A New Method for Estimating the Dykstra-Parsons Coefficient To Characterize Reservoir Heterogeneity

Jerry L. Jensen, SPE, and **Iain D. Currie,** Heriot-Watt U.

Summary. We propose a new method to estimate the Dykstra-Parsons coefficient that leads to a more statistically reliable indication of the true heterogeneity level. The new method extracts more information from the data to produce an estimate that gives half the error of the traditional approach. Several cases are considered to demonstrate the effects of more reliable Dykstra-Parsons coefficients on predicted reservoir performance.

Introduction

The heterogeneity of a petroleum reservoir is a vital factor when the performance of an enhanced recovery project is considered. Numerous studies[1-5] have shown that permeability variations in the reservoir can be important in determining the amount of petroleum recovered. These variations are also influential in determining how the petroleum is recovered; performance factors such as time to breakthrough and peak hydrocarbon production have important economic implications for a recovery process. Unfortunately, the influence of heterogeneity on process performance is usually quite complex; detailed studies involving substantial amounts of data and extensive analyses are needed to predict hydrocarbon recovery accurately.

The petroleum engineer, however, may have to make a preliminary assessment of the reservoir performance when extensive data sets are unavailable or detailed analyses are not justified. In such cases, one is forced to use a simple statistic (e.g., the Dykstra-Parsons coefficient[6] or Koval's heterogeneity factor[3]) to represent the level of heterogeneity. Such situations can arise, for example, during screening studies. Basic models[1-3] may then be applied to estimate the recovery behavior and to determine whether a detailed study is justified. The use of simple heterogeneity measures can also arise when results from different field trials[7,8] or different process models are compared. Thus, despite the complex nature of the heterogeneity/performance relationship, a need for simple, representative, heterogeneity measures still exists.

In the past, several simple statistical measures of reservoir heterogeneity have been used.[1,3,6,8] The most popular appears to be the Dykstra-Parsons coefficient,[6] K_{DP}. This coefficient has been found to be a good indicator of the level of heterogeneity[9] and has been used in a variety of enhanced recovery studies.[2,4,5,10,11] Jensen and Lake,[12] however, have shown that estimates of the true reservoir K_{DP} can suffer from substantial statistical errors because the estimated value of K_{DP}, $(K_{DP})_{est}$, is computed on the basis of a limited number of samples. These statistical variations in $(K_{DP})_{est}$ can lead to significant errors in performance predictions.[12] To make the most of this useful heterogeneity measure, the statistical variations of $(K_{DP})_{est}$ should be reduced as much as possible.

Three factors influence the variability of K_{DP} estimates (assuming random sampling). Two factors are the size of the data set and the true value of K_{DP} for the reservoir.[12] Only the first of these two can be changed for a given reservoir. Not surprisingly, as the number of samples increases, the variations in $(K_{DP})_{est}$ decrease. Typically, a four-fold increase in the number of samples will halve the error.[12] On this basis, obtaining an acceptably accurate $(K_{DP})_{est}$ could prove rather expensive in terms of the number of permeability measurements required. Also, we already observed that there are times when predictions must be made with small data sets. The remaining factor that influences $(K_{DP})_{est}$ errors is the method used to compute $(K_{DP})_{est}$. A prudent choice in the Dykstra-Parsons coefficient estimator could also reduce $(K_{DP})_{est}$ variability.

The method currently used in the industry to compute $(K_{DP})_{est}$ has not substantially changed from the original description by Dykstra and Parsons.[6,9] Lambert[13] examined six estimators of K_{DP} by comparing, on a well-by-well basis, the values of $(K_{DP})_{est}$ for data from five different fields. Her study compared the estimators for agreement of values (bias). She did not, however, examine the statistical variability of the estimates (efficiency). The study by Jensen and Lake[12] proposed a new heterogeneity measure, U, that is related to K_{DP} and gives estimates 30% less variable than $(K_{DP})_{est}$.

This paper proposes a method, based on the maximum likelihood technique, that improves on the traditional method of calculating $(K_{DP})_{est}$. We show that, by obtaining more information from the data at hand, we can halve the variability of $(K_{DP})_{est}$ compared with the traditional method of estimation. This performance gives a further 30% improvement over that achieved by Jensen and Lake.[12] The implications of this improvement on performance predictions are considered in several examples. Because these results depend on estimating the underlying probability density function (PDF) of the permeability data, we discuss and compare some methods to do this efficiently. The results of a Monte Carlo study suggest two equally efficient methods.

Estimation of the Dykstra-Parsons Coefficient

This section begins by defining some statistical terms used in the analysis. The traditional method of estimating K_{DP} and its variants are next reviewed with an emphasis on statistical properties. The proposed estimation method is then described and its properties developed. The section finishes with several examples where the proposed method gives substantially improved performance estimates.

Definition of Statistical Terms. To compare estimators, the properties of their estimates must be determined. Suppose we have a set of data, containing n values (e.g., core-plug permeabilities), coming from a population that has a parameter with value u. Each datum is a random number whose behavior is dictated, in part, by the parameter u. We apply an estimator to the data to obtain an estimate of the value u, u_{est}. u_{est} is also a random variable because it is based on the n samples taken from the population. For the cases we consider (i.e., $n \geq 20$), the PDF of u_{est} is approximately normal with mean $\bar{u}$ and variance s_u^2 [written as $N(\bar{u}, s_u^2)$]. The difference $\bar{u}-u$ is the bias of the estimator and, if $\bar{u}-u=0$, the estimator is unbiased. The quantity s_u is the standard error of the estimator and gives a measure of how variable the u_{est} can be. The normal distribution of u_{est} implies that there is a 68% chance that $|u_{est}-\bar{u}| \leq s_u$ and a 95% chance that $|u_{est}-\bar{u}| \leq 2s_u$. When two estimators, A and B, are compared, if A produces an estimate with a lower standard error than that of B, then A is more efficient than B.

