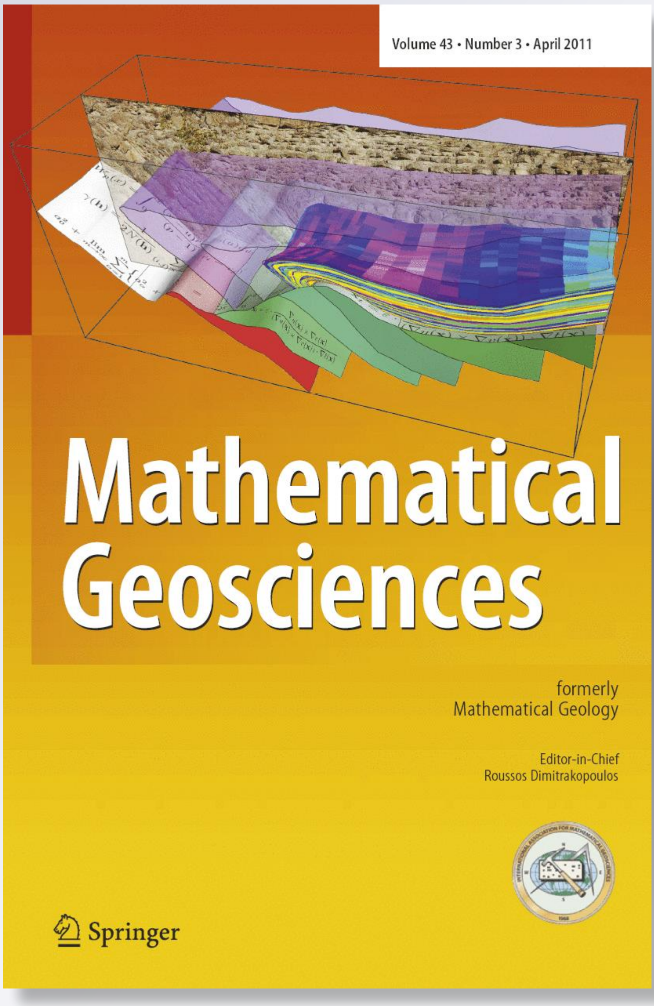
Asymptotic Dykstra–Parsons Distribution, Estimates and Confidence Intervals

Mathematical Geosciences

ISSN 1874-8961
Volume 43
Number 3

Math Geosci (2011) 43:329-343
DOI 10.1007/
s11004-011-9327-8



Math Geosci (2011) 43: 329–343
DOI 10.1007/s11004-011-9327-8

Asymptotic Dykstra–Parsons Distribution, Estimates and Confidence Intervals

Salvador Pintos · Claritza Bohorquez · Nestor V. Queipo

Received: 9 February 2009 / Accepted: 29 August 2010 / Published online: 22 March 2011

Abstract The sample Dykstra–Parsons (DP) coefficient, the most popular heterogeneity static measure among petroleum engineers, may exhibit significant sampling errors. Moreover, approximations of its probability distributions (uncertainty estimates) are only available for specific families of permeability models (e.g., log-normal). The cited probability distributions allow for the specification of confidence intervals and other inferences for the theoretical DP, which will be useful for reservoir screening purposes, or to establish if a more detailed study is justified. This paper presents the development of an asymptotic approximation of the distribution of the sample Dykstra–Parsons coefficient, which is independent of the permeability probability distribution. The effectiveness (bias and confidence intervals) of the proposed approach is demonstrated using analytical and field case studies and by comparing the results gleaned with those obtained using a well-known parametric approximation, under different scenarios of reservoir maturity levels (i.e., number of wells) and different degrees of deviation from the log-normal probability density function assumption. The results show that, in the vast majority of the case studies, the proposed approach outperformed the parametric approximation; in particular, our approach resulted in a significant reduction of the bias and the confidence intervals always including the theoretical DP coefficient. In addition, an excellent agreement was observed between the asymptotic cumulative distribution of the DP coefficient and the corresponding empirical distribution for sample sizes as small as one hundred, which suggests that high success rates can be obtained when reservoirs are classified according to the asymptotic DP coefficient.

S. Pintos · C. Bohorquez · N.V. Queipo (✉)
Applied Computing Institute, Faculty of Engineering, University of Zulia, Maracaibo, Venezuela
e-mail: nqueipo@ica.luz.ve

S. Pintos
e-mail: spintos@ica.luz.ve

C. Bohorquez
e-mail: claritza@ica.luz.ve

Keywords Dykstra–Parson coefficient · Reservoir heterogeneity characterization · Asymptotic order statistics

1 Introduction

In the context of enhanced oil recovery projects, heterogeneity (the spatial variation of properties) has long been recognized as a key factor in the prediction of reservoir performance. The latter is measured in terms of the amount of petroleum recovered, the time to breakthrough, and the peak hydrocarbon production (Jensen et al. 1987; Jensen and Lake 1988; Lake and Jensen 1991; Jensen and Currie 1990). While the complexity of the heterogeneity/performance relationship is well documented, for the purpose of screening or to establish whether a more detailed study is justified, the sample Dykstra–Parsons (DP) coefficient remains the most popular static measure of heterogeneity static measure among petroleum engineers. Recent applications include its use in sensitivity analysis (McCoy and Rubin 2009; Alajmi et al. 2009; Bossie-Codreanu and Le Gallo 2004; Adewusi 2002), reservoir heterogeneity classification (Mergany 2007) and upscaling techniques (Maschio and Schiozer 2003). The DP coefficient estimates may, however, be at a significant error, which may lead to both unrealistic reservoir performance predictions and unsuccessful development plans. As a result, assessing the DP coefficient's bias and confidence intervals is a project of considerable interest.

