERM	
If *OIP* equals:	Then the uncorrected recovery efficiency, *RE'*, is:	If mobility $\frac{k}{\mu_{ob}}$ equals:	Then the multiplier is:	If water saturation S_w equals:	Then the multiplier is:	If P_b/P_a equals:	Then the multiplier is:
		0.001	0.509	0.05	0.328	1.50	1.07
3%	23.8%	0.01	0.637	0.10	0.424	2.00	1.13
5%	25.8%	0.03	0.709	0.15	0.494	3.00	1.21
10%	28.9%	0.10	0.798	0.20	0.549	5.00	1.32
15%	30.8%	0.30	0.889	0.30	0.639	10.00	1.49
20%	32.3%	1.00	1.000	0.40	0.711	20.00	1.68
30%	34.4%	3.00	1.114	0.50	0.773	30.00	1.81

Example:

$\phi = 0.174$) $S_w = 0.340$) $B_{ob} = 1.40$) *OIP* = 8.20% *RE'* = 27.9%

$k = 0.020$ d) $\mu_{ob} = 0.50$ cp) $\frac{k}{\mu_{ob}} = 0.040$ Multiplier = 0.729

$S_w = 0.340$ Multiplier = 0.669

$P_b = 3660\#$) $P_a = 580\#$) $\frac{P_b}{P_a} = 6.31$ Multiplier = 1.38

Recovery Factor (BAF) = 0.279 × 0.729 × 0.669 × 1.38 × 0.082 × 7758 = 119 Barrels per Acre Foot
(vs. 112 observed)

---

variables, viz., increased mobility, lower water saturation, and a pressure drawdown ratio closer to one. According to the data on Table 11, each of these three additional factors under optimum conditions may be responsible for an improvement of around 1.5 times as compared to minimum conditions. The combined effect of these four factors, which are all in the same direction when the reservoir rock and its fluids approach ideal conditions, would cause a maximum improvement of better than four times over the poorest conditions.

In the solution gas drive case a maximum variation of the oil-in-place variable may result in an improvement in recovery efficiency of 1.5 times. This improvement expresses essentially the beneficial effect of better reservoir rock. When the permeability and, therefore, the mobility increases, the water saturation usually decreases. According to Table 12, the results of such changes in the mobility factor and the water saturation factor are in opposite directions and approximately cancel each other. Table 12 also shows that a maximum pressure drawdown may cause an improvement in recovery efficiency of around 75 percent. The combined effect of the oil-in-place factor and this pressure ratio factor, together with the cancellation of the mobility and water saturation factors, would cause a total improvement of slightly less than three times over the minimum conditions.

VII — NOMENCLATURE

Petroleum Engineering Symbols

α	= angle of formation dip; degrees
BAF	= recovery factor; ultmate recovery in barrels of stock-tank oil per acre foot of net pay
B_o	= oil formation volume factor; a dimensionless factor representing the volume of oil under reservoir conditions per unit volume of stock-tank oil
D_{bs}	= depth below the surface; feet
g_o	= gravity of stock-tank oil; degrees API
h_n	= net pay thickness; feet
k	= arithmetic average of absolute permeability; darcys
μ_o	= viscosity of reservoir oil; centipoises
μ_w	= viscosity of reservoir water; centipoises
p	= pressure; psig
RE	= recovery efficiency; recoverable oil as a percentage of stock-tank oil initially in place in the reservoir
RE'	= recovery efficiency yet to be corrected for the effect of mobility, water saturation, and pressure ratio terms
R_s	= gas solubility factor; the number of standard cubic feet of gas liberated under specific separator conditions which, under reservoir conditions, are in solution in one stock-tank barrel of oil
ρ_o	= density of reservoir oil; grams per cubic centimeter
ρ_w	= density of reservoir water; grams per cubic centimeter
S_{gr}	= residual gas saturation under reservoir conditions; a fraction of the total pore space
S_{or}	= residual oil saturation under reservoir conditions; a fraction of the total pore space
S_w	= interstitial water content; a fraction of the total pore space
T	= reservoir temperature; degrees Fahrenheit
ϕ	= effective porosity; a fraction of rock bulk volume