Several studies[13-15] found permeability populations that are not log-normally distributed. Jensen *et al.*[15] proposed that permeability, k, is "p-normally distributed" (p-ND). That is, permeability data from a population will have a PDF such that $(k)^p$ is approximately normally distributed for some p, $-1 \leq p \leq +1$. When $p=0$, the PDF is the log-normal distribution.[15] We assume that permeability is p-ND in this study.

Review of the Conventional K_{DP} Estimator. The technique of computing $(K_{DP})_{est}$ described by Dykstra and Parsons[6] requires

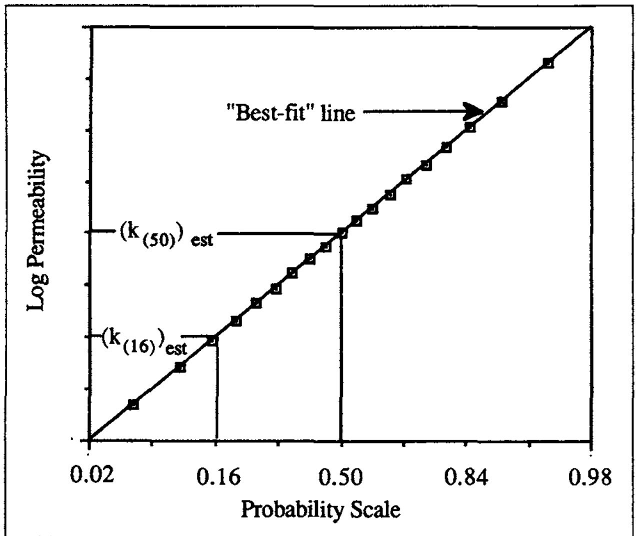


Fig. 1—Percentile estimates based on best-fit line.

estimating the 16th and 50th percentiles, $k_{(16)}$ and $k_{(50)}$ with $k_{(16)} \leq k_{(50)}$, from a set of ordered permeability data. The data are assumed to be log-normally distributed. The method calls for the data to be plotted on a log-normal probability plot and a best-fit line to be drawn and used to establish $[k_{(16)}]_{est}$ and $[k_{(50)}]_{est}$ (**Fig. 1**). The authors stipulate that if the data do not lie approximately on a straight line, the best-fit line is to be drawn by weighting the central portion more than the tails. The two percentile values are then used to define the heterogeneity measure as

$$K_{DP}=1-k_{(16)}/k_{(50)} \quad \text{(1)}$$

or, in terms of the estimated quantities,

$$(K_{DP}^{T})_{est}=1-[k_{(16)}]_{est}/[k_{(50)}]_{est}, \quad \text{(2)}$$

where T denotes the traditional method. This procedure, while calling for a log-normal probability plot, does not require that the data be log-normally distributed; a best-fit line may still be drawn and the percentiles estimated. Part of the success of the Dykstra-Parsons measure, however, appears to result from the fact that many reservoirs have nearly log-normal permeability distributions.

Subsequent descriptions of this technique vary in the details concerning the best-fit line. For example, some authors[16,17] do not mention the different weightings of the central and tail portions of the plot. Other descriptions[1,18,19] of the Dykstra-Parsons method omit mention of the best-fit line altogether; the 16th and 50th percentiles are estimated directly from the data. This approach eliminates the subjectivity of best-fit lines. Willhite[9] discusses an example where the fitting of such a line can produce unrepresentative models and results. The problem, however, appears to be caused by applying the procedure to data that are not log-normally distributed.

In the case where permeability is log-normally distributed [i.e., $\log_e(k)$ is $N(m_k, s_k^2)$], Jensen and Lake[12] give the following results for a sample size $n \geq 20$.

1. The bias of $(K_{DP})_{est}$ is slightly negative (<2%) with

$$m_{DP}^{T}=-0.7488 s_k^2 \exp(-s_k)/n. \quad \text{(3)}$$

2. The standard error of $(K_{DP})_{est}$ may be substantial and is given by

$$s_{DP}^{T}=1.486 s_k \exp(-s_k)/\sqrt{n}. \quad \text{(4)}$$

Because $K_{DP}=1-\exp(-s_k)$ ......(5)

for the log-normal distribution, Eqs. 3 and 4 can also be expressed in terms of the true K_{DP} of the permeability distribution:

$$m_{DP}^{T}=-0.7488[\log_e(1-K_{DP})]^2(1-K_{DP})/n \quad \text{(6)}$$

and $s_{DP}^{T}=-1.486 \log_e(1-K_{DP})(1-K_{DP})/\sqrt{n}$. ......(7)

Figs. 2 and 3 show the behavior described by Eqs. 6 and 7, respectively, when $K_{DP}=0.70$.

The errors in $(K_{DP}^{T})_{est}$ become particularly important for reservoirs with high levels of heterogeneity ($K_{DP} \geq 0.7$). Unfortunately, these high values are very common.[9] Variations of 10% in K_{DP} at high heterogeneity levels can make significant differences in predicted recovery performance, as we will see in the examples. Some investigators[1,12] have suggested new heterogeneity measures because of this sensitivity. An alternative approach, taken here, is to use a more efficient estimation technique.

Presentation of the Proposed K_{DP} Estimator. We begin with the case of log-normal permeability data. Using the additional information afforded by knowing the PDF explicitly in the estimation procedure results in a more efficient estimator. Eq. 5 suggests that, if we can estimate s_k efficiently by use of a maximum likelihood (ML) approach, an estimator for K_{DP} with lower error will result:

$$(K_{DP}^{ML})_{est}=1-\exp[-(s_k)_{est}]. \quad \text{(8)}$$

In Appendix A, we show that for the estimate

$$(s_k)_{est}=[1+0.25/(n-1)]\{[1/(n-1)]\Sigma[\log_e(k_i)-(m_k)_{est}]^2\}^{1/2}, \quad \text{(9)}$$

where $(m_k)_{est}=(1/n)\ \Sigma \log_e(k_i)$ and the summations are over the n data, $k_1, k_2 \ldots k_n$, the estimator of Eqs. 8 and 9 has the following behavior:

$$m_{DP}^{ML}=-0.250 s_k^2 \exp(-s_k)/n \quad \text{(10)}$$

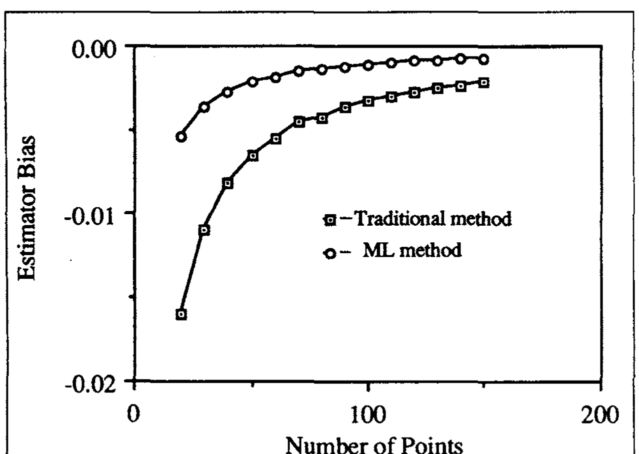


Fig. 2—Theoretical performance of estimator bias for K_{DP}=0.70.

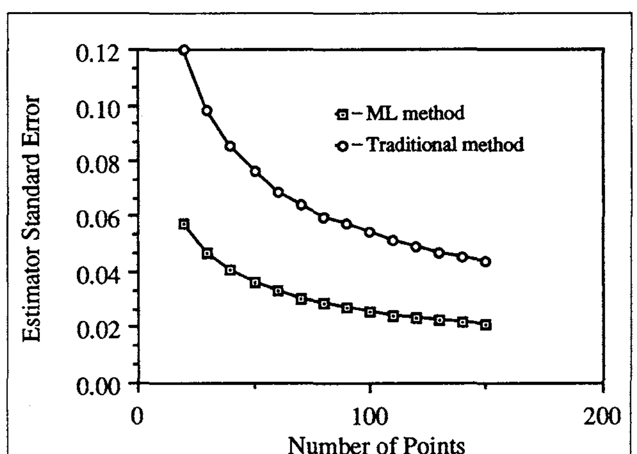


Fig. 3—Theoretical performance of estimator standard error for K_{DP}=0.70.

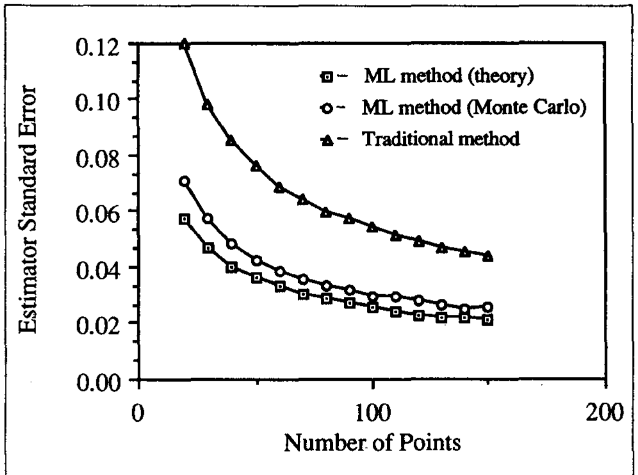


Fig. 4—Observed and theoretical estimator performance for $K_{DP}=0.70$.

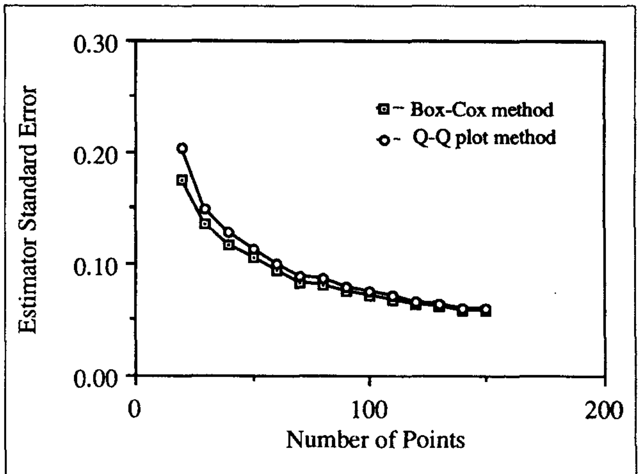


Fig. 5—Observed estimator performance for $p=0$ and $K_{DP}=0.70$.

and $s_{DP}^{ML}=0.707 s_k \exp(-s_k)/\sqrt{n}$. .... (11)

By use of Eq. 5, Eqs. 10 and 11 can be re-expressed in terms of the population K_{DP}:

$$m_{DP}^{ML}=-0.25[\log_e(1-K_{DP})]^2(1-K_{DP})/n \qquad (12)$$

and $s_{DP}^{ML}=-0.707 \log_e(1-K_{DP})(1-K_{DP})/\sqrt{n}$. .... (13)

These results are also shown in Figs. 2 and 3 for the case $K_{DP}=0.70$.

A comparison of Eqs. 3 and 4 with Eqs. 10 and 11 reveals the following features.

1. The ML estimator has a slightly smaller bias but still underestimates (on average) the heterogeneity level.
2. The ML estimator has about one-half the standard error of the traditional estimator. Alternatively, this means that the sample size for the proposed estimator is one-quarter that required by the Eq. 2 estimator for the same standard error.

Furthermore, the estimator of Eqs. 8 and 9 has a standard error that is 30% smaller than the heterogeneity estimator proposed by Jensen and Lake.[12] The proposed method puts the efficiency of heterogeneity estimation (à la Dykstra-Parsons) on par with that of the Lorenz coefficient estimator, which Jensen and Lake[12] showed to have superior ability in predicting waterflood performance.