The classic form of the DP coefficient represents a robust estimate of the well-known coefficient of variation σ/μ, a normalized measure of dispersion of a normal random variable that is used for describing reservoir permeability heterogeneity. However, uncertainty estimates of the classical DP are only available on the assumptions that the reservoir property of interest (typically permeability) exhibits a log-normal probability density function, or that there is a transformation (Box–Cox) that can lead to normal behavior (Jensen and Lake 1988). These assumptions frequently do not hold. This paper presents a novel development of an asymptotic approximation of the sample DP distribution which is independent of the permeability probability distribution. The development is based on a result from order statistics: the asymptotic normal behavior of the joint probability distribution of sample quantiles (Cramer 1999; David and Nagaraja 2003).

Section 2 discusses the theoretical and sample DP coefficient; Sect. 3 presents a frequently used parametric approach for approximating the sample DP distribution; Sect. 4 develops an asymptotic approximation of the sample DP distribution; and Sect. 5 includes a description of the case studies used for evaluating the relative performance (bias and confidence intervals) of the proposed approach under different scenarios, of reservoir maturity levels (i.e., number of wells), and degrees of deviation from the log-normal probability density function assumption. Finally, Sects. 6 and 7 discuss the results obtained and the most significant conclusions derived from them, respectively.

2 The Dykstra–Parsons Coefficient

The Dykstra–Parsons coefficient (Dykstra and Parsons 1950) is a variability measure that overcomes the limitations (such as sensitivity to extreme values) of the classical coefficient of variation for asymmetric probability distributions by substituting the statistics σ and μ by analog quantities calculated using order statistics (quantiles).

More specifically, the coefficient of variation—that is, $(\mu - (\mu - \sigma))/\mu$—can also be written as $[\phi^{-1}(p_2) - \phi^{-1}(p_1)]/\phi^{-1}(p_2)$, if ϕ^{-1} denotes the inverse cumulative distribution function (cdf) of a normal probability distribution and p_1 and p_2 represent $\phi(\mu - \sigma) \approx 0.159$ and $\phi(\mu) = 0.5$, respectively. Substituting ϕ by F—that is, the cumulative probability distribution of the property of interest (random variable X)—the theoretical Dykstra–Parsons ($\mathrm{DP_T}$) coefficient can then be expressed as

$$\mathrm{DP_T} = \frac{F^{-1}(p_2) - F^{-1}(p_1)}{F^{-1}(p_2)} = \frac{x_2 - x_1}{x_2}, \tag{1}$$

where x_p is the quantile of the probability distribution of X associated with probability p, such that $F(x_p) = p$. Note that x_2 is the median of the population and that, for positive random variables such as permeability, $0 < \mathrm{DP_T} < 1$.

Since only a sample (size n) $X_1, \ldots, X_n$ of the random variable X is available, it is ordered such that $X_{(1)} \le X_{(2)} \cdots \le X_{(n)}$, where each element $X_{(i)}$ represents the ith order statistic. The sample DP_n coefficient (2) is then calculated using sample quantiles $q_n(p)$ where, for a given probability p, $q_n(p)$ is the hth order statistics $X_{(h)}$, with $h = [np] + 1$ corresponding to the sample size n. The symbol [.] denotes the integer part operator. This is expressed as follows

$$\mathrm{DP} = \frac{q_n(0.5) - q_n(0.159)}{q_n(0.5)} = 1 - \frac{q_n(0.159)}{q_n(0.5)}. \tag{2}$$

Please note that the sample DP coefficient (1) is a variability measure not limited to log-normal scenarios (Jensen and Currie 1988; Lake and Jensen 1991).

3 Dykstra–Parsons Coefficient Estimator: Parametric Scenario

Jensen and Currie (1990) show that if the probability distribution of the permeability (random variable X) is assumed to be log-normal, it is possible to obtain a consistent DP coefficient estimator that outperforms (in terms of bias and variance) the corresponding sampled DP_n coefficient. More specifically, if the random variable X is log-normal (μ, σ^2), the theoretical DP can be expressed as

$$\mathrm{DP_T} = 1 - \frac{e^{\mu - \sigma}}{e^{-\mu}} = 1 - e^{-\sigma}.$$

As expected, the theoretical DP grows with increasing values of σ. Since σ is unknown in the expression above, it is substituted by an unbiased estimator $w = \frac{s}{C4}$, where $C4 \cong 1 - \frac{1}{4n} - \frac{7}{32n^2}$ and s is the sample (log of sample values) standard deviation.