Subscripts

i	= initial conditions
a	= abandonment conditions
b	= bubble point conditions
d	= differential separation
f	= flash separation
o	= oil
w	= water
g	= gas
r	= residual at abandonment conditions

Probability and Statistical Symbols

N	= number of elements in a sample from a population
$m_1, m_2, \ldots m_n$	= elements or sets of data
x	= independent variable
y	= dependent variable
$\hat{y}$	= y value computed by regression analysis
$\bar{y}$	= mean of y values
$\bar{x}$	= mean of x values
$a, b, c, \ldots$	= coefficients in regression equation
$\alpha, \beta, \Upsilon \ldots$	= exponents in regression equation

C_1, C_2,	= constants in logistic equation
s_y^2	= variance of y with respect to $\bar{y}$ before regression
s_x^2	= variance of x with respect to $\bar{x}$ before regression
s_{yx}	= covariance of y and x
s_y	= standard deviation of y = square root of variance s_y^2
s_x	= standard deviation of x = square root of variance s_x^2
n	= number of degrees of freedom = $N-1$, $N-2$, etc.
w	= Weight factor; $\bar{w}$, mean of w values
S_y^2	= variance remaining after regression
p	= number of independent variables
Σ	= summation symbol
r	= coefficient of multiple correlation
F	= "F" test probability function (Ref. 9, page 240)
t	= "student" t probability function (Ref. 9, page 225)
S_y	= standard error of estimate in y
σ	= standard deviation of a normal distribution
ln	= natural logarithm to the base e
e	= base of natural logarithm

Subscripts

i	= any number
s	= significant
o	= first regression
1	= second regression
α	= significance level
obs	= observed
comp	= computed

VIII — REFERENCES

[1]Craze, R. C. and Buckley, S. E: A Factual Analysis of the Effect of Well Spacing on Oil Recovery, *API Drilling and Production Practice,* 144 (1945).

[2]Guthrie, R. K. and Greenberger, M. H: The Use of Multiple Correlation Analysis for Interpreting Petroleum Engineering Data, *API Drilling and Production Practice,* 130 (1955).

[3]Arps, J. J. and Roberts, T. G: The Effect of Relative Permeability Ratio, the Oil-Gravity and the Solution Gas-Oil Ratio on the Primary Recovery from a Depletion Type Reservoir, *Trans. Am. Inst. Mining Met. Engrs. (Petroleum Development and Technology,* **204**, 120 (1955).

[4]Leverett, M. C: Capillary Behaviour in Porous Solids, *Trans. Am. Inst. Mining Met. Engrs. (Petroleum Development and Technology)* **142**, 152 (1941).

[5]Ezekiel, M: *Methods of Correlation Analysis,* 211, John Wiley & Sons, Inc., New York, 1941, 2nd Ed.

[6]Dickey, P. A. and Bossler, R. B: Role of Connate Water in Secondary Recovery of Oil, *Trans. Am. Inst. Mining Met. Engrs. (Petroleum Development and Technology)* **155**, 175 (1944).

[7]Muskat, M. and Taylor, M. O: Effect of Reservoir Fluid and Rock Characteristics on Production Histories of Gas Drive Reservoirs, *Trans. Am. Inst. Mining Met. Engrs. (Petroleum Development and Technology)* **165**, 78 (1946).

[8]Arps, J. J: Estimation of Primary Oil and Gas Reserves, Frick *Petroleum Production Handbook,* **II**, 37, McGraw, Hill Book Company, New York, 1962.

[9]Beyer, W. H: *Handbook of Tables for Probability and Statistics,* The Chemical Rubber Company, 1966.