The preceding analysis assumed that permeability is log-normally distributed ($p=0$). In Appendix B, we develop corresponding expressions for the case when $p \neq 0$. For both cases, Appendix C gives a comparison of the traditional and proposed estimators for data sets from two fields. Whatever the value of p, it must be estimated to use the ML technique proposed here. A good technique for evaluating p_{est} is an important element in the procedure; a bad estimate for p (e.g., $|p_{est}-p|>0.1$) could degrade the benefits of the new technique. We discuss two ways of estimating p efficiently later. Using these estimators in Monte Carlo simulations, we found that the actual performance of the ML approach is closely matched by the theory (**Fig. 4**).

In the context of estimating K_{DP}, a knowledge of the exponent p nearly quadruples the effective number of data on hand. More information is being extracted from the available data by estimating both the standard deviation, s_k, and the exponent p. The merits of estimating p may go beyond the purposes of this study, however. Jensen and Lake[12] showed that the parameter p can be as important as mobility ratio in waterflood performance. This parameter may have equally important performance implications for enhanced recovery methods.

Impact of Estimator Errors on Recovery Performance. To demonstrate the importance of efficient estimation of K_{DP} to predicted recovery performance, we chose three examples that use enhanced recovery models described in the literature. In all cases, we suppose that 40 samples are available from a reservoir having a log-normal permeability distribution and with $K_{DP}=0.70$. From Eq. 7, $s_{DP}^{T}=0.085$, and from Eq. 13, $s_{DP}^{ML}=0.040$. Consequently, the 68% confidence limits (neglecting bias) are $0.615 \leq (K_{DP}^{T})_{est} \leq 0.785$ and $0.660 \leq (K_{DP}^{ML})_{est} \leq 0.740$ for the traditional and ML estimators, respectively.

Claridge[1] considered the effects of heterogeneities on enhanced recovery processes by using computer simulations to model a graded-bank polymer flood in a 10-layer reservoir with crossflow. The drive water was 10 times more mobile than the polymer front. With this model at 2 PV of throughput, the 68% confidence limits of the traditional estimator translate to a range in oil recovery from 35 to 60%. For the ML estimator, the oil recovery ranges from 44 to 53%.

Paul *et al.*[2] presented a simplified model for predicting micellar/polymer performance. The effects of heterogeneity are accounted for in the vertical sweep, E_V, and mobility buffer, E_{MB}, efficiencies by stipulating a value for K_{DP}. For a dimensionless slug size of 1.3 and a 0.1-PV mobility buffer, the predicted sweep efficiency $(E_S)_{est}=[(E_V)_{est}(E_{MB})_{est}]$ varies by $\pm 27\%$ $[0.15 \leq (E_S)_{est} \leq 0.26]$ when $(K_{DP}^{T})_{est}$ is used, while $(E_S)_{est}$ varies by $\pm 12\%$ $[0.18 \leq (E_S)_{est} \leq 0.23]$ when $(K_{DP}^{ML})_{est}$ is used. The predicted breakthrough time, which is inversely proportional to the effective mobility ratio, is also affected by the variations in $(K_{DP})_{est}$. The 68% limits on the traditional and ML estimators correspond to errors of ± 37 and $\pm 17\%$, respectively, on the breakthrough times.

For unstable miscible displacements, we consider Koval's[3] model. The breakthrough time, t_{bt}, predicted by this model is inversely proportional to a heterogeneity factor K_K. Koval shows that K_K is related to K_{DP} and provides a curve (his Fig. 20) for which Claridge[1] gives a polynomial curve fit. When Claridge's curve fit on the above standard errors for $(K_{DP})_{est}$ is used with an assumed viscosity ratio of 20, the 68% points for the traditional approach represent the range $0.097 \leq (t_{bt})_{est} \leq 0.145$, while the ML approach gives $0.107 \leq (t_{bt})_{est} \leq 0.126$. Hence, an error band has been reduced from 40 to 17% with the more efficient ML approach.

The above examples all relate to enhanced recovery processes. Similar results pertain for waterflooding. When a unit mobility ratio and a 50% water cut are used in the Fassihi's[20] correlation, variations of one standard error in $(K_{DP}^{T})_{est}$ translate to $0.20 \leq (E_V)_{est} \leq 0.49$. For one-standard-error variations in $(K_{DP}^{ML})_{est}$, the sweep variation is $0.27 \leq (E_V)_{est} \leq 0.41$, reducing the range by one-half.

Taking the 95 instead of the 68% points would put the ML technique in a yet more favorable light for all the preceding examples because a reservoir with $(K_{DP}^{T})_{est}=K_{DP}+2s_{DP}^{T}=0.87$ is considerably more heterogeneous than a reservoir with $(K_{DP}^{ML})_{est}=K_{DP}+2s_{DP}^{ML}=0.78$. A similar situation exists when the heterogeneity is underpredicted. The relationship between K_{DP} and performance is generally nonlinear.[1,12]

Exponent Estimation

This section concerns finding an efficient technique for evaluating p_{est}, the exponent in the ML approach to estimating K_{DP}. We begin by describing and comparing several possible estimators. The results of a Monte Carlo study are then presented.

Review of Exponent Estimators. Several methods to estimate the exponent p from data exist. Each method was designed to satisfy particular criteria. For example, the Box and Cox[21] method is an ML technique, while Hinkley's[22] method selects the exponent that renders the transformed PDF as symmetrical as possible. Emerson and Stoto's[23] technique is also a symmetry-oriented method, but they use a different approach from Hinkley. In any event, the application of these estimators to a data set would likely result in three different estimates of p because each estimator has a different bias, efficiency, and criterion that it is meant to satisfy.

In the past, Emerson and Stoto's[23] method was used by Jensen *et al.*,[15] who found that it gave good results. For their problem, however, they needed a method that did not assume the result that they were trying to demonstrate—i.e., that permeability is p-ND. Because we are assuming that permeability is p-ND for this study, we can take advantage of this knowledge to establish a better exponent estimator.