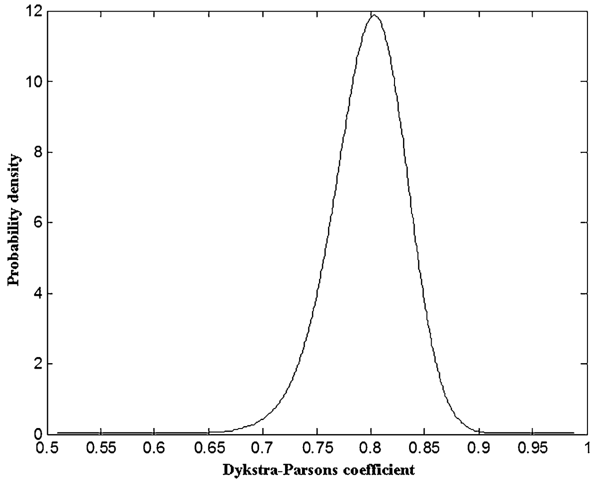


Fig. 1 Density function for the DP_{200} coefficient for 10,000 samples of size 200 from a log-normal distribution with $\mu = 800$ and theoretical $\mathrm{DP} = 0.8$

The Jensen and Currie (1990) estimator denoted as DPJ_n is given by

$$\mathrm{DPJ}_n = 1 - e^{-w}.$$

Hence, its expected value can be approximated using the Delta-Method (Casella and Berger 2002), by

$$E(\mathrm{DPJ}_n) \cong 1 - e^{-\sigma} - e^{-\sigma}\frac{\sigma^2}{4n}.$$

Note that the expression above underestimates the theoretical DP coefficient (DP_T) and that the bias $e^{-\sigma}\frac{\sigma^2}{4n}$ is proportional to the inverse of the sample size (n). The standard deviation of the Jensen and Currie estimator can be approximated as

$$\sigma(\mathrm{DPJ}_n) \cong e^{-\sigma}\frac{\sigma}{\sqrt{2n}}.$$

Assuming the DPJ_n estimator is normally distributed, Jensen and Currie set up a 95% confidence interval of the DP_T that is equal to

$$\mathrm{DPJ}_n \pm 2e^{-\sigma}\frac{\sigma}{\sqrt{2n}}.$$

Note that this may not be aligned with the observed asymmetric distribution of the sample DP_n for high DP coefficient values (Fig. 1), even when using log-normal data. More importantly, while the log-normal assumption has been useful, it is well-known that it frequently does not hold (Lambert 1981; Goggin et al. 1988; Jensen and Lake 1988).

4 Dykstra–Parsons Coefficient Estimator: Non-parametric Scenario

The true reservoir permeability probability distribution is unknown and often non-log-normal. However, it can be mistaken for log-normal based on normality tests (e.g., the Lilliefors test) (Lilliefors 1967), considering the small data sets typically available. DP estimates for improving the sample DP coefficient for non-log-normal scenarios can be obtained using the so-called Box–Cox of power transformations (Jensen and Currie 1990); this transformation assumes the permeability distribution as p-normally distributed (k^p is normally distributed). This section presents the development of an asymptotic distribution of the Dykstra–Parsons coefficient that overcomes the above referenced limitation; that is, it is independent of the permeability probability distribution. The cited development is based on two factors: (i) the sample DP coefficient is a function of two sample central quantiles, and (ii) a result in order statistics related to the asymptotic joint normal distribution of central quantiles. Given an ordered sample of size n of a random variable X denoted as $X_{(1,n)} \leq X_{(2,n)} \cdots \leq X_{(n,n)}$, a sample quantile $X_{(i,n)}$ is called central if as the sample size n grows, the ratio $X_{(i,n)}/n$ converges to a value different from 0 or 1.

The joint distribution of two central quantiles, denoted as $q_n(p_1)$ and $q_n(p_2)$ with $(p_1 < p_2)$, is asymptotically normal with (i) a mean μ equal to a vector with the corresponding quantiles of the population (David and Nagaraja 2003) as components; that is,

$$\mu = \begin{bmatrix} xp_1 \\ xp_2 \end{bmatrix},$$

and (ii) a covariance matrix COV$/n$, where

$$\mathrm{COV} = \begin{bmatrix} \frac{p_1(1-p_1)}{f^2(xp_1)} & \frac{p_1(1-p_2)}{f(xp_1)f(xp_2)} \\ \frac{p_1(1-p_2)}{f(xp_1)f(xp_2)} & \frac{p_2(1-p_2)}{f^2(xp_2)} \end{bmatrix}. \tag{3}$$

The symbol f denotes the probability density function associated with the random variable X. More precisely, the result states that when $n \to \infty$, the following expression converges to a normal probability distribution

$$n^{1/2} \begin{pmatrix} q_n(p_1) - xp_1 \\ q_n(p_2) - xp_2 \end{pmatrix} \quad \Rightarrow \quad N(0, \mathrm{COV}). \tag{4}$$

Note that (i) the symmetric and positive definite nature of matrix COV (3), and (ii) the asymptotic normal distribution of the individual quantiles that results from the bivariate normal distribution specified in (4).

The probability distribution of a linear combination of quantiles (Z)

$$Z = a \cdot q_n(p_1) + b \cdot q_n(p_2) \tag{5}$$

can be derived. Following the previous result, Z is asymptotically normal, with an expected value of

$$E(Z) = a \cdot xp_1 + b \cdot xp_2 \tag{6}$$

and a variance $Var(Z) = [a \quad b]\mathrm{COV}(Z)\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$. More precisely,

$$Var(Z) = a^2 \frac{p_1 q_1}{nf^2(xp_1)} + 2ab \frac{p_1 q_2}{nf(xp_1)f(xp_2)} + b^2 \frac{p_2 q_2}{nf^2(xp_2)}. \tag{7}$$