Appendix

REGRESSION ANALYSIS AND TESTS OF SIGNIFICANCE

Consider a population of N elements $m_1, m_2, \ldots m_N$, (which may be N cores, oilfields, etc.) and let y_i and x_i be properties measured on element m_i. The purpose of regression analysis is to establish if there is a significant relation between y and x. This relation is expressed by the regression equation. More specific, a linear relation of y to x is expressed by $y = a + bx$. This means that for every observed value x_i a calculated value $\hat{y}_i$ is determined by the equation:

$$\hat{y}_i = a + bx_i \quad . \; . \; . \qquad (13)$$

Before carrying out the process of determining a and b the following statistical properties are defined:

Mean of y: $\bar{y} = \frac{\Sigma\, y_i}{N}$

Mean of x: $\bar{x} = \frac{\Sigma\, x_i}{N}$

Variance of y with respect to $\bar{y}$: $s_y^2 = \frac{\Sigma\,(y_i - \bar{y})^2}{N-1}$

Variance of x with respect to $\bar{x}$: $s_x^2 = \frac{\Sigma\,(x_i - \bar{x})^2}{N-1}$

Covariance of y and x: $s_{yx} = \frac{\Sigma\,(y_i - \bar{y})(x_i - \bar{x})}{N-1}$

The standard deviations s_y and s_x are the square roots of the variances s_y^2 and s_x^2.

The factor $N-1$ appearing above signifies the number of independent observations (or degrees of freedom n) which is one less than the number of elements N, because one independent observation was lost in determining $\bar{y}$ and $\bar{x}$. If the measurements on all elements do not have the same validity (or accuracy) then this may be remedied by applying weight factors w_i which gives:

$$\bar{y} = \frac{\Sigma\, w_i y_i}{\Sigma\, w_i} \qquad s_y^2 = \frac{\Sigma\, w_i (y_i - \bar{y})^2}{\Sigma w_i - \bar{w}}, \text{ etc.}$$

However, the number of independent observations remains N.

In Equation (13) the constants a and b are determined by least square analysis, which is so well known that it is omitted here:

$$\left.\begin{aligned} b &= \frac{s_{yx}}{s_x^2} \\ a &= \bar{y} - b\bar{x} \end{aligned}\right\} \quad . \; . \; . \qquad (14)$$

At this point it may be desirable to note that the correlation between y and x may be either an *"apriori"* or a *"nonapriori"* correlation. If the correlation is "apriori", then the physical relation between y and x and the mathematical equation expressing this relation is largely known. The purpose of the regression analysis is here only to determine the constants in the equation which give the best fit. If, however, the correlation is "nonapriori", then the physical relation is as yet unknown; and, therefore, the purpose of the regression is to find the best form of both the equation and the size of the constants in it, which gives the best statistical fit. The physical relation not being known, there is a danger of constructing by regression analysis a curve which waves through all observed points causing *overfitting*.[2] Overfitting can largely be avoided first, by using the simplest equation possible, if the number of observations is small; and second, by using any apriori information available to select the equation which best agrees with this information. For instance, if it is known that y increases with increasing x, then the equation expressing this relation must be unidirectional. An equation $y = a + bx + cx^2 + \ldots$ where some of the constants are of opposite sign would not meet this requirement.

Coming back to the regression in Equation (13), the observed values y_i are distributed around the value $\hat{y}_i$. Only if this distribution is independent of x_i, i.e., the position of $\hat{y}_i$ on the regression line, will the values of a and b computed (14) give the best estimate of the linear regression. This is clarified in the following example. Let RE be the recovery efficiency of a field and $\frac{k}{\mu}$ the mobility ratio of its oil. In Fig. 8a the result is shown of a regression of $y = RE$ on $x = \frac{k}{\mu}$ by $\hat{y}_i = a + bx_i$. The observed values y_i are plotted vs. the computed values $\hat{y}_i$. It shows that the spread of y_i around $\hat{y}_i$ increases from left to right. Now instead the logarithms of RE and $\frac{k}{\mu}$ are regressed by taking $y = log\ RE$ and $x = log\ \frac{k}{\mu}$. Fig. 8b shows that the spread of y_i around $\hat{y}_i$ in this case is independent of the position on the regression line. Therefore, this last regression is preferred over the first.