The ML approach is used to establish a likelihood function[21]

$$L(\tilde{p})=-n/2\ \log_e(\tilde{s}_k^2)+(\tilde{p}-1)\ \Sigma\ \log_e(k_i),$$

where $\tilde{s}_k=[1/(n-1)]\Sigma(y_i-\tilde{m}_k)^2$, $\tilde{m}_k=(1/n)\Sigma y_i$, and $y_i=[(k_i)^{\tilde{p}}-1]/\tilde{p}$. The tilde signifies that the quantity is neither an estimated nor a true quantity of the population; it is a parameter of the calculation. The function L is evaluated for various values of $\tilde{p}$. In the present case, $\tilde{p}$ could vary from -1 to $+1$ in increments of 0.1. The value $\tilde{p}_{max}$ at which L is a maximum is the exponent that is most likely to be the true population exponent, so $p_{ML}=\tilde{p}_{max}$.

Box and Cox[21] discuss the theory and application of the ML method. They advocate making a plot of $L(\tilde{p})$ vs. $\tilde{p}$ to examine the behavior of the function before deciding on $\tilde{p}_{max}$. Our experience is that, while such a plot can help, one does not get a "feel" for the changes in the distribution of the data as $\tilde{p}$ is varied. Also, there is no absolute value that L must exceed; hence, there is a risk of selecting an exponent on the basis of the behavior of a few anomalous values. We found that making probability plots for selected trial values of $\tilde{p}$ is a more meaningful technique.

Our method consists of making a series of normal probability plots (QQ plots),[15] and estimating p on the basis of which exponent gives the "best" behavior. In this context, "best" usually equates with the straightest line. The line straightness can be numerically assessed by use of the correlation coefficient, r. If, however, a few "rogue" values are influencing the plot, that will be apparent and further investigation will be required. Usually, when no value of $\tilde{p}$ is appropriate, the maximum correlation coefficient obtained for the data set, r_{max}, behaves in one of two ways:

1. The r_{max} value is obtained for some $\tilde{p}$, $-1<\tilde{p}<+1$, but r_{max} is smaller than the significance test value for the size of data set being studied.[24]
2. The r_{max} value is obtained for some $\tilde{p}$ that is well outside the expected range (e.g., $\tilde{p}=3$).

In either case, the data should be examined for possible causes of non-p-ND behavior (e.g., data-entry error, fractured core plug, or low permeability value assigned a zero value).

Unmeasurably low permeabilities cannot be treated as 0 md with either the Box-Cox approach or the probability plots (when $\tilde{p}\leq 0$). A small [e.g., $<0.1\ k_{(50)}$] positive constant may be added to all the data, however, without changing the PDF.

Assessment of Exponent Estimators. A Monte Carlo computer program was written to test the properties of four exponent estimators: the Box-Cox[21] (BC) method, the normal probability plot (QQ) method, Hinkley's[22] (H) method, and Emerson and Stoto's[23] (ES) method. For each of nine exponents, $\tilde{p}=-1.0$, -0.75, -0.50 ... $+1.0$, the program generated 1,000 data sets of n points ($20\leq n\leq 150$) with which the estimators were tested. From a few preliminary trials, it became evident that all the estimators had fairly low bias $[|p-(\overline{p_{est}})|\leq 0.10]$. The ES method had the largest and the QQ method had the smallest bias. For efficiency, the BC method was best, followed by QQ, ES, and H. The standard error of $(p_H)_{est}$ was about twice the standard error of $(p_{BC})_{est}$. Because of the relatively poor efficiency of the H method, it was not included in the extended testing. Similarly, the ES method was not included because, while its efficiency was only slightly lower than the QQ method, it was more biased. Consequently, we selected only the BC and QQ method for extensive testing.

Fig. 5 is typical of the Monte Carlo test results obtained. For the case $\tilde{p}=0$, the standard error of the QQ method marginally exceeded that of the BC method by about 6%. When $p\neq 0.0$, the margin decreased to about 5%. From the preceding analysis and these test results, we chose the QQ method for the Monte Carlo results shown in Fig. 4. Either estimator, however, appeared to give good results.

The value of K_{DP} enters into these tests because it is a measure of the variability of the distribution. As K_{DP} increases, the data become more variable in value and, hence, it becomes easier to predict the true p value for the population. It is perhaps some consolation to know that, when we're working with a reservoir for which $K_{DP}=0.90$, the exponent estimate p_{est} will be very accurate.

Observations and Conclusions

A method based on maximum likelihood has been proposed as an estimator of the Dykstra-Parsons coefficient, K_{DP}. The suggested technique extracts more information from the available data than the conventional estimator. As a result, the ML approach produces estimates with half the error of estimates produced the traditional way. Equivalently, the proposed approach requires one-fourth the number of data points used by the conventional method to produce an estimate with a given accuracy.

The proposed technique requires the efficient estimation of the permeability distribution. Four methods were examined and two were found to give good results. The two methods differ in the amount of computational effort needed and interpretability of the result.

There is a continuing need for simple, representative heterogeneity measures in reservoir studies. Whenever the Dykstra-Parsons coefficient is used, the statistical errors of performance predictions can be reduced with the proposed technique. The additional computational effort required is readily justified by the improvement in performance prediction.

Nomenclature

E = efficiency
k = permeability, md
K = heterogeneity coefficient
m = bias or first moment of transformed data
n = number of data points in a set
N = normally distributed
p = exponent parameter of a distribution
r = correlation coefficient
s = standard deviation
t_{bt} = breakthrough time, dimensionless

Subscripts

BC = Box-Cox method
DP = Dykstra-Parsons
est = estimated quantity
H = Hinkley method
K = Koval
max = maximum
s = sweep
T = traditional
(16),(50) = percentile value of a distribution

Superscripts

E = extended definition
MB = mobility buffer
ML = maximum likelihood
$\sim$ = equation parameter
$^{-}$ = average value