Hence, independently of the probability distribution of origin for the random variable X (permeability), the linear combination of quantiles Z follows a normal probability distribution with mean and variance as specified by (6) and (7). Now, the cumulative probability distribution of the sample Dykstra–Parsons coefficient $F_{\mathrm{DP}}(x)$ can be estimated. The $F_{\mathrm{DP}}(x)$ is represented by

$$F_{\mathrm{DP}}(x) = \mathrm{Prob}(\mathrm{DP}_n < x) = \mathrm{Prob}\left(1 - \frac{q_n(0.1587)}{q_n(0.5)} < x\right) \quad \text{for } 0 < x < 1,$$

which can be rearranged as

$$\mathrm{Prob}(\mathrm{DP}_n < x) = \mathrm{Prob}\big(0 < q_n(0.1587) + (x-1)q_n(0.5)\big). \tag{8}$$

The expression in parentheses on the right-hand side of (8) is a linear combination of quantiles, as specified in (6) with $p_1 = 0.1587$, $p_2 = 0.5$, $a = 1$ and $b = x - 1$. Hence, (8) can be rewritten as

$$F_{\mathrm{DP}}(x) = \mathrm{Prob}(\mathrm{DP}_n < x) = \mathrm{Prob}(0 < Z).$$

The variable Z asymptotically is normally distributed with an expected value of

$$E(Z) = x_1 + (x-1)x_2, \tag{9}$$

where $x_1 = F^{-1}(0.1587)$ and $x_2 = F^{-1}(0.5)$ are the quantiles required by the theoretical Dykstra–Parsons, and where the $\mathrm{DP_T}$ and the variance are equal to

$$\mathit{Var}(Z) = \frac{0.1587 * 0.841}{nf^2(x_1)} + 2(x-1)\frac{0.1587 * 0.5}{nf(x_1)f(x_2)} + (x-1)^2\frac{0.5 * 0.5}{nf^2(x_2)}. \tag{10}$$

If the expected value $E(Z)$ and square root of $\mathit{Var}(Z)$ in (9) and (10) are denoted by μ_z and σ_z, respectively, the asymptotic cumulative probability distribution of the Dykstra–Parsons coefficient can be written as

$$F_{\mathrm{DPA}}(x) = \mathrm{Prob}(0 < Z) = 1 - \mathrm{Prob}(Z < 0) = 1 - \Phi\left(\frac{-\mu_z}{\sigma_z}\right), \tag{11}$$

where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. Note that this expression depends only on the theoretical quantiles x_1 and x_2, the density function values $f(x_1)$ and $f(x_2)$, and the sample size n. The cited theoretical quantiles and density function values could be approximated using the corresponding sample values.

Once the F_{DPA} distribution is available, the expected value of the asymptotic DP coefficient can be calculated using

$$E(\mathrm{DP_A}) = \int_0^1 \big(1 - F_{\mathrm{DPA}}(x)\big)\,dx.$$

Since the $F_{\mathrm{DPA}}(x)$ is available for every x, it is possible to numerically estimate the above-referenced integral. The density function $f_{\mathrm{DPA}}(x)$, however, can be analytically established directly by differentiating (11). Asymptotic confidence intervals

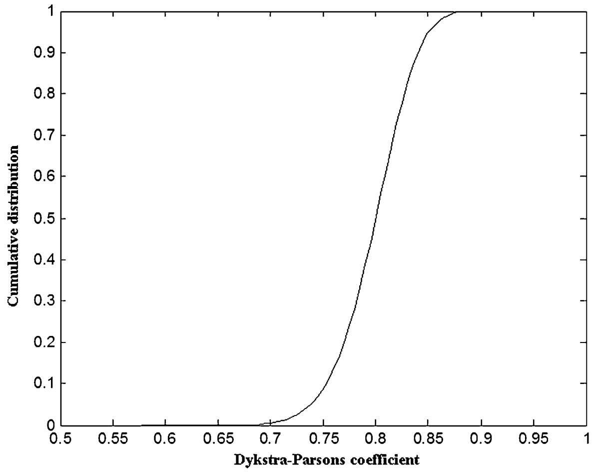


Fig. 2 Cumulative density function of the asymptotic DP coefficient $F_{\mathrm{DPA}}(x)$ (illustrative example)

$[L, U]$ with a $1-\alpha$ associated probability can also be obtained as $L = F_{\mathrm{DPA}}^{-1}(\alpha/2)$ and $U = F_{\mathrm{DPA}}^{-1}(1-\alpha/2)$. Note that the median of the asymptotic DP coefficient is equivalent to the $\mathrm{DP_T}$, since $F_{\mathrm{DPA}}(x) = 0.5$ corresponds to $\mu_z = 0$ and since from $x_1 + (x-1)x_2 = 0$, the $\mathrm{DP_A}$ coefficient is given by

$$x = \frac{x_2 - x_1}{x_2} = \mathrm{DP_T}.$$

4.1 Illustrative Example

Let us assume that a given reservoir has a permeability that is log-normally distributed, with a mean equal to 800 md, and a $\mathrm{DP_T}$ coefficient equal to 0.8; the sample size is equal to two hundred. The associated log-normal distribution parameters μ and σ, are 5.390 and 1.609, respectively, as calculated below.

$$\mu = Log(mean) - 0.5 * \sigma^2 = 5.390, \qquad \sigma = -Log(1 - \mathrm{DP}) = 1.609$$

The following five steps establish the cumulative density function of the asymptotic DP coefficient $F_{\mathrm{DPA}}(x)$ for a sample size n.