The average of the fluctuations of the observed y values around the calculated y values can be expressed by the *remaining variance:*

$$S_y^2 = \frac{\Sigma (y_i - \hat{y}_i)^2}{(N-2)} \ldots \tag{15}$$

where $N - 2$ are the remaining degrees of freedom, because two independent observations are lost in determining a and b. More general, if a multiple regression is carried out with p independent variables ($p = 1$ for single regression) then (15) becomes:

$$S_y^2 = \frac{\Sigma (y_i - \hat{y}_i)^2}{(N-p-1)} \ldots \tag{15'}$$

To investigate how significant the regression is, the variance remaining after regression, S_y^2, must be compared with either the variance before regression s_y^2 or with the part removed by regression $s_y^2 - S_y^2$. This is achieved as follows:

Because $a = \bar{y} - b\bar{x}$ the regression equation $\hat{y}_i = a + bx_i$ may also be written as:

$$y_i - \hat{y}_i = y_i - \bar{y} - b\,(x_i - \bar{x}) \quad ,$$

which gives

$$\Sigma (y_i - \hat{y}_i)^2 = \Sigma (y_i - y)^2 + b^2 \Sigma (x_i - \bar{x})^2 - 2b\, \Sigma (y_i - \bar{y})(x_i - \bar{x})$$

or

$$(N-2)\, S_y^2 = (N-1)\, [s_y^2 + b^2 s_x^2 - 2b s_{yx}]$$

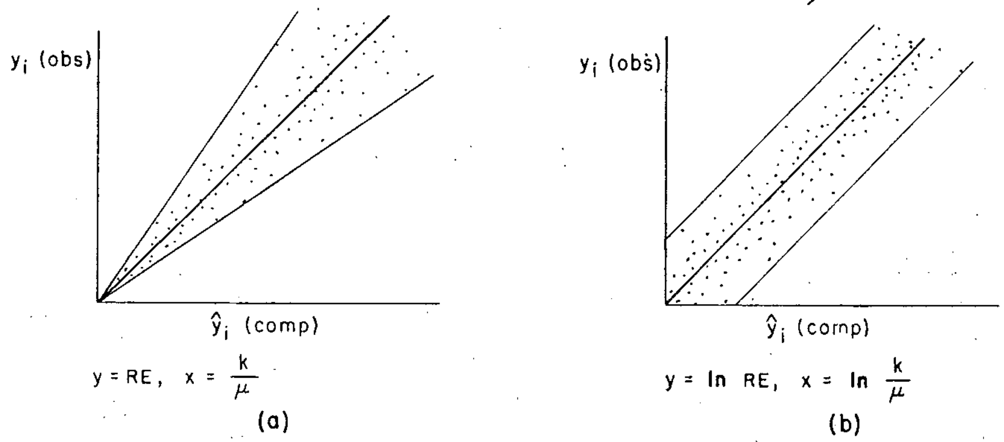


Fig. 8 — Correlation of Observed and Computed Dependent Variables

For large enough values of N only a negligible error is made by writing instead:

$$S_y^2 = s_y^2 + b^2 s_x^2 - 2b s_{yx} \quad ,$$

or, using $b = \frac{s_{yx}}{s_x^2}$

$$S_y^2 = s_y^2 - \frac{s_{yx}^2}{s_x^2}$$

Introducing the coefficient of correlation

$$r = \frac{s_{yx}}{s_y s_x}$$

gives:

$$S_y^2 = s_y^2 - r^2 s_y^2$$

remaining variance — original variance — part removed by regression (16)

degrees of freedom: $(N-2) = (N-1) - (1)$

The coefficient of correlation r ranges from zero (no correlation) to plus or minus one (perfect correlation).