References

1. Claridge, E.L.: "Control of Viscous Fingering in Enhanced Oil Recovery Processes: Effect of Heterogeneities," paper SPE 7662 available at SPE, Richardson, TX.
2. Paul, G.W. *et al.*: "A Simplified Predictive Model for Micellar/Polymer Flooding," paper SPE 10733 presented at the 1982 SPE California Regional Meeting, San Francisco, March 24-26.
3. Koval, E.J.: "A Method for Predicting the Performance of Unstable Miscible Displacement in Heterogeneous Media," *SPEJ* (June 1963) 145-54; *Trans.*, AIME, **228**.
4. Kossack, C.A. and Bilhartz, H.L.: "The Sensitivity of Micellar Flooding to Certain Reservoir Heterogeneities," paper SPE 5808 presented at the 1976 SPE Improved Oil Recovery Symposium, Tulsa, March 22-24.
5. Arya, A. *et al.*: "Dispersion and Reservoir Heterogeneity," *SPERE* (Feb. 1988) 139-48.
6. Dykstra, H. and Parsons, R.L.: "The Prediction of Oil Recovery by Water Flood," *Secondary Recovery of Oil in the United States,* second edition, API, Dallas (1950) 160-74.
7. Claridge, E.L.: "CO_2 Flooding Strategy in a Communicating Layered Reservoir," *JPT* (Dec. 1982) 2746-56.
8. Lake, L.W. and Pope, G.A.: "Status of Micellar-Polymer Field Tests," *Pet. Eng. Intl.* (Nov. 1979) 38-60.
9. Willhite, G.P.: *Waterflooding,* Textbook Series, SPE, Richardson, TX (1986) 172-74, 182.
10. Doepel, G.W. and Sibley, W.P.: "Miscible Displacement—A Multilayer Technique for Predicting Reservoir Performance," *JPT* (Jan. 1962) 73-80; *Trans.*, AIME, **225**.
11. Orr, F.M. Jr. and Sageev, A.: "Reservoir Characterization for the CO_2 Enhanced Oil Recovery Process," annual report, Contract No. AC21-85MC22042, U.S. DOE (March 1987).
12. Jensen, J.L. and Lake, L.W.: "The Influence of Sample Size and Permeability Distribution on Heterogeneity Measures," *SPERE* (May 1988) 629-37.
13. Lambert, M.E.: "A Statistical Study of Reservoir Heterogeneity," MS thesis, U. of Texas, Austin (1981).
14. Goggin, D.J. *et al.*: "Patterns of Permeability in Eolian Deposits," *SPEFE* (June 1988) 297-306.
15. Jensen, J.L., Hinkley, D.V., and Lake, L.W.: "A Statistical Study of Reservoir Permeability: Distributions, Correlations, and Averages," *SPEFE* (Dec. 1987) 461-68.
16. Pirson, S.J.: *Oil Reservoir Engineering*, McGraw-Hill Book Co., New York City (1958) 88-89.
17. Smith, C.R.: *Mechanics of Secondary Oil Recovery,* Robert E. Krieger Publishing Co., Huntington, NY (1975) 227-28.
18. Arps, J.J.: "Estimation of Primary Oil and Gas Reserves," *Petroleum Production Handbook,* T.C. Frick (ed.), SPE, Richardson, TX (1962) **2**, 37-29.
19. Mungan, N.: "Enhanced Oil Recovery Using Water as a Driving Fluid," *World Oil* (Sept. 1981) 155-67.
20. Fassihi, M.R.: "New Correlations for Calculation of Vertical Coverage and Areal Sweep Efficiency," *SPERE* (Nov. 1986) 604-06.
21. Box, G.E.P. and Cox, D.R.: "An Analysis of Transformations," *J. Royal Statistical Soc., Series B* (1964) **26**, No. 2, 211-52.
22. Hinkley, D.V.: "On Quick Choice of Power Transformation," *Applied Statistics* (1977) **26**, No. 1, 67-69.
23. Emerson, J.D. and Stoto, M.A.: "Exploratory Methods for Choosing Power Transformations," *J. Am. Stat. Assn.* (March 1982) **77**, No. 377, 103-08.
24. Looney, S.W. and Gulledge, T.R. Jr.: "Use of the Correlation Coefficient With Normal Probability Plots," *Am. Statistician* (Feb. 1985) **39**, No. 1, 75-79.
25. DeGroot, M.H.: *Probability and Statistics*, Addison-Wesley Publishing Co., Reading, MA (1986) 338-49.
26. Johnson, N.L. and Kotz, S.: *Continuous Univariate Distributions*, John Wiley and Sons Inc., New York City (1970) 62-63.
27. Webb, M.G.: "Reservoir Description, K_f Sandstone, Redwash Field, Utah," paper SPE 7046 presented at the 1978 SPE Improved Oil Recovery Symposium, Tulsa, April 16-19.

Appendix A—ML Properties

The ML technique is a common estimation technique when the underlying PDF of randomly chosen data is known. The theory for normally distributed variables is particularly well established.[25,26] Because the permeability data $k_1, k_2 \ldots k_n$ are log-normally distributed, $\log_e(k_1), \log_e(k_2) \ldots \log_e(k_n)$ are $N(m_k, s_k^2)$ for some m_k and s_k. The ML estimator for s_k^2 is[26]

$$(s_k^2)_{est}=(1/n)\Sigma[\log_e(k_i)-(m_k)_{est}]^2, \quad \text{(A-1)}$$

where $(m_k)_{est}=(1/n)\ \Sigma \log_e(k_i)$ and the summations are over the n samples in the data set. The quantity $n(s_k^2)_{est}/s_k^2$ has a χ^2 distribution with $n-1$ degrees of freedom.[25] Hence, the estimate $(s_k^2)_{est}$ has the properties $\mathrm{E}[(s_k^2)_{est}]=[(n-1)/n]s_k^2$ and $\mathrm{var}[(s_k^2)_{est}]=2[(n-1)/n^2]s_k^4$, where E and var represent the expectation and variance operators, respectively. Clearly, the estimator is biased and we can correct for that by using $1/(n-1)$ instead of $1/n$ in Eq. A-1. We also want to obtain $(s_k)_{est}$, not $(s_k^2)_{est}$ or $[(s_k)_{est}]^2$, so a correction will be required for taking the square root.