(1) Establish the quantiles corresponding to x_1 and x_2 associated with the 0.159 and 0.5 probabilities, as specified in the DP formula, namely: $x_1 = \exp(\mu - \sigma) = 43.817$; $x_2 = \exp(\mu) = 219.087$.
(2) Compute the log-normal density values at x_1 and x_2: that is, $f(x_1) = 0.00343$, $f(x_2) = 0.00113$.
(3) Calculate $E(Z) = x_1 + x_2 * (x - 1) = -175.27 + 219.09 * x$ (from (9)).
(4) Calculate $Var(Z) = 828.84 - 1748.6 * x + 976.49 * x^2$ (from (10)).
(5) Since there is no analytical expression for the cdf of a normal probability distribution, the $F_{\mathrm{DPA}}(x)$ is constructed point-wise using a set of values for x in

the interval [0, 1] (Fig. 2). In this example, for $x = 0.75$, the value of F_{DPA} is calculated as

$$E(Z) = \mu_z = -10.955, \qquad Var(Z) = 66.639, \qquad \sigma(Z) = 8.163,$$

and from (11),

$$F_{\mathrm{DPA}}(0.75) = 1 - \Phi\left(-\frac{\mu_z}{\sigma_z}\right) = 1 - \Phi(1.3419) = 0.090.$$

The reservoir permeability distribution is often unknown. In this case, sample quantiles are used in Step 1, and the density function in Step 2 must be estimated (kernel density estimation). No further changes are necessary for estimating the F_{DPA}.

5 Case Studies

The case studies correspond to two reservoirs, one whose permeability is analytically specified as a two-component mixture of log-normal probability distributions, and one with real permeability sample data. In both instances, the relative performance of the parametric and asymptotic nonparametric approaches is established firstly by comparing the DP estimation bias, and secondly by considering different degrees of deviation from the log-normal assumption, regardless of whether the confidence intervals contain the theoretical DP coefficient. In addition, in the case where permeability is analytically specified, the effectiveness of the proposed asymptotic DP estimator for decision-making (e.g., classifying a reservoir as low, medium, or highly heterogeneous) is also evaluated.

5.1 Two-component Mixture of Log-normal Probability Distributions

This case study is designed to model situations where an oil reservoir is represented by two lithofacies and where a mixture of log-normal probability distributions may provide a reasonable approximation to characterize its permeability. The log-normal assumption is evaluated (p-value of Lilliefors test) (Lilliefors 1967) using different samples of the mixture distribution under alternative scenarios of reservoir maturity levels (i.e., 100 and 200 wells). A Lilliefors test evaluates the hypothesis (null) that the data come from a normally distributed population, when the null hypothesis does not specify its parameters (μ, σ^2). Hence, the log-normality hypothesis can be tested by applying a Lilliefors test on the logarithm of the permeability data, where all data with a p-value lower than 0.05 are considered non-log-normal. Note that non-log-normal cases can be misinterpreted as log-normal based on normality tests (e.g., the Lilliefors test) in cases where only a small sample of the permeability distribution is available.

The parameters of the two-component mixture of log-normals are shown in Table 1, with the theoretical DP coefficient being equal to 0.667 (medium heterogeneity); see Lake and Jensen (1991) for a classification scheme based on the sample DP

Table 1 Parameters of the two-way mixture of log-normal probability distributions (case studies)

Parameter	Distribution 1	Distribution 2
Permeability mean (md)	600	145
Dykstra–Parsons coefficient	0.4	0.4
Weight in the mixture (%)	60	40

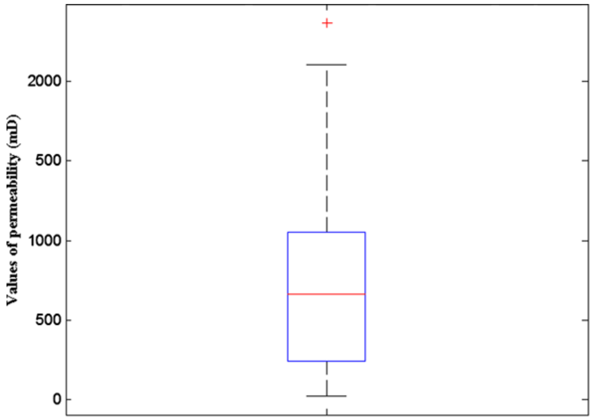


Fig. 3 Empirical distribution of permeabilities (field case study)

coefficient. The sampled density function for this mixture constructed using Parzen windows (Parzen 1962) is, in general, similar to single log-normal probability distributions. This may mislead practitioners insofar as the true nature of the permeability statistical characterization is concerned.

5.2 Real Permeability Sample Data

The data correspond to a 2596-acre Miocene oil reservoir in western Venezuela, which lies at a depth of 4500 feet and has an average thickness of 500 feet: as such, it belongs to a fluvial sedimentary environment (three lithofacies) with braided channels. The data come from 93 wells, with two randomly selected permeability sample values per well, for a total of 186 samples. All the available permeability samples are taken as the population, and the corresponding DP coefficient is labeled as the theoretical one. The results correspond to 20 sets of size 100. Figure 3 shows the empirical distribution of permeability data.