To investigate the meaning of an r between zero and plus or minus one, the variance removed by regression $r^2 s_y^2$ (with one degree of freedom) is compared with the part remaining after regression $S_y^2 = s_y^2 (1 - r^2)$ (with $N-2$ degrees of freedom). For that purpose the F test can be used, where F is defined as:

$$F = \frac{\textit{variance removed/degrees of freedom}}{\textit{variance remaining/degrees of freedom}}$$

$$F_{1,n} = \frac{r^2 s_y^2/1}{(1-r^2)\, s_y^2/(n)} = n \frac{r^2}{1-r^2} \quad . \; . \; . \qquad (17)$$

$(n = N - 2 = \textit{number of degrees of freedom})$

The subscript $1, n$ of F signifies the number of degrees of freedom in the numerator and denominator and is used for reading F values from tables.

The F function has a known probability function which can be read from tables.[9] However, it is usually more convenient to use the "*student*" t function whose absolute value is related to F by:
or

$$|t_n| = \sqrt{F_{1,n}} = \sqrt{\frac{nr^2}{1-r^2}} \quad , \qquad (18)$$

where n is the number of degrees of freedom of the remaining variance. Note that for multiple regression with $p > 1$ independent variables, the F function becomes

$$F_{p,n} = \frac{r^2 s_y^2/p}{(1-r^2)\, s_y^2/(n)} = \frac{n \cdot r^2}{p\,(1-r^2)} \quad . \; . \; . \qquad (19)$$

where $n = N - p - 1$, the degrees of freedom of the remaining variance. In this case F cannot be replaced by a t-function.

The reasoning by which the significance of r is determined goes as follows: r was determined on a sample of limited size N. To draw conclusions from this sample valid for a much larger population from which N was taken, we should consider that another sample taken from the large population might have given a somewhat different value of r. Now suppose that in the large population y and x are unrelated. The expected value of r in N would then be zero, but there is still a chance that it fluctuates around zero. The corresponding fluctuations of the corresponding t around zero are described by the "*student*" t function, which can be found in tables and which is sketched in Fig. 9.

Fig. 9 shows that even if y and x are not related there is still 1/2 α% chance that t as calculated from the regression analysis is larger than t_α, and 1/2 α% chance that it is smaller than $-t_\alpha$. Therefore, there is α% chance that $|t| > |t_\alpha|$ even if no correlation exists. Suppose we take $\alpha = 5\%$ and let n be 60. Then from tables[9] it is read that $|t_{n,\alpha}| = 2.0$. If from regression analysis a value of r is found corresponding to a value of $|t| = 3$, and we accept this

r as a significant correlation coefficient, then there is less than 5% chance that we are wrong in our decision. Actually the value of α corresponding to $|t| = 3$ is less than 1%. Therefore, we will accept the correlation between y and x to be significant with the chance being less than 1% that our decision was wrong. If, as is conventional in certain industries, the significant level α is taken as 5%, with a corresponding value of $|t_\alpha|$, then we will only accept those r values as being significant, for which the corresponding $|t|$ value is larger than $|t_\alpha|$. The minimum value of r which is significant is r_s which is related to $|t_\alpha|$ as found from (18):

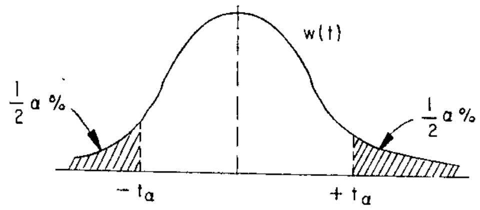


Fig. 9 — *"student"* t Distribution

$$r_s = \sqrt{\frac{t_\alpha^2}{n + t_\alpha^2}} \quad \ldots \qquad (20)$$

with $t_\alpha = t_{n,5\%}$ read from tables and $n = N - 2$ for single regression $(p = 1)$.