Corrected for both bias effects, Eq. A-1 becomes[26]

$$(s_k)_{est}=[1+0.25/(n-1)]\{[1/(n-1)]\Sigma[\log_e(k_i)-(m_k)_{est}]^2\}^{1/2}. \quad \text{(A-2)}$$

For this estimator (Eq. A-2), $\mathrm{E}[(s_k)_{est}]=s_k$ and $\mathrm{var}\ [(s_k)_{est}]=s_k^2/2n$, neglecting terms of order n^2 and higher. Using the estimate $(s_k)_{est}$ of Eq. A-2 in Eq. 8 and recognizing that $\mathrm{var}\{\exp[-(s_k)_{est}]\}$ is related to the moment-generating function of a normal variate, we obtain $\mathrm{E}[(K_{DP}^{ML})_{est}]=K_{DP}-s_k^2\exp(-s_k)/(4n)$ and $\mathrm{var}[(K_{DP}^{ML})_{est}]=s_k^2\exp(-2s_k)/2n$ (neglecting terms of order n^2 and higher) for the proposed estimator, $(K_{DP}^{ML})_{est}=1-\exp[-(s_k)_{est}]$.

Appendix B—K_{DP} Estimation for Non-Log-Normal Permeability Distributions

For the case where the permeability distribution is not log-normal (i.e., any PDF except the log-normal distribution), there is no theoretical relationship similar to Eq. 5 to relate the distribution to a value for K_{DP}. For the p-ND case, the curve on the log-normal probability plot will be concave downward or concave upward for $p>0$ or $p<0$, respectively.[12]

The procedure advocated by Dykstra and Parsons[6] includes instructions for drawing a best-fit line on the probability plot if the curve is not straight. This procedure thus takes a non-log-normal distribution (the curve) and produces an "equivalent" curve (the best-fit line) with a log-normal distribution. The estimate for K_{DP} is based on that equivalent curve and, as Willhite[9] has observed (his Example 5.6), may overstate the level of heterogeneity in the reservoir.

We could adopt either of two approaches to define K_{DP} for non-log-normal PDF's: (1) use the percentile estimates $[k_{(16)}]_{est}$ and $[k_{(50)}]_{est}$ obtained from some best-fit line procedure (which we would have to define mathematically) or (2) use the estimates $[k_{(16)}]_{est}$ and $[k_{(50)}]_{est}$ obtained directly from the data. We selected the second approach because it (1) is a simple method for obtaining $(K_{DP})_{est}$, (2) avoids such poorly defined techniques as drawing the best-fit line, (3) is consistent with the no-line approach already used by some investigators,[1,18,19] and (4) provides an ML estimator for p-ND populations for all p that is mathematically consistent with the case of $p=0$.

Thus, we extend the standard definition of K_{DP},

$$K_{DP}^E=1-k_{(16)}/k_{(50)}, \quad \text{(B-1)}$$

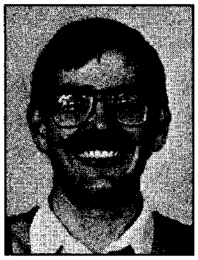
recognizing that $k_{(16)}$ and $k_{(50)}$ could be the 16th and 50th percentiles of any permeability distribution. From Eq. B-1, we obtain the estimator

$$(K_{DP}^E)_{est}=1-[k_{(16)}]_{est}/[k_{(50)}]_{est}, \quad \text{(B-2)}$$

which, while resembling Eq. 2, has no best-fit-line aspect. When permeability is p-ND, $[(k)^p-1]/p$ is approximately $N(m_k, s_k^2)$. By knowing the PDF, we can substitute into Eq. B-1 the exact expressions for the two percentiles, $k_{(16)}=[1+p(m_k-s_k)]^{1/p}$ and $k_{(50)}=(1+pm_k)^{1/p}$, to give

$$K_{DP}^E=\begin{cases}1-[1-ps_k/(1+pm_k)]^{1/p}, & p\neq 0 \quad \text{(B-3)}\\ 1-\exp(-s_k), & p=0. \quad \text{(B-4)}\end{cases}$$

Authors




Jensen Currie

Jerry L. Jensen is a lecturer in the Petroleum Dept. at Heriot-Watt U. in Scotland. He previously worked for Services Techniques Schlumberger and Gearhart Industries. He holds a BS degree from Birmingham U., an degree from the U. of Houston, and a PhD degree from the U. of Texas. Jensen is SPE Student Chapter faculty sponsor at Heriot-Watt U. Iain Currie is senior lecturer in the Dept. of Actuarial Mathematics and Statistics at Heriot-Watt U. His main research interests are in statistical modeling, particularly as applied to problems in insurance. He holds a degree in pure mathematics from St. Andrews U. and a PhD degree from Heriot-Watt U.

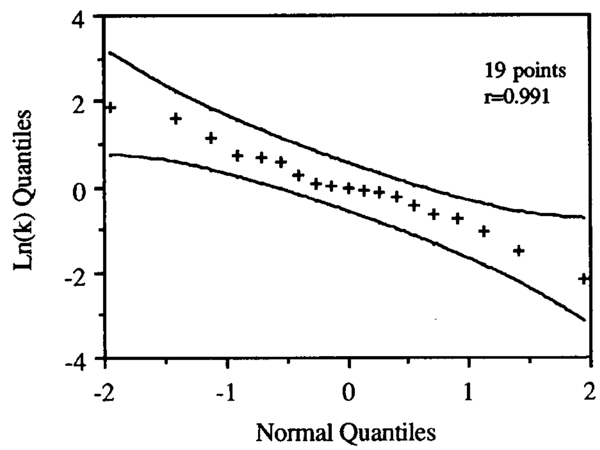


Fig. C-1—QQ plot for Set A.