6 Results and Discussion

6.1 Two-component Mixture of Log-normal Probability Distributions

Table 2 shows the number of times a DP estimate exhibits the smallest bias for one hundred samples of different sizes (100 and 200) and the results of the Lilliefors test. In this context, the bias is the difference between the parametric/asymptotic DP and the theoretical DP (known in the analytical case study); ideally, a DP estimator should

Table 2 Number of times a DP estimate exhibits the smallest bias for one hundred samples of different sizes (100 and 200) and results of the Lilliefors test with a theoretical DP coefficient equal to 0.667

Lilliefors Test–Null hypothesis ($\alpha = 5\%$)	Sample size					
	100			200		
	Total	DP-parametric	DP-asymptotic	Total	DP-parametric	DP-asymptotic
Reject	58	4	54	93	2	91
Do not reject	42	1	41	7	0	7
Total	100	5	95	100	3	97

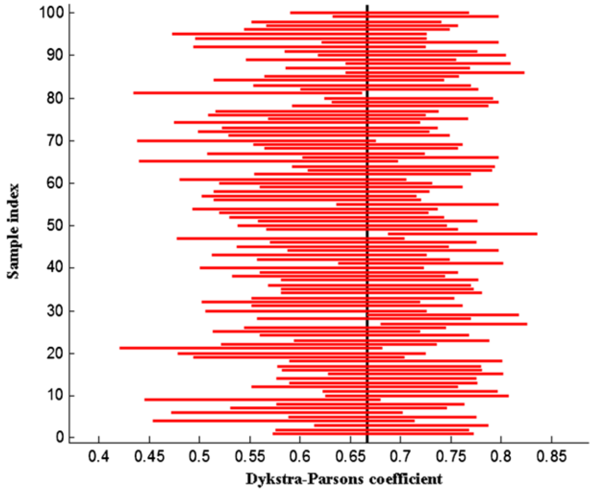
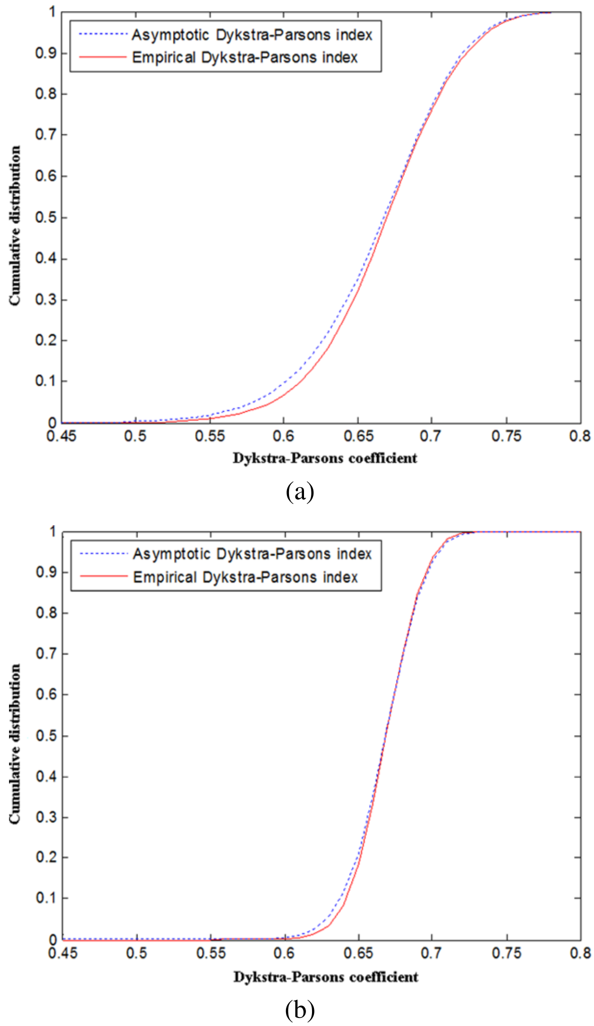


Fig. 4 Confidence intervals associated with the asymptotic DP (**a**) and parametric (**b**) estimates for a sample of size 100. The *vertical line* represents the theoretical Dykstra–Parsons coefficient. Case study: two-way mixture of log-normal probability distributions

be unbiased. In general, for samples of size 100, the DP_A estimates were close to the theoretical one, with the median of the differences between the theoretical DP value and the DP-parametric and asymptotic estimates being approximately 0.08 (biased) and 0 (unbiased), respectively. Similar results were obtained for the larger sample size (200). Even though the permeability distribution of interest is not log-normally distributed, in several instances the log-normal hypothesis could not be rejected; in such instances the DP_A estimates were always closer to the theoretical one than the parametric DP. Hence, not rejecting the log-normal hypothesis is not a sufficient condition for using the parametric DP approach: the permeability distribution may not be log-normally distributed, in which case and the asymptotic DP estimator may be a better choice.

For a sample of size 100, Figs. 4(a) and 4(b) exhibit the 95% confidence intervals associated with the DP asymptotic and parametric estimates, respectively. The latter only included the theoretical value of the DP coefficient (0.667) in three instances, while the former did on all but three occasions. Increasing the sample size to 200 yielded DP parametric estimates that were further away from the theoretical value. Sample independent results were also obtained using population parameters for computing the DP estimates; that is, quantiles and density function values (asymptotic),

Fig. 5 Asymptotic cumulative probability distribution of the Dykstra–Parsons coefficient and the corresponding empirical CDF for 1000 simulations of sample sizes of 100 (**a**) and 400 (**b**). Case study: two-way mixture of log-normal probability distributions



and standard deviation (parametric). The lower and upper limits for the asymptotic DP confidence interval were 0.557 to 0.746, which was close to the theoretical one (0.573 to 0.747). In contrast, the confidence interval associated with the parametric DP approach (0.528 to 0.631) is biased to the left and does not include the theoretical DP coefficient (0.667). Similar results were obtained for the larger sample size.