For multiple regression with p independent variables the F function must be used, giving

$$r_s = \sqrt{\frac{F_\alpha}{n + F_\alpha}} \quad \ldots \qquad (21)$$

with $F_\alpha = F_{p,\,n(5\%)}$

and $n = N - p - 1$

In Fig. 10 values of r_s are shown for $p = 1$ and $p = 2$ based on (20) and (21).

For example, on a sample $N = 60$, the r of the equation $y = a + bx$ $(p = 1)$ is only significant if it is larger than 0.25 and for the equation $y = a + bx + cz$, r must be larger than 0.32. For $N = 10$ these values become 0.65 and 0.75 respectively. Next consider a regression analysis giving

$$y = a + bx + cz + \ldots \qquad (p \text{ independent variables})$$

Let the remaining variance be S_{y0}^2 and the coefficient of correlation r_0. Now add another independent variable to it:

$$y = a + bx + cz + eu + \ldots \qquad (p + 1 \text{ independent variables})$$

Let the remaining variance be S_{y1}^2 and the coefficient of correlation r_1, which is larger than r_0.

To investigate if the second correlation is significantly better than the first, compare S_{y1}^2 with S_{y0}^2:

$$S_{y1}^2 = S_{y0}^2 - (S_{y0}^2 - S_{y1}^2)$$

variance after second regression | variance after first regression | variance removed between first and second regression

degrees of freedom: $(N - p - 2) = (N - p - 1) - (1)$

Again the F test is used, defined as

$$F_{1,n} = \frac{\textit{variance removed/degrees of freedom}}{\textit{variance remaining/degrees of freedom}}$$

or

$$F_{1,n} = (n)\,\frac{S_{y0}^2 - S_{y1}^2}{S_{y1}^2}$$

But

$$S_{y0}^2 = s_y^2\,(1 - r_0^2)$$

$$S_{y1}^2 = s_y^2\,(1 - r_1^2)$$

which gives

$$F_{1,n} = n\,\frac{r_1^2 - r_0^2}{1 - r_1^2}$$

with $n = N - p - 2$ = degrees of freedom after second regression.