Some properties of this expression have been described elsewhere.[12] Perhaps the most important feature is that K_{DP}^{E} values for two reservoirs, A and B, with significantly different exponents, p_{A} and p_{B}, are not comparable. For example, suppose $K_{\mathrm{DP}}^{EA}=K_{\mathrm{DP}}^{EB}=0.6$, $k_{(50)}{}^{\mathrm{A}}=k_{(50)}{}^{\mathrm{B}}=100$ md, $p_{\mathrm{A}}=0$, and $p_{\mathrm{B}}=0.2$. For Reservoir A, $k_{(84)}=270$ md, while for Reservoir B, $k_{(84)}=225$ md. Reservoir A is more heterogeneous because wider permeability variations are possible than with Reservoir B.

With estimated quantities, Eqs. B-3 and B-4 provide an alternative expression for estimating K_{DP}^{E}:

$$(K_{\mathrm{DP}}^{EML})_{\mathrm{est}}=$$

$$\begin{cases} 1-\{1-p_{\mathrm{est}}(s_k)_{\mathrm{est}}/[1+p_{\mathrm{est}}(m_k)_{\mathrm{est}}]\}^{1/p_{\mathrm{est}}}, & p_{\mathrm{est}}\neq 0 \quad \ldots\ldots(\text{B-5}) \\ 1-\exp[-(s_k)_{\mathrm{est}}], & p_{\mathrm{est}}=0, \quad \ldots\ldots(\text{B-6}) \end{cases}$$

where $(m_k)_{\mathrm{est}}=(1/n)\Sigma y_i$, ... (B-7)

$$(s_k)_{\mathrm{est}}=[1+0.25/(n-1)]\{[1/(n-1)]\Sigma[y_i-(m_k)_{\mathrm{est}}]^2\}^{1/2}, \quad \ldots\ldots(\text{B-8})$$

and $y_i=[(k_i)^{p_{\mathrm{est}}}-1]/p_{\mathrm{est}}$. Eqs. B-7 and B-8 are the same ML estimators described in Appendix A. These ML estimators have been found to be insensitive to errors in p_{est}.[15] Comparing the estimators (Eq. B-2 with Eq. B-5 or B-6) with Monte Carlo simulations shows that, when $p\neq 0$, the advantages of the ML technique are not quite as pronounced as they are for $p=0$. For equal standard errors of $(K_{\mathrm{DP}}^{E})_{\mathrm{est}}$ and $(K_{\mathrm{DP}}^{EML})_{\mathrm{est}}$, the ML method requires one-third the number of data required by the percentile method.

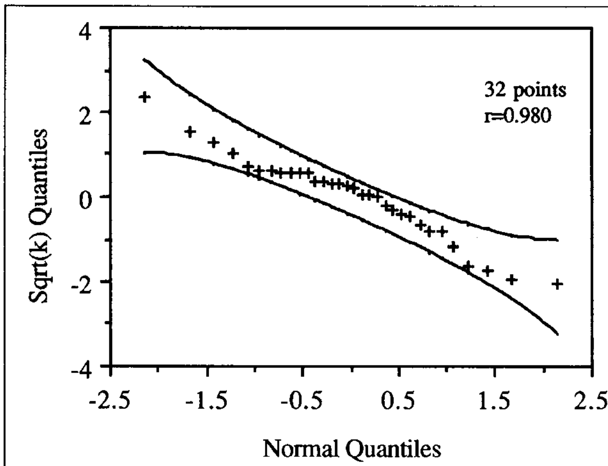


Fig. C-2—QQ plot for Set B.

TABLE C-1—TRADITIONAL AND PROPOSED ESTIMATOR PROPERTIES

	Set A ($p_{\mathrm{est}}=0$)	Set B ($p_{\mathrm{est}}=0.5$)
$(K_{\mathrm{DP}}^{E})_{\mathrm{est}}$	0.62	0.78
$(K_{\mathrm{DP}}^{EML})_{\mathrm{est}}$	0.68	0.72
$(m_{\mathrm{DP}}^{E})_{\mathrm{est}}$	−0.014	−0.028
$(m_{\mathrm{DP}}^{EML})_{\mathrm{est}}$	−0.005	−0.033
$(s_{\mathrm{DP}}^{E})_{\mathrm{est}}$	0.13	0.11
$(s_{\mathrm{DP}}^{EML})_{\mathrm{est}}$	0.06	0.07

Appendix C—Examples

We present two examples (Sets A and B) from field data to demonstrate the differences between the traditional and proposed techniques. **Figs. C-1 and C-2** are probability plots (as discussed in Refs. 14 and 15) based on data from the Redwash[27] (Set A) and the El Dorado[9] (Set B) fields, respectively. The upper and lower lines are the 95% confidence lines.

Fig. C-1 indicates that the data come from a log-normally distributed population—i.e., $p_{\mathrm{est}}=0$. A similar analysis of Fig. C-2 suggests that $p_{\mathrm{est}}=0.5$ for Set B. On the basis of these p_{est} values, **Table C-1** lists the results for the traditional and proposed methods for the reservoir K_{DP} estimates. The biases and standard errors listed were calculated from the relevant equations in the text for Set A, while Monte Carlo simulation was used for Set B.

For both cases, the bias and standard error performances of the ML approach give the K_{DP} estimates superior statistical properties. The ML estimator for Set A gives a more pessimistic indication of the reservoir heterogeneity level. The opposite is the case for Set B. The bias and standard error performances of both estimation methods is apparently poorer for Set B than for Set A, particularly when the relative sizes of the data sets are considered. As discussed in Appendix B, however, the heterogeneity levels represented by the two sets are significantly different. Depending on the process to be implemented, the standard errors for Set A could represent much larger variations in reservoir performance than the approximately equal errors listed for Set B. In any case, the ML approach has evidently reduced the uncertainty associated with the heterogeneity estimates.

SI Metric Conversion Factor

md × 9.869 233 E−04 = μm^2

SPERE

Original SPE manuscript received for review March 22, 1988. Paper accepted for publication Feb. 26, 1990. Revised manuscript received Oct. 13, 1989. Paper (SPE 17364) first presented at the 1988 SPE/DOE Enhanced Oil Recovery Symposium held in Tulsa, April 17–20.