Figure 5 shows the excellent agreement between the asymptotic cumulative probability distribution of the DP coefficient and the corresponding empirical cdf, for 1,000 simulations of sample sizes of 100 (Fig. 5(a)) and 400 (Fig. 5(b)). This agreement opens the possibility of using the asymptotic $F_{\mathrm{DPA}}(x)$ to help support decisions related to reservoir classification. For example, the cdf could be used for screening purposes; to establish if a more detailed study is justified; or to classify the reservoir as low, medium, or highly heterogeneous according to the DP estimate value (Lake and Jensen 1991). A common reservoir classifying scheme labels a reservoir

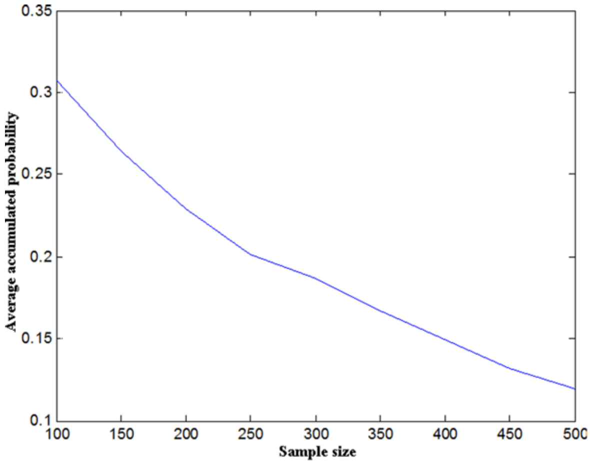


Fig. 6 Average probability of an asymptotic Dykstra–Parsons coefficient higher than 0.7 corresponding to 10,000 simulations. Case study: two-way mixture of log-normal probability distributions

as medium or highly heterogeneous according to whether the DP estimate is within the $0.5 < \mathrm{DP} < 0.7$, or $\mathrm{DP} > 0.7$ intervals, respectively. For this case, the DP coefficient is known (0.667) and is close to the classification boundary value of 0.7: accordingly, it would be relevant to investigate the probability of an asymptotic DP coefficient larger than 0.7 (hence, wrongly classifying it as highly heterogeneous), in the case that the DP coefficient is 0.667. Figure 6 shows the results corresponding to 10,000 simulations and different sample sizes.

Even though the theoretical DP coefficient is close to the classification frontier (0.667 in comparison with 0.7), the average probability of an asymptotic Dykstra–Parsons coefficient larger than 0.7 is about 0.3 for a sample size of 100 and, as expected, it significantly decreases with larger sample sizes. The cited average includes the probability of the asymptotic DP coefficient of each of the samples being higher than 0.7; this probability is computed using the asymptotic cumulative probability distribution (11). On a separate note, among the 10,000 simulations, the proportion of samples with asymptotic Dykstra–Parsons coefficients actually exceeding 0.7 was only about 20% (classification error) for the sample size equal to 100, and it decreased to about 9% and 4% for the sample sizes of 300 and 500, respectively.

6.2 Real permeability sample data

The DP-parametric estimator exhibits a significant bias with a median value of approximately 0.13 and with the ideal value of zero well outside the sample values; in contrast, the DP non-parametric estimator shows a bias with a median close to zero and with the interquartile interval including the zero value (Fig. 7). In all instances, the DP-asymptotic was closer to the theoretical DP coefficient (0.667) and the log-normal hypothesis was rejected using the Lilliefors test with a 5% significance level, confirming the potential of the DP-Asymptotic estimator for real case scenarios.

None of the DP-parametric confidence intervals include the theoretical DP coefficient (Fig. 8(b)) while the opposite result is observed for the asymptotic DP estimates (Fig. 8(a)). In the latter case, note how centered the theoretical DP coefficient is with respect to the confidence intervals.

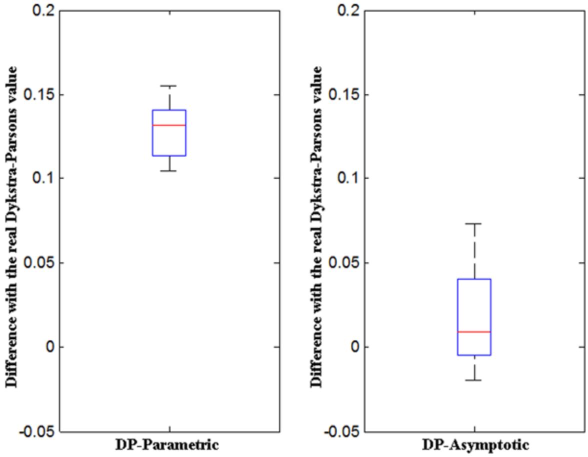


Fig. 7 Differences between the true DP coefficient and the estimators. Sample size: 20