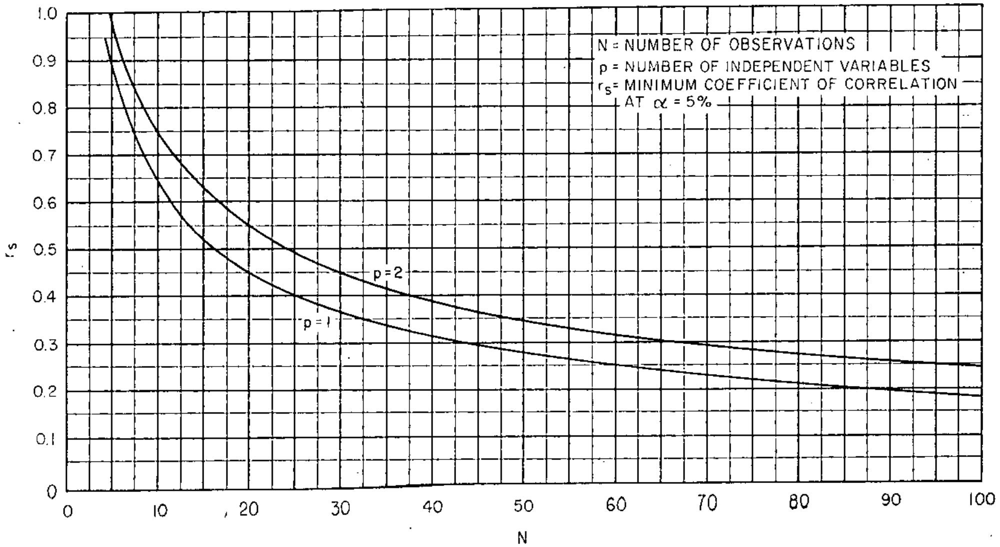


Fig. 10 — Significant Coefficient of Correlation

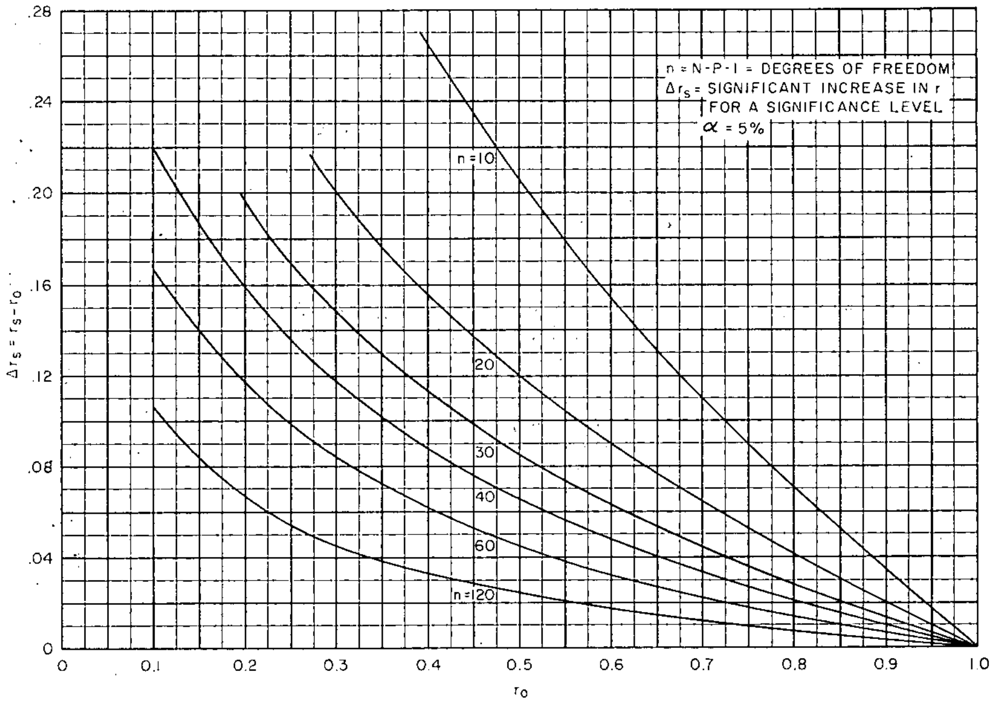


Fig. 11 — Significant Increase in Coefficient of Correlation

Again as before

$$F_{1,n} = t_n^2 = n\frac{r_1^2 - r_0^2}{1 - r_1^2} \tag{22}$$

The minimum value of t_n needed to be a significant improvement is t_α. The minimum value r which is significantly higher than r_0 is found from (22):

$$r_s = \sqrt{\frac{nr_0^2 + t_\alpha^2}{n + t_\alpha^2}}$$

and a significant increase over r_0 is

$$\Delta r_s = r_s - r_0 = \sqrt{\frac{nr_0^2 + t_\alpha^2}{n + t_\alpha^2}} - r_0 \tag{23}$$

Equation (23) is graphically represented in Fig. 11 for $\alpha = 5\%$. As an example of the use of Fig. 11 consider a sample of $N = 80$. First a single regression analysis is made $(p = 1)$ giving $r_0 = 0.80$. Then a variable is added $(p = 2)$ giving $r_1 = 0.82$. Is this a significant improvement?

Solution: $n = N - p - 2 = 76$. The minimum increase in r_0 to be significant as read from Fig. 11 is about 0.01. Therefore this is a significant improvement. The problem can also be considered in a reverse manner:

$$r_1 = 0.82, \quad r_2 = 0.80, \quad n = 76$$

From Equation (22) it follows that

$$t_n = \sqrt{n\frac{r_1^2 - r_0^2}{1 - r_1^2}} = 2.75$$

The corresponding value of α is found to be less than 1 percent, therefore, there is less than 1 percent chance that the increase in r from 0.80 to 0.82 is insignificant.