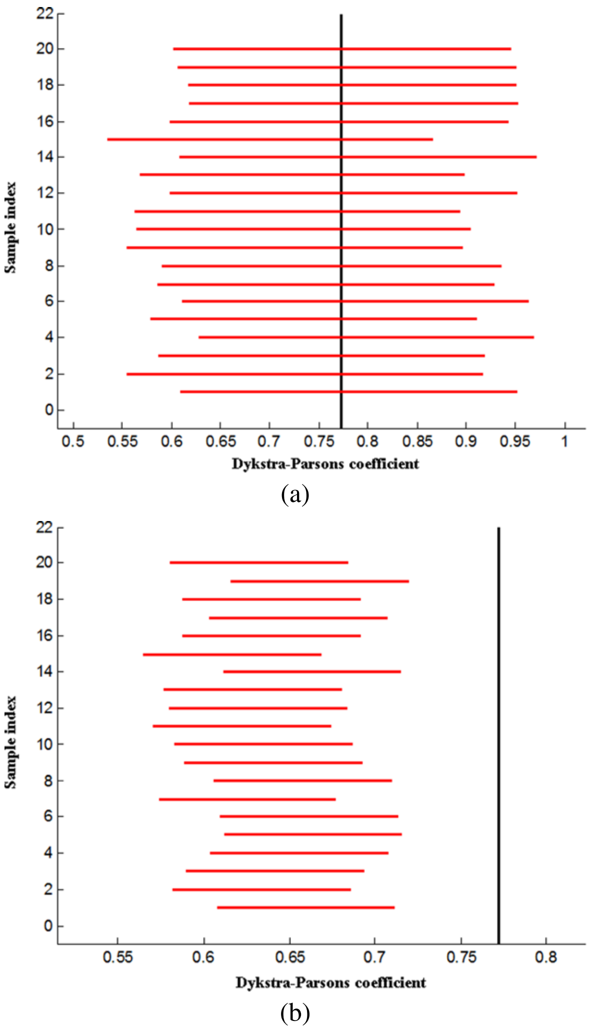


Fig. 8 Confidence intervals associated with the asymptotic DP (**a**) and parametric (**b**) estimates for a sample of size 20. The *vertical line* represents the theoretical DP coefficient

7 Conclusions

This paper presented the development of an asymptotic distribution of the Dykstra–Parsons coefficient that is independent of the permeability probability distribution. The effectiveness (through bias and confidence intervals) of the proposed asymptotic DP coefficient is demonstrated by comparing the results with those obtained using the DP parametric approach (log-normal distribution) under different scenarios of reservoir maturity levels (i.e., number of wells).

The results show that in the vast majority of the case studies, the asymptotic DP coefficient outperformed the parametric counterpart, independently of whether or not the log-normal probability density function assumption held (with $\alpha = 5\%$); in particular, the asymptotic DP coefficient resulted in a significant reduction of the bias and confidence intervals when including the theoretical DP coefficient. In addition, an excellent agreement was observed between the asymptotic cumulative distribution of the DP coefficient and the corresponding empirical distribution for sample sizes as low as 100, which makes it possible to classify reservoirs according to their DP coefficient with high success rates.

Real permeability data may frequently reject the log-normal hypothesis (e.g., Lilliefors test with a 5% significance level), which confirms the potential of the DP-asymptotic estimator for real case scenarios. The DP-asymptotic estimator coefficient can easily be implemented as a computational aid and has the potential to be successfully incorporated in the workflow of reservoir engineers in view of quantifying/classifying reservoir heterogeneity without making any assumptions about the permeability probability distribution.

References

Adewusi V (2002) EWAN C-2/C-3 polymer flood: performance predictions. Pet Sci Technol, 20(5–6):441–463

Alajmi A, Gharbi R, Algharaib M (2009) Investigating the performance of hot water injection in geostatistically generated permeable media. J Pet Sci Eng 66(3–4):143–155

Bossie-Codreanu D, Le Gallo Y (2004) A simulation method for the rapid screening of potential depleted oil reservoirs for CO2 sequestration. Energy 29(9–10):1347–1359

Casella G, Berger RL (2002) Statistical inference. Duxbury, Scituate

Cramer H (1999) Mathematical methods of statistics. Princeton University Press, Princeton

David HA, Nagaraja HN (2003) Order statistics. Wiley, Hoboken

Dykstra H, Parsons RL (1950) The prediction of oil recovery by water flood. In: American Petroleum Institute, Secondary Recovery of Oil in the United States, 2nd edn. API, Dallas, pp 160–174

Goggin DJ, Chandeler MA, Kocurek G, Lake LW (1988) Patterns of permeability in Eolian deposits. SPE Form Eval 3(2):297–306

Jensen JL, Currie I (1988) Improving performance prediction by more accurate heterogeneity assessment. SPE 17364, presented at the SPE/DOE Enhanced Oil Recovery Symposium held in Tulsa, OK, April 17–20, 1988

Jensen JL, Currie I (1990) A new method for estimating the Dykstra-Parsons coefficient to characterize reservoir heterogeneity. SPE Reserv Eng 3(7):369–374

Jensen JL, Lake LW (1988) The influence of sample size and permeability distribution on heterogeneity measures. SPE Reserv Eng 3(2):629–637

Jensen JL, Lake LW, Hinkley DV (1987) A statistical study of reservoir permeability: distributions, correlations and averages. SPE Form Eval 2(4):461–468

Lake LW, Jensen JL (1991) A review of heterogeneity measures used in reservoir characterization. In Situ, 15(4):409–440

Lambert ME (1981) A statistical study of reservoir heterogeneity. MS Thesis, University of Texas at Austin

Lilliefors H (1967) On the Kolmogorov–Smirnov test for normality with mean and variance unknown. J Am Stat Assoc 62:399–402

Maschio C, Schiozer D (2003) A new upscaling technique based on Dykstra–Parsons coefficient: evaluation with streamline reservoir simulation. J Pet Sci Eng 40(1–2):27–36

McCoy S, Rubin E (2009) The effect of high oil prices on EOR project economics. Energy Procedia 1(1):4143–4150

Mergany M (2007) Geological and statistical reservoir characteristics of the Late Carboniferous-Permian. Masters Abstr Int 46(1)

Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 33:1065–